Beyond the Label: Regulatory Arbitrage in Pharmaceutical Market Entry*

Charu N. Gupta
Harvard University

Jennifer Kao
UCLA Anderson School of Management

August 16, 2025

Abstract

In regulated markets, high entry barriers create incentives for firms to circumvent regulation or selectively avoid approval for certain products. We provide novel evidence of such regulatory arbitrage in U.S. pharmaceuticals: Firms conduct clinical trials in large patient populations but initially seek approval in smaller, lower-risk markets to reduce regulatory burden. They then rely on off-label drug use to expand demand beyond approved uses. While this approach expedites market entry, it raises policy concerns about treatment adoption without strong supporting evidence. Our findings highlight how regulatory design may induce strategic entry by firms and alter the diffusion of new technologies.

JEL Codes: I11; I18; L51; O31; O33
Keywords: Regulation; Innovation; Pharmaceuticals; Off-Label Use; Genetics

*charu_gupta@hms.harvard.edu, jennifer.kao@anderson.ucla.edu. We thank Victor Bennett, David Bradford, Amitabh Chandra, Rena Conti, Wes Greenblatt, Matt Grennan, Matt Higgins, Katya Khmelnitskaya, Josh Krieger, Daniel Levinthal, Luca Maini, Alex Olssen, Chris Poliquin, Olav Sorenson, Noam Yuchtman, and seminar and conference participants at the Academy of Management (AOM) Annual Meeting, American Society of Health Economists (ASHEcon) Annual Conference, CEPR-Toulouse Health Economics Conference, Food & Drug Administration, SoCal Strategy and OT Workshop, Society for Institutional & Organizational Economics (SIOE) Conference, Strategy Science Conference, Technology & Policy Research Initiative (TPRI) IP Day, UCLA Strategy Seminar, Utah Winter Strategy Conference, and Wharton Health Care Management Workshop for helpful comments. We acknowledge the Sorenson Center for Discovery and Innovation Studies at the David Eccles School of Business and the Levy-Rosenblum Institute at the Tulane University Freeman School of Business for data access. Financial support for this research came from the Price Center for Entrepreneurship & Innovation at UCLA Anderson and the National Institute for Health Care Management Foundation. Sam Lin and Jordan Youhanaie provided outstanding research assistance. Any errors are our own.

1 Introduction

In many industries, government regulation determines which products may enter the market and under what conditions (Joskow, 2005; Maini and Pammolli, 2023; Vives, 2019). While such entry regulation is intended to ensure product safety and efficacy, it can also create incentives for firms to respond strategically. For instance, ride-sharing apps such as Uber and Lyft have been able to bypass city taxicab regulations by arguing that they are technology platforms and not transportation providers (Posen, 2015). Similarly, financial technology ("fintech") companies like PayPal refrain from certain activities, including holding customer funds or making loans, to avoid being regulated as banks (Douglas, 2016).

The decision to circumvent regulatory approval is particularly salient in health care, where entry regulation by the U.S. Food and Drug Administration (FDA) creates a unique set of incentives and disincentives for manufacturers to seek regulatory approval for new uses of approved products (Friedman, 1996). Firms must obtain separate FDA approval to legally market a drug for each distinct disease market. Although this structure incentivizes firms to pursue supplemental approvals for a drug's additional therapeutic uses, the high costs and risks of conducting rigorous clinical trials necessary for approval may deter them. Further, an alternative, non-regulatory pathway exists where approved drugs can be used by consumers for unapproved ("off-label") uses (Eisenberg, 2005).

In this paper, we provide new empirical evidence that pharmaceutical firms engage in regulatory arbitrage—the practice of exploiting gaps or differences in regulation across jurisdictions or within a single regulatory framework—in their market entry strategies. Rather than seek approval in the largest potential disease market (or indication), firms may initially pursue approval in smaller, more favorable indications with lower regulatory risks. Once approved, firms can rely on off-label use to access broader patient populations without incurring the full costs of formal approval across all indications.

In choosing to engage in regulatory arbitrage, firms face a central trade-off between regulatory risk and market reach: Initial approvals in narrow indications can lower the upfront risk of market entry, but expanding demand later through supplemental approvals or off-label use introduces other costs and uncertainties. Further, off-label uses cannot be directly promoted, which may limit uptake and expose firms to reputational or legal risks. As a result, the prevalence of off-label use—and thus regulatory arbitrage—is likely to vary across contexts.

This behavior also has important implications for the regulation of new medical technologies. Regulatory arbitrage by pharmaceutical firms may speed market entry, but it also raises concerns about the diffusion of treatments without sufficient clinical evidence in broader patient populations (Kao, 2017). Although off-label use can expand patient access to promising treatments, it also risks weakening the evidence base underpinning therapeutic use. These findings contribute to ongoing policy debates over the role of off-label use, the design of approval pathways, and the potential trade-offs between flexibility and rigor in drug regulation. Although policymakers have highlighted the existence of such practices among pharmaceutical firms (Congressional Research Service, 2021), systematic evidence remains limited.

While prior studies suggest that firms have incentives to seek approval in a drug's largest possible disease market for a new drug (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013), these analyses typically treat a drug as having a single approval. In practice, many drugs have multiple potential uses, and the possibility of off-label prescribing can shift firm incentives. In settings with high off-label use, firms may expand demand without pursuing formal supplemental approvals, while low off-label use may limit the feasibility of such strategies. Using a comprehensive dataset linking clinical trials and regulatory approvals across cancer types, we show that firms often conduct trials in large disease markets but seek initial FDA approval in smaller ones (i.e., diseases where the expected probability of approval is higher) (Chandra et al., 2019; Michaeli et al., 2023). To systematically measure a drug's off-label potential, we construct a novel index based on information from large-scale genomic sequencing studies. Our findings suggest that regulatory arbitrage is particularly prevalent among drugs with high off-label potential and that this strategy enables earlier market entry while avoiding or delaying more risky approval processes in broader disease markets.

The idea that the potential for regulatory arbitrage may shift firms' decisions, while conceptually clear, is difficult to examine empirically. There are two main issues. First, variation in regulation requirements is rare, with regulatory changes typically anticipated and similar across product markets. As a result, traditional empirical methods—which may rely on exogenous changes over time or across product markets—are ill-suited to assess how regulatory design may shift firms' approval decisions. Second, to understand how regulation shapes these decisions, it is necessary to identify the benchmark set of indications in which a firm has conducted clinical trials of its products and might plausibly seek regulatory approval. In most industries, the set of plausible markets a firm

might have entered (but chose not to) is unobserved, limiting researchers' ability to characterize how firms alter their entry strategies in response to regulation.

Our empirical focus on drug development allows us to make progress on both of these issues. To address the first challenge, identifying meaningful variation in regulation, we build on prior work that leverages unique features of the health care sector to identify useful variation (see, e.g., Budish et al., 2015; Chandra et al., 2016; Cohen et al., 2021; Stern, 2017). In particular, we exploit variation across disease indications in both the likelihood of regulatory approval and the potential for off-label use. Some indications, particularly those with smaller patient populations, offer firms a less risky route to approval. For instance, trials for smaller indications require fewer participants (Sertkaya et al., 2016), and advances in genetic sequencing and biomarker testing allow firms to selectively enroll patients with higher treatment responsiveness, thereby increasing the probability of approval (Chandra et al., 2019; Michaeli et al., 2023). These differences, combined with the potential to expand demand through off-label prescribing, create scope for regulatory arbitrage. To address the second challenge, we construct a comprehensive dataset linking a drug's regulatory approvals to its full set of clinical trials. These data allow us to distinguish between disease indications in which a firm invests through clinical testing (the benchmark) and those in which it ultimately seeks FDA approval.

To illustrate, Figure 1 explores market entry decisions across a drug's lifecycle, plotting mean market size by indication order—defined as the sequence of disease indications in which a given drug is clinically tested or approved—for cancer drugs first approved from 1978 to 2016. Panel A indicates that firms prioritize large indications for their clinical trial testing, consistent with incentives to gather information about larger market sizes. However, Panel B indicates that firms prioritize smaller indications for regulatory approval.[1] While such patterns could be due to differences in scientific opportunities (Kao, 2025), market conditions (Acemoglu and Linn, 2004), or intellectual property (IP) protection (Budish et al., 2015), the indication order-market size relationships persist even after controlling for these factors. The patterns are consistent with regulatory arbitrage: Firms pursue early approval in smaller, lower-risk indications, while delaying formal approval in broader markets or relying on off-label diffusion to expand demand.

[1]These patterns are not simply driven by variation in indication labeling (e.g., "cancer") or trial phase: Our trial data are specific to the cancer site (e.g., "ovarian cancer"), and analyses restricted to phase III trials yield similar results (Appendix Figure A1).

Motivated by this finding, we examine how approval sequencing varies with the size of the total market a drug can access, including both approved *and* off-label indications. Although off-label use is not directly observable ex ante, we exploit two features of the cancer market—the largest pharmaceutical market in terms of spending (IQVIA Institute, 2018)—for these analyses. First, multiple uses are common, and estimates of off-label use range from 50 to 75% (Pfister, 2012).[2] Second, the rise of cancer genome sequencing enables us to use genetic data to identify biologically-likely off-label disease indications. Cancer sequencing is an advance in medical technology that systematically catalogs the genetic aberrations underlying different types of cancer. By comparing the DNA sequences of cancer cells to those of normal tissue, researchers can characterize the genetic mutations likely driving the progression and growth of specific cancers and determine similarities across different cancer types (Weinstein et al., 2013). For example, cancer mapping efforts have revealed the occurrence of the same genetic mutations underlying both ovarian and breast cancer (TCGA Research Network, 2011). This suggests that ovarian cancer may be an off-label site for a drug approved for breast cancer (Pleasance et al., 2022). We use these data to construct a novel genomic similarity index between cancer sites, allowing us to approximate each drug's total potential market size, including both approved and off-label uses.

After accounting for off-label potential, we uncover a meaningful reversal in the relationship between approval order and market size: Indications that receive earlier approval are often linked to larger total potential market sizes. Although we caution that our estimates are not causal, this pattern is consistent with regulatory arbitrage: Firms seek early approval in smaller indications that offer a lower-risk path to market and then rely on off-label use to access broader markets without formal approval. We consider alternative explanations—such as differences in trial duration, regulatory review times, and firm selection—and show that they are unlikely to explain our results.

Finally, we examine the policy implications of our findings. Back-of-the-envelope calculations suggest that seeking initial approval in a smaller indication shortens time to market by approximately 7.9 months, yielding an estimated $99.6 million in value to firms through earlier revenues and reduced trial costs. However, this acceleration comes with a potential cost: Off-label uses are often not backed by randomized or high-quality evidence and are associated with higher rates of adverse events (Eguale et al., 2016; Richardson, 2016; Wittich et al., 2012). Our findings underscore the importance of understanding how firms respond to the structure of regulatory incentives—and raise

[2]This is due to favorable off-label reimbursement policies in cancer treatment, as well as other factors (e.g., high disease severity and fewer treatments for rare cancers) encouraging physicians to experiment beyond formally approved uses.

questions about whether existing approval pathways strike the appropriate balance between speed and quality.

This paper contributes to a broad literature on the relationship between regulation and innovation incentives, particularly in the context of medical technologies (Acemoglu and Linn, 2004; Berger et al., 2021; Blume-Kohout and Sood, 2013; Danzon and Keuffel, 2014; Dubois et al., 2015, 2025; Grennan and Town, 2020; Maini and Pammolli, 2023; Malani and Philipson, 2012; Stern, 2017). First, our paper improves understanding of the consequences of entry regulation, both intended and unintended. We highlight how regulation affects not only whether a drug is approved, but also how firms prioritize among multiple potential markets. When firms can access broader markets through off-label prescribing, regulation creates incentives to pursue approval in smaller indications, while avoiding or deferring approval in larger ones. This dynamic has important implications for how policymakers allocate oversight and set evidentiary standards across therapeutic areas.

Second, prior research has documented regulatory arbitrage in various contexts, such as in the choice of bankruptcy filing jurisdictions (Ellias, 2018), states of incorporation (Bebchuk and Cohen, 2003; Cain and Davidoff, 2012), litigation courts (Cohen et al., 2019; Sytch and Kim, 2021), global sourcing (Berry et al., 2021), and store locations (Holmes, 1998; Rao et al., 2011). However, despite growing policy attention to its implications for consumer welfare and drug access, empirical evidence on regulatory arbitrage in the pharmaceutical sector remains limited. The prior work generally focuses on how firms selectively choose among different jurisdictions, with different regulations, for their activities (also called forum shopping). We add to the existing literature by demonstrating that, even when governed by a single regulatory framework, pharmaceutical firms strategically circumvent this regulation by selectively choosing therapeutic uses for formal approval.

Finally, this paper contributes to a growing set of papers on the measurement (Radley et al., 2006; Stafford, 2008), drivers (Dubois et al., 2025; Larkin et al., 2014; McKibbin, 2023; Shapiro, 2018), and consequences of off-label drug use (Bradford et al., 2018; Tunçel, 2025). Most closely related to our research, Dubois et al. (2025) show how policy changes on off-label promotion and prescribing influence firms' decisions to submit drug uses for formal approval. We build on this research by introducing a novel, ex ante measure of off-label market potential—one that does not require access to health care claims data but instead leverages biologically-based measures of similarity across disease indications to identify functional similarities.

The paper proceeds as follows. Section 2 describes the institutional background. Section 3 outlines the data, including the construction of our novel disease similarity index, and provides summary statistics. Section 4 explains our research design and gives the empirical results. Section 5 discusses the firm and policy implications of regulatory arbitrage. Section 6 concludes.

2 Background

2.1 Pharmaceutical Development and Entry Regulation

The development and approval process for a novel drug is long, costly, and risky. Drug development typically begins with preclinical testing in animals and human cells. If successful, the manufacturer submits an Investigational New Drug (IND) application to the FDA and begins the most expensive aspect of drug development: clinical trials. These trials advance through three primary phases, with costs increasing at each subsequent phase. Phase I trials assess safety and dosage in healthy volunteers; phase II trials examine safety and efficacy in 100 to 300 patients; and phase III trials test the drug in larger populations of 300 to 3,000 patients (FDA, 2024). Upon successful completion of phase III trials, the sponsor submits a New Drug Application (NDA) for FDA review. The full process typically spans 8 to 12 years, with estimated costs ranging from $300 million to over $2.6 billion and a success rate of only 9% for drugs that enter clinical testing (Adams and Brantner, 2006; Aryal et al., 2024; CSDD, 2014; Danzon and Keuffel, 2014; DiMasi, 2001; DiMasi et al., 2003; Wouters et al., 2020).

Importantly, FDA approval is indication-specific, meaning a drug receives regulatory approval to treat a specific disease in a specific patient population. To gain approval for an additional use, the manufacturer must submit a new IND, conduct supporting clinical trials, and file a supplemental New Drug Application (sNDA). The cost and duration of this process depend on the similarity between the original and new indications (FDA, 1998a). For example, if the original and new uses are closely related, manufacturers seeking approval for the new use may skip phase I trials and conduct fewer phase II trials.[3] While some drug uses are discovered after the drug is first approved, others are known prior to approval. Indeed, industry reports highlight that "indication sequencing"—including the choice of a drug's "lead" indication—is an important decision for pharmaceutical firms (IQVIA, 2023). Beyond the resource and time costs, there can be additional risks to the firm associated with seeking supplemental approvals, for instance, if subsequent clinical trials threaten the evidence base for approved indications or lead to the discovery of significant adverse events.

[3]Examples include a new stage of the same disease or the same disease in a new population.

2.2 Off-Label Drug Use

The practice of using approved health care technologies for unapproved uses is legal and common, particularly among treatments for cancer, cardiovascular diseases, and psychiatric diseases (Stafford, 2008). An estimated 90% of approved drugs have multiple therapeutic uses (Gelijns et al., 1998). Off-label use occurs for various reasons, including the lack of FDA-approved therapies for specific populations, the practice of using the same treatment for different but similar conditions, and insurance coverage (Wittich et al., 2012). For example, off-label use of psychiatric medicines is common in children because children are rarely included in clinical trials for drug approval (Lee et al., 2012). Many mental illnesses share the same or similar symptoms, motivating physicians to use a drug approved for one condition to treat another.[4]

The FDA recognizes that off-label use can be clinically appropriate under some circumstances, but it has expressed concern about its widespread adoption without rigorous supporting evidence (FDA, 2014). Indeed, prior research has found that off-label use is associated with higher rates of adverse events (Eguale et al., 2016). To mitigate potential public health risks, the FDA prohibits the direct promotion of off-label uses to physicians and patients. Nevertheless, evidence suggests that manufacturers still generate clinical evidence to encourage off-label use.[5,6] Additionally, the FDA's policy forbidding off-label advertising has gradually loosened and been challenged over time (FDA, 2014). Current regulation permits firms to respond to unsolicited questions from health care professionals and to disseminate information from peer-reviewed literature and clinical practice guidelines (Avorn et al., 2015). Given this regulatory landscape, firms may strategically rely on off-label use—rather than pursue the costly, risky, and time-intensive process of formal approval for supplemental indications—to expand demand for their drugs.

2.3 Regulatory Arbitrage in the Pharmaceutical Industry

Regulatory arbitrage has been defined as "structuring activity to take advantage of gaps or differences in regulations or laws" (Pollman, 2019, p. 567). Since FDA approval is specific to a given use, firms

[4]The prevalence of off-label use depends on a physician's propensity to prescribe a drug with limited evidence of safety and efficacy. In practice, physicians who engage in off-label use are rarely accused of medical malpractice. The process of informed consent does not require physicians to disclose that a drug is being used off-label. Further, off-label use is not necessarily negligent if the off-label use is included in the current standard of practice.

[5]The U.S. Department of Justice has charged and fined several major companies with illegal off-label promotion, including Eli Lilly ($1.4 billion in 2009), Pfizer ($2.3 billion in 2009), GlaxoSmithKline ($3 billion in 2012), and Abbott ($1.6 billion in 2012).

[6]Off-label drug use could still occur without supporting clinical evidence, driven instead by physician experience and experimentation.

developing drugs with multiple uses face variation in the likelihood of approval across indications. At the same time, off-label prescribing provides a mechanism for market diffusion beyond formally approved uses. This regulatory structure allows firms to respond strategically: They can pursue initial approval in small, lower-risk indications to minimize regulatory burden and expedite market entry, while relying on off-label drug use to access larger markets without formal approval.

The approach presents both benefits and costs for pharmaceutical firms. On the one hand, obtaining approval in smaller disease indications is often less costly than in larger markets (Jayasundara et al., 2019). Trials for smaller indications typically require fewer participants, reducing recruitment costs (Sertkaya et al., 2016). Moreover, advances in genetic sequencing and biomarker testing allow firms to target patient subgroups with higher treatment responsiveness, further increasing the likelihood of approval (Chandra et al., 2019; Michaeli et al., 2023). Consistent with this view, Appendix Figure A2 shows that drugs tested in smaller indications are associated with higher FDA approval rates and faster regulatory review times (i.e., the number of months between application submission and approval). Additionally, because drug prices in the U.S. are typically not indication-specific (Chandra and Garthwaite, 2017), firms can set higher prices based on inelastic demand in the small indication and apply the same price for its broader (off-label) use.

On the other hand, smaller indications often have limited patient populations, which can slow enrollment and restrict early revenues. Relying on off-label use for market expansion introduces uncertainty, as insurers may decline coverage and physician adoption can be uneven or unpredictable (Shapiro, 2018). Regulatory risk also remains as off-label promotion is restricted and enforcement has varied over time. Further, these off-label uses will typically be supported by weaker scientific evidence, and new adverse events may be discovered among patient populations not formally approved for the drug, threatening existing approved uses.

Taken together, FDA regulation creates asymmetric incentives across therapeutic markets. Regulatory arbitrage by pharmaceutical firms reflects a deliberate trade-off between regulatory risk and market reach. The extent to which firms pursue this strategy in practice is an empirical question we examine in the remainder of the paper.

3 Data

To empirically investigate whether pharmaceutical firms engage in regulatory arbitrage when making market entry decisions, we construct a novel dataset combining regulatory approvals, clinical trials,

market size, and genetic similarity across cancer indications. Our main analytic dataset consists of 129 cancer drugs first approved by the FDA between 1978 and 2016. We consider two types of market entry decisions: regulatory approvals and clinical trials. We obtain regulatory approval data from the Clarivate Cortellis Competitive Intelligence Global database and the FDA's Drugs@FDA database, which together provide information on both original NDAs and sNDAs. We classify approved indications at the cancer site-stage level (e.g., "breast–metastatic").[7] We link these regulatory data to clinical trial records from the Clarivate Cortellis Clinical Trials Intelligence database.[8] Data limitations prevent us from categorizing clinical trial indications to the cancer stage level; as a result, our clinical trial analyses are conducted at the drug-cancer site level.

To proxy for market size, we collect data on the number of new diagnoses associated with each cancer's site and stage (Budish et al., 2015). These data come from the Surveillance, Epidemiology, and End Results (SEER) database, available from the National Cancer Institute (NCI). We focus on five-year lagged averages of market size, where we calculate these lags relative to either the year of indication approval (for regulatory approval analyses) or trial start (for clinical trial analyses).

3.1 Cancer Genome Mapping Data

To capture the potential for regulatory arbitrage, we construct a novel proxy for each drug's off-label market size using data from large-scale cancer genome mapping studies. Cancer genome sequencing is an advance in medical technology that systematically catalogs the genetic mutations underlying different cancer types. Genome sequencing compares the DNA sequences of cancer cells to those of normal tissue, so that researchers can determine the mutations behind specific cancers and understand similarities across different cancer types (Weinstein et al., 2013). Using genetic sequencing data, we characterize the similarity between different diseases to define a drug's expected off-label cancer sites. For example, cancer mapping efforts have shown that if a cancer site (e.g., breast cancer) shares genetic mutations (e.g., BRCA1) with another cancer site (e.g., ovarian cancer), drugs approved for one site may be used off-label to treat the other (Pleasance et al., 2022; TCGA Research Network, 2011).

[7] FDA-approved indications are typically more granular than our site-stage coding. For instance, letrozole was initially approved for "advanced breast cancer in postmenopausal women with disease progression following anti-estrogen therapy," and later for "first-line treatment of postmenopausal women with hormone receptor-positive or hormone receptor-unknown locally advanced or metastatic breast cancer." Both approvals are classified as "breast–metastatic."

[8] The clinical trials used in our main analyses consist of those linked to an approved drug. In robustness checks, we confirm that our results are robust to including all clinical trials, regardless of whether they are linked to an approved or unapproved drug.

We obtain information on gene-cancer pairings uncovered in large-scale cancer mapping efforts from the publicly accessible Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) database (Sondka et al., 2018; Tate et al., 2018).[9] The COSMIC team curates cancer genome data from hundreds of genetic sequencing studies and literature to catalog the set of genes containing mutations that are causally associated with cancer. In the CGC, each gene (e.g., BRCA1) is linked to the set of cancers (e.g., breast and ovarian cancers) where mutations in that gene are likely contributors to the disease's development.[10]

Our use of genome mapping data to quantify biologically-driven off-label potential is conceptually related to recent efforts in the bioinformatics literature to use molecular and genomic similarity to guide drug repurposing and discovery (Cheng et al., 2019; Tanoli et al., 2021). This includes leveraging data from disease-specific sequencing efforts (e.g., the Cancer Genome Atlas) and genome-wide association studies (GWAS) to uncover biological linkages across diseases that are then used to identify new uses for existing drugs (Uffelmann et al., 2021). As a point of validation, genetically similar cancer sites are likely to be treated with the same drugs (Chakravarty et al., 2017; Suehnholz et al., 2024). We build on these insights by exploiting genetic similarity to study how regulation shapes firms' innovation and entry decisions, using it to estimate the ex ante set of disease markets where a drug may plausibly be used off-label.[11]

3.1.1 Gene-Based Disease Similarity

Using the CGC data, we then estimate the similarity between cancer sites using the extent of overlap between sets of genetic mutations associated with each cancer type. In the spirit of Krieger et al. (2022), we quantify the similarity between two different cancer sites by calculating the Tanimoto distance (Jaccard coefficient). This measure calculates the distance between each set of genetic mutations associated with each cancer site. For example, the similarity index s between cancer sites A and B is the intersection of A and B's genetic mutations divided by the union of these mutations:

$$s_{A,B} \equiv \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{1}$$

[9] For more details, see `https://cancer.sanger.ac.uk/census`.

[10] We consider both somatic and germline mutations. Somatic mutations occur in any cells of the body except the germ cells (sperm and egg) and are thus not passed on from parents to children. Germline mutations are changes in DNA inherited from parents.

[11] We are unable to exploit the staggered release of information from large-scale cancer genome mapping studies (as in Kao (2025)) because that information disclosure occurs at the gene-cancer site level, which is often too granular to generate meaningful variation in our setting.

A similarity index of 0 implies that a pair of cancer sites are not closely related and have no common mutations, while a measure of 1 implies that they are closely related and have exactly the same set of mutations. Figure 2 provides a heat map of our similarity index across all cancer sites. The mean index across these cancer sites is 0.074 (SD = 0.198).[12]

3.1.2 Calculating Expected Off-Label Market Size

We use the similarity index to determine the expected off-label market size for each indication. This calculation is based on the market size for cancer sites related to indication i and the corresponding similarity indices:

$$\textit{Off-Label Market Size}_{i,t} = \sum_{j \neq i} s_{i,j} \textit{Market Size}_{j,t} \tag{2}$$

Here, indication i has an off-label market size at time t that is defined as the weighted sum of the market sizes of all other cancer sites j ($j \neq i$) at time t, where the weight is the similarity index between cancer sites i and j. To illustrate, consider a simple example between three cancer sites, A, B, and C. Assume that cancer site B has a similarity with cancer site A ($s_{A,B} = 0.09$) and a similarity with cancer site C ($s_{B,C} = 0.01$). Suppose that at time t, the market size for cancer site A is 100 and the market size for cancer site C is 20. At time t, the off-label market size for a drug initially approved in cancer site B is $(0.09 \times 100) + (0.01 \times 20) = 9.2$.

Before continuing, we note some limitations of this measure. First, this measure is aimed at reflecting off-label drug use that is supported by scientific evidence. Since off-label drug use can be driven by other factors, such as costs and drug side effects (Stafford, 2008), it is likely to be an underestimate of the overall off-label market size (Radley et al., 2006). We address this by including a robustness check where we directly incorporate *actual* measures of off-label drug use. Second, evidence from the CGC may not fully capture the true level of overlap across cancer sites: The creators of the CGC describe it as being a "conservative but high-confidence list" of genes associated with cancer, raising questions about the possibility of false negatives when deciding which gene-cancer associations to include in the database.[13] As a robustness check, we also construct a less restrictive disease similarity measure based on a broader range of gene-cancer associations identified in large-scale

[12] As an example, Appendix B provides the mutations for breast and ovarian cancer, including their overlapping mutations (Appendix Table B1), and shows the calculation of their similarity index. Breast and ovarian cancer, which have been known to be treated by similar therapies (e.g., PARP inhibitors), have a similarity index of 0.118.

[13] The CGC's cancer experts apply strict criteria when determining the set of gene-cancer associations to include in the database. For more information, see `https://www.sanger.ac.uk/data/cancer-gene-census/`.

cancer sequencing studies. As we show in Section 4.2, the results using this less restrictive disease similarity measure are largely the same.

3.2 Controls

We collect data on controls to help address the concern that the relationship between indication order and market size may be confounded by factors unrelated to the practice of regulatory arbitrage. As a proxy for the level of competition faced by each drug in a given indication, we count the cumulative number of drug approvals in the same indication in the year prior to the indication approval year (for regulatory approval analyses) or the trial start year (for clinical trial analyses). To capture regulatory incentives associated with subsequent market entry decisions, we compile data on whether the drug ever received an orphan drug designation. Established in 1983, the Orphan Drug Act (ODA) act offers targeted regulatory incentives (e.g., tax benefits and regulatory exclusivities) to firms developing drugs for "orphan" diseases (i.e., rare diseases affecting less than 200,000 individuals).[14] Since the passage of the ODA, there has been a significant increase in the number of approved orphan drugs (from 10 in 1990 to 77 in 2017) (Miller and Lanthier, 2018). Since orphan designation increases firm incentives for entry in small indications, failing to control for it may conflate entry decisions that are motivated by targeted regulatory incentives, rather than firms' decisions to engage in regulatory arbitrage.

Next, we construct controls for intellectual property protection, which may also influence the relationship between indication order and market size. For instance, Budish et al. (2015) show that firms have reduced incentives to pursue research for early-stage cancers relative to late-stage ones because of the former's longer development times and shorter resulting patent terms post-launch. If early-stage cancers tend to have larger market sizes, then observed patterns in indication order may partly reflect patent-related incentives rather than strategic regulatory behavior. In light of this, we combine information on pharmaceutical IP protection from the United States Patent and Trademark Office (USPTO) and the FDA's Orange Book.[15] We create two controls for IP protection: *Primary IP protection* measures the months from each indication approval (or trial start) date to when the drug's primary IP expires. We consider the primary IP expiration to be the latter of either the molecule patent or the new chemical entity exclusivity expiration. Primary IP protection is generally

[14] For more details, see `https://www.fda.gov/industry/medical-products-rare-diseases-and-conditions/designating-orphan-product-drugs-and-biological-products`.

[15] Drugs are protected by two types of IP rights: patents granted by the USPTO and regulatory exclusivities granted by the FDA.

considered the strongest form of IP protection, with almost certain enforcement.[16] *Potential IP protection* gives the maximum possible term of monopoly protection, if all IP on a drug were upheld. It measures the months from each indication approval (or trial start) to when the final IP on the drug expires.[17]

Table 1 presents descriptive statistics for our sample of oncology drugs first approved between 1978 and 2016. The average drug is tested in 47 different trial indications (cancer sites) and receives FDA approval in 4 unique approval indications (cancer site-stage observations).[18] Within our sample, 64% of drugs receive an orphan drug designation. Across all drug-indications, the mean primary IP protection remaining at the time of approval is 124 months (10.3 years) and mean potential IP protection is 174 months (14.5 years). The maximum primary IP protection remaining is 277 months (23.1 years) and maximum potential IP protection is 346 months (28.8 years). Mean off-label market size (11,179 diagnoses) is more than 11 times the mean market size of approval indications (985 diagnoses). Appendix Figure A4 shows that there are meaningful time lags between indications: The average time between trial indications is more than 1 year and the average time between approval indications is nearly 3 years.

4 Empirical Approach and Results

To examine whether firms sequence clinical trials and regulatory approvals in ways consistent with regulatory arbitrage, we analyze how the size of targeted markets varies with indication order. We aggregate our data on market entry decisions, market size, and drug characteristics into drug-indication order (e.g., letrozole-first indication) observations, where indication order is defined as the sequence of indications in which a given drug is clinically tested (or approved).[19] For drug d and indication order i, we estimate the following:

$$MarketSize_{d,i} = \alpha + \beta_1 IndicationOrder_{d,i} + \gamma X_{d,i} + \epsilon_{d,i} \tag{3}$$

[16] For those trials taking place before a drug's initial launch, we assume the firms have an ex ante expectation of what this primary IP expiration will be. For trials or approvals taking place after the primary IP has expired, we consider this measure to be zero.

[17] For trials occurring before a drug's initial launch, we consider the patents and exclusivities in effect at launch to calculate this measure. For all trials after launch, we consider the patents and exclusivities in effect at the trial start date.

[18] Appendix Figure A3 shows the complete distribution of trial and approval indications per drug.

[19] When a single indication order corresponds to multiple indications (e.g., a drug is initially clinically tested in both breast and ovarian cancers in trials starting on the same date), we take the average of their market sizes.

Our outcome variable $MarketSize$ is the natural log of the lagged five-year average market size associated with indication order i for drug d. The coefficient on $IndicationOrder$ is our main estimate of interest. While our theoretical argument is that firms choose the sequencing of indications as a function of indication market size, we set the outcome and explanatory variables as such for ease of interpretation of our estimates. As presented, our estimates allow the reader to easily discern how earlier indications systematically differ in market size from later ones.[20] Further, our goal here is not to obtain a causal effect but rather to descriptively document how regulation may shape firms' market entry decisions.

We include a comprehensive set of controls X to account for observable factors that may influence firms' indication order decisions, including detailed measures of intellectual property protection (i.e., primary and potential IP); the drug's initial approval year; its cancer group or type (e.g., stomach and esophageal cancers would belong to the "digestive system" cancer group); the level of competition (i.e., the cumulative number of drug approvals in the same indication as of the prior year); and regulatory incentives (i.e., whether it has an orphan designation). All estimates are from ordinary least squares (OLS) models with robust standard errors.

4.1 Trends in Market Entry Decisions

We begin by examining descriptive trends in firms' market entry decisions. Figure 3 plots the relationship between indication order and market size for our sample of drugs, separately for clinical trials (Panel A) and regulatory approvals (Panel B). Consistent with Figure 1, Panels A and B illustrate key differences between the order of clinical trials and regulatory approvals. Panel A shows a strong negative correlation between trial indication order and market size: Firms tend to initiate clinical trials in their largest potential markets. In contrast, Panel B reveals no discernible relationship between approval indication order and market size, suggesting that firms do not systematically prioritize larger markets when seeking FDA approval. These patterns are suggestive of regulatory arbitrage—firms may initially target smaller, more favorable indications for formal approval while using off-label pathways to access larger markets.

Importantly, the fact that firms tend to initiate trials in larger potential markets but seek initial FDA approval in smaller markets suggests that scientific and logistical constraints, such as availability of technological resources and trial participants, are unlikely to be driving the observed patterns. For example, if firms' market entry decisions were primarily driven by patient availability, we would

[20] For the interested reader, we also present estimates where the outcome and explanatory variables are reversed.

expect similar patterns across both trials and approvals. Instead, the divergence suggests that firms may delay approval in larger markets, despite having trial data, to reduce their regulatory burden.

Appendix Table C1 formalizes these relationships. Column 1 reports the raw correlation between trial indication order and market size. The estimated coefficient implies that a 1-unit increase in trial indication order is associated with an 8% ($\approx (exp(-0.0823) - 1) \times 100$) decrease in market size. Column 2 shows that this negative relationship between trial indication order and market size persists once all controls are included, and the relationship remains significant at the 1% level. Columns 3 and 4 repeat these same regressions for approvals instead of trials. Column 3 confirms that approval indication order has no significant relationship with market size. Once all controls are included in Column 4, the relationship remains quantitatively small and insignificant. These findings suggest that observable firm and market characteristics do not explain the divergence between the ordering of clinical trials and regulatory approvals.[21]

While we do not claim a causal interpretation—our setting does not allow for identification of the direct effects of regulatory structure on entry behavior—these results provide novel descriptive evidence consistent with regulatory arbitrage. Specifically, the findings suggest that firms may strategically pursue approval in smaller disease markets where entry barriers are lower, relying on off-label use to access larger markets. In the next section, we test this mechanism more directly by incorporating our proxy for off-label market potential.

4.2 Incorporating Off-Label Market Potential

To assess whether the patterns observed above are consistent with regulatory arbitrage, we incorporate our measure of total market size, which is defined as the sum of the focal indication's market size and its genetically similar off-label markets. Table 2 presents results using this expanded outcome measure.

For reference, Columns 1 and 3 repeat the baseline specifications from Appendix Table C1, where the primary outcome variable is the market size of the focal indication only, and all controls are included. Column 1 shows the clinical trial specification, and Column 3 shows the approval specification. In both cases, the regressions use indication order as the main explanatory variable. Columns 2 and 4 replace the outcome variable with total market size, incorporating potential off-label opportunities

[21] Appendix Table D1 presents the specification where we reverse the outcome and explanatory variables. Appendix Table D2 shows that we observe similar trends when expanding our set of trials to include those associated with both approved and unapproved cancer drugs whose first trial is between 1998 and 2001, corresponding to the most complete coverage of our trials dataset.

based on disease similarity. Column 2 shows that for clinical trials, the negative relationship between indication order and market size persists even when off-label potential is included: A one-unit increase in trial indication order is associated with a 6% ($\approx (exp(-0.0610) - 1) \times 100$) decrease in total market size. Column 4 reveals a noticeable shift in approvals. Whereas Column 3 showed no relationship between approval indication order and focal market size, Column 4 shows that when total market size is used, the relationship becomes negative and statistically significant. A one-unit increase in approval indication order is associated with a 15% ($\approx (exp(-0.163) - 1) \times 100$) decline in total market size.

These results suggest that firms systematically prioritize larger markets when initiating clinical trials, including those with potential for off-label expansion, but seek initial FDA approval in smaller focal indications. This pattern is consistent with regulatory arbitrage: Firms may enter through smaller, more favorable regulatory pathways while anticipating subsequent expansion into biologically-similar markets via off-label use.

As a robustness check, we construct alternative measures of total market size based on broader gene-cancer similarity thresholds. Specifically, we follow the bioinformatics literature in focusing on high-frequency mutations within each cancer site, and define "high frequency" as the top 10%, 20%, or 30% of mutations identified in large-scale sequencing studies. Table 3 shows that the main results are robust to these alternative definitions of genetic similarity, supporting the validity of our findings across a range of biologically-similar off-label markets. Under these alternative definitions, we find that a one-unit increase in approval indication order is associated with a decline in total market size of between 18% ($\approx (exp(-0.203) - 1) \times 100$) and 21% ($\approx (exp(-0.236) - 1) \times 100$).

4.3 Differences in Actual Off-Label Drug Use

In the previous analyses, we estimated off-label market potential by linking our cancer dataset to cancer genome mapping information. This allowed us to approximate a drug's *expected* off-label market size as supported by scientific evidence. In practice, off-label use may occur in the absence of supporting scientific evidence (e.g., due to physician experience, experimentation, the lack of FDA-approved treatments, and reimbursement policies). To complement our prior approach, we now incorporate measures of *actual* off-label drug use. One challenge is that available measures of actual off-label drug use are generally restricted to survey and health insurance claims data that cover specific patient populations (e.g., Medicare enrollees), limiting our ability to capture total

off-label drug use. We use therapeutic-level off-label prescribing rates from Radley et al. (2006), which estimates these rates based on physician surveys, as a proxy for actual off-label activity.

For this exercise, we extend our analyses beyond cancer by using a "cross-disease dataset" consisting of cancer and non-cancer drugs first approved between 1998 and 2021. We categorize approved indications using International Classification of Disease (ICD-9) codes. As a proxy for the market size of drugs in this dataset, we use the number of diagnoses from the Medical Expenditure Panel Survey (MEPS). We calculate yearly average market size at the ICD-9 level, using data from 1996 to 1997. We distinguish disease categories with high versus low opportunities for regulatory arbitrage, as measured by their off-label propensity. Diseases with low off-label propensity (low opportunities for regulatory arbitrage) include antidiabetics, antihypertensives, and antihyperlipidemics; those with high off-label propensity (high opportunities for regulatory arbitrage) include oncology, anticonvulsants, psychiatry, and antiasthmatics (Radley et al., 2006). Our final cross-disease sample includes 615 drugs (of which 482 drugs are initially approved in diseases with high off-label propensity and 133 drugs in diseases with low propensity) mapped into 802 drug-indication observations.

Figure 4 shows that drugs in high off-label propensity disease categories receive more total approvals, on average, than those in low-propensity categories. Notably, the relationship between indication order and market size differs across the two groups. For high off-label drugs, each additional indication is associated with a 23% ($\approx (\exp(-0.265) - 1) \times 100$) decrease in market size. For low off-label drugs, the decline is steeper—an 89% ($\approx (\exp(-2.202) - 1) \times 100$) decrease in market size. These results are consistent with the hypothesis that firms in high off-label propensity areas are more willing to pursue initial approval in smaller markets, anticipating later expansion through off-label prescribing. In contrast, firms in low off-label propensity areas face fewer opportunities for informal diffusion of their pharmaceuticals and therefore have stronger incentives to prioritize larger markets in their initial approval decisions.

We supplement these cross-disease findings with several robustness checks. First, we control for observable drug and market characteristics (Appendix Table D3). Second, we include controls for alternative disease measures (Anatomical Therapeutic Chemical (ATC) classifications) in Appendix Table D4.[22] Finally, we verify the robustness of our categorization using an alternative measure of off-label use derived from IMS prescribing data (Appendix Table D5).

[22]We are grateful for the ATC-drug crosswalk provided by the authors of Kakani et al. (2022).

4.4 Alternative Explanations

Our findings are consistent with the practice of regulatory arbitrage by pharmaceutical firms in their entry decisions. Through the inclusion of our controls, we have addressed several alternative explanations. For instance, one concern may be that the indication sequencing choices we observe reflect firms' abilities to obtain regulatory incentives or exclusivities for more narrow indications; our controls for orphan drug designation, intellectual property protection, and indication group account for such opportunities. In this section, we consider several alternative explanations for the patterns we document.

4.4.1 Differences in Expected Trial Duration

One alternative explanation is that firms sequence their clinical trials based on expected trial duration rather than market size. For example, firms may give priority to indications with longer anticipated trials if these involve more complex diseases requiring extended trial periods and closer monitoring. If this were the case, the negative relationship between trial indication order and market size, as well as the absence of a corresponding relationship for approvals, could be explained by variation in trial length rather than by regulatory arbitrage.

To evaluate this alternative explanation, we return to the cancer dataset used in our main analyses and redefine trial indication order using trial end dates rather than start dates. As shown in Appendix Figure D1, the strong negative relationship between trial order and market size remains. This result suggests that the observed patterns are unlikely to be driven by differences in trial duration.

4.4.2 Differential FDA Review Times

A second alternative explanation is that the absence of a strong relationship between approval indication order and market size could be driven by differential FDA review times. If firms tend to submit applications for larger markets earlier, but those applications are subject to longer regulatory review periods, then the approval order we observe may reflect agency timing rather than firm strategy. To address this, we collect original application submission dates from FDA review letters and redefine approval indication order using submission rather than approval timing. As shown in Appendix Figure D2, the relationship between approval order and market size remains statistically insignificant, suggesting that variation in FDA review times does not explain our findings.

4.4.3 FDA Selection

A third alternative explanation is that the observed approval patterns reflect direct FDA selection rather than firm strategy. Because publicly available FDA data include only successful approvals, it is possible that we observe more approvals in smaller markets simply because applications for larger indications are more likely to be rejected or delayed. If so, the lack of a strong relationship between approval order and market size could reflect direct FDA selection (e.g., where the FDA systematically denies or slows approval for drugs targeting larger indications) rather than firms' indication order decisions.

Although we cannot directly observe unsubmitted or rejected applications, several factors suggest this explanation is unlikely to account for our findings. First, our results are robust to controls for regulatory incentives, such as orphan drug designation, which are closely linked to approval likelihood. Second, the costs associated with filing an FDA application are substantial, making it unlikely that firms would routinely submit applications they expect to be denied. Taken together, these considerations support the interpretation that the observed sequencing of approvals reflects firm strategy (which may be potentially informed by regulatory expectations) rather than a pattern driven primarily by FDA selection.

4.4.4 Firm selection

Finally, a separate concern is that these patterns are driven not by strategic rationales but by firm selection (e.g., an sNDA is not sought since the drug is found to be unsafe or ineffective). In Section 5.1, we assess the clinical trial quality of a drug's indications for which the firm seeks regulatory approval and those for which it does not. If differences in the pattern of clinical trials and regulatory approvals were primarily due to scientific rationales, we would expect to see little variation in trial quality between indications with and without regulatory approval. Contrary to this expectation, we find statistically significant differences, suggesting that firm selection is not a primary factor behind the sequencing of indications that we observe.

5 Managerial and Policy Implications

Our results suggest that firms engage in regulatory arbitrage by initially seeking approval in smaller, more favorable markets and relying on off-label use to expand into others. In considering the effects of regulatory design, prior papers have examined how regulation affects R&D costs through changes in trial design and time to market (Berndt et al., 2005; Chandra et al., 2024), as well as how it

shapes the number of markets firms choose to enter (Dubois et al., 2025; Maini and Pammolli, 2023). Building on this work, this section presents a set of back-of-the-envelope calculations to assess how regulatory arbitrage affects three outcomes: firms' trial design choices, the speed of market entry, and the number of markets entered. These findings raise important questions for regulators, who must weigh the benefits of accelerated access to new treatments against the need for sufficient evidence on drug safety and efficacy. While our calculations are intentionally simple (e.g., they do not account for the actual uptake or costs of promoting off-label use, payer coverage, or legal risks), they offer insight into the trade-offs that arise when regulatory structure shapes firm incentives.

5.1 Impact on Clinical Trial Design Quality

Given the structure of FDA regulation, a key question is whether firms respond to less stringent regulatory requirements by relying on weaker forms of evidence when expanding off-label use. To assess this, we examine the design quality of clinical trials that have a primary purpose to increase off-label use (e.g., through dissemination of results in scientific publications) and not to support a future sNDA. For firms, a key advantage of such "non-regulatory" trials is that they, by definition, do not require the high costs or rigor associated with trials supporting regulatory approval. For example, the FDA considers randomized controlled trials the "gold standard," but such trials are typically more costly to conduct. Although publicly available cost estimates are limited, Moore et al. (2018) find that pivotal trials with a placebo or active comparator cost between \$28.8 to \$48.9 million, on average, while those without a comparator cost \$13.5 million.[23] However, trials aimed at expanding off-label use often do not use randomized controlled designs (Bothwell et al., 2016). As a result, we compare the quality of trials likely conducted for regulatory purposes with those likely conducted for non-regulatory purposes.

For each drug in our sample, we classify the indications tested in clinical trials as either "regulatory" (for which the firm eventually receives FDA approval) or "non-regulatory" (for which the firm does not seek approval). Figure 5 compares the share of high-quality trials for regulatory and non-regulatory indications, with quality defined as trials that are randomized and controlled. Looking at all trials, we find a significantly higher share of high-quality trials for regulatory indications than for non-regulatory indications (48% vs 39%, respectively; $p < 0.01$).

The likelihood of high trial quality is likely to vary across different trial phases, which vary in length and cost. Firms are less likely to conduct late-stage clinical trials (i.e., phases II/III and

[23] Pivotal trials are trials of safety and/or efficacy used to support regulatory approval.

III) that are high quality if they do not intend to pursue regulatory approval. Consistent with this view, we find little difference in trial quality among late-stage trials. In contrast, we find these differences are driven by early-stage trials (i.e., phases I and II), with a higher share of high-quality early-stage trials among regulatory indications than among non-regulatory indications (35% vs 31%, respectively; $p < 0.01$). Overall, through this exercise, we see that the ability to engage in regulatory arbitrage appears to shift the quality of firms' clinical trials.

5.2 Impact on the Speed of Entry Into New Markets

Long development timelines are a major driver of high drug development costs (Wong et al., 2014). By seeking initial regulatory approval in smaller indications, firms can potentially bring a drug to market more quickly by conducting smaller clinical trials with a segment of participants more likely to respond to treatment (Chandra et al., 2019). In this section, we carry out a back-of-the-envelope calculation to quantify the dollar value to firms of expediting market entry, focusing specifically on the speed with which cancer drugs enter the market when pursuing initial approval in small versus large indications.

We restrict our sample to each drug's first approval and consider "small" indications as those within the lowest quartile of market size (e.g., soft tissue including heart cancer) and "large" indications as the rest (e.g., breast cancer). Measuring from pivotal trial start date to approval date, drugs with small initial indications reach the market in 44.8 months on average, while those with large initial indications reach the market in 52.7 months, for a difference of 7.9 months. To translate the time savings into dollar savings, we make use of a recent study of clinical trial costs, which finds that each additional month in late-stage trials amounts to a median $671,000 spent (Martin et al., 2017). Multiplying this figure by 7.9 months suggests that pharmaceutical firms can save more than $5.3 million in trial costs alone by engaging in regulatory arbitrage and seeking initial approval in a smaller market.

In addition to trial cost savings, firms also benefit from revenues obtained earlier. To determine per-drug revenues over this time, we turn to Schuhmacher et al. (2022), who examine new drugs launched and their total sales from 2011 to 2020. Given that we are considering drugs launched in small indications, we take a conservative approach by excluding blockbusters (with mean annual sales of >$1 billion) and high-selling ($0.5–0.999 billion) drugs from our revenue calculations. Looking only at low- (<$0.1 billion) and medium-selling ($0.1–0.499 billion) drugs, we determine these drugs

have an average annual revenue of $143.2 million, equating to $94.3 million over 7.9 months.[24] We consider this to be a conservative estimate of total revenues because drugs initially launched in small indications may still generate large or even blockbuster-level sales, due to both the potential for off-label use and competitive factors allowing monopoly pricing to be high.

Summing the costs saved from clinical trials and the revenues obtained via earlier market entry, we obtain a total value to firms of $5.3 million + $94.3 million = $99.6 million per drug from seeking initial approval in a small market relative to a large one. While these calculations are purposefully simple (e.g., they do not take into account variation in clinical trial costs or the costs and uncertainty associated with off-label promotion), they do provide insight into the potential financial incentives firms face when selecting initial markets for approval.

5.3 Impact on the Number of Markets Entered

For policymakers, an important issue is how regulatory policy influences the amount of innovation developed. Here, we analyze how regulatory arbitrage in the pharmaceutical industry affects the number of markets a firm enters for a single drug. The high barriers to entry associated with regulatory approval for larger markets suggest that if off-label use were banned, we might expect firms to pursue regulatory approval for a portion of these indications, provided the market expansion benefits exceed the costs of obtaining sufficient scientific evidence for approval. For the remaining indications, for which the benefits of formal approval do not exceed the costs, these drug-indication pairs would become "missing" in the sense that firms would no longer pursue them. As such, a natural question is, how many "additional" indications do firms explore because of the regulatory structure that allows them to enter new markets through off-label drug use? Stated differently, how many off-label indications are associated with a given drug?[25]

To determine this, we begin with each drug's set of non-regulatory indications (i.e., those tested in clinical trials but never submitted for formal approval). Non-regulatory indications include both ones pursued for off-label purposes and ones discontinued for scientific reasons. To isolate "off-label" indications, we use two methods to remove indications likely discontinued due to scientific concerns. Our first method is based on the idea that firms typically terminate trials early when they encounter safety or efficacy issues, making it unlikely they will seek regulatory approval for the indication or

[24] We calculate annual revenues from Figure 1 of Schuhmacher et al. (2022) as follows: ($82.1 billion total revenues for medium-selling drugs + $4.7 billion for low-selling drugs)/(51 medium-selling drugs + 50 low-selling drugs)/6 years average commercialization period over 2011 to 2020 = $143.2 million annual per-drug revenue.

[25] For simplicity, this exercise does not consider the "size" of the associated indications.

disclose trial outcomes to facilitate off-label drug use (Cook et al., 2014). We therefore exclude unapproved indications with an above-median share of early trial terminations, suggesting negative findings. Our second method is motivated by the idea that firms conducting high-quality (i.e., randomized controlled) trials without seeking regulatory approval likely do so because of inadequate safety or efficacy findings. We therefore also exclude indications with an above-median share of high-quality trials that did not lead to approval, reflecting potentially unfavorable results. The remaining non-regulatory indications are classified as off-label.

Using these data, we document the total number of regulatory and off-label indications associated with a drug. The average drug is associated with a total of 9.7 such indications, of which only 3 are approved and a mean of 6.7 are off-label.[26] Figure 6 gives the full distribution of off-label indications across our sample. Our findings support the idea that the opportunity for regulatory arbitrage likely increases the overall number of indications that firms focus on in their clinical trial efforts. A key caveat to interpreting these results is that we are examining equilibrium outcomes. If off-label drug use were banned, the total number of indications tested in clinical trials may decline. However, it is not clear that the average number of indications for a given drug would fall to 3, as firms might pursue regulatory approval for some indications initially intended for off-label use. Therefore, we consider the mean number of off-label indications (6.7) as an upper bound on the number of additional indications that firms may still pursue for a given drug if off-label use were prohibited.

6 Discussion and Conclusion

This paper provides new evidence on how pharmaceutical firms engage in regulatory arbitrage in their market entry decisions. By combining data on cancer drug approvals with information from large-scale genomic sequencing efforts, we show firms often seek initial approval in smaller disease indications, where approval likelihood is more favorable, and subsequently expand use through off-label prescribing into larger, related disease areas. These relationships hold even after accounting for key drug and market characteristics, such as regulatory incentives targeted towards rare diseases and intellectual property protections. The patterns are most pronounced for drugs with

[26] Our main analyses consider trial indications at the site level and approval indications, for which we have more granular data, at the site-stage level. For this analysis of the number of markets, because we explicitly match trial to approval indications, all indications are at the site level. As such, we have a mean of 3 approval indications per drug at the site level versus, as indicated in Table 1, 4 approval indications per drug at the site-stage level.

high expected off-label potential, as measured by the genetic similarity of approved and unapproved cancer indications.

Our results offer several important insights for future research. First, we provide evidence consistent with the view that differences in regulatory burden (e.g., approval rates) may drive the observed patterns (e.g., Appendix Figure A2). However, other explanations remain possible. For example, smaller indications may allow firms to set higher prices due to reduced levels of competition (Chandra and Garthwaite, 2017) or involve lower commercializaton, marketing, and manufacturing costs. Initial approval in such markets may enable firms to generate revenue to finance the production and sale of additional indications. Investigating these mechanisms, using detailed data on pricing and drug costs, remains an important area for future work.

Second, we document a new fact about the order in which drug applications are obtained. Like many studies on the pharmaceutical industry (see, e.g., Acemoglu and Linn, 2004; Ellison and Ellison, 2011), our analysis is limited to successful regulatory applications (i.e., approved NDAs and sNDAs). This restricts our ability to fully characterize firms' strategic entry decisions. This concern is mitigated by the fact that firms generally do not submit applications to the FDA that they anticipate will be rejected. However, researchers with access to such rejected application data may aim to more precisely identify the set of indications for which firms are likely to seek regulatory approval. This will broaden our understanding of how regulatory incentives shape firms' market entry strategies.

Third, our findings not only speak to how regulatory design shapes a firm's approval sequencing for an individual product, but also raise questions about the broader—often unintended—effects of that design. For example, the potential for regulatory arbitrage may shift the types of drugs a firm chooses to develop in the first place. It may encourage investment in drugs that target small indications with relatively little patient burden, potentially at the expense of drugs aimed at larger, unmet needs. Similarly, understanding how off-label expansion compares to formal supplemental approvals, in terms of prescribing behavior and patient outcomes, is critical for evaluating the welfare implications of these market entry strategies.

Nevertheless, our findings provide new empirical evidence on an understudied dimension of regulatory design in shaping the diffusion of new medical technologies. Rather than treating approval as a one-time decision, FDA regulation requires firms to navigate a sequence of market entry choices, each with its own regulatory burden. The ability to rely on off-label use introduces an informal, lower-risk

pathway to market expansion, fundamentally altering firm incentives regarding where and when to seek formal approval. Our work suggests that regulatory flexibility can promote earlier access to new therapies but may also lead to uneven evidence generation across patient populations. These trade-offs raise important policy questions about how to structure pre-market approval regulation and post-market surveillance in a way that balances access to medical innovations with the need for rigorous evidence.

References

Acemoglu, D. and J. Linn (2004). Market size in innovation: Theory and evidence from the pharmaceutical industry. *Quarterly Journal of Economics* 119(3): 1049–1090.

Adams, C. P. and V. V. Brantner (2006). Estimating the cost of new drug development: Is it really $802 million? *Health Affairs* 25(2): 420–428.

Aryal, G., F. Ciliberto, L. E. Farmer, and E. Khmelnitskaya (2024). Valuing pharmaceutical drug innovations. SSRN Working Paper No. 4300678.

Avorn, J., A. Sarpatwari, and A. S. Kesselheim (2015). Forbidden and permitted statements about medications—Loosening the rules. *New England Journal of Medicine* 373(10): 967–973.

Bebchuk, L. A. and A. Cohen (2003). Firms' decisions where to incorporate. *Journal of Law and Economics* 46(2): 383–425.

Berger, B., A. Chandra, and C. Garthwaite (2021). Regulatory approval and expanded market size. NBER Working Paper No. 28889.

Berndt, E. R., A. H. Gottschalk, T. J. Philipson, and M. W. Strobeck (2005). Industry funding of the FDA: Effects of PDUFA on approval times and withdrawal rates. *Nature Reviews Drug Discovery* 4(7): 545–554.

Berry, H., A. Kaul, and N. Lee (2021). Follow the smoke: The pollution haven effect on global sourcing. *Strategic Management Journal* 42(13): 2420–2450.

Blume-Kohout, M. E. and N. Sood (2013). Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development. *Journal of Public Economics* 97: 327–336.

Bothwell, L. E., J. A. Greene, S. H. Podolsky, and D. S. Jones (2016). Assessing the gold standard—Lessons from the history of RCTs.

Bradford, W. D., J. L. Turner, and J. W. Williams (2018). Off-label use of pharmaceuticals: A detection controlled estimation approach. *Journal of Industrial Economics* 66(4): 866–903.

Budish, E., B. N. Roin, and H. Williams (2015). Do firms underinvest in long-term research? Evidence from cancer clinical trials. *American Economic Review* 105(7): 2044–2085.

Cain, M. D. and S. M. Davidoff (2012). Delaware's competitive reach. *Journal of Empirical Legal Studies* 9(1): 92–128.

Chakravarty, D., J. Gao, S. Phillips, R. Kundra, H. Zhang, J. Wang, J. E. Rudolph, R. Yaeger, T. Soumerai, M. H. Nissan, and others (2017). OncoKB: A precision oncology knowledge base. *JCO Precision Oncology* 1: 1–16.

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016). Health care exceptionalism? Performance and allocation in the U.S. health care sector. *American Economic Review* 106(8): 2110–2144.

Chandra, A. and C. Garthwaite (2017). The economics of indication-based drug pricing. *New England Journal of Medicine* 377(2): 103–106.

Chandra, A., C. Garthwaite, and A. D. Stern (2019). Characterizing the drug development pipeline for precision medicines. In E. R. Berndt, D. P. Goldman, and J. W. Rowe (Eds.), *Economic Dimensions of Personalized and Precision Medicine*, pp. 115–157. Chicago: University of Chicago Press.

Chandra, A., J. Kao, K. L. Miller, and A. D. Stern (2024). Regulatory incentives for innovation: The FDA's breakthrough therapy designation. *Review of Economics and Statistics*: 1–46.

Cheng, F., W. Lu, C. Liu, J. Fang, Y. Hou, D. E. Handy, R. Wang, Y. Zhao, Y. Yang, J. Huang, D. E. Hill, M. Vidal, C. Eng, and J. Loscalzo (2019). A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nature Communications* 10(1): 3476.

Cohen, L., U. G. Gurun, and S. D. Kominers (2019). Patent trolls: Evidence from targeted firms. *Management Science* 65(12): 5461–5486.

Cohen, L., U. G. Gurun, and D. Li (2021). Internal deadlines, drug approvals, and safety problems. *American Economic Review: Insights* 3(1): 67–82.

Congressional Research Service (2021). Drug pricing and the FDA: How the approval process affects the cost of prescription drugs.

Cook, D., D. Brown, R. Alexander, R. March, P. Morgan, G. Satterthwaite, and M. N. Pangalos (2014). Lessons learned from the fate of AstraZeneca's drug pipeline: A five-dimensional framework. *Nature Reviews Drug Discovery* 13(6): 419–431.

CSDD (2014). Costs to develop and win marketing approval for a new drug is $2.6 billion. `http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study`, Tufts Center for the Study of Drug Development.

Danzon, P. M. and E. L. Keuffel (2014). Regulation of the pharmaceutical-biotechnology industry. In N. L. Rose (Ed.), *Economic Regulation and Its Reform: What Have We Learned?*, pp. 407–484. Chicago: University of Chicago Press.

DiMasi, J. A. (2001). New drug development in the United States from 1963 to 1999. *Clinical Pharmacology & Therapeutics* 69(5): 286–296.

DiMasi, J. A., R. W. Hansen, and H. G. Grabowski (2003). The price of innovation: New estimates of drug development costs. *Journal of Health Economics* 22(2): 151–185.

Douglas, J. L. (2016). New wine into old bottles: Fintech meets the bank regulatory world. *North Carolina Banking Institute* 20: 17.

Dubois, P., O. De Mouzon, F. Scott Morton, and P. Seabright (2015). Market size and pharmaceutical innovation. *RAND Journal of Economics* 46(4): 844–871.

Dubois, P., E. Gentry, and T. Tunçel (2025). Strategic submissions: A cross-country analysis of supplemental drug approvals. TSE Working Paper No. 23-1459.

Eguale, T., D. L. Buckeridge, A. Verma, N. E. Winslade, A. Benedetti, J. A. Hanley, and R. Tamblyn (2016). Association of off-label drug use and adverse drug events in an adult population. *JAMA Internal Medicine* 176(1): 55–63.

Eisenberg, R. S. (2005). The problem of new uses. *Yale Journal of Health Policy, Law, and Ethics* 5: 717–740.

Ellias, J. A. (2018). What drives bankruptcy forum shopping? Evidence from market data. *Journal of Legal Studies* 47(1): 119–149.

Ellison, G. and S. F. Ellison (2011). Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration. *American Economic Journal: Microeconomics* 3(1): 1–36.

FDA (1998). Guidance for industry: FDA approval of new cancer treatment uses for marketed and biological products.

FDA (2014). Guidance for industry: Distributing scientific and medical publications on unapproved new uses—recommended practices.

FDA (2024). The drug development process—Step 3: Clinical research. `https://www.fda.gov/patients/drug-development-process/step-3-clinical-research`, accessed on 2024-05-04.

Friedman, M. (1996). Testimony on supplemental indications for approved prescription drugs. House Committee on Government Reform and Oversight, Subcommittee on Human Resources and Intergovernmental Relations.

Gelijns, A. C., N. Rosenberg, and A. J. Moskowitz (1998). Capturing the unexpected benefits of medical research. *New England Journal of Medicine* 339(10): 693–698.

Grennan, M. and R. J. Town (2020). Regulating innovation with uncertain quality: Information, risk, and access in medical devices. *American Economic Review* 110(1): 120–161.

Holmes, T. J. (1998). The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy* 106(4): 667–705.

IQVIA (2023). Success multiplied: Launch excellence for multi-indication assets. `https://www.iqvia.com/library/white-papers/success-multiplied-launch-excellence-for-multi-indication-assets`, accessed on 2024-05-07.

IQVIA Institute (2018). Global oncology trends 2018. `https://www.iqvia.com/insights/the-iqvia-institute/reports/global-oncology-trends-2018`, accessed on 2021-01-13.

Jayasundara, K., A. Hollis, M. Krahn, M. Mamdani, J. S. Hoch, and P. Grootendorst (2019). Estimating the clinical cost of drug development for orphan versus non-orphan drugs. *Orphanet Journal of Rare Diseases* 14**:** 1–10.

Joskow, P. L. (2005). Regulation and deregulation after 25 years: Lessons learned for research in industrial organization. *Review of Industrial Organization* 26**:** 169–193.

Kakani, P., M. Chernew, and A. Chandra (2022). The contribution of price growth to pharmaceutical revenue growth in the United States: Evidence from medicines sold in retail pharmacies. *Journal of Health Politics, Policy and Law* 47(6)**:** 629–648.

Kao, J. (2017). Pharmaceutical regulation and off-label uses. White paper.

Kao, J. (2025). Charted territory: Mapping the cancer genome and R&D decisions in the pharmaceutical industry. Working paper.

Krieger, J., D. Li, and D. Papanikolaou (2022). Missing novelty in drug development. *Review of Financial Studies* 35(2)**:** 636–679.

Larkin, I., D. Ang, J. Avorn, and A. S. Kesselheim (2014). Restrictions on pharmaceutical detailing reduced off-label prescribing of antidepressants and antipsychotics in children. *Health Affairs* 33(6)**:** 1014–1023.

Lee, E., A. R. Teschemaker, R. Johann-Liang, G. Bazemore, M. Yoon, K.-S. Shim, M. Daniel, J. Pittman, and A. K. Wutoh (2012). Off-label prescribing patterns of antidepressants in children and adolescents. *Pharmacoepidemiology and Drug Safety* 21(2)**:** 137–144.

Maini, L. and F. Pammolli (2023). Reference pricing as a deterrent to entry: Evidence from the European pharmaceutical market. *American Economic Journal: Microeconomics* 15(2)**:** 345–383.

Malani, A. and T. Philipson (2012). The regulation of medical products. In *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, pp. 101–140. Oxford: Oxford University Press.

Martin, L., M. Hutchens, C. Hawkins, and A. Radnov (2017). How much do clinical trials cost? *Nature Reviews Drug Discovery* 16(6)**:** 381–382.

McKibbin, R. (2023). The effect of RCTs on drug demand: Evidence from off-label cancer drugs. *Journal of Health Economics* 90**:** 102779.

Michaeli, T., H. Jürges, and D. T. Michaeli (2023). FDA approval, clinical trial evidence, efficacy, epidemiology, and price for non-orphan and ultra-rare, rare, and common orphan cancer drug indications: Cross sectional analysis. *BMJ* 381**:** e073242.

Miller, K. L. and M. Lanthier (2018). Investigating the landscape of U.S. orphan product approvals. *Orphanet Journal of Rare Diseases* 13**:** 1–8.

Moore, T. J., H. Zhang, G. Anderson, and G. C. Alexander (2018). Estimated costs of pivotal trials for novel therapeutic agents approved by the U.S. Food and Drug Administration, 2015–2016. *JAMA Internal Medicine* 178(11)**:** 1451–1457.

Pfister, D. G. (2012). Off-label use of oncology drugs: The need for more data and then some. *Journal of Clinical Oncology* 30(6)**:** 584–586.

Pleasance, E., A. Bohm, L. Williamson, J. Nelson, Y. Shen, M. Bonakdar, E. Titmuss, V. Csizmok, K. Wee, S. Hosseinzadeh, and Others (2022). Whole-genome and transcriptome analysis enhances precision cancer treatment options. *Annals of Oncology* 33(9)**:** 939–949.

Pollman, E. (2019). Tech, regulatory arbitrage, and limits. *European Business Organization Law Review* 20**:** 567–590.

Posen, H. A. (2015). Ridesharing in the sharing economy: Should regulators impose Über regulations on Uber. *Iowa Law Review* 101**:** 405.

Radley, D. C., S. N. Finkelstein, and R. S. Stafford (2006). Off-label prescribing among office-based physicians. *Archives of Internal Medicine* 166(9)**:** 1021–1026.

Rao, H., L. Q. Yue, and P. Ingram (2011). Laws of attraction: Regulatory arbitrage in the face of activism in right-to-work states. *American Sociological Review* 76(3)**:** 365–385.

Richardson, E. (2016). Off-label drug promotion. *Health Affairs* Health Policy Brief, DOI: 10.1377/hpb20160630.920075.

Schuhmacher, A., M. Hinder, N. Boger, D. Hartl, and O. Gassmann (2022). The significance of blockbusters in the pharmaceutical industry. *Nature Reviews Drug Discovery*.

Sertkaya, A., H.-H. Wong, A. Jessup, and T. Beleche (2016). Key cost drivers of pharmaceutical clinical trials in the United States. *Clinical Trials* 13(2): 117–126.

Shapiro, B. T. (2018). Informational shocks, off-label prescribing, and the effects of physician detailing. *Management Science* 64(12): 5925–5945.

Sondka, Z., S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, and S. A. Forbes (2018). The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nature Reviews Cancer* 18(11): 696–705.

Stafford, R. S. (2008). Regulating off-label drug use—Rethinking the role of the FDA. *New England Journal of Medicine* 358(14): 1427–1429.

Stern, A. D. (2017). Innovation under regulatory uncertainty: Evidence from medical technology. *Journal of Public Economics* 145: 181–200.

Suehnholz, S. P., M. H. Nissan, H. Zhang, R. Kundra, S. Nandakumar, C. Lu, S. Carrero, A. Dhaneshwar, N. Fernandez, B. W. Xu, and others (2024). Quantifying the expanding landscape of clinical actionability for patients with cancer. *Cancer Discovery* 14(1): 49–65.

Sytch, M. and Y. H. Kim (2021). Quo vadis? From the schoolyard to the courtroom. *Administrative Science Quarterly* 66(1): 177–219.

Tanoli, Z., M. Vähä-Koskela, and T. Aittokallio (2021). Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opinion on Drug Discovery* 16(9): 977–989.

Tate, J. G., S. Bamford, H. C. Jubb, Z. Sandra, D. M. Beard, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jute, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, and S. A. Forbes (2018). COSMIC: The catalogue of somatic mutations. *Nucleic Acids Research* 47(D1): D941–47.

TCGA Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474(7353): 609–615.

Tunçel, T. (2025). Should we prevent off-label drug prescriptions? Empirical evidence from France. *Review of Economic Studies* 92(3): 2027–2070.

Uffelmann, E., Q. Q. Huang, N. S. Munung, J. De Vries, Y. Okada, A. R. Martin, H. C. Martin, T. Lappalainen, and D. Posthuma (2021). Genome-wide association studies. *Nature Reviews Methods Primers* 1(1): 59.

Vives, X. (2019). Digital disruption in banking. *Annual Review of Financial Economics* 11: 243–272.

Weinstein, J. N., E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* 45(10): 1113–1120.

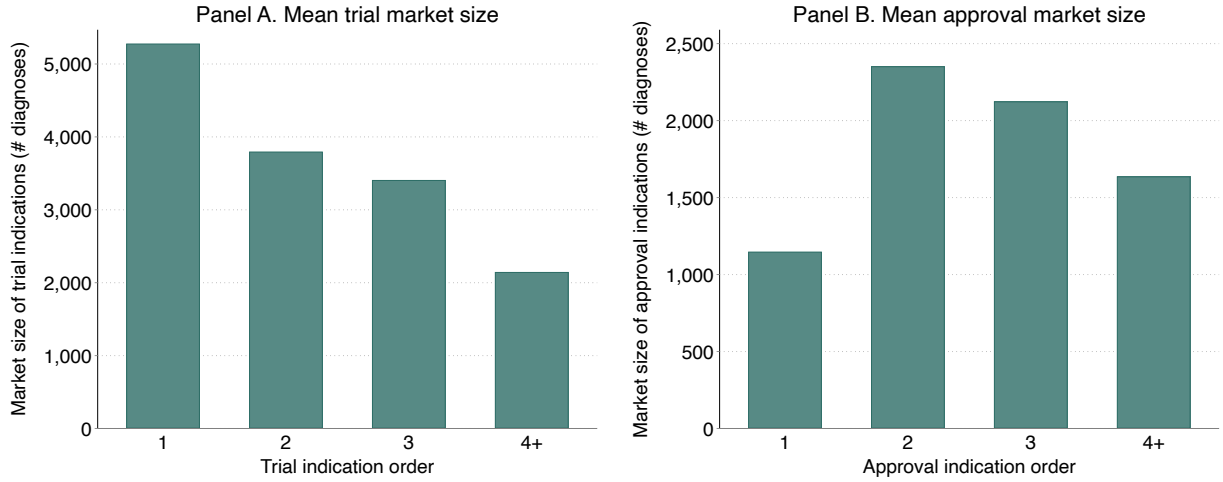
Wittich, C. M., C. M. Burkle, and W. L. Lanier (2012). Ten common questions (and their answers) about off-label drug use. *Mayo Clinic Proceedings* 87(10): 982–990.

Wong, H.-H., A. Jessup, A. Sertkaya, A. Birkenbach, A. Berlind, and J. Eyraud (2014). Examination of clinical trial costs and barriers for drug development final. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services.

Wouters, O. J., M. McKee, and J. Luyten (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA* 323(9): 844–853.

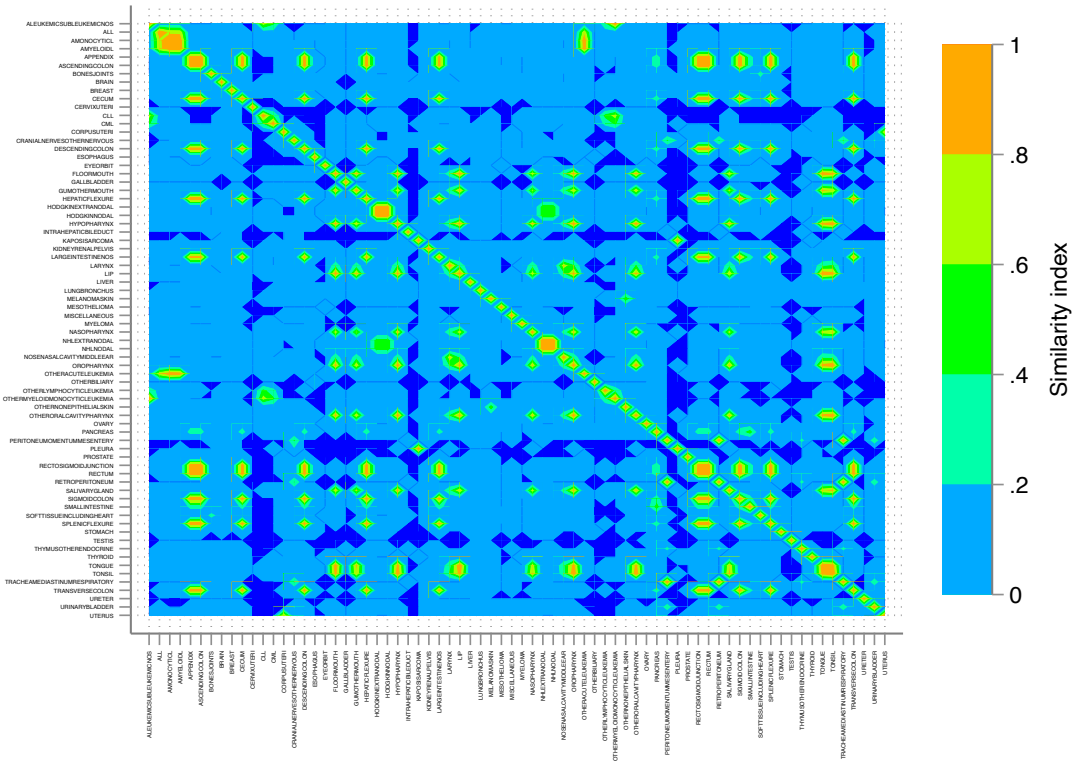
Figures and Tables

FIGURE 1: MARKET ENTRY DECISIONS



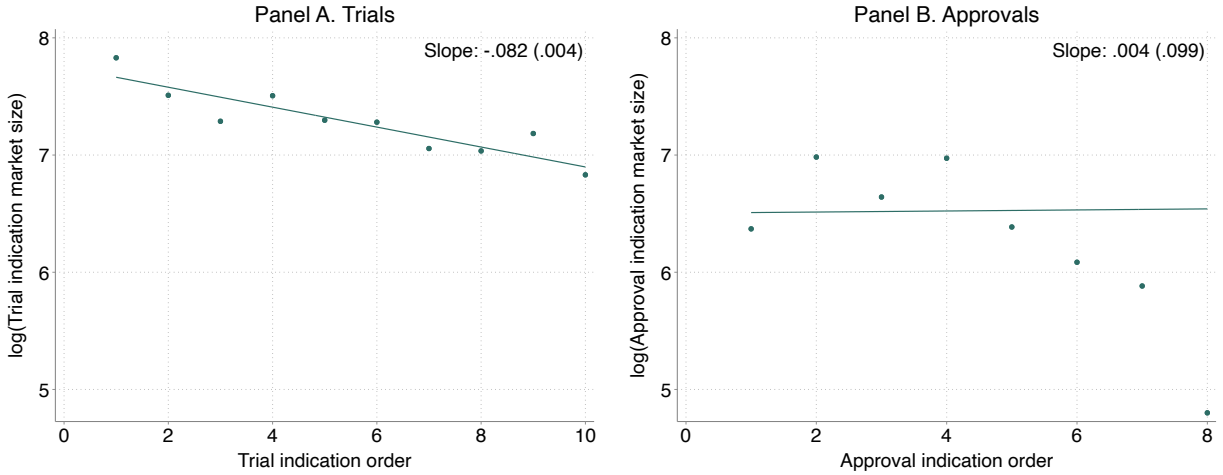
NOTES: This figure shows the mean market size by indication order for both clinical trials (Panel A) and regulatory approvals (Panel B) for cancer drugs first approved from 1978 to 2016. Market size is measured by new diagnoses for an indication in the SEER data. Trial indications are measured at the cancer site level and approval indications are at the cancer site-stage level.

FIGURE 2: SIMILARITY INDEX HEAT MAP



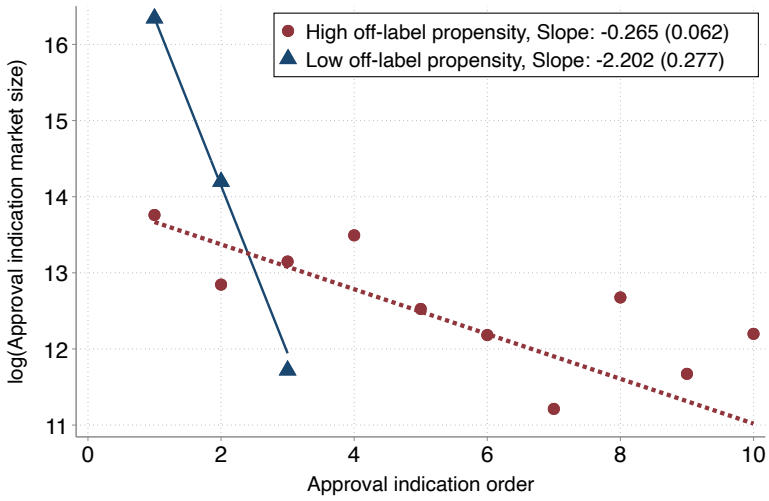
NOTES: This figure shows a heat map of our similarity index across all cancer sites in the COSMIC Cancer Gene Census (CGC). Cancer site pairs associated with lighter colors are more similar.

FIGURE 3: TRENDS IN MARKET ENTRY DECISIONS



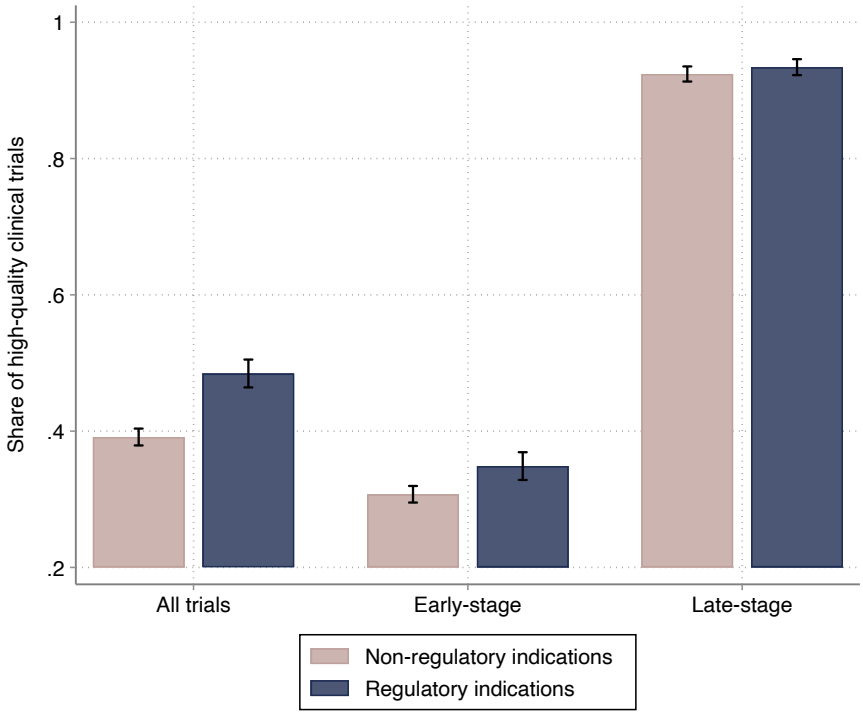
NOTES: This figure shows trends in market entry decisions for cancer drugs first approved from 1978 to 2016. The level of observation is the drug-indication order. Panel A shows the relationship between trial indication order and market size; the number of observations is 1,980. Panel B shows the relationship between approval indication order and market size; the number of observations is 187. Market size is measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size relative to either the trial start year (Panel A) or approval year (Panel B). Each marker represents binned averages for a given indication order. For ease of interpretation, we display up to the 10th trial indication.

FIGURE 4: TRENDS IN REGULATORY APPROVALS ACROSS DISEASES, BY PROPENSITY FOR OFF-LABEL DRUG USE



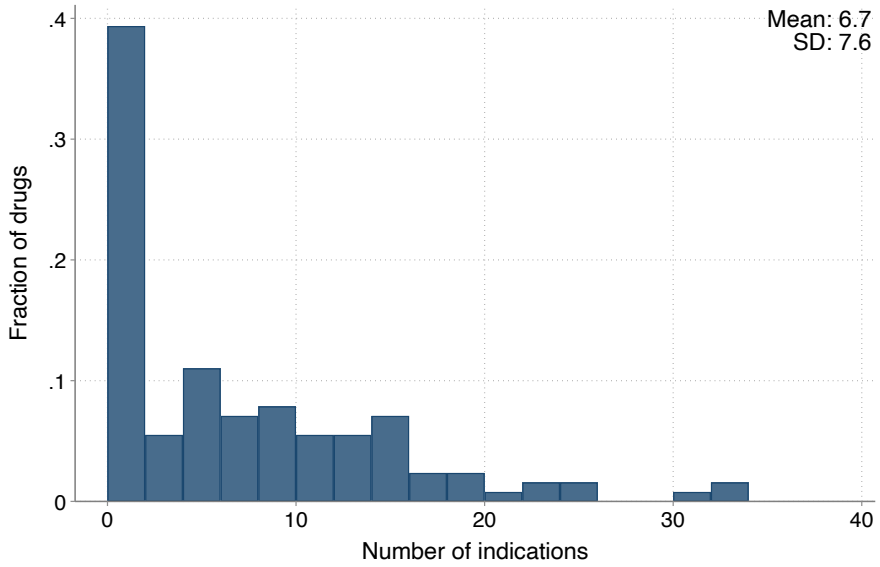
NOTES: This figure shows the relationship between regulatory approvals and market size for approved drugs across several disease categories from 1998 to 2021, by propensity for off-label drug use. The level of observation is the drug-indication order, where an indication corresponds to an ICD-9 code. Market size is measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with indication order averaged over 1996 and 1997. Each marker represents binned averages for a given indication order. For ease of interpretation, we display up to the 10th indication.

FIGURE 5: QUALITY OF CLINICAL TRIALS: NON-REGULATORY VS. REGULATORY INDICATIONS



NOTES: This figure shows differences in the quality of clinical trials for non-regulatory versus regulatory indications for cancer drugs first approved from 1978 to 2016. High-quality clinical trials refer to trials that are randomized and controlled. Bars give shares, while capped ranges provide 95% confidence intervals.

FIGURE 6: DISTRIBUTION OF OFF-LABEL INDICATIONS



NOTES: This figure plots the distribution of off-label indications for cancer drugs first approved from 1978 to 2016. Number of drugs is 127.

TABLE 1: DESCRIPTIVE STATISTICS

	Mean	SD	Min	Max
Drug level				
Number of Unique Approval Indications	4	6	1	35
Number of Unique Trial Indications	46.7	23.6	1	78
Share with Orphan Disease Designation	0.64	0.48	0	1
Drug-indication level				
Market Size: Approval Indications (Diagnoses)	985	1,749	2.4	9,104
Market Size: Trial Indications (Diagnoses)	1,616	3,218	4.6	17,915
Market Size: Potential Off-Label (Diagnoses)	11,179	13,344	0	47,515
IP Protection: Primary (Months)	123.5	55.3	0	277
IP Protection: Potential (Months)	173.9	50.6	44	346
Competition: Approval Indications (Total Approvals)	3.1	3.6	0	20
Competition: Trial Indications (Total Approvals)	3.3	3.8	0	21

NOTES: This table shows summary statistics for our dataset of cancer drugs first approved from 1978 to 2016. Statistics are at the cancer site level for trial indications and the cancer site-stage level for approval indications and potential off-label market size. Market size is measured by new diagnoses for an indication in the SEER data, while off-label market size is measured by the proportion of new diagnoses for any potential off-label indications, with the proportions given by our disease similarity index.

Table 2: Ordering of Market Entry Decisions, Incorporating Off-Label Opportunities

Dependent variable:	Clinical trials		Regulatory approvals	
	Focal market size (1)	Total market size (2)	Focal market size (3)	Total market size (4)
Indication order	-0.0471***	-0.0610***	-0.0265	-0.163**
	(0.00576)	(0.00608)	(0.103)	(0.0815)
Mean of dep. var	7.040	8.282	6.522	8.990
Observations	1,523	1,523	182	182
Controls:				
Initial approval year	yes	yes	yes	yes
Indication group	yes	yes	yes	yes
Competition	yes	yes	yes	yes
Regulatory incentives	yes	yes	yes	yes
Intellectual property	yes	yes	yes	yes

Notes: This table shows the relationship between indication order and market size for cancer drugs first approved from 1978 to 2016. The level of observation is the drug-indication order. Columns 1 and 2 examine clinical trials, and Columns 3 and 4 examine regulatory approvals. The outcome variable in Columns 1 and 3 is focal market size, while that in Columns 2 and 4 is total market size, including potential off-label markets. Focal market size is measured by new diagnoses for an indication in the SEER data. Total market size is measured by new diagnoses for the focal indication plus a proportion of new diagnoses for any potential off-label indications, with the proportions given by our disease similarity index. For both variables, we use the log of the five-year average market size associated with indication order. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

TABLE 3: REGULATORY APPROVALS AND TOTAL MARKET SIZE, USING ALTERNATIVE SIMILARITY MEASURES

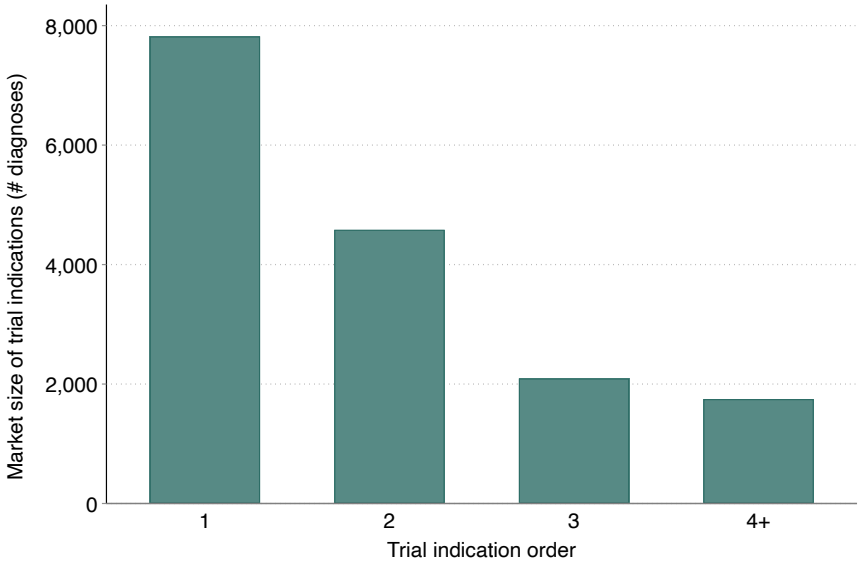
	Dependent variable: Total market size					
	All genes			**Cancer Gene Census**		
	Top 10% (1)	Top 20% (2)	Top 30% (3)	Top 10% (4)	Top 20% (5)	Top 30% (6)
Indication order	-0.203**	-0.215**	-0.223**	-0.206**	-0.233**	-0.236**
	(0.0943)	(0.0970)	(0.0971)	(0.0974)	(0.0984)	(0.102)
Mean of dep. var	10.049	10.199	10.246	10.283	10.561	10.650
Observations	182	182	182	182	182	182
Controls:						
Initial approval year	yes	yes	yes	yes	yes	yes
Indication group	yes	yes	yes	yes	yes	yes
Competition	yes	yes	yes	yes	yes	yes
Regulatory incentives	yes	yes	yes	yes	yes	yes
Intellectual property	yes	yes	yes	yes	yes	yes

NOTES: This table shows the relationship between indication order for regulatory approvals and total market size for cancer drugs first approved from 1978 to 2016, using alternative similarity measures. The level of observation is the drug-indication order. Alternative measures are generated by directly using cancer genome sequencing data from 168 large-scale mapping studies. Genetic mutations are restricted to those that occur at a high frequency within each mapping study, where a genetic mutation is "high frequency" within a cancer if it occurs in the top 10% (Columns 1 and 4), top 20% (Columns 2 and 5), or top 30% (Columns 3 and 6) of most frequently occurring mutations. Columns 1 to 3 focus on genetic mutations occurring among all genes. Columns 4 to 6 focus on the set of genetic mutations occurring among genes found in the COSMIC Cancer Gene Census (CGC). The outcome variable is total market size, including potential off-label markets. Total market size is measured by new diagnoses for the focal indication plus a proportion of new diagnoses for any potential off-label indications, with the proportions given by our disease similarity index. We use the log of the five-year average market size associated with indication order. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Appendices

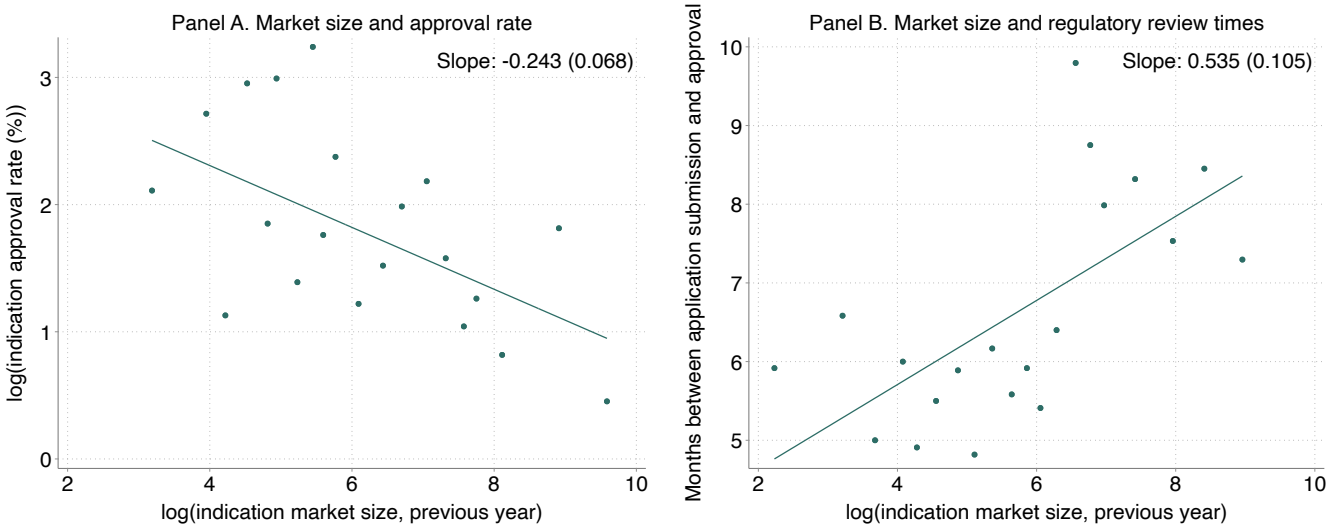
A Trends in Market Entry Decisions and Regulatory Review Processes

FIGURE A1: PHASE III TRIAL MARKET ENTRY DECISIONS



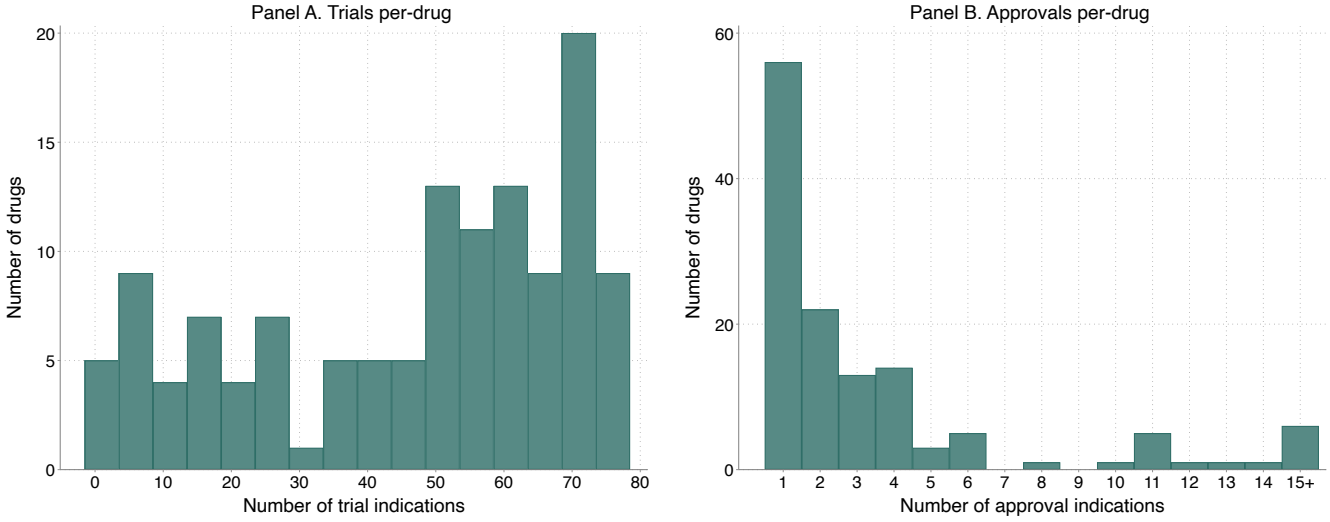
NOTES: This figure shows the mean market size by indication order for phase III clinical trials for cancer drugs first approved from 1978 to 2016. Market size is measured by new diagnoses for an indication in the SEER data. Trial indications are measured at the cancer site level.

FIGURE A2: APPROVAL RATES AND REGULATORY REVIEW TIMES ACROSS INDICATIONS



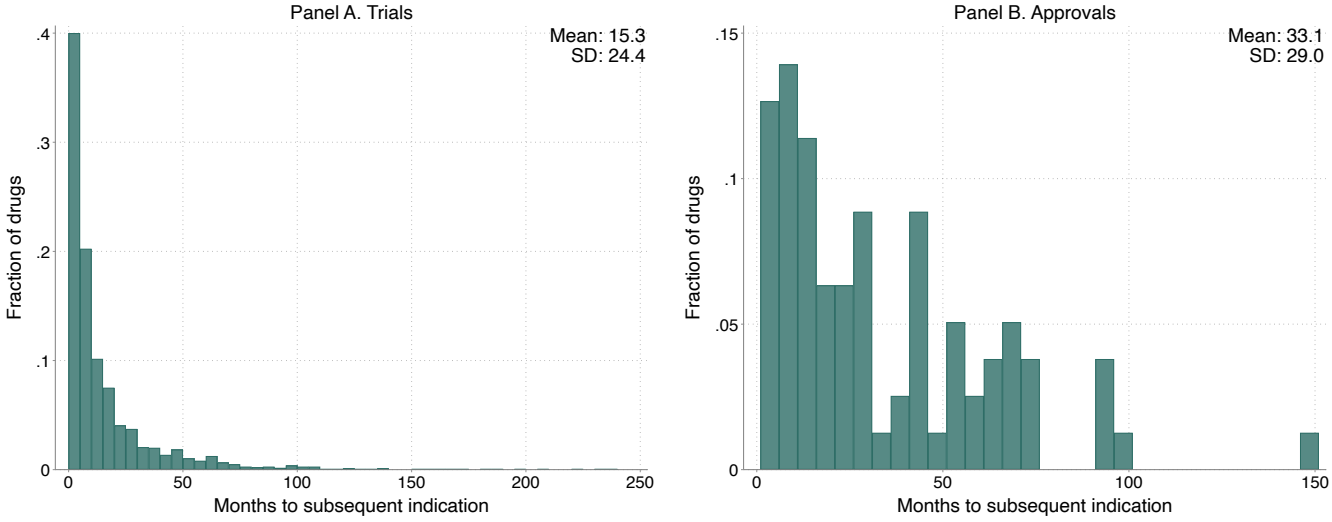
NOTES: This figure shows binned scatter plots examining indication approval rates and regulatory review times across cancer sites, for cancer drugs first approved from 1978 to 2016. The level of observation is the cancer site-year. Indication approval rate is defined as the probability that a drug tested in pre-clinical research in a given year advances to approval. Regulatory review time is the number of months between application submission and approval. Panel A compares the log of the indication approval rate to the log of the market size in the year prior to pre-clinical testing. Panel B compares the regulatory review time to the log of the market size in the year prior to pre-clinical testing.

FIGURE A3: PER-DRUG DISTRIBUTION OF MARKET ENTRY DECISIONS



NOTES: This figure shows the distribution of market entry decisions for cancer drugs first approved from 1978 to 2016. Panel A gives the per-drug distribution of trial indications; number of drugs is 127. Panel B gives the per-drug distribution of approval indications; number of drugs is 129.

FIGURE A4: TIME BETWEEN MARKET ENTRY DECISIONS



NOTES: This figure shows the months between market entry decisions for cancer drugs first approved from 1978 to 2016. Panel A gives the distribution of months between successive trial indications. Panel B gives the distribution of months between successive approval indications.

B Similarity Index Example: Breast and Ovarian Cancers

Table B1 provides the genetic mutations associated with breast and ovarian cancers from the CGC database. We calculate their similarity index as follows: $s_{Breast,Ovarian} = \frac{|B \cap O|}{|B \cup O|} = 10/85 = 0.118$.

Table B1: Genetic Mutations for Breast and Ovarian Cancers

Breast gene mutations	**Ovarian gene mutations**	**Breast & ovarian gene mutations**
ALK	AKT2	AKT1
APOBEC3B	ATR	ARID1A
ASPM	BRAF	ARID1B
BAP1	CASP3	BARD1
BRIP1	CCNE1	BRCA1
CASP8	CDK12	BRCA2
CCND1	COL3A1	ERBB2
CDH1	CREB1	GOLPH3
CDKN1B	CSMD3	PPM1D
CHEK2	CTNNB1	RAD50
CTCF	EIF1AX	
DCTN1	EWSR1	
EP300	FES	
ESR1	FOXL2	
ETV6	GOPC	
FADD	LRP1B	
FBLN2	MAPK1	
FEN1	MLH1	
FLCN	MSH2	
FLNA	MSH6	
FOXA1	PIK3R1	
GATA3	PLAG1	
HGF	PMS1	
IKZF3	PMS2	
IRS4	PPP2R1A	
MAP2K4	PRDM2	
MAP3K1	PTK6	
MAP3K13	RNF43	
MED12	ROS1	
NCOR1	STK11	
NOTCH1		
NTRK3		
PALB2		
PBRM1		
PIK3CA		
POLQ		
PPFIBP1		
RANBP2		
RB1		
SALL4		
SMARCD1		
TBX3		
TP53		
VHL		
ZMYM3		

C Ordering of Market Entry Decisions

TABLE C1: ORDERING OF MARKET ENTRY DECISIONS

	Dependent variable: Focal market size			
	Clinical trials		Regulatory approvals	
	(1)	(2)	(3)	(4)
Indication order	-0.0823***	-0.0471***	0.00448	-0.0265
	(0.00371)	(0.00576)	(0.0995)	(0.103)
Mean of Dep. Var.	6.829	7.040	6.513	6.522
Observations	1,980	1,523	187	182
Controls:				
Initial approval year	no	yes	no	yes
Indication group	no	yes	no	yes
Competition	no	yes	no	yes
Regulatory incentives	no	yes	no	yes
Intellectual property	no	yes	no	yes

NOTES: This table shows the relationship between indication order and market size for cancer drugs first approved from 1978 to 2016. The level of observation is the drug-indication order. Columns 1 and 2 examine clinical trials, and Columns 3 and 4 examine regulatory approvals. The outcome variable is market size, measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size associated with indication order. The number of observations differs between Columns 1 and 2 (and between Columns 3 and 4) due to missing data on control variables. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D Robustness Checks

D.1 Indication Order as the Outcome Variable

TABLE D1: ORDERING OF MARKET ENTRY DECISIONS WITH INDICATION ORDER AS THE OUTCOME VARIABLE

	Dependent variable: Indication order			
	Clinical trials		Regulatory approvals	
	(1)	(2)	(3)	(4)
Focal market size	-2.315***	-0.974***	0.00411	-0.0316
	(0.106)	(0.118)	(0.0910)	(0.124)
Mean of Dep. Var.	10.963	10.454	1.743	1.764
Observations	1,980	1,523	187	182
Controls:				
Initial approval year	no	yes	no	yes
Indication group	no	yes	no	yes
Competition	no	yes	no	yes
Regulatory incentives	no	yes	no	yes
Intellectual property	no	yes	no	yes

NOTES: This table shows the relationship between indication order and market size for cancer drugs first approved from 1978 to 2016. The level of observation is the drug-indication order. Columns 1 and 2 examine clinical trials, and Columns 3 and 4 examine regulatory approvals. The outcome variable is the indication order. Market size is measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size associated with indication order. The number of observations differs between Columns 1 and 2 (and between Columns 3 and 4) due to missing data on control variables. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.2 Incorporating Trials Linked to FDA-Approved and Unapproved Drugs

TABLE D2: ORDERING OF MARKET ENTRY DECISIONS WITH ALL CLINICAL TRIALS

	Dependent variable: Focal market size	
	(1)	(2)
Indication order	-0.0000426***	-0.0000708***
	(0.000000622)	(0.000000516)
Mean of dep. var	6.752	8.897
Observations	1,012,583	1,012,583
Controls:		
Initial approval year	no	yes
Indication group	no	yes
Competition	no	yes
Regulatory incentives	no	yes

NOTES: This table shows the relationship between indication order and market size for FDA-approved and unapproved cancer drugs with an initial trial between 1998 and 2021 (corresponding to the most complete coverage of our trials dataset). The level of observation is the drug-indication order. Both columns examine clinical trials, where the sample of clinical trials consists of all trials (not just those that are linked to FDA-approved drugs). The outcome variable is market size, measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size associated with indication order. We omit controls for intellectual property protection since this sample includes drugs developed both within and outside the U.S., where intellectual property protections vary. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.3 Cross-Disease Analysis

D.3.1 Controlling for Drug and Market Characteristics

TABLE D3: ORDERING OF REGULATORY APPROVALS ACROSS DISEASES, BY PROPENSITY FOR OFF-LABEL DRUG USE

	Dependent variable: Focal market size			
	Low off-label diseases		High off-label diseases	
	(1)	(2)	(3)	(4)
Indication order	-2.202***	-1.230***	-0.265***	0.0452
	(0.277)	(0.230)	(0.0617)	(0.0532)
Mean of Dep. Var.	16.044	16.124	13.524	13.575
Observations	152	140	650	598
Controls:				
Initial approval year	no	yes	no	yes
Indication group	no	yes	no	yes
Competition	no	yes	no	yes
Regulatory incentives	no	yes	no	yes

NOTES: This table shows the relationship between indication order for regulatory approvals and market size for approved drugs across several disease categories from 1998 to 2021, by propensity for off-label drug use. The level of observation is the drug-indication order. The outcome variable is market size, measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with focal indication order averaged over 1996 and 1997. The number of observations differs between Columns 1 and 2 (and between Columns 3 and 4) due to missing data on control variables. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.3.2 Controlling for ATC Classification

TABLE D4: ORDERING OF REGULATORY APPROVALS ACROSS DISEASES, BY PROPENSITY FOR OFF-LABEL DRUG USE, WITH ALTERNATIVE DISEASE CONTROLS

	Dependent variable: Focal market size	
	Low off-label diseases (1)	High off-label diseases (2)
Indication order	-1.212***	0.0532
	(0.157)	(0.0552)
Mean of Dep. Var.	16.124	13.575
Observations	140	598
Controls:		
Initial approval year	yes	yes
ATC classification	yes	yes
Competition	yes	yes
Regulatory incentives	yes	yes

NOTES: This table shows the relationship between indication order for regulatory approvals and market size for approved drugs across several disease categories from 1998 to 2021, controlling for each drug's ATC classification. The level of observation is the drug-indication order. There are 14 mutually exclusive ATC categories: alimentary tract and metabolism, anti-infectives for systemic use, antineoplastic and immunomodulating agents, antiparasitic products, insecticides and repellents, blood and blood forming clots, cardiovascular system, dermatologicals, genitourinary system and sex hormones, musculoskeletal system, nervous system, respiratory system, sensory organs, systemic hormonal preparations, and various. The outcome variable is market size, measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with focal indication order averaged over 1996 and 1997. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.3.3 Incorporating an Alternative Measure of Actual Off-Label Drug Use

We incorporate into the cross-disease analysis an alternative measure of actual off-label drug use using physician prescribing data from the IMS Health National Disease and Therapeutic Index^{TM} (NDTI^{TM}) from 2004 to 2009. The NDTI^{TM} provides data based on surveys of practicing physicians, allowing us to measure both on-label and off-label drug use. For each drug, we calculate the share of prescriptions that are off-label in a given year. To examine how off-label use shapes the indication order-market size relationship, we restrict our analysis to drug approvals over a given period and control for the share of off-label prescriptions in the year prior to the start of the analysis period. For example, Panel A in Table D5 focuses on the set of all drugs approvals starting in 2005. Column 1 shows the correlation between indication order and market size. Column 2 controls for the share of off-label prescriptions in 2004 and shows that the indication order-market size correlation is more negative after accounting for off-label drug use. We observe a similar pattern in subsamples beginning in subsequent years (2006, 2007, 2008, 2009, and 2010).

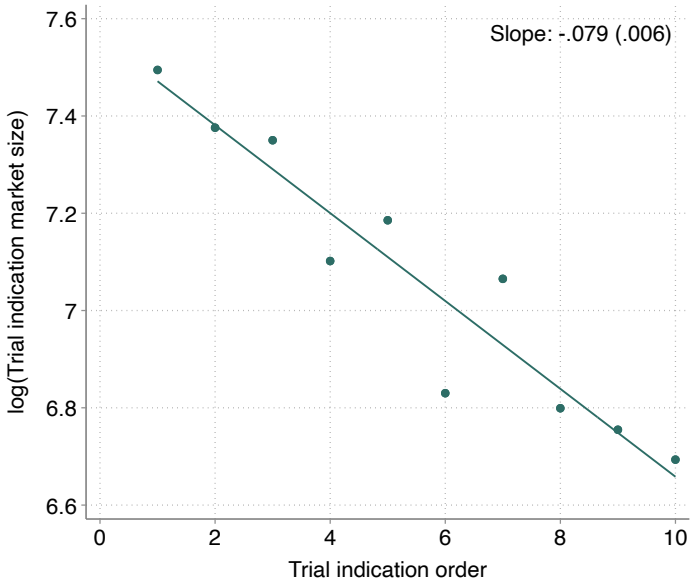
TABLE D5: ORDERING OF REGULATORY APPROVALS ACROSS DISEASES, ACCOUNTING FOR ACTUAL OFF-LABEL DRUG USE

	Dep. Var.: Total Market Size	
	(1)	(2)
Panel A. Drugs approved 2005+		
Indication order	-0.312***	-0.325***
	(0.0583)	(0.0594)
Mean of Dep. Var.	13.838	13.838
Observations	716	716
Panel B. Drugs approved 2006+		
Indication order	-0.308***	-0.332***
	(0.0591)	(0.0598)
Mean of Dep. Var.	13.796	13.796
Observations	680	680
Panel C. Drugs approved 2007+		
Indication order	-0.296***	-0.323***
	(0.0589)	(0.0595)
Mean of Dep. Var.	13.766	13.766
Observations	650	650
Panel D. Drugs approved 2008+		
Indication order	-0.317***	-0.337***
	(0.0610)	(0.0615)
Mean of Dep. Var.	13.724	13.724
Observations	618	618
Panel E. Drugs approved 2009+		
Indication order	-0.322***	-0.337***
	(0.0620)	(0.0623)
Mean of Dep. Var.	13.694	13.694
Observations	584	584
Panel F. Drugs approved 2010+		
Indication order	-0.297***	-0.318***
	(0.0614)	(0.0622)
Mean of Dep. Var.	13.619	13.619
Observations	547	547

NOTES: This table shows the relationship between indication order for regulatory approvals and market size for approved drugs across several disease categories. Each coefficient is from a separate regression. Regressions in Column 2 control for the share of off-label prescriptions in the year prior to the start of the analysis period, while those in Column 1 do not. The outcome variable is market size, measured by new diagnoses for an indication (ICD-9) in the MEPS data. We use the log of the market size associated with focal indication order averaged over 1996 and 1997. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

D.4 Trial End Dates

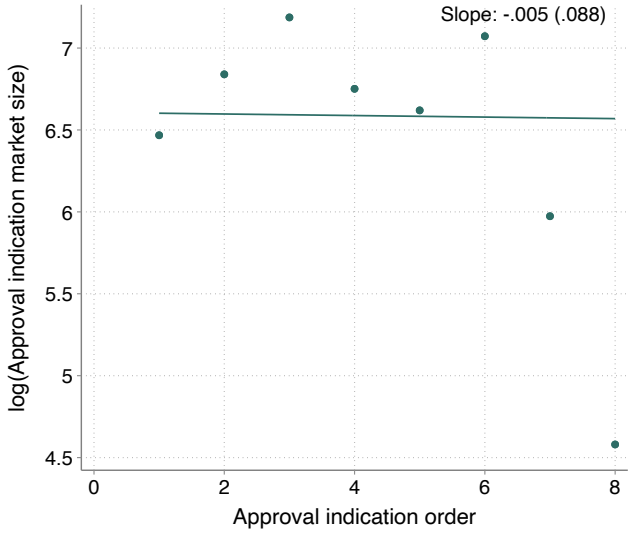
FIGURE D1: TRENDS IN CLINICAL TRIALS, WITH INDICATION ORDER DETERMINED BY TRIAL END DATES



NOTES: This figure shows trends in clinical trials for cancer drugs first approved from 1978 to 2016, with indication order determined by trial end dates. The level of observation is the drug-indication order. Market size is measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size. Each marker represents binned averages for a given indication order. For ease of intrepretation, we display up to the 10th indication.

D.5 Submission Dates

FIGURE D2: TRENDS IN REGULATORY APPROVALS, WITH INDICATION ORDER DETERMINED BY SUBMISSION DATES



NOTES: This figure shows trends in regulatory approvals for cancer drugs first approved from 1978 to 2016, with indication order determined by FDA submission dates. The level of observation is the drug-indication order. Market size is measured by new diagnoses for an indication in the SEER data. We use the log of the five-year average market size. Each marker represents binned averages for a given indication order.