The Health Impact of a Flexible Drug Approval Policy

MASTER'S THESIS ACTUARIAL STUDIES

Supervisor: Dr. W. Zhu
Co-assessor: Dr. N. D. van Foreest

Kris Rama

Jan 29, 2024

Abstract

Low productivity and opaque investment undermine the public trust in pharmaceutical companies. We propose an alternative approval policy based on a flexible significance threshold using a queuing network and following the path of a candidate drug. By controlling the key parameters of the evaluation process we determine the health impact associated with such a policy. We extend the current understanding by constructing a simulation environment from the historical data of individual drugs. A significant difference in disease area emerges in the optimal threshold which ranges from 1.6% in Metabolic/Endocrinology to 10.5% in Oncology. Our simulation shows that the health impact can be maximized by allowing for a flexible significance level. The framework we proposed can be used as a practical tool in the hands of policymakers, particularly in the context of regulatory convergence. Reimbursement agencies and insurers could benefit from the cost-benefit analysis intrinsic to this approach.

Keywords: Drug Approval, Efficacy, Significance, Queuing Network.

1 INTRODUCTION

The process of introducing a new drug as a Biologics License Application (BLA) or a New Molecular Entity (NME) is increasingly perceived as an uncertain and expensive venture by pharmaceutical companies, the only ones that can pursue it. The life of a drug unfolds through a series of rigorous steps, starting from the research in a lab and culminating in the final administration to the patients that benefit from it. A promising compound is subject to intense scrutiny to demonstrate its qualities against the current standard of care; a period spanning decades is dedicated to conducting clinical trials with the main intent of demonstrating that the drug is fit for the market. Only once a drug is approved, it can be sold exclusively for the remaining time of its patent.

Governments incentivize pharmaceutical companies through a combination of financial and regulatory measures, such as tax credits, direct funding, expedited drug approval processes, and granting extended periods of market exclusivity. The inherent intricacy of the sifting process, however, necessitates substantial investments of time and financial resources. Only a few of the thousands of compounds currently researched will eventually become available for patients. Governments are striving to align the health needs of the population with the incentives of the industries. The EU recently introduced the Supplementary Protection Certificate (SPC), which extends the patent life of new drugs by up to five years to offset the time lost during the research phase. The policy tool is set to encourage innovation by allowing companies to recoup their investments with a longer time in the market. Additionally, in the evolving EU regulatory landscape, the upcoming Joint Clinical Assessment (JCA) in 2025, is set to significantly influence the assessment of European health technology assessment (HTA) bodies, inducing a convergence in the criteria used and a unified framework for assessment. The different institutions have come up with a set of rules shaped by their medical history. There is some overlap but the difference remains substantial for particular health measures or patient groups. Uncertainties remain over the specifics of a unified framework and the strategies that will be enacted to implement it.

Patients looking for care, on the other hand, would welcome a policy that increases viable treatment options. The current standard requires an efficacy for candidate drugs to be significant at 2.5% compared to the alternative, a one-size-fits-all, that may not be the optimal level for all disease areas. In this contribution, we investigate the welfare effects of a flexible approval policy by constructing a queuing network that simulates the drug development paths. We search for the significance level across different disease areas that maximizes some quantitative health outcomes. The objective is to extend previous work in the field by offering more realistic results through a simulation-based approach and drug-level empirical values. An approval policy peculiar to the biological and market characteristics would increase drug production for areas that need more treatment options and reduce it for congested ones. Additionally, this method proposes to increase transparency and allow more flexibility in the approval process by rendering the current approval evaluation explicit and reproducible, something effective in the current regulatory

scene. In the rest of this section, we introduce the most important players involved, provide an overview of the challenges they face, and report the relevant findings from the literature.

1.1 *Pharma*

Pharmaceutical companies are dedicated to understanding the underlying causes of a medical condition or disease, creating suitable drugs, and persuading customers of their value. By investing in research and development (R&D) they identify new molecules or compounds that have the potential to become effective treatments that address unmet medical needs. The exact cost associated with developing a new drug, however, is often covered in a veil of secrecy. As Morgan et al. (2011) points out, undisclosed data provided by anonymous pharmaceutical companies often about unidentified products or samples, hinders the ability to evaluate the precision, or susceptibility of these estimates. This lack of transparency is a significant obstacle to understanding the trends and market dynamics for a reliable forecast. Nonetheless, it has been clear in recent years that the industry has faced challenges in converting scientific breakthroughs and fundamental research into viable therapeutic options. As a result, there has been a decline in the number of drugs receiving approval, despite an overall increase in R&D expenditure. Less than 10% of drug candidates that enter clinical trials ultimately secure market authorization. For those that do succeed, the median duration to obtain market approval is 8.3 years, with a median capitalized R&D investment of $1.14 billion, as reported by Brown et al. (2021).

1.2 *Unmet Medical Needs*

Unmet medical needs refer to conditions or patient populations for which no current treatment options produce satisfactory outcomes. Examples include the absence of efficacious therapies, limited treatment, and deficiencies in current treatment modalities. This principle is an essential instrument for both policy creation and the commercial sector, assisting in pinpointing critical public health demands and ultimately steering choices regarding investments. Identifying unmet medical needs is often initiated by healthcare providers due to their proximity to patients. They observe the limitations of current treatments and identify discrepancies between academic evidence and the practical experiences of healthcare practitioners. These findings are then communicated to research centers for further investigation. Currently, one of the main challenges in this area is the lack of a standardized method, leading to variations in what is considered an unmet need.

1.3 *Risk Sharing*

Clinical research is typically a collaborative endeavor involving pharmaceutical companies, academic institutions, research organizations, and government agencies. These partnerships allow for risk sharing, shared initial investment, and access to complementary capabilities and technologies. Governments play an active role in all stages of the

development process, agilely changing roles as needed, acting as an investor, a regulator, or a consumer. Nonetheless, the public and private investment that fuels this research, is increasingly subject to debate. Questions remain about the allocation of funds, the balance between public and private contributions, and the actual return on public investment.

Like other businesses, pharmaceutical companies participate in the equity market, where they must maintain the trust and confidence of their shareholders. Even so, the market is particularly delicate. The presence of competing drugs, expiring patents, and potential market entry from generic or biosimilar alternatives creates pressure for innovation, while investments are of a high-risk nature due to the uncertainty of the drug discovery and development process. To protect these investments and ensure a return, pharmaceutical companies heavily rely on patent protection for their marketed products. Patents provide a period of exclusivity, typically 20 years, during which the company has the sole right to manufacture and sell the drug (Caves et al. (1991)). This allows the company to recoup its investment and make a profit before generic versions can enter the market.

1.4 *Cost-Benefit Analysis*

Reimbursement agencies or insurers perform cost-benefit analyses to decide if covering a drug is economically viable and cost-saving in the long term. Upon market approval, HTA agencies scrutinize the drug's benefits and cost implications to inform healthcare decisions and policy. A high level of trust is required by the final consumers, and the meticulous and transparent validation process serves as a guarantee of its therapeutic value. Any biologically active compound carries the risk of unexpected outcomes and potential side effects that cannot be observed in the limited sample, although well-designed, of a clinical trial. In this work, we consider the cost-benefit analysis as an explicit component of the approval process.

1.5 *Medical Agencies*

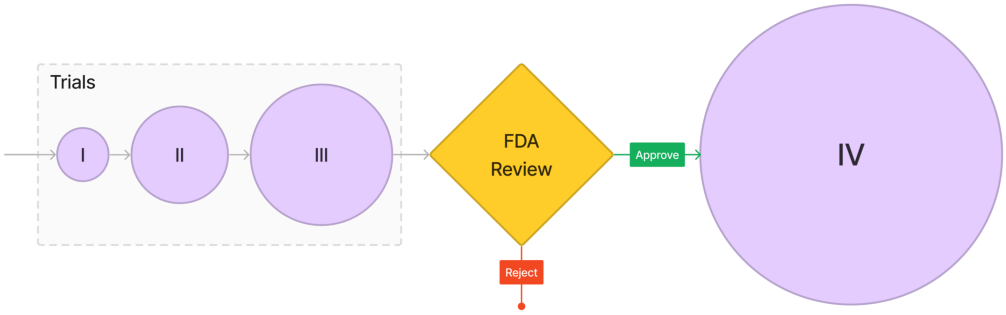
In the United States, the Federal Drug Administration (FDA), and in Europe, the European Medical Agencies (EMA), serve as the principal agencies overseeing the authorization of new pharmaceuticals and their subsequent monitoring. While there are variations in the approval guidelines between the two authorities, these differences are not substantial. A study by Kashoki et al. (2020) examined the initial and final decisions for drug applications submitted to both the FDA and EMA, highlighting the marketing approvals granted for an indication, a term that refers to the specific disease or condition and patient group. The disparities were classified into six areas: clinical data, efficacy assessments, safety conclusions, regulatory judgments, quality standards, and data integrity. The investigation identifies factors that contribute to the differing outcomes between the EMA and FDA, with the primary determinant being the clinical data presented at the time of application. The overarching principles guiding both agencies align closely. The benefit-risk evaluations

conducted by both the FDA and EMA are inherently qualitative and tailored to each drug, reflecting a shared approach in considering the severity of the condition and the availability of other therapeutic options in their decision-making processes. Given these similarities in objectives, methodologies, and considerations, it is conceivable to consider FDA and EMA in a unified framework.

1.6 *The Approval Process*

Current policy requires pharmaceutical companies to initially demonstrate the safety of the compound with evidence of no adverse effects, and subsequently verify its efficacy in improving the health outcomes of the target condition. The primary evidence used for deliberation is the clinical data from the trials. The final decision revolves around a delicate balance between providing patients with a wide choice of potential positive treatments while safeguarding consumers from harmful adverse events.
The approval process evaluates both the efficacy and safety of a drug managing the risk of approving an ineffective drug (*Type I* error) or rejecting an effective one (*Type II* error) using statistical hypothesis testing. Current guidelines, shared by both agencies, tolerate efficacy significant at level $\alpha = 2.5\%$ for all diseases. The fixed threshold makes it impartial, but the exact value is somewhat arbitrary. A patient with few alternatives or who has not responded to any previous treatment is willing to accept a higher risk than one who has plenty of alternatives. We capture the increase in patient population before and after the review in Figure 1.

Figure 1: Patients Treated Before and After Market Approval



Notes: Phase I involves testing the new drug on a small group of healthy volunteers to observe its pharmacokinetics, how the drug is metabolized. In Phase II, the drug is given to a larger group of people who have the condition the drug is designed to treat. During Phase III the drug is tested on a large group of patients to confirm its effectiveness. Success in the Trials leads to an IND submission for the FDA to process.

1.6.1 *The Case for a Flexible Approval Policy*

Here, we move within the quantitative research of the optimal design, considering the role of the significance level threshold on approval outcomes. A recent study by Bravo et al. (2022) critically analyzed the FDA's fixed threshold for novel drug approval. They argued that while the FDA often uses regulatory discretion when interpreting these standards, the current guidelines consider factors such as disease severity, prevalence, and availability of existing therapies in approval decisions, but no quantitative elements are publicly disclosed. The study proposed a queuing framework to analyze the FDA's drug approval process, incorporating disease-specific factors and obsolescence—newer drugs replacing older formulas—through a set of pre-emptive queues. Using public data encompassing all registered U.S. clinical trials and FDA-approved drugs, they estimated model parameters for three high-burden diseases (breast cancer, human immunodeficiency virus (HIV), and hypertension) and solved for the optimal policy to maximize net life-years gained following FDA approval. The results indicated that the optimal policy relaxes approval standards for diseases with long trial duration, high attrition, or low research and development intensity. This suggests 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. A study of this type 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. Other research streams have focused on flexible policies, the main approaches can be summarized by the work of:

- Bravo et al. (2022) maximized expected net health outcomes using a queuing model quantitatively including disease severity, prevalence, and alternative treatment in the decision process.
- Montazerhodjat et al. (2017) focused on minimizing the expected harms associated with enrolling patients in a trial, including the costs of treatment with a toxic drug or missing treatment with effective therapy, as well as the post-approval impact on disability-adjusted life expectancy.
- Isakov et al. (2019) introduced a Bayesian decision analysis (BDA) to minimize the expected health cost of drug approval.
- Delshad et al. (2023) developed an analytical tool to assess the direct impact of the FDA's approval policies on the research level of pharmaceutical companies in a game with three players: the FDA and two competing companies.
- Patriarca et al. (2018) proposes a quantitative risks/benefits ration assessment based on *QALY* in a logistic regression framework.

The previous works have contributed valuable insights into the field of flexible approval, but they all come with a limited empirical grounding. In contrast, our approach addresses these concerns by simulating all relevant stages from a rich set of empirical distributions and historical data. This method offers a more realistic and pragmatic framework for

assessing the impact of drug approval policies. This empirical foundation not only enhances the granularity of the results but also makes the methodology more appealing for practical application. Regulators can see the direct implications of their decisions, and pharmaceutical companies can better understand how their research efforts might be affected by the current policy. A more accessible approval process can improve the treatment options for patients with few alternatives and guide investment decisions with clear targets.

1.7 *Overview*

Section 2 describes the techniques and assumptions used in our analysis. Section 3 provides an overview of the sources or methods used to estimate the parameters used in our simulation exercise. Section 4 showcases the results obtained by our modeling approach in the differences per disease areas. Section 5 summarizes the findings while reiterating the need for prudence. Section 6 concludes.

2 METHOD

2.1 *Trial Types*

The series of clinical trials is designed to minimize harm to patients while generating enough statistical power for evaluation with the primary quantitative endpoint depending on the targeted condition. For a cancer drug, the primary endpoint could be the five-year survival rate of patients in the group that received the experimental therapy compared to the group that received the current treatment if it exists, or a placebo, if not. The main types of trials are:

SUPERIORITY TRIALS conducted to establish that a new treatment is better than an existing one or a control. The null hypothesis in these trials posits that the new treatment is not superior by some treatment measure, which is predefined based on clinical significance. Sample sizes for these trials depend on various factors including the expected difference in treatment effects, variance, significance level, and statistical power. A larger margin of clinical significance makes it more challenging to demonstrate superiority as it increases the required sample size.

NON-INFERIORITY TRIALS that aim to show that a new treatment is not significantly worse than the standard treatment by more than a small margin. These trials are useful when the new treatment may offer other benefits such as reduced cost or side effects. The null hypothesis here suggests that the new treatment is inferior by more than a specified value, while the alternative hypothesis indicates non-inferiority within this margin. Establishing non-inferiority typically requires smaller sample sizes compared to superiority trials because the margin of clinical significance is usually set lower.

EQUIVALENCE TRIALS designed to demonstrate that two treatments produce similar effects, falling within a tolerance. In these trials, the null hypothesis states that the difference between treatment effects lies outside the acceptable range, whereas the alternative hypothesis asserts equivalence within this range. Equivalence trials can be imagined as intersecting two non-inferiority trials, where neither treatment is considered inferior to the other. Similar to non-inferiority trials, the sample size increases with a larger tolerance margin.

In our modeling approach, we consider a balanced two-armed superiority randomized controlled trial (RCT) design where drug efficacy is based on a single quantitative (primary) endpoint. The assumption is common but limiting. Modern trial designs include more the two arms, involve a dynamic patient allocation and include composite endpoints. The approach used here can extrapolate to more sophisticated trial designs, but we limit ourselves to the simple case for tractability.

2.2 *Statistics of a Trial*

The n patients are randomly assigned to each arm and individual responses are statistically independent of each other. We let the i.i.d response variables in the treatment group be denoted by $\{T_1,\dots,T_n\}$, where T_i is drawn from a distribution with mean μ_t, and variance σ^2, and similarly, we let $\{C_1,\dots,C_n\}$ denote the response variables in the control group where C_i is drawn from a distribution with mean μ_c, and variance σ^2, for $i = 1,\dots,n$. The variance is assumed to be known and the same for both arms, an assumption that can be relaxed in a more realistic approach that would not change the nature of the results obtained.

The treatment effect of the drug is defined as the difference in response means of the two groups $\delta := \mu_t - \mu_c$. A candidate drug in such a trial wants to show a positive treatment effect compared to the competitor. The hypothesis test can be formulated as follows:

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

The drug is considered effective if the treatment effect is positive and ineffective otherwise. The Wald statistic from the sample is given by

$$Z_n = \sum_{i-1}^{n} (T_i - P_i) \frac{\sqrt{I_n}}{n}.$$

If the assumptions of the Central Limit Theorem hold, Z_n follows a normal random variable, $Z_n \sim \mathcal{N}(\sigma\sqrt{I_n}, 1)$ where $I_n = \frac{n}{2\sigma^2}$ is the information of the sample from which the associated p-value can be computed.

2.3 *Approval Policy*

Let α be the approval threshold if p-value $< \alpha$, H_0 is rejected and the drug is considered effective, while if p-value $> \alpha$, H_0 cannot be rejected and the drug is considered ineffective. For a given α an approval policy is set to approve a candidate drug if the p-value $< \alpha$ and reject it otherwise. Let p denote the prior probability that a candidate is effective under the alternative hypothesis H_1. For a give α and p We obtain the join probability of the four possible scenarios as

$$\begin{aligned}
\pi_{AE}(\alpha) &= \left(1 - \Phi\left(\Phi^{-1}(1-\alpha) - \delta\sqrt{I_n}\right)\right) p, \\
\pi_{AI}(\alpha) &= \alpha(1-p), \\
\pi_{RE}(\alpha) &= \Phi\left(\Phi^{-1}(1-\alpha) - \delta\sqrt{I_n}\right) p, \\
\pi_{RI}(\alpha) &= (1-\alpha)(1-p),
\end{aligned} \tag{2.1}$$

where Φ and Φ^{-1} are the cumulative and inverse cumulative distribution of the standard normal. The subscript (AE) refers to approving an effective drug, (AI) to approving an ineffective drug, (RE) rejecting an effective drug and (RI) rejecting an ineffective drug. These probabilities determine the ending states in our simulation.

2.3.1 *Flexible Threshold*

Let us assume there is no cost associated with making the correct decisions (AE and RI) if the cost-benefit analysis of each group of patients is not the same, we must have asymmetric costs associated with making a *Type I* or *Type II* errors, which we denote by Q_1 and Q_2. Consider a life-threatening condition, a patient would hugely benefit from an effective treatment. The cost associated with a *Type II* error (mistakenly rejecting an effective drug) is larger than the cost of a *Type I* error (mistakenly approving an ineffective drug), such that $Q_2 > Q_1$. Consider a mild condition instead, the potential manifestation of adverse events is a more prominent factor than the treatment effect, in this case, $Q_2 > Q_1$. In other words, the risk a patient is willing to take by deciding to participate in a trial (or accept an approved treatment) depends on the severity and urgency of the condition. This represents the theoretical foundation when advocating a flexible approval policy.

2.4 *Health Impact in QALYs*

Quantifying health benefits from different clinical domains requires some form of economic measure. Economic impact analyses are regularly conducted during the decision process evaluating health impacts alongside monetary values[1]. A commonly used tool is the Quality-Adjusted Life Years (QALYs), popular among organizations like the UK National

[1] Economic Impact Analyses of FDA Regulations. Food and Drug Administration. (2018f).

Institute for Health and Care Excellence (NICE), the measure of QALY provides a way to quantify health outcomes by weighting quantity and quality of life, see Devlin and Parkin (2004) for a primer.
A year of perfect health is considered equal to 1 QALY, while a year of less than perfect health < 1 QALY. This allows for a comparison of the effectiveness of different treatments across conditions. The measure has some drawbacks, it does capture psychological and emotional well-being, Nord et al. (2009). Here, we denote them by Q.

2.5 *Modeling the Approval Process*

We model the approval process using a queuing network, similarly to Bravo et al. (2022), with some adjustments. They model trials in one single queue, we treat each phase separately in a Markov chain for the clinical testing. The choice allows us to produce an estimate of the time to submission by considering the trial transition probabilities by disease area in (Wong et al. (2019)) and consider the idiosyncratic trial paths and partially treated population.

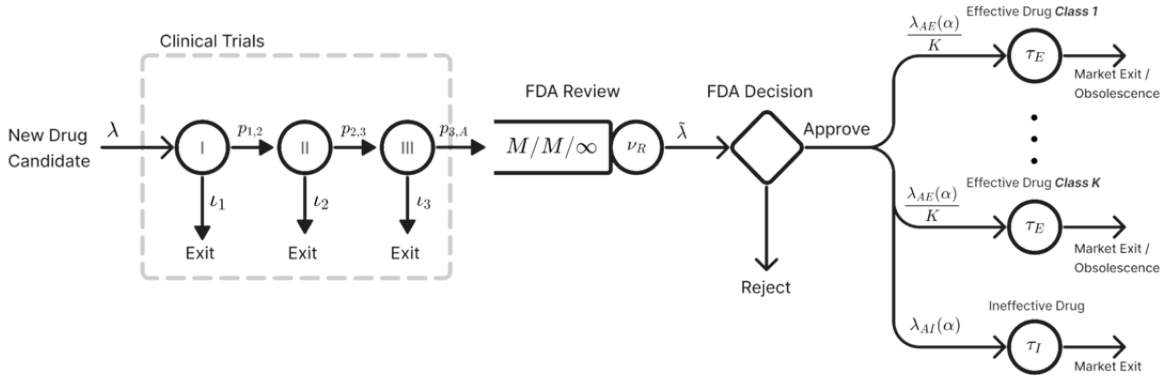
A drug showing positive results during preclinical analysis submits an Investigational New Drug application (IND) to the FDA, which supervises the trial. The sponsor is set to conduct clinical testing in an increasingly large population (and expense), sequentially transitioning from one phase to the next or departing from the system at exponential rates λ. We assume the investigation continues until there is enough power to prove the drug efficacy, usually 90%, or no treatment effect has been shown, and the manufacturer goes back to the drawing board. A common approximation used for the initial arrival rates is using exponential distributions. We don't include repurposing for simplicity, the drug can enter the system with a new name. Let a promising drug enter the system in *Phase I* trial according to a Poisson process with rate λ. If a drug completes the trial stage successfully it transitions to the next stage, otherwise, it leaves the system with abandonment probabilities ι_1, ι_2, and ι_3 at the end of each phase. It advances to the next stages with a transition probability $p_{1,2}$, $p_{2,3}$, $p_{3,A}$, serving an increasing number of patients and exhausting the overall demand in the system.

The Review stage represents a queue in the system after the successful completion of the *Phase III* trial. The FDA acts as a referee before a larger amount of patients is treated. We model the review stage as a priority $M/M/\infty$ queue. The choice allows us to estimate the effect of a flexible approval policy and an experiment with various policy interventions in a single framework. The agency receives applications for drugs that have collected enough evidence in the trials and processes them at exponential service rate ν_R. Drugs exit the review state at rate $\tilde{\lambda}$, rejected drugs exit the system, and approved ones enter the market. In the steady state, the output of the stage constitutes a thinning of the Poisson process

into four independent processes as described in Figure 2. Depending on the probabilities of Equation (2.1), the arrival rates of the ending states are:

$$\lambda_{AE}(\alpha) = \tilde{\lambda}\pi_{AE}(\alpha), \qquad \lambda_{AI}(\alpha) = \tilde{\lambda}\pi_{AI}(\alpha), \qquad \lambda_{RE}(\alpha) = \tilde{\lambda}\pi_{RE}(\alpha), \qquad \lambda_{RI}(\alpha) = \tilde{\lambda}\pi_{RI}(\alpha).$$

Figure 2: Diagram of The Queuing Network Model of the Drug Development and Approval Process



Notes: The filled arrow represent transition probabilities while the non-filled ones are arrival rates. A drug enters the trials with rate λ and passes through a Markov chain that filters them before the FDA Review. The queue of the review process has a service rate υ_R that outputs in four different states, two effective, AE, AI, and two ineffective, RE, and RI. The K classes among effective drugs are meant to capture the degree of market concentration.

The ending state is conditioned on the efficacy of the treatment, while the time to obsolescence, depends on the rate of discovery in that class. Ineffective drugs are administered for a short amount of time as the patients are not satisfied while effective ones stay longer and only depart the system when they become obsolete. Effective drugs are assigned to their corresponding ending state among the K available. We borrow the definitions of therapeutic classes to determine market shares. The factors that determine them are the body interaction, the chemical structure, and the therapeutic recommendations. For simplicity, we allow only one drug per class, when a new drug is available it substitutes the existing one. The change in net benefit to patients would marginally be affected.

Let Q_E and Q_I denote the health benefit of an effective drug measured in QALYs, and N_E and N_I be the number of drugs approved within a period, for effective and ineffective drugs respectively. The optimal approval policy α^* is chosen to maximize the expected net benefit $V(\alpha)$

$$\alpha^* = \underset{\alpha \in [0,1]}{\arg\max}\ V(\alpha),$$

where

$$V(\alpha) = Q_E \mathbb{E}[N_E(\alpha)] - Q_I \mathbb{E}[N_I(\alpha)]. \tag{2.2}$$

The choice of α determines the expected number drugs approved drugs of each type. Let $\frac{1}{\kappa_E}$ and $\frac{1}{\kappa_I}$ be the average time in the market of an effective and ineffective drug respectively. We can adjust the two rates that lead to the market by the average market life and classes obtaining

$$\psi_E(\alpha) = \frac{\lambda_{AE}}{K \cdot \kappa_E} \quad \text{and} \quad \psi_I(\alpha) = \frac{\lambda_{AI}}{\kappa_I},$$

where K adjusts for the degree of market concentration, we can write the elements of the expected net benefit $V(\alpha)$ as

$$\mathbb{E}[N_E(\alpha)] = \frac{K \cdot \psi_E(\alpha)}{1 + \psi_E(\alpha)}, \qquad \mathbb{E}[N_I(\alpha)] = \psi_I(\alpha).$$

Both $\psi_I(\alpha)$ and $\psi_E(\alpha)$ are concave functions in α and $\mathbb{E}[N_E(\alpha)]$ and $\mathbb{E}[N_I(\alpha)]$ are concave functions in $\psi_I(\alpha)$ and $\psi_E(\alpha)$, which are in turn concave in α. We show that $V(\alpha)$, as a sum of concave functions of α, is also concave. We summarize all the parameters used in the model in Table 1.

Table 1: Summary of the Parameters Used in the Queuing Network

Parameter	**Symbol**	**Description**
Trial Size	n	Assumed to be the same for the two arms.
Treatment responses	T_i, C_i	Random treatment response in the treatment and control groups, for $i \in \{1, \ldots, n\}$.
Variance	σ^2	Assumed known and same for both arms, represents the variability of responses.
Treatment effect	δ	Difference in mean responses between treatment and control groups ($\mu_t - \mu_c$).
Approval threshold	α	Significance level for the hypothesis test; if $p < \alpha$, the drug is considered effective.
Prior probability	p	Prior probability that a candidate drug is effective under the alternative hypothesis H_1.
Arrival rate	λ	Rate at which drugs enter the clinical trial phases.
Abandonment probabilities	μ_i	Probabilities of a drug leaves the system after each phase, for $i = 1, 2, 3$.
Transition probabilities	p_i	Probability of advancing to the next phase, for $i = 1, 2, 3$.
Service rate	ν_R	Rate at which the applications are processed in the review stage.
Health Measures	Q_E, Q_I	Health benefit, and cost, of a drug (in QALYs).
Approvals	N_E, N_I	The number of approved drugs.

The optimal policy α^* is unique and satisfies the first-order condition:

$$\alpha^* = 1 - \Phi\left(\frac{1}{\delta\sqrt{I_n}} \cdot \ln\left(\frac{1-p}{p} \cdot \frac{Q_I}{Q_E} \cdot \frac{\tau_E}{\tau_I} \cdot \left(1 + \psi_E\left(\alpha^*\right)\right)^2\right) + \frac{\delta\sqrt{I_n}}{2}\right).$$

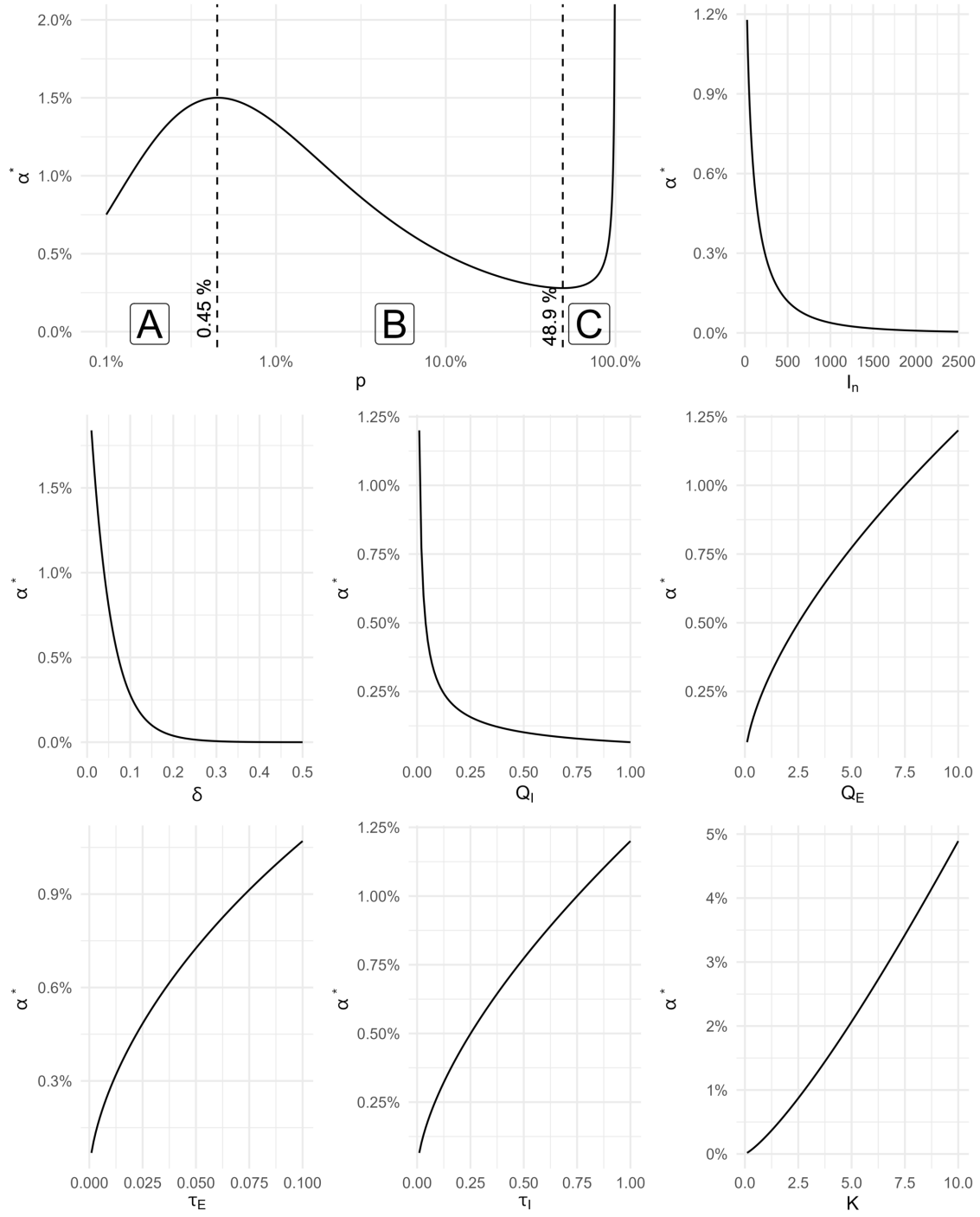
In the steady state, the optimal approval policy balances the health benefits/costs of approving an effective/ineffective drug. Using the first order condition, it is possible to show, however, that α^* is

- increasing in Q_E, τ_I, and K,
- increasing in p and decreasing in τ_E if $\psi_E(\alpha^*) < 1$,
- decreasing in I_n, δ, and Q_I.

We anticipate that approval outcomes will reveal variations in the ideal policy across different disease areas based on the existing data. Specifically, we expect the results to depend on the trial design, efficacy results, and market dynamics. The optimal policy is more stringent for diseases with higher drug production λ and more lenient for drugs with a larger benefit Q_E. As the prior probability p that drugs are effective increases, or as the average time effective drugs remain on the market $\frac{1}{\kappa_E}$ increases, one may expect more drugs to be approved. Extending the time ineffective drugs remain on the market $\frac{1}{\kappa_I}$ increases patient exposure to harm, thereby disincentivizing their approval. However, this approach is contingent upon the condition that $\psi_E(\alpha^*)$, implying that the rate $\frac{\lambda_{AE}(\alpha^*)}{K}$ at which effective drugs within a class are approved is lower than the rate τ_E. This condition is not necessary for market stability but is designed to prevent excessive market crowding.

No closed-form expression for the optimal policy can be derived because the equation contains $\Phi_E(\alpha)$, on the right-hand side. Consequently, we employ the Newton-Raphson method, a well-established numerical optimization technique, to converge upon a solution for α iteratively. Building upon the numerical solutions obtained, we proceed with a theoretical sensitivity analysis which is designed to elucidate how changes in each input parameter independently affect the optimal policy. The results are shown in Figure 3. The optimal policy exhibits significant variability in response to modifications in trial design, average treatment effect, and market duration. This variability is consistent with the signs of the slopes assumed in the previous section. The non-monotonic behavior of the optimal approval policy is explained by dividing the first panel into three regions based on the effectiveness probability p. We define drugs with low effectiveness probability $p < 48.9\%$ as long shots, and those with high effectiveness probability $p > 48.9\%$ as safe bets. Region A deals with neglected markets and long-shot drugs, where increasing p leads to more approvals because of the substantial health benefits and scarcity of treatments. In Region B the market is crowded, so fewer drugs are approved as p increases due to obsolescence and diminishing marginal health benefits. Conversely, in Region C, although the market is crowded, high-probability effective drugs warrant more approvals as p increases, since each new drug is expected to yield net health benefits.

Figure 3: Sensitivity Analysis of the Input Parameters



Note: We follow Bravo et al. (2022)'s initial inputs to be able to compare the theoretical results. The parameters are set to: $\sigma = 1$, $\delta = 0.10$, $n = 500$, $\tilde{\lambda} = 8$, $K = 1$, $Q_E = 1$, $Q_I = 0.1$, $\tau_E = 0.01$, and $\tau_I = 0.10$. Region A corresponds to $0 < p < 0.0045$, Region B to $0.0045 < p < 0.489$, and Region C to $0.489 < p < 1$.

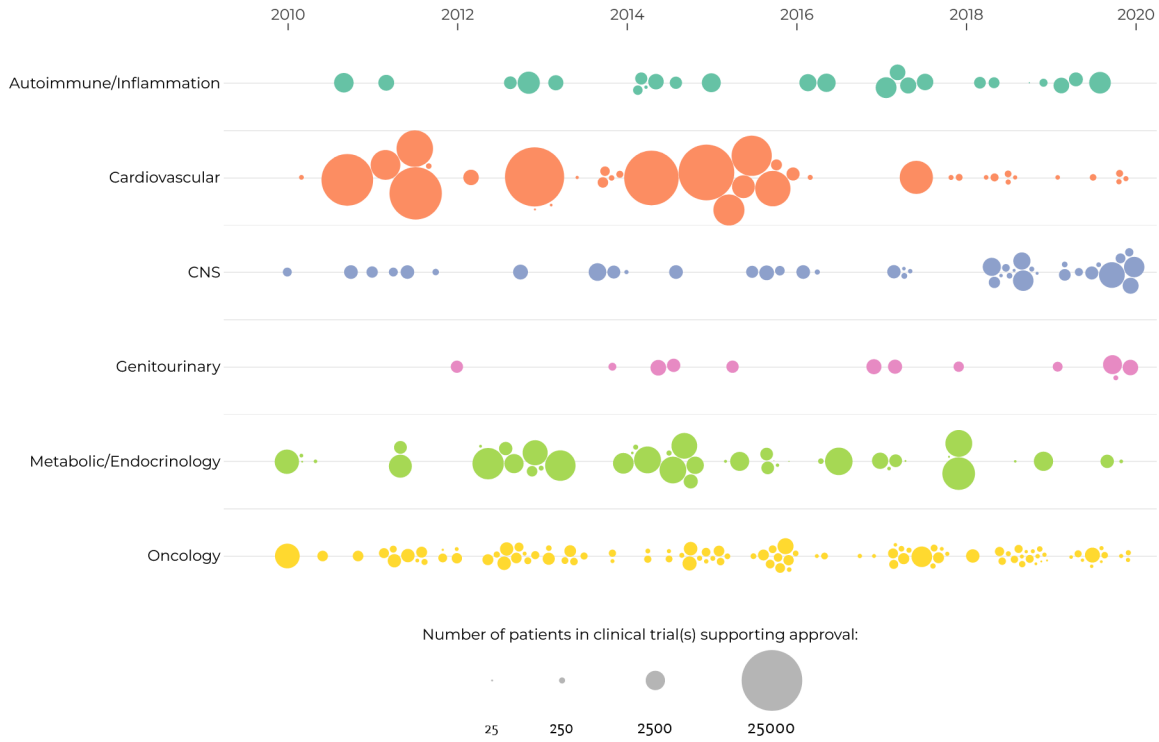
3 DATA

3.1 *Drug Paths*

Submitted, approved, and discontinued Drugs come from the National Drug Code (NDC) Directory, a database from the FDA describing unfinished and unapproved MNEs and compounds. Manufacturers need to disclose the complete list of all drugs manufactured or compounded for sale in the US. The NDC Directory, updated daily, includes data on active and certified drugs, finished or unfinished. For details on clinical trials, we rely on the Drug Trials Snapshots to collect the relevant information from 2010 to 2019, included. The tool provides concise information about the key clinical trials that supported the original FDA approval of New Molecular Entities (NMEs) and original biologics.

We group by disease areas according to the Anatomical Therapeutic Chemical (ATC) classification, obtaining the overview of the drugs trial population in Figure 4. The difference in approval frequency among areas is compared against the FDA submission rates of the NDC registry, providing the per-area prior approval probabilities p. The trial transition probabilities come from Wong et al. (2019) and are collected in Table 2. We compare the approval paths of each disease area, including trial transition probabilities, priors, and population sizes to estimate the chances of success of a candidate drug.

Figure 4: Approved Drugs by Disease Areas



Source: FDA@Snapshot
Notes: The size of the circles represent the total trial population of the major disease areas among newly approved drugs. The observations have been grouped in disease areas according to the Anatomical Therapeutic Chemical (ATC) classification system, a standard that groups active substances according to the organ or system on which they act upon, biochemical properties.

Table 2: Arrivals and Trial Transtion Probabilities

Disease Area	n_1	$p_{1,2}$ (SE)	n_2	$p_{2,3}$ (SE)	n_3	$p_{3,\text{A}}$ (SE)
Autoimmune/Inflammation	2900	78.9 (0.8)	1862	48.7 (1.2)	659	68.6 (1.8)
Cardiovascular	1599	71.1 (1.1)	1002	64.9 (1.5)	473	72.3 (2.1)
CNS	2777	75.0 (0.8)	1695	54.5 (1.2)	648	63.0 (1.9)
Genitourinary	568	73.4 (1.9)	382	59.2 (2.5)	176	69.3 (3.5)
Metabolic/Endocrinology	2012	75.2 (1.0)	1273	57.0 (1.4)	535	62.8 (2.1)
Oncology	3107	78.7 (0.7)	1601	53.9 (1.2)	431	48.5 (2.4)

Source: Wong et al. (2019).
Notes: The values shown apply to the lead indications, the treatment and population the drug was originally targeted for; $p_{i,i+1}$: Probability of Success in Trial i (%); SE: Standard Error

TRIAL DURATION A common assumption is to assume exponential trial duration. Here we sample from the empirical distributions, with the average trial duration summarized in Table 3.

Table 3: Trial Duration by Disease Area

Disease Area	T1	T2	T3	T4
Autoimmune/Inflammation	335	980	979	1207
Cardiovascular	379	1025	1208	1174
CNS	334	932	1034	1068
Genitourinary	378	787	1005	913
Metabolic/Endocrinology	325	946	976	1036
Oncology	1216	1490	2080	1394

Source: Wong et al. (2019)

TREATMENT EFFECT The trial information $\delta\sqrt{I_n}$ used in our simulated paths is constructed relying on two assumptions. We bootstrap the population sizes I from from the approved drugs, here we assume the trial designs to be the same between candidates and approved. We then consider *Phase III* trial completed when a statistical power, $1-\beta = 90\%$, is achieved, and the drug is ready for submission. For a sampled trial population I, we let the treatment effect δ satisfy:

$$1-\beta = 1-\Phi\left(\Phi^{-1}(1-\alpha)-\delta\sqrt{I_n}\right).$$

PRIORS We obtain the probability that a drug is effective upon submission to the FDA by the proportion of approved drugs among submitted ones from the NDA's registry.

HEALTH IMPACT Let Q_E represent the annual health benefits, and Q_I denote the annual health costs, each scaled by the market size. These are associated with a unit increase in the expected number of drugs entering the market for AE and AI drug responses, respectively. We construct Q_E from the mortality gains following the sequence of drug approvals in the disease area and We assume that the total health costs of approving an ineffective drug Q_I/κ_I are proportional to the total health benefits of approving an effective drug Q_E/κ_E, with a constant ratio:

$$\frac{Q_I}{Q_E}\cdot\frac{\kappa_I}{\kappa_E} = c.$$

For the ratio we rely on the baseline estimate of $c = 2$ by Bravo et al. (2022). They use the response of the stock prices of listed companies upon approval or withdrawal to come up with a ratio, postulating the equivalence between health benefits and sale prices as the sole consideration in the reimbursement negotiations.

QALYS The health gains by disease area of Table 4 were obtained from the work of Shafrin et al. (2023). This study evaluated the health benefits of US-approved drugs from 2011 to 2021, focusing on the marginal QALY they provide.

Table 4: Disease Areas with Total Treatments, CEA Studies Found, and Average QALYs

Disease Area	**Studies**	**Treatments**	Q
Autoimmune/Inflammation	4	7	0.47
Cardiovascular	8	12	1.10
CNS	30	59	1.24
Genitourinary	2	5	0.03
Metabolic/Endocrinology	25	54	1.30
Oncology	81	139	1.20

Source: Shafrin et al. (2023).
Notes: Q: mean QALYs gained. Out of 483 new therapies, 252 had a relevant cost-effectiveness analysis, with an average gain of 1.04 QALYs across all disease areas. CNS and Oncology treatments showed the highest benefits.

MARKET REGIMES The role of patented and generic regimes is a fundamental aspect of drug trademark policy, which governs how a drug is marketed and sold after it has been approved for use. When a new drug enters the market, it typically does so under a patent regime. This regime grants the manufacturer exclusivity over the drugs, effectively creating a temporary monopoly. The duration of this patent protection is represented by the time $(1/\kappa_P)$, during which the company can recoup research and development costs and earn profits without competition from generic versions. Once the patent expires, the drug enters the generic regime. During this phase, other manufacturers are allowed to produce and sell generic versions of the original drug, which are bioequivalent but often sold at lower prices. The period that a drug spends as a generic is denoted by $(1/\kappa_G)$. The introduction of generics increases market competition, typically leading to more affordable prices and wider accessibility for patients.

TIME IN MARKET AI drugs spend a period $1/\kappa_I$ on the market calculated as the average time until withdrawal from the drugs listed in Table 7 of the Appendix. For the market concentration and market time of AE drugs we rely on the figures for the Canadian market provided by Lexchin (2017), shown in Table 5. The study observed that 58 drugs without any competition had been on the market for a median duration of 5357 days, or approximately 14.7 years, with an interquartile range from 3291 to 6679 days, indicating a non-normal distribution of data. The study also highlighted variations in the presence of competition across different disease areas. The cardiovascular area had the highest competition at 94% for its 18 drugs, while the Autoimmune/Inflammation area had only 24% competition among its 21 drugs. There was no significant correlation between the percentage of biologics in these groups and the presence of competition. However, there was a statistically significant difference in the average market exclusivity times across different disease areas. Market protection is assumed to be 20 years.

Table 5: Concentration levels of top Selling Drugs by Therapeutic Group

Disease Area	k	$K(K > 1)$	**Excl. (days)**
Metabolic/Endocrinology	15	7 (47%)	4390
Autoimmune/Inflammation	21	5 (24%)	4544
Cardiovascular	18	17 (94%)	4843
Genitourinary	12	10 (83%)	4717
CNS	22	14 (64%)	3517
Oncology*	21	5 (24%)	4544

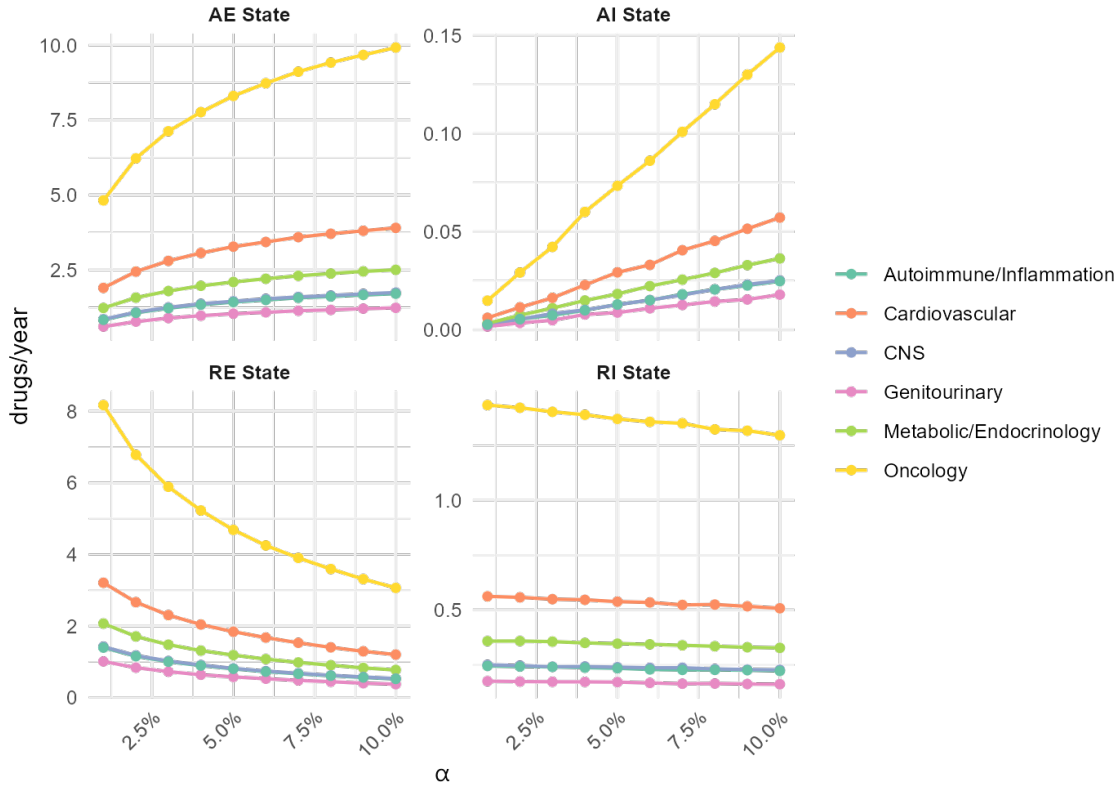
Source: Lexchin (2017).
Notes: k is the total number of products in a group while K is the number of drugs with competition. The observed differences across therapeutic areas may be attributed to factors such as the complexity of treatment development, patient population size, and the required investment for research and development. *Oncology and Autoimmune/Inflammation have the same values as they were combined in one group by the authors.

4 RESULTS

4.1 *Simulation Environment*

Our simulation runs on the stochastic discrete-event simulation (DES) framework of the **`simmer`** R package, see Ucar et al. (2017) for technical details. We define the prior approval probabilities for each of the disease areas analyzed based on historical data and we let the individual paths be simulated according to the processes described in Figure 1 where the time in each stage is drawn from an exponential distribution with the rate given by the reciprocal of the historical average time spent in that stage. Branching logic is incorporated to model the probabilistic nature of trial outcomes and the review process with transition probabilities defined in Table 2. Our simulation accounts for the effectiveness of the drugs by incorporating statistical calculations into the trajectory, determining the likelihood of regulatory approval or rejection based on the efficacy shown in the clinical trials. We let the simulation run long enough for the values to converge to their asymptotic values and we collect the average long-term number of drugs that end up in each of the ending states in Figure 5.

Figure 5: Ending States of the Simulation Environment

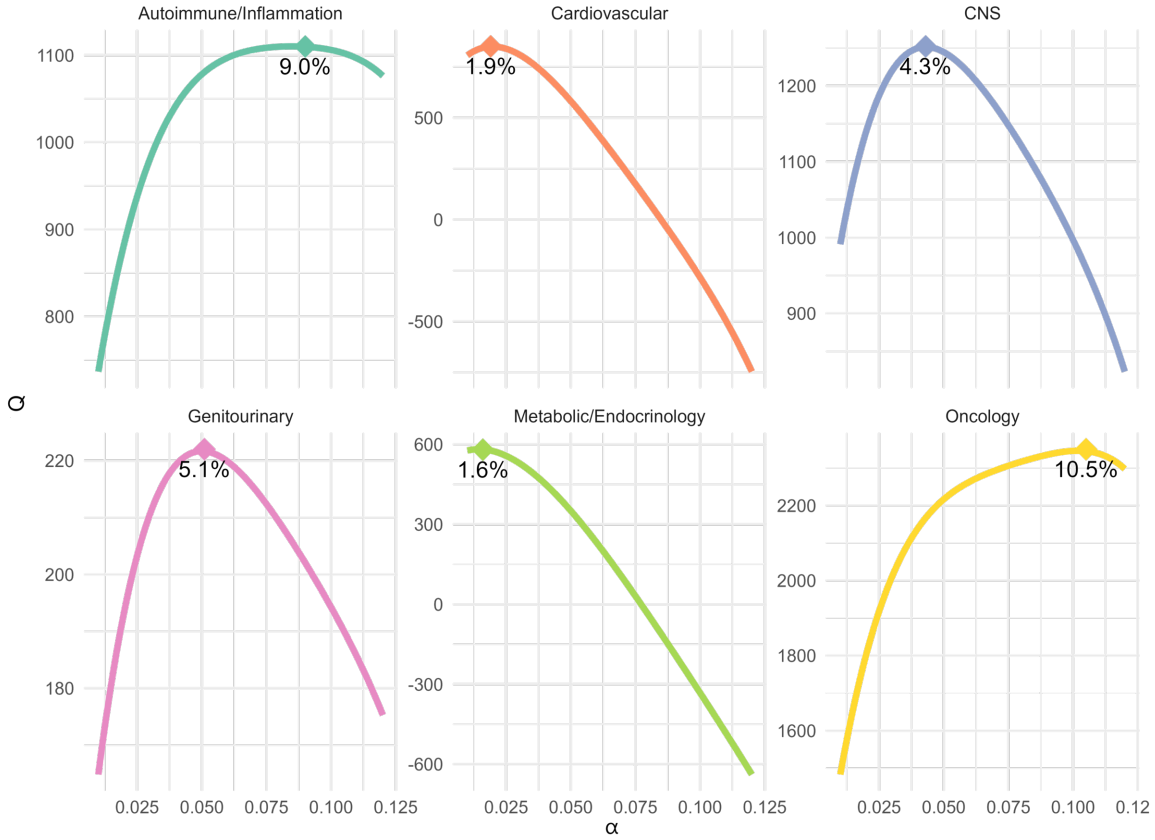


Notes: The number of drugs that and up in any of the four states is proportional to the approval threshold α, inversely proportional in the case of RE and RI, and directly proportional in the case of AE and AI.

4.2 *Optimal Approval Policy*

By adjusting this significance threshold used in the review process and multiplying the resulting approved drug by the marginal increase in QALY we can obtain the overall change in QALY associated with its threshold. The resulting curves and their maxima are shown in Figure 6.

Figure 6: Optimal Significance Level by Disease Area



Notes: The optimal approval policy is chosen to be the apex of the curves shown here, the value of α that confers the highest gain in QALYs.

The results show that the optimal approval threshold varies among disease areas. In particular, the α^* is well above the current threshold of 2.5% for Oncology (10.5%) and Autoimmune drugs (9.0%), due to the sporadic production rates and low market concentration in the disease area. The same applies to CNS (4.3%) and Genitourinary (5.1%) but the effect is less pronounced. For Cardiovascular (1.9%) and Metabolic/Endocrinology (1.6%), the optimal policy is lower than the current one due to the high production rates and more pronounced market concentration. We collect the optimal policies for the selection of disease areas in Table 6.

4.3 *Sensitivity Analysis*

The influence that some of the assumptions have on the final results can be investigated by a sensitivity analysis of the simulation results. In particular, we determine the impact on the most important parameters used in our simulation, the health impact ratio c and the market concentration K. We can observe how the optimal approval policy α^* changes when varying c from 1 to 5 and K from 50% to 150% of the baseline in Figure 7. The analysis confirms our initial expectations regarding the relationship between the health impact ratio

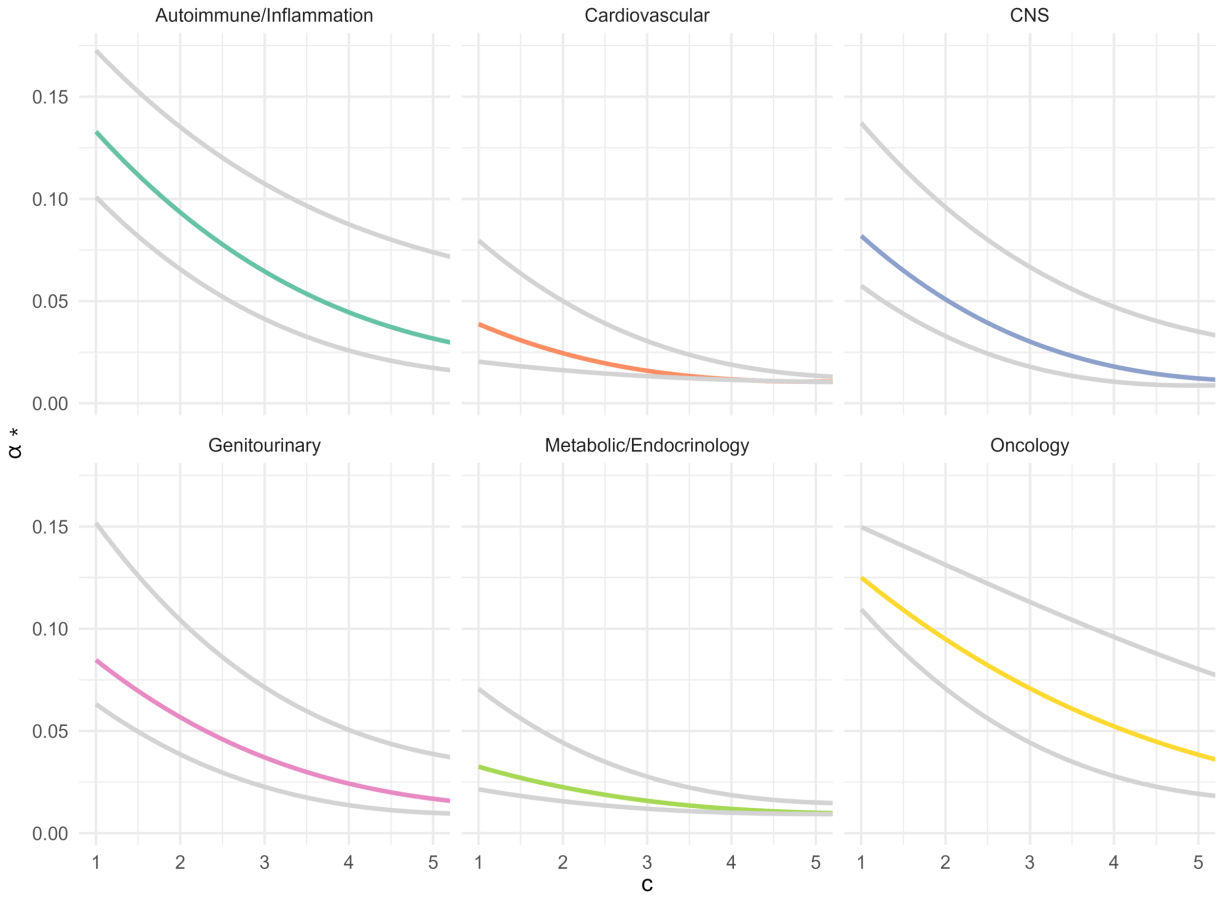
Table 6: Optimal Policies for Disease Areas

Disease Area	Q	ΔQ	α^*
Autoimmune/Inflammation	1109	172	9.0%
Cardiovascular	850	12	1.9%
CNS	1252	64	4.3%
Genitourinary	222	20	5.1%
Metabolic/Endocrinology	587	13	1.6%
Oncology	2351	438	10.5%

Notes: Q refers to the QALYs per year generated by the newly approved drugs, ΔQ refers to the change in QALYSs induced by the flexible approval policy.

c and the market concentration K, and their influence on the optimal approval policy α^*. A larger health impact ratio c tends to result in a stricter approval policy. This is because, with higher values of c, AI drugs are considered more harmful, thus necessitating more stringent regulations to prevent adverse effects on public health. On the other hand, the market concentration parameter K affects the marginal health benefits of an approved drug due to its crowding-out effect. In markets with high concentrations, the introduction of new drugs has a lesser incremental benefit because of the saturation effect, which could lead to a more lenient approval policy to encourage competition and innovation. The optimal approval policy α^* undergoes significant changes as we vary the parameters c and K. The differences underscores the complexity of the decision-making process for drug approval policies. The sensitivity of α^* to these parameters makes it challenging to draw conclusions without precise estimates for each disease area.

Figure 7: Sensitivity Analysis for K and c



Notes: We let the value of c to vary between 1 (health benefit/harm from AE/AI drugs is the same) and 5 (harm much higher than benefit) in the x axis while we gray line represent a range of 50% (lower) to 150% (higher) of the baseline values of K.

5 DISCUSSION

In this work, we evaluated a methodology designed to establish a transparent and adaptable framework for the drug approval process. Agencies such as the FDA and EMA conduct cost-effectiveness analyses that likely include similar considerations, yet the specifics of their metrics remain unclear. We found that the optimal significance level for the efficacy of a drug varies according to the disease area and that allowing for a flexible significance level produces an improvement in health outcomes compared to a fixed one. Our approach closely relates to the work of Montazerhodjat et al. (2017) which uses a Bayesian Decision Analysis to estimate the approval policy that minimizes expected harm to patients. We build upon the work of Bravo et al. (2022) by allowing for different paths in the clinical trials while disregarding the financial market component they considered. For simplicity, we do not consider other constraints that may influence the approval decision process, like manufacturing constraints or trial designs. We, however, allow for a more realistic review

stage where the time for the review is taken from historical data. By including transition times and probabilities in the trial phase, we estimate the number of patients treated and the time needed to complete each phase. Moreover, we performed the exercise on a larger set of disease areas compared to previous work, all the ones for which we could gather enough information.

5.1 *Limitations*

The creation of drugs involves a sophisticated synergy between scientific research, policy formulation, and the ongoing pursuit of medical innovations. Additionally, the confidentiality of industrial data and the intricacies of published clinical trials limit the generalizability of our findings. We found the optimal policy to be heavily dependent on the input parameters, particularly the level of market concentration and the health benefit/harm ratio. One of the major constraints we faced was the collection of comparable information coming from different studies. The disease areas considered do not perfectly overlap. A more robust approach would necessitate an extensive data collection process with interviews and nonpublic data. The use of a fixed marginal increase in QALY may also not reflect the change in biotechnologies. The use of historical data also comes with problems as it may not reflect trends and advances in the fields.

5.2 *Future Work*

We suggest future work to combine elements of the Bayesian Decision Framework into the pipeline we constructed. We took the perspective of a policy-maker, game strategies induced by different regulations can also be included. Even so, A method that fully captures the dynamics among all the players involved is yet to be found. Any attempt to construct a solid policy recommendation would require a high degree of data quality. We advocate for the need for a centralized repository where the data used in the review is collected. As a corollary research, one could also investigate the exact parameters needed to capture the difference in approval policy in the EU, supporting the convergence of HTA agencies. However, we urge caution in the universal application of our approach. The new policy could serve as a framework for decision-making where the multifaceted nature of a drug approval process can be evaluated quantitatively. We posit that a known and flexible framework could expedite decision-making for regulators and enhance pharmaceutical companies' productivity by allowing them to modulate clinical trials in response to preliminary results and defined objectives. As the EMA strives to harmonize European Health Technology Assessment (HTA) practices, a flexible framework like ours could cater to individual country needs and foster broader acceptance of EU healthcare policies.

5.3 *The Case for Prudence*

We conclude the discussion with an excursus about one of the most emblematic cases that formed the current drag approval status quo. In the early 50s, Chemie Grunenthal, a soap manufacturer, while researching to address the increasing need for antibiotics, discovered a substance that seemed to possess extraordinary sedative and antiemetic properties. The substance went by the name of Thalidomide. Quickly introduced in Germany, the UK, Australia, and New Zealand by 1957, the drug was marketed as a sedative to treat a wide range of symptoms and was frequently prescribed in pregnancy to alleviate morning sickness. The drug had never been tested in pregnant women because at the time it was believed that medications could not cross the placental barrier. Nobody anticipated the tragedy that followed. Within a few years, physicians in Germany (Lenz (1962)) and Australia (McBride (1961)) reported a shocking increase in phocomelia, a rare malformation, in thousands of kids born with missing or deformed limbs. They were able to attribute the cases to the use of Thalidomide in the first and second trimesters of pregnancy. Unprecedented public outrage forced the withdrawal from the market of a drug that became the case study for imprudence in clinical research. According to Vargesson (2015), more than 10000 children were affected worldwide in what many consider the largest man-made medical disasters in history.

At that time, the US was the only country requiring manufacturers to seek government approval before launching to market. Despite the preliminary approval, Doctor Frances Kelsey was not convinced by the evidence proposed to attest to its safety and pressured the Food and Drug Administration (FDA) to investigate. Thanks to her skepticism, and independence, only a few were affected. During the same period, US Senator Estes Kefauver was investigating the pharmaceutical industry and the escalating expenses of prescription drugs. After 17 months of hearings, he proposed a bill in favor of a stricter regulation which was received by fierce opposition and fading support. The news of Thalidomide persuaded the Kennedy administration to support the bill. What became the Kefauver-Harris Amendment for the first time granted the FDA the power to demand proof of efficacy, rather than just safety, when approving a new drug. A piece of legislation which guidelines are, to a certain extent, still in vigor today. Within a few years, many countries around the world adopted the American system and the requirements converged since. After a global ban, public outrage, expensive legal compensations, and the promise to never develop the drug again, something unexpected happened. Thalidomide was found to be particularly effective against myeloma, a type of blood cancer, by Singhal et al. (1999). The drug increasingly became the primary treatment for multiple myeloma Stewart et al. (2009). A plot twist of sinister irony in what became the most surprising case of drug repositioning. This story shaped the public perception of drug-related risk and its tragic consequences marked the approach of Regulatory Agencies towards drug approval.

The story also elucidated the skepticism and prudence that should surround any change related to drug approval. If the lack of data disclosure is to be understood as the original

sin, however, we should attempt to make the drug decision process as understandable, and reproducible, as possible. More flexibility in the decision-making process, guided by a quantitative approach, could reinforce public trust in both the pharmaceutical industry and the regulator.

6 CONCLUSION

In this study, we navigated the complex stages of a drug's life cycle. We scrutinized the roles of regulatory bodies and investigated the benchmarks they use to evaluate new pharmaceuticals. We constructed a framework to assess the effect of an approval policy on patient health outcomes. Assuming the role of regulators responsible for approving novel drugs, we illustrate how healthcare impact could be maximized by adopting a flexible threshold for approval. We show that the optimal threshold adjusts to the specific disease area depending on production rates, health impacts, and market concentration levels, becoming more lenient or stringent as necessary. The generalizability of our approach allows for the construction of estimates at different levels of granularity depending on data availability. Future research could evaluate the ability of such a model to reconcile the differences in the current EU approval landscape. With satisfactory data quality and validation, this framework can be adopted by policymakers as a reliable tool for a quantitative approval process, guiding investment decisions, and improving productivity in the pharmaceutical industry.

REFERENCES

Bravo, Fernanda, Taylor C Corcoran, and Elisa F Long (2022). Flexible drug approval policies. *Manufacturing & Service Operations Management 24*(1), 542–560.

Brown, Dean G, Heike J Wobst, Abhijeet Kapoor, Leslie A Kenna, and Noel Southall (2021). Clinical development times for innovative drugs. *Nat. Rev. Drug Discov 21*(11), 793–794.

Caves, Richard E, Michael D Whinston, Mark A Hurwitz, Ariel Pakes, and Peter Temin (1991). Patent expiration, entry, and competition in the us pharmaceutical industry. *Brookings papers on economic activity. Microeconomics 1991*, 1–66.

Delshad, Saeid, Hamed Rahimian, and Amin Khademi (2023). Experimentation levels and social welfare under fda's flexible approval standards. *Available at SSRN 4453707*.

Devlin, Nancy and David Parkin (2004). Does nice have a cost-effectiveness threshold and what other factors influence its decisions? a binary choice analysis. *Health economics 13*(5), 437–452.

Isakov, Leah, Andrew W Lo, and Vahid Montazerhodjat (2019). Is the fda too conservative or too aggressive?: A bayesian decision analysis of clinical trial design. *Journal of econometrics 211*(1), 117–136.

Kashoki, Mwango, Zahra Hanaizi, Stella Yordanova, Richard Veselỳ, Christelle Bouygues, Jordi Llinares, and Sandra L Kweder (2020). A comparison of ema and fda decisions for new drug marketing applications 2014–2016: concordance, discordance, and why. *Clinical Pharmacology & Therapeutics 107*(1), 195–202.

Lenz, Wolfgang (1962). Thalidomide and congenital abnormalities. In *Problems of Birth Defects: From Hippocrates to Thalidomide and After*, pp. 199–199. Springer.

Lexchin, Joel (2017). Market exclusivity time for top selling originator drugs in canada: a cohort study. *Value in Health 20*(8), 1139–1142.

McBride, William Griffith (1961). Thalidomide and congenital abnormalities. *Lancet 2*(1358), 90927–8.

Montazerhodjat, Vahid, Shomesh E Chaudhuri, Daniel J Sargent, and Andrew W Lo (2017). Use of bayesian decision analysis to minimize harm in patient-centered randomized clinical trials in oncology. *JAMA oncology 3*(9), e170123–e170123.

Morgan, Steve, Paul Grootendorst, Joel Lexchin, Colleen Cunningham, and Devon Greyson (2011). The cost of drug development: a systematic review. *Health policy 100*(1), 4–17.

Nord, Erik, Norman Daniels, and Mark Kamlet (2009). Qalys: some challenges. *Value in health 12*, S10–S15.

Patriarca, Peter A, R Michael Van Auken, and Scott A Kebschull (2018). Analysis of the risks and benefits of new chemical entities approved by the us food and drug administration (fda) and subsequently withdrawn from the us market. *Therapeutic Innovation & Regulatory Science 52*(5), 649–655.

Shafrin, Jason, Sabiha Quddus, Moises Marin, and Dennis Scanlon (2023). A decade of health innovation: The impact of new medicines on patient health and the implications for nice's size of benefit multiplier. *Value in Health 26*(10), 1435–1439.

Singhal, Seema, Jayesh Mehta, Raman Desikan, Dan Ayers, Paula Roberson, Paul Eddlemon, Nikhil Munshi, Elias Anaissie, Carla Wilson, Madhav Dhodapkar, et al. (1999). Antitumor activity of thalidomide in refractory multiple myeloma. *New England Journal of Medicine 341*(21), 1565–1571.

Stewart, A Keith, Paul G Richardson, and Jesus F San-Miguel (2009). How i treat multiple myeloma in younger patients. *Blood, The Journal of the American Society of Hematology 114*(27), 5436–5443.

Ucar, Iñaki, Bart Smeets, and Arturo Azcorra (2017). Simmer: discrete-event simulation for r. *arXiv preprint arXiv:1705.09746*.

Vargesson, Neil (2015). Thalidomide-induced teratogenesis: History and mechanisms. *Birth Defects Research Part C: Embryo Today: Reviews 105*(2), 140–156.

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

7 APPENDIX

Table 7: The collection of withdrawn drugs by disease area used in our estimation

Area	Drug	Description
Cardiovascular	Eliquis (apixaban) 5 mg Tablets	Anticoagulant to prevent blood clots
	Irbesartan and Hydrochlorothiazide Tablets USP	Used for high blood pressure and may also be used in heart failure management.
	Dabigatran Etexilate Capsules, USP	Anticoagulant to prevent stroke in patients with atrial fibrillation.
Nervous System	Lorazepam Oral Concentrate, USP 2mg/mL	Anxiolytic for anxiety disorders.
	Levetiracetam Injection	Antiepileptic drug for seizures.
Endocrinology	Glucagon Emergency Kit	Used to treat severe hypoglycemia.
	Insulin glargine injection, 100 units/ml	Long-acting insulin used for diabetes management.
	Metformin Hydrochloride Extended-Release Tablets	Medication for type 2 diabetes.
Infectious Disease	Cefazolin	Antibiotic used for various bacterial infections.
	Veklury® (remdesivir 100 mg for injection)	Antiviral drug that has been used to treat COVID-19.
Metabolic Disorders	Sodium Bicarbonate in 5% Dextrose Injection	Can be used to treat metabolic acidosis, which may occur in severe renal disease, uncontrolled diabetes, circulatory insufficiency due to shock or severe dehydration, cardiac arrest and severe primary lactic acidosis.
	0.9% Sodium Chloride Injection, USP	Used for extracellular fluid replacement, e.g., dehydration.
	70% Dextrose Injection (2000 mL) USP	High concentration dextrose solution used for caloric supply and fluid replenishment.
Autoimmune	SYMJEPI (epinephrine) Injection	Used for emergency treatment of allergic reactions (Type I) including anaphylaxis.
Ophthalmology	Artificial Tears Lubricant Eye Drops	Used to relieve dry eyes.
	Brimonidine Tartrate Ophthalmic Solution, 0.15%	Used to lower increased intraocular pressure in glaucoma or ocular hypertension.
Genitourinary	Drospirenone and Ethinyl Estradiol Tablets, USP	Combination hormonal contraceptive pills.
Dermatology	Sunscreen	Protects against sunburn and other skin damage due to sun exposure.
	Clobetasol Propionate	Topical corticosteroid used for various skin disorders.