NBER WORKING PAPER SERIES

THE MARKET-EXPANDING ROLE OF REGULATORY APPROVAL IN MEDICINE

Benjamin Berger
Amitabh Chandra
Craig Garthwaite

Working Paper 28889
http://www.nber.org/papers/w28889

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2021

We are grateful to David Cutler, Luca Maini, Ljubica Ristovska, Mark Shepard, Ariel Stern, Heidi Williams, and seminar participants at the Harvard Kennedy School for their helpful comments. We would also like to thank Tanya Natwick and Gerd Schweinitz immensely for their assistance with data access. We benefited from the expert research assistance of Lauren Mostrom, Kendra Singh, and Richard Sweeney. Amitabh Chandra is an OptumLabs® Visiting Fellow. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Market-Expanding Role of Regulatory Approval in Medicine
Benjamin Berger, Amitabh Chandra, and Craig Garthwaite
NBER Working Paper No. 28889
June 2021
JEL No. I1, I18, O3

ABSTRACT

Regulatory review is often seen as a barrier to innovation, increasing costs and delaying new medicines. Yet approval may also expand markets by certifying quality and reducing uncertainty. We test this by studying FDA approval for follow-on indications—uses that physicians could already prescribe "off-label"—and find approval raises use in newly approved diseases by 25 percent within a year, with larger increases in smaller off-label markets. Subsequent approvals in the same disease yield smaller gains. Our results suggest regulatory approval of medicines expands market size by increasing demand, rather than easing insurer-imposed restrictions, revealing limits to market-based learning.

Benjamin Berger
Harvard University
Kennedy School of Government
benjamin_berger@hks.harvard.edu

Amitabh Chandra
Harvard University
Department of Economics
and NBER
amitabh_chandra@harvard.edu

Craig Garthwaite
Northwestern University
Kellogg School of Management
and NBER
c-garthwaite@kellogg.northwestern.edu

Regulation is often framed as a trade-off between product quality and innovation: stricter standards raise the quality bar but can reduce innovation by increasing production costs and delaying market entry through lengthy approval processes (Peltzman 1973). Consequently, policy debates typically focus on whether the benefits of improved quality justify the costs of delay and higher prices. This framing, however, overlooks the potential for regulation to expand markets by serving as a credible signal of quality. For experience goods — products whose quality is difficult to assess before use — such certification can be essential. Without it, these products may generate limited revenue and support only small markets. The issue is particularly salient in the pharmaceutical industry, where manufacturers possess more information about drug quality than patients, physicians, or payers, but often struggle to communicate this credibly (Katz 2007, Dranove & Jin 2010). The lack of a trusted signal is especially problematic when market learning, through physician experience and experimentation, is slow. Given that pharmaceutical firms operate under time-limited market exclusivity, delays in demand growth can substantially reduce expected profits. Since R&D investments are based on expected profitability, smaller markets may deter innovation (Acemoglu & Linn 2004, Finkelstein 2004, Blume-Kohout & Sood 2013, Dubois et al. 2015, Dranove et al. 2020). These considerations suggest that regulatory approval may stimulate demand by providing a credible certification of quality, but the extent to which regulatory approval by agencies like the Food and Drug Administration (FDA) expands markets remains an empirical question.

Anecdotal evidence from the development of SARS-CoV-2 vaccines illustrates the potential market-expanding role of regulatory approval. In the summer of 2020, concerns arose that the FDA might succumb to political pressure and approve a vaccine without requiring the usual level of evidence on safety and efficacy (LaFraniere et al. 2020). Under a view of FDA review as a pure burden, firms should have welcomed reduced regulatory standards, for these would allow greater vaccine utilization.[1] Instead, leading

[1]Peltzman (1973) argues that market forces can sort out quality and that regulation can do

vaccine developers publicly pledged to seek the standard, more rigorous FDA review process (Lupkin 2020), consistent with an understanding that uncertainty about quality could dampen demand for pharmaceutical innovations, and that regulatory approval would increase market demand through a certification effect.

Ideally, we would estimate the demand-expanding effect of regulatory approval by examining primary approvals — those granted when a drug is first authorized for sale. But because drugs cannot be marketed before such approval, this counterfactual is unobservable. We address this challenge by studying follow-on indications — subsequent FDA approvals for additional uses of a drug after its initial launch. Follow-on approvals require rigorous clinical evidence similar to that of primary approvals. In our data, diseases covered by a drug's primary indication account for only 52 percent of prescriptions two quarters after approval, suggesting that nearly half of early use involves off-label prescribing. Because physicians are permitted to prescribe drugs off-label for any medically appropriate use, market learning can still occur without regulatory approval. If such learning were complete, we would observe little or no change in use following a follow-on indication. Many drugs gain follow-on indications over time (see Figure A.1), making this setting well-suited for studying whether regulatory approval expands utilization.

These follow-on indications come in two forms: some expand approved uses within a previously approved disease (e.g., approval for use as a second-line treatment for lung cancer), while others authorize use for entirely new diseases (e.g., approval for colon cancer after initial approval for lung cancer).[2] If learning is slower for diseases not previously

more harm than good by slowing innovation. Grabowski & Vernon (1983) suggest that post-market learning and physician expertise could substitute for pre-market review.

[2]Indications include information about the disease being treated, prevented, or diagnosed, and potentially additional circumstances relevant to use including disease stage, severity, subtype, biomarkers, age group, sex, genetic background, and previous or concomitant treatment.

associated with the drug, we would expect follow-on approvals in new diseases to have a larger effect on demand. Consider the case of Humira (adalimumab). In December 2002, it received primary approval for rheumatoid arthritis. Physicians could thereafter prescribe it off-label for other conditions. In October 2005, Humira received a follow-on approval for treating psoriatic arthritis — a distinct disease. Then in November 2006, it was approved for a second psoriatic arthritis indication, expanding its approved use within that disease. While both follow-on approvals concerned psoriatic arthritis, only the first extended the drug to a new disease. Our framework allows us to distinguish between these cases and assess how novelty affects utilization.

We begin by estimating the effect of follow-on approvals in new diseases using a stacked difference-in-differences model. We compare changes in drug use for newly approved diseases to use in control diseases for which the same drug receives approval later. For example, we compare the change in Humira fills for psoriatic arthritis following the October 2005 approval to a control group of later-approved indications. This within-drug, across-disease design has several advantages: it controls for drug-specific trends and avoids contamination from spillovers that might occur if other drugs were used as controls. For instance, if Humira receives approval for a disease, fills for competing drugs such as Remicade may decline, biasing estimates upward if used as controls.

We find that follow-on approvals in new diseases increase first fills by 25 percent in the year after approval. Most of this increase occurs within the first four quarters, with little further growth thereafter. While approval expands use dramatically in small off-label markets, it has more limited effects in larger markets, where prior learning may already be substantial.[3] Our identification strategy — using yet-to-be-approved diseases as controls — is likely conservative. Approval for one disease may inform perceptions of drug efficacy in related, unapproved conditions, raising use in the control group and biasing

[3]Ristovska (2024) similarly finds little expansion in use after approval for widely used pediatric indications.

our estimates toward zero. We discuss this issue further in Section 2.

Our second analysis examines whether additional approvals within a disease increase use. We find that the marginal effect of a second indication for a drug-disease pair is less than half the effect of the first, suggesting diminishing returns to additional information. Importantly, the effect of the first indication aligns closely with our difference-in-differences estimates, providing reassurance about the robustness of our approach. These results show that FDA approval expands pharmaceutical markets and highlight the importance of the novelty of information conveyed. We consider two possible mechanisms: increased demand from patients and physicians, or reduced insurer restrictions on off-label use (e.g., prior authorization or step therapy). We find that utilization increases regardless of whether prior utilization management was in place, suggesting that greater demand — rather than lower prescribing barriers — drives the observed effects.

In the following section, we describe our primary data assets. We combine detailed indication approval histories with health plan claims data to measure drug use by disease. Section 2 outlines our empirical approach. Section 3 presents our findings, and Section 4 discusses their broader economic and policy implications.

1 Data

1.1 Regulatory Approval of Indications

Our first data source encompasses all indications for new molecular entities and biologic drugs first approved by the FDA between 1995 and 2019, as listed on the FDA's Drugs@FDA website. We constructed a dataset by reviewing several thousand drug labels and FDA letters to manufacturers — publicly available through Drugs@FDA — to identify the clinical content and approval dates of each new indication. While documentation is generally complete for drugs originally approved after 1995, earlier approvals often lack sufficient information. Because our empirical analysis depends on accurately identifying

approval histories, we restrict our sample to indications for new molecular entities and biologics approved from 1995 onward. For each drug, we read all FDA letters to manufacturers to determine whether and when a new indication was granted. Through this process, we identified 1,552 indications across 784 drugs.

While follow-on indications are fairly common, they are concentrated among a subset of drugs. Only 39 percent of drugs have at least one follow-on indication, and the drugs with many such approvals are often blockbuster therapies. For example, Keytruda (pembrolizumab) alone received 21 follow-on indications by 2019, just five years after its initial approval.[4]

1.2 Mapping Indications to Diseases

Ideally, we would conduct our analysis at the level of drug-indication pairs. However, (to our knowledge) no dataset that measures utilization reports the clinical indication for which a drug is prescribed. We therefore perform our analysis at the drug-disease level because claims data permit no finer level of granularity. To do this, we manually assigned diseases to each indication by reviewing the associated text. We then mapped each disease to one or more diagnosis codes using the WHO's ICD code descriptions (see Section B.1 for details). We excluded indications that did not represent therapeutic use, those that could not be mapped to diseases, and diseases for which we could not assign a valid ICD-10 code. The resulting dataset contains 1,137 drug-disease pairs across 706 drugs.

1.3 Measuring Use of Medicines

Our second data asset is de-identified administrative claims from the OptumLabs® Data Warehouse, which allows us to observe drug use across patient groups with different diseases. This dataset includes medical and pharmacy claims, laboratory results, and enroll-

[4]See Figure A.2.

ment records for commercially insured and Medicare Advantage enrollees from 1993 to 2019. It offers longitudinal health information on a diverse U.S. population, spanning a range of ages, ethnicities, and geographic regions.

To measure utilization, we count the number of first prescription fills — defined as a patient's first recorded claim for a given drug — by drug-disease pair, year, and quarter. This focus on first fills captures how prescribing to new patients evolves following indication approval and avoids confounding from inertia in ongoing medication use. We identify relevant diagnoses using all ICD codes recorded in the 180 days prior to each beneficiary's first fill and assign each fill to all diseases with matching diagnoses. We exclude beneficiary-drug pairs if the beneficiary was not continuously enrolled throughout this lookback window.

Our data-use agreement requires us to censor all cells with between 1 and 10 individuals. In our primary specification, we assume censored cells contain 4 fills — the average value of such cells. We demonstrate that this assumption is reasonable in Appendix Table A.1 and Figure A.3.

Table 1 presents summary statistics at the drug-disease level, our core unit of analysis. Of the 1,137 pairs, 821 (72 percent) were first approved via primary indications, while 316 were first approved through follow-on indications. Because primary indications mark a drug's entry to market, there is no pre-approval data on their use, and they are excluded from our identification strategy. As expected, all primary approvals are for new diseases. Follow-on indications are similar in most characteristics, though they are more likely to pertain to cancer treatments. Most drug-disease pairs (81 percent) have only one approved indication, but 14 percent gain a second within-disease indication and 5 percent gain two or more. Table 1 also summarizes characteristics relevant to our difference-in-differences analysis: 20–25 percent of drugs were subject to utilization management, and over 40 percent were cancer therapies. We describe the construction of the estimation sample in the next section.

2 Identifying the Effect of Regulatory Approvals on Use

We identify the effect of FDA indication approvals on drug use with a difference-in-differences design. The approach compares changes in first fills for drug-disease pairs that receive a follow-on approval (treated pairs) to changes in matched control pairs that are not affected by the approval. The key challenge for identification lies in selecting control pairs that plausibly represent the counterfactual trajectory of treated pairs had the approval not occurred.

2.1 Defining Control Groups

One intuitive approach might compare different drugs within the same disease — for example, using Stelara or Cimzia as controls for Humira's Crohn's disease approval. However, this strategy has several limitations. First, control drugs may not yet be on the market at the time of the approval. For instance, Stelara and Cimzia were not available when Humira was approved for Crohn's disease in Q1 2007. Second, even when such drugs are available, using them as controls risks bias due to substitution. If patients switch from Remicade to Humira after the latter receives an indication, both the increase in Humira fills and the decline in Remicade use would reflect the approval's effect, overstating the treatment effect. Comparisons across drugs also raise concerns about unobserved differences in market dynamics or prescribing behavior.

We instead adopt a within-drug, across-disease design. For each treated drug-disease pair, we use the same drug in diseases that have not yet — but eventually will — receive follow-on approval. For example, to evaluate Humira's Crohn's disease approval in Q1 2007, we use Humira in ulcerative colitis (approved Q3 2012) and other not-yet-approved uses as controls. This approach ensures that controls were prescribable at the time of treatment and reduces bias from spillovers or market entry. Because control pairs involve the same drug, they avoid differences in product characteristics. Because they involve

different diseases, they are less likely to experience cross-disease substitution.

Figure 1 illustrates our procedure using Humira's approval history. Panel A shows the selection of controls for its Q3 2006 approval in ankylosing spondylitis (AS), and Panel B shows the analogous process for its Crohn's disease approval. In each case, we analyze utilization from 8 quarters before to 4 quarters after the approval (shaded region). Diseases already approved by the start of the window (e.g., rheumatoid arthritis and psoriatic arthritis in Panel A) are excluded from the control group, while future approvals (e.g., ulcerative colitis, uveitis) are eligible. Similarly, we exclude diseases with approvals during the post-treatment window, as in Panel B, where plaque psoriasis and juvenile idiopathic arthritis are excluded due to overlapping timing.

Our approach is not without issue. Drug-disease pairs that lack eligible controls — either because they are primary indications or because no suitable future approvals exist — are dropped. This greatly reduces our sample size and reduces the generalizability of our results because all included drugs must have at least three approved diseases: one primary disease, one treated disease that receives the first follow-on indication, and one or more control diseases that also receive follow-on indications. Moreover, if approval in one disease spills over to related diseases for the same drug, such within-drug information spillovers would bias estimates. However, the bias would attenuate estimates toward zero rather than inflating them.

The most significant identification challenge this study faces is that it cannot fully disentangle the effects of FDA review and the release of clinical trial information. We cannot fully separate these channels because: (a) clinical trial results often become public around the time of FDA submission/approval, and (b) we lack comprehensive data on trial publication and readout dates.[5] We note that this means our estimates represent an upper

[5]Moreover, the effects of publication likely vary substantially over time. Controlling for fully dynamic effects of trial publication would absorb many of our remaining degrees of freedom, which are already limited due to our small sample.

bound on the pure certification effect of FDA review. However, we argue this combined effect is still policy-relevant because the FDA review process itself influences the quality and design of trials that manufacturers conduct, so the two are not truly separable in practice. Additionally, our event studies show substantial increases after approval which are unlikely to be driven by new clinical study results in isolation because those results are submitted at the beginning of FDA review which typically lasts 6–12 months (Downing et al. 2012).

Our final sample includes 70 treated drug-disease pairs (22% of all follow-on indications) matched to valid controls. The characteristics of this estimation sample are reported in the final two columns of Table 1. Treated pairs tend to be for drugs of earlier vintages than those in either the full sample or the sample of all pairs approved as follow-on indications. Treated pairs are also more likely to be for cancer drugs than pairs in the full sample (40% vs. 27%) but no more likely than pairs in the sample of all follow-on pairs (40%). The certification effect of FDA approval could be especially impactful for cancer drugs if patients are reluctant to try unapproved cancer therapies, but given widespread off-label use of cancer drugs, market learning could diminish the certification effect of approval. Lastly, treated pairs tend to have more first and total fills in the quarter of their approval than the sample of all follow-on pairs. This could reflect greater off-label market learning in the estimation sample than in the larger sample, suggesting that the effect we estimate may be smaller than what we would expect in a more general sample of indications.

2.2 Estimation

We estimate stacked difference-in-differences and event study models, following the regression-based method described by Wing et al. (2024). We rely on stacked models because they allow for precise and flexible definition of control units and are not susceptible to the negative weight bias of two-way fixed effects. To estimate these models, we index the $k = 1, \ldots, 70$ indicated drug-disease pairs in our sample, treating each as a separate approval

event to study (these approval events are conceptually identical to the "sub-experiments" referenced in the econometric literature). For each approval event, we create a panel dataset comprising first fills of the approved pair and its controls in each of the 13 quarters beginning 8 quarters prior to the approval and ending 4 quarters after.[6]

$$\text{First Fills}_{idtk} = \alpha_{idk} + \lambda_{itk} + \beta_{DD}\text{Treat}_{idk} \times \text{Post}_{tk} + \beta_0 \text{Treat}_{idk} \times \mathbb{1}\{E_k = t\} + X'_{idtk}\Gamma + u_{idtk} \quad (1)$$

We stack the datasets together to create a single stacked dataset, and model $\text{First Fills}_{idtk}$, the number of first fills of drug i for disease d in year and quarter t and approval event k in Equation 1.[7] $\text{Treat}_{idk} = 1$ if drug-disease (i, d) is the approved pair in approval event k and 0 if it serves as a control pair. E_k is the calendar year and quarter of approval event k, so that $\text{Post}_{tk} = \mathbb{1}\{E_k > t\}$ indicates whether quarter t falls after E_k. The difference-in-differences regression includes drug-disease fixed effects for each dataset α_{idk}, drug-specific time trends for each dataset λ_{itk}, and possibly one or more controls X_{idtk}.[8] The coefficient of interest β_{DD} corresponds to the additional change in first fills of newly approved drug-disease pairs relative to the change in their controls in the 4 quarters after approval. Under the assumption that indicated diseases would have followed the same trend as their controls had they not been approved, β_{DD} identifies the average effect of receiving an indication. We also include a coefficient β_0 that corresponds to the same difference in the quarter of approval, but do not report our estimates of this parameter because approved pairs are only partially treated in the quarter of approval (i.e. some

[6] See Figure A.4 for a case study of Humira that illustrates this approach.

[7] We discuss our choice to model fills in levels rather than logarithms in Section B.3.2, and present estimates of a $\log(\text{Fills} + 1)$ specification in the appendix.

[8] Notably, we do not explicitly control for drug age fixed effects. This is because they are perfectly collinear with λ_{itk} and thus provide no additional information to the regression.

indications occur at the end of the quarter of approval).

$$\text{First Fills}_{idtk} = \alpha_{idk} + \lambda_{itk} + \sum_{\ell \neq -1} \beta_\ell \text{Treat}_{idk} \times \mathbb{1}\{t - E_k = \ell\} + u_{idtk} \tag{2}$$

To test the parallel trends assumption, we also estimate the event study specification in Equation 2, which includes a coefficient for each relative time period so that β_ℓ can be interpreted as the average additional change in fills from the quarter prior to approval to quarter ℓ among approved drug-disease pairs relative to the change in their controls.

As shown by Wing et al. (2024), ordinary least squares estimates of Equations 1 and Equations 2 do not consistently estimate average causal effects because treated and control trends receive different implicit weights. We follow their recommendation and estimate both equations by weighted least squares with corrective sample weights.[9] In all models we cluster standard errors at the drug level.

2.3 Effect of Subsequent Follow-on Indications on Use

Next, we develop a framework for assessing whether additional indications increase use at a decreasing rate. The difference-in-differences model abstracts away from this possibility and reports the effect of only the first indication for each drug-disease pair. However, drugs are often approved for multiple uses within a single disease, and the effect of an additional indication on a drug-disease pair may depend on the cumulative information provided by all indications for that pair.[10]

To examine how the marginal effects of additional indications vary, we use our full sample of 1,137 drug-disease pairs to estimate Equations 3–5. These regressions model fills of drug i for disease d in year-and-quarter t as functions of the number of approved indications for that drug-disease pair at that time (Indications_{idt}), drug-disease pair fixed

[9] See Section B.3.4 for details on calculating these weights.

[10] This is also supported by graphical evidence. See Figure A.5.

effects, drug-specific time trends, and covariates X_{idt}, which allow us to control for time trends within disease area. Like the difference-in-differences model, these models use within-drug variation to identify the impact of receiving an indication, but assume that indications enter with particular parametric forms. We consider three ways in which the subsequent indications might affect fills:

$$\text{First Fills}_{idt} = \alpha_{id} + \lambda_{it} + X'_{idt}\Gamma + \gamma \mathbb{1}\{\text{Indications}_{idt} \geq 1\} + u_{idt} \tag{3}$$

$$\text{First Fills}_{idt} = \alpha_{id} + \lambda_{it} + X'_{idt}\Gamma + \delta(\text{Indications}_{idt} - 1) + u_{idt} \tag{4}$$

$$\text{First Fills}_{idt} = \alpha_{id} + \lambda_{it} + X'_{idt}\Gamma + \beta_1(\text{Indications}_{idt} - 1) + \beta_2(\text{Indications}_{idt} - 1)^2 + u_{idt} \tag{5}$$

In Equation 3, indications enter through a dummy variable that equals 1 if the drug-disease pair has any approved indications at time t and 0 otherwise. In this model, γ captures the association between first fills and having any approved use, controlling for fixed effects and covariates. In Equation 4, indications enter linearly so that δ captures the average association of an additional indication with first fills, controlling for the same factors (subtracting 1 from Indications_{idt} has no effect on δ but is useful for interpreting the coefficients in Equation 5). Lastly, in Equation 5, we estimate a quadratic specification that allows the marginal effect of an additional indication to vary. In this model, the marginal effect of an additional indication is given by $\frac{\partial \text{First Fills}_{idt}}{\partial \text{Indications}_{idt}} = \beta_1 + 2\beta_2(\text{Indications}_{idt} - 1)$. Therefore, the marginal effect of the first indication is given by β_1 and the marginal effect of the second indication is given by $\beta_1 + 2\beta_2$. A t-test of the null hypothesis $\beta_2 = 0$ is sufficient to test whether the first and second indications have the same marginal effect.

3 Results

Figure 2 presents estimates from the event study regression. Trends in first fills for treated and control pairs do not diverge significantly before approval, suggesting that the parallel

trends assumption is likely satisfied and that bias from its violation is minimal.[11] Table 2 then reports difference-in-differences results. Column 1 assumes that treated and control pairs would have followed identical trends in the absence of treatment. Column 2 allows for differential linear trends between the two groups by including an interaction between treatment status and calendar time. The results are similar: we estimate that approval increases first fills by 88 per quarter, or 353 in the year after approval — a 25 percent increase over the counterfactual.[12] Extending the post-treatment window to eight quarters instead of four does not change the magnitude of the estimated effect.[13]

Columns 3–5 of Table 2 report estimates from Equations 3–5, which use different specifications to examine how the marginal effect of each additional indication varies. In each, we control for sub-chapter-specific time trends by including interactions between ICD-10 sub-chapter indicators and calendar quarter. Column 3 shows that having any approved indication is associated with 91.7 additional first fills per quarter ($p < 0.1$), controlling for drug-disease pair and time fixed effects. Column 4 estimates a slightly smaller but still

[11] The imprecise negative coefficients observed in quarters -4 through -2 are consistent with potential anticipatory effects in which utilization first decreases prior to release of clinical trial results and then increases as those results are revealed. While we could potentially control for clinical trial publications or readouts, we do not because we expect the impact of trials on utilization to be highly dynamic. Allowing for flexible trial timing controls in addition to the existing interactive fixed effects would likely leave us with too little identifying variation given our sample size.

[12] This 25 percent figure reflects the percent increase in average fills, not the average percent increase across pairs. If approvals have smaller effects in high-volume off-label markets and larger effects in low-volume ones, the overall average percent increase may understate the typical effect.

[13] Figure A.6 shows that nearly all learning through regulatory approval occurs within the first year, with little change in use thereafter.

significant coefficient on the number of approved indications, consistent with diminishing returns to subsequent indications. Column 5 reports the quadratic specification from Equation 5. The estimated effect of the first indication — 94 additional fills — is nearly identical to the difference-in-differences estimate of 88. The coefficient on the quadratic term (β_2) is negative and statistically significant, indicating decreasing marginal returns. The estimated marginal effect of the second indication is 44 fills — less than half that of the first.

3.1 Robustness and Alternative Specifications

We assess robustness through several alternative specifications, reported in the appendix. First, we show that our difference-in-differences estimates are consistent under alternative transformations of the dependent variable (Table A.2). Using $\log(\text{First Fills} + 1)$ or the inverse hyperbolic sine transformation yields an estimated effect of about 16 percent, which, while smaller than our preferred 25 percent estimate, is economically similar.[14]

To assess the effect of censoring cells with fewer than 11 observations, we conduct sensitivity analyses. In our main specification, we impute censored cells as 4, the average value in those cells. In Table A.1, we show that results are unchanged when using imputed values of 1 or 10. In Figure A.3, we present results from 1,000 simulations replacing censored cells with random values drawn from a conditional Poisson distribution. The distribution of estimates is tight, supporting the robustness of our approach.

We also find that the percent increase in fills after approval is inversely related to the size of the off-label market. Figure A.7 shows that small off-label markets often expand severalfold, while large ones expand very little. This pattern suggests that market learning can substitute for regulatory approval in larger markets, whereas in smaller markets, the certification function of approval plays a more important role.

[14]These specifications are widely used despite critiques (see e.g. Chen & Roth (2024)).

3.2 Mechanisms: More Demand or Less Utilization Management?

Two mechanisms could explain the observed increase in utilization after approval: increased demand from patients and physicians, or decreased utilization management by insurers.[15] Both imply a market-expanding role for regulatory approval, but they operate through different channels. To distinguish between them, we compare utilization responses across drugs with differing levels of access restrictions.[16]

We use formulary data from Managed Markets Insight & Technology (MMIT) to measure the extent of utilization management. The data cover self-administered drugs with at least one follow-on indication approved between 2013 and 2019. For each drug, we compute the share of insured individuals subject to utilization restrictions and classify drugs as having "high" utilization management if this share exceeds 50 percent.

Figure 3 shows that first fills increase substantially in the four quarters following approval, regardless of whether a drug faces high or low utilization management. While baseline utilization is lower for high-restriction drugs, the proportional increase in use after approval is similar across both groups. The magnitude of these changes exceeds pre-approval trends, indicating that they are not driven by underlying differences in growth trajectories. These findings suggest that regulatory approval increases demand, even in the presence of access barriers, by providing valuable information to market participants.[17]

[15] Increased demand could result from greater use by existing patients or expansion of the patient population by greater diagnosis. The latter channel is interesting and may be impactful in certain areas of medicine but is unlikely to substantially drive our results due to our cancer-heavy sample.

[16] Prior authorization requires providers to obtain insurer approval before prescribing. Step therapy mandates that patients first try and fail other therapies before coverage is granted.

[17] Table A.3 shows that approval is associated with a 5 percentage point increase in utilization management share, which is modest and may reflect reverse causality (i.e., strategic

4 Discussion

Our central contribution is to demonstrate that regulatory approval — traditionally understood as a barrier to entry — can also serve a market-expanding function by credibly certifying product quality, especially in pharmaceutical markets with substantial informational frictions about the quality of a product. Even in a setting where physicians could have learned the value of a medicine through off-label prescribing, we find that receiving a follow-on indication increases utilization by 25 percent in the year following approval, and that subsequent approvals have diminishing marginal effects within the same disease. The observed increases in fills are immediate and complete within one year — faster than would be expected if driven solely by marketing efforts like detailing or advertising. Moreover, the effects are larger for approvals in new diseases, where the information is likely to be more novel. This pattern implies that the effect is not merely due to representatives informing the same physicians about expanded use, but instead reflects new knowledge in new clinical contexts. This suggests that loosening evidentiary requirements may reduce time to market, but could also dilute the informational content of approval, reducing its value as a quality signal.

Our results further suggest that physician financial incentives—though documented to influence prescribing behavior (Clemens & Gottlieb 2014, Iizuka 2012, Carey et al. 2021)—are insufficient to drive market-based discovery of new therapeutic uses. Even pharmaceutical detailing, which can expand off-label prescribing when physicians receive new information (Shapiro 2018), does not substitute for formal regulatory review. The substantial increase in utilization only after formal approval indicates that information frictions dominate financial incentives in determining adoption, underscoring the FDA's distinct value in certifying quality.

These findings build on a growing literature examining how FDA regulatory decisions

indication-seeking by makers of heavily restricted drugs).

influence adoption and utilization. Prior work has shown that FDA actions — such as safety warnings, label expansions, and approvals — can meaningfully shape clinical behavior. For example, Ristovska (2024) finds limited changes in pediatric drug use following label changes, suggesting that when off-label use is already widespread, the marginal informational value of approval may be limited. Other studies have focused on regulatory restrictions and shown that narrowing opioid labels significantly reduced prescribing (Ody & Schmitt 2019). These findings highlight how label changes can influence utilization even when products remain legally available, but in this case, regulation constrained demand.

By contrast, McKibbin (2023) examines the adoption of FDA-approved medical devices and finds that approval facilitates hospital uptake, especially when safety concerns are initially high. Like our study, McKibbin highlights the informational role of regulation in settings with expert users, but his focus is on initial approval rather than follow-on indications. Similarly, Grennan & Town (2020) explore how regulatory pathways and uncertainty affect the adoption of new medical devices. They show that diffusion patterns are shaped by both regulatory risk and downstream provider incentives, emphasizing the interplay between uncertainty and access. Our contribution is distinct in that we study drugs that were already legally available for off-label use, allowing us to isolate the market-expanding effects of certification in a setting where access is held constant.

Our paper also contributes to a broader economic literature on quality certification and signaling in markets with asymmetric information. Spence (1975) developed foundational theory on how firms may use costly signals to communicate quality under monopoly, a framework that underlies much of the subsequent literature on disclosure and certification. Dranove & Jin (2010) provides a comprehensive review of disclosure and certification mechanisms, arguing that when signals are credible and costly, they can meaningfully influence demand. As an example, Jin & Leslie (2003) shows that restaurant hygiene grades affect consumer behavior. Our context is related to this literature but differs from other

markets such as cybersecurity audits, energy-efficiency labels, and financial product ratings in several ways: pharmaceutical products are experience or credence goods where learning about quality is often slow or incomplete; consumers are in a high-stakes situation, where their willingness to pay for extra quality is high (reflecting the high willingness to pay for life) but also distorted by insurance and the prescribing decisions of doctors; and the FDA serves as a centralized, widely trusted certifier, whose imprimatur is necessary for a medicine to be sold and marketed. These features amplify the value of certification in our setting, and explain the substantial increases in utilization we observe following approval, relative to restaurants, energy-labels, or ISO/ANSI certification, where consumers may respond differently depending on their ability to assess quality independently and the credibility of the certifier. Still, our findings reinforce the core insight from this broader literature: that certification can function not merely as a regulatory barrier, but as a powerful demand-side intervention.

References

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

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

Carey, C. M., Lieber, E. M. & Miller, S. (2021), 'Drug firms' payments and physicians' prescribing behavior in Medicare Part D', *Journal of Public Economics* **204**, 104532.

Chen, J. & Roth, J. (2024), 'Logs with zeros? Some Problems and Solutions*', *The Quarterly Journal of Economics* **139**(2), 891–936.

Clemens, J. & Gottlieb, J. D. (2014), 'Do physicians' financial incentives affect medical treatment and patient health?', *American Economic Review* **104**(4), 1320–1349.

Downing, N. S., Aminawung, J. A., Shah, N. D., Braunstein, J. B., Krumholz, H. M. & Ross, J. S. (2012), 'Regulatory review of novel therapeutics—comparison of three regulatory agencies', *New England Journal of Medicine* **366**(24), 2284–2293.

Dranove, D., Garthwaite, C. & Hermosilla, M. I. (2020), Expected profits and the scientific novelty of innovation, Technical report, National Bureau of Economic Research.

Dranove, D. & Jin, G. Z. (2010), 'Quality disclosure and certification: Theory and practice', *Journal of Economic Literature* **48**(4), 935–63.

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

Finkelstein, A. (2004), 'Static and dynamic effects of health policy: Evidence from the vaccine industry', *The Quarterly Journal of Economics* **119**(2), 527–564.

Grabowski, H. G. & Vernon, J. M. (1983), The regulation of new drug introductions in the united states, *in* D. F. Bradford, ed., 'Regulation', Brookings Papers on Economic Activity: Microeconomics, Brookings Institution, Washington, DC, pp. 363–407.

Grennan, M. & Town, R. J. (2020), 'Regulating innovation with uncertain quality: Information, risk, and access in medical devices', *American Economic Review* **110**(1), 60–84.

Iizuka, T. (2012), 'Physician agency and adoption of generic pharmaceuticals', *American Economic Review* **102**(6), 2826–2858.

Jin, G. Z. & Leslie, P. (2003), 'The effect of information on product quality: Evidence from restaurant hygiene grade cards', *Quarterly Journal of Economics* **118**(2), 409–451.

Katz, A. (2007), 'Pharmaceutical lemons: Innovation and regulation in the drug industry', *Mich. Telecomm. & Tech. L. Rev.* **14**, 1.

LaFraniere, S., Thomas, K., Weiland, N., Baker, P. & Karni, A. (2020), 'Scientists Worry About Political Influence Over Coronavirus Vaccine Project', *The New York Times* .

Lupkin, S. (2020), '9 Companies Pledge Not To Submit Coronavirus Vaccine To FDA Until It's Proved Safe', https://www.npr.org/2020/09/08/910788743/9-companies-pledge-not-to-submit-coronavirus-vaccine-to-fda-until-its-proved-saf.

McKibbin, C. (2023), The role of regulatory approval in the diffusion of medical technology: Evidence from FDA review. Working paper.
URL: *https://callummckibbin.com/research*

Ody, C. & Schmitt, M. (2019), 'The effect of fda regulation of abuse-deterrent opioids on opioid utilization', *Journal of Health Economics* **65**, 20–33.

Peltzman, S. (1973), 'An evaluation of consumer protection legislation: The 1962 drug amendments', *Journal of Political Economy* **81**(5), 1049–1091.

Radley, D. C., Finkelstein, S. N. & Stafford, R. S. (2006), 'Off-label prescribing among office-based physicians', *Archives of internal medicine* **166**(9), 1021–1026.

Ristovska, L. (2024), 'Drug adoption under uncertain quality and impact of FDA (de)certification'.

Shapiro, B. T. (2018), 'Informational shocks, off-label prescribing and the effects of physician detailing', *Management Science* **64**(12), 5925–5945.

Spence, M. (1975), 'Monopoly, quality, and regulation', *Bell Journal of Economics* **6**(2), 417–429.

Wing, C., Freedman, S. M. & Hollingsworth, A. (2024), Stacked difference-in-differences, Technical report, National Bureau of Economic Research.

Table 1: Indicated Diseases for Drugs Approved 1995–2019

	All Drug-Disease Pairs			Diff-in-Diff Pairs	
	All Pairs (1)	Approved as Primary (2)	Approved as Follow-on (3)	Treated (4)	Control (5)
Characteristic	**N = 1,137**	**N = 821**	**N = 316**	**N = 70**	**N = 130**
First approved as primary	821 (72%)	821 (100%)	0 (0%)	0 (0%)	0 (0%)
First approved as follow-on	316 (28%)	0 (0%)	316 (100%)	70 (100%)	130 (100%)
Indications					
1	916 (81%)	654 (80%)	262 (83%)	58 (83%)	120 (92%)
2	161 (14%)	122 (15%)	39 (12%)	11 (16%)	9 (6.9%)
3+	60 (5.3%)	45 (5.5%)	15 (4.7%)	1 (1.4%)	1 (0.8%)
Drug cohort					
1995-1999	324 (28%)	223 (27%)	101 (32%)	28 (40%)	37 (28%)
2000-2004	216 (19%)	140 (17%)	76 (24%)	17 (24%)	44 (34%)
2005-2009	154 (14%)	109 (13%)	45 (14%)	13 (19%)	16 (12%)
2010-2014	210 (18%)	142 (17%)	68 (22%)	12 (17%)	33 (25%)
2015-2019	233 (20%)	207 (25%)	26 (8.2%)	0 (0%)	0 (0%)
New ICD-10 sub-chapter	1,037 (91%)	821 (100%)	216 (68%)	55 (79%)	71 (55%)
New ICD-10 chapter	928 (82%)	821 (100%)	107 (34%)	26 (37%)	27 (21%)
Cancer drug	311 (27%)	185 (23%)	126 (40%)	28 (40%)	66 (51%)
Utilization management					
High util. mgmt.	96 (8.4%)	42 (5.1%)	54 (17%)	15 (21%)	33 (25%)
Low util. mgmt.	96 (8.4%)	35 (4.3%)	61 (19%)	15 (21%)	38 (29%)
Unknown	945 (83%)	744 (91%)	201 (64%)	40 (57%)	59 (45%)
First fills	79 (635)	17 (123)	241 (1,174)	463 (1,917)	228 (1,169)
Total fills	500 (5,974)	20 (148)	1,746 (11,246)	3,371 (15,739)	2,844 (18,022)

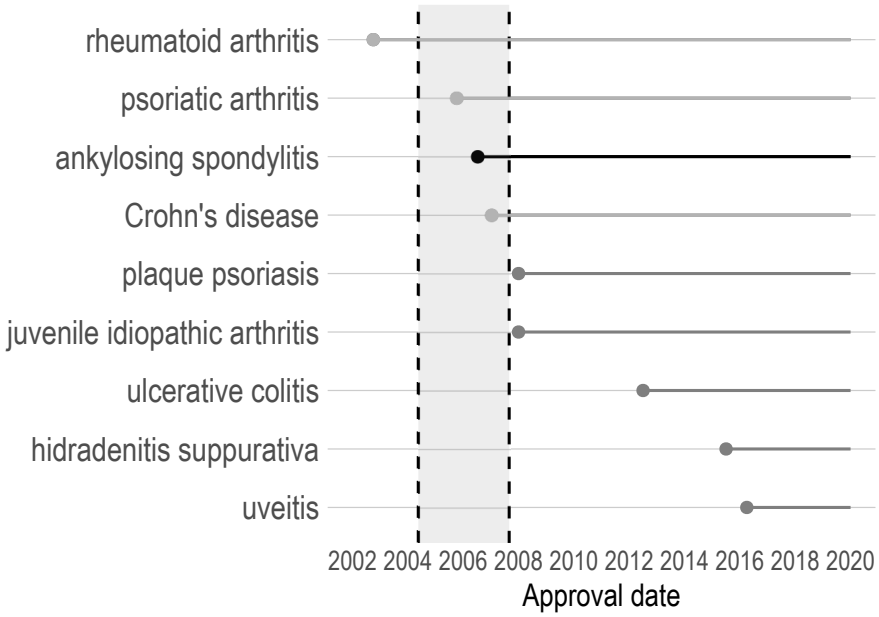
This table reports summary statistics for the universe of approved drug-disease pairs for drugs initially approved between 1995 and 2019. It reports totals and percentages for categorical variables and means and standard deviations for continuous variables. We include only indications that are approved for a therapeutic use in the sense of treating an underlying condition. Control pairs in the differences-in-differences sample are counted multiple times if they serve as controls to multiple treatment pairs. Utilization management is flagged as "high" for all indications of self-administered drugs that have an average of 50% or higher beneficiary-weighted share of prior authorization or step therapy between 2013 and 2019. Fills are reported in the quarter of approval for each drug-disease pair.

Table 2: Regression Estimates

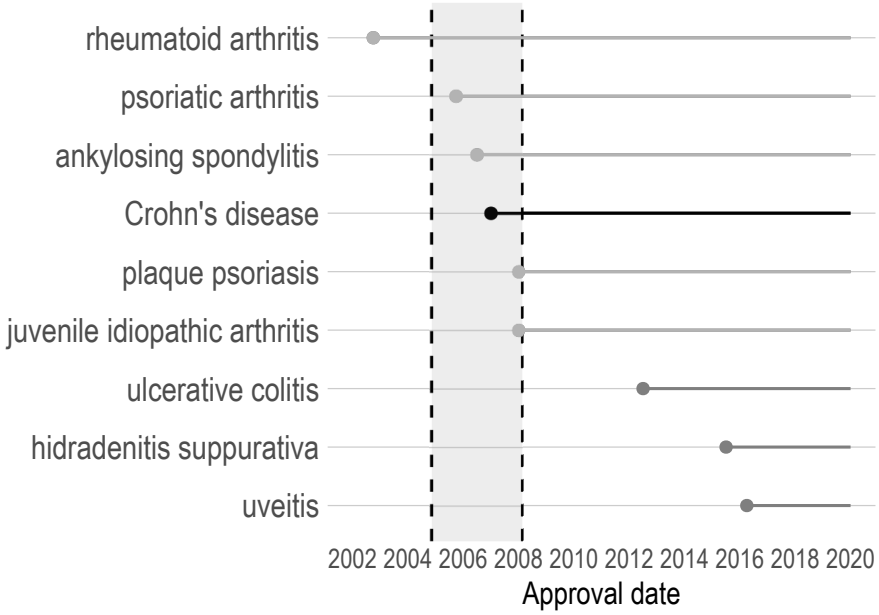
	DiD		Fixed Effects		
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Treat × Post	78.3†	88.1*			
	(40.1)	(35.0)			
Treat × Quarters relative to approval		-1.41			
		(4.50)			
Has approved indication			91.7†		
			(48.5)		
Indications - 1				78.0*	93.6*
				(39.4)	(41.9)
(Indications - 1)^2					-24.8*
					(12.2)
Fixed-effects					
Event-Drug-Disease	Yes	Yes			
Event-Year-Quarter	Yes	Yes			
Drug-Disease			Yes	Yes	Yes
Drug-Year-Quarter			Yes	Yes	Yes
ICD-10 Sub-chapter-Year-Quarter			Yes	Yes	Yes
Fit statistics					
Observations	2,600	2,600	60,643	60,643	60,643
Indicated pairs	70	70			
Control pairs	130	130			
Unique pairs	121	121	1,137	1,137	1,137
Dependent variable mean	294.1	294.1	268.5	268.5	268.5
Counterfactual mean	365.8	355.9			
R^2	0.996	0.996	0.933	0.933	0.933

Clustered (Drug) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, †: 0.1*

This table reports estimates of Equation 1 and Equations 3–5. In all models, the dependent variable is the number of first fills of drug i for disease d in calendar quarter t. Columns 1 and 2 report estimates of β_{DD}, the additional change in first fills among approved diseases in the 4 quarters after approval, relative to the change among control diseases. We estimate this coefficient using the weighted least squares method described by Wing et al. (2024), which reweights control observations to match the distribution of event times among treated units. Column 2 adds a control that allows fills for approved diseases to differ from control diseases up to a linear trend. Columns 3–5 report estimates of the fixed effects models coefficients.



(A) Ankylosing Spondylitis (black) and Controls (dark gray)



(B) Crohn's Disease (black) and Controls (dark gray)

Figure 1: Control Disease Selection for Humira (adalimumab)

This figure depicts control pair selection for the drug Humira for two different diseases, ankylosing spondylitis (Panel A) and Crohn's disease (Panel B). Control pairs are selected only out of other diseases later approved for Humira. Control pair candidates are further omitted from the set of controls if they are approved any time within the shaded 13-quarter window. In Panel A, Crohn's disease is omitted as a control for ankylosing spondylitis. In Panel B, plaque psoriasis and juvenile idiopathic arthritis are omitted as controls for Crohn's disease.

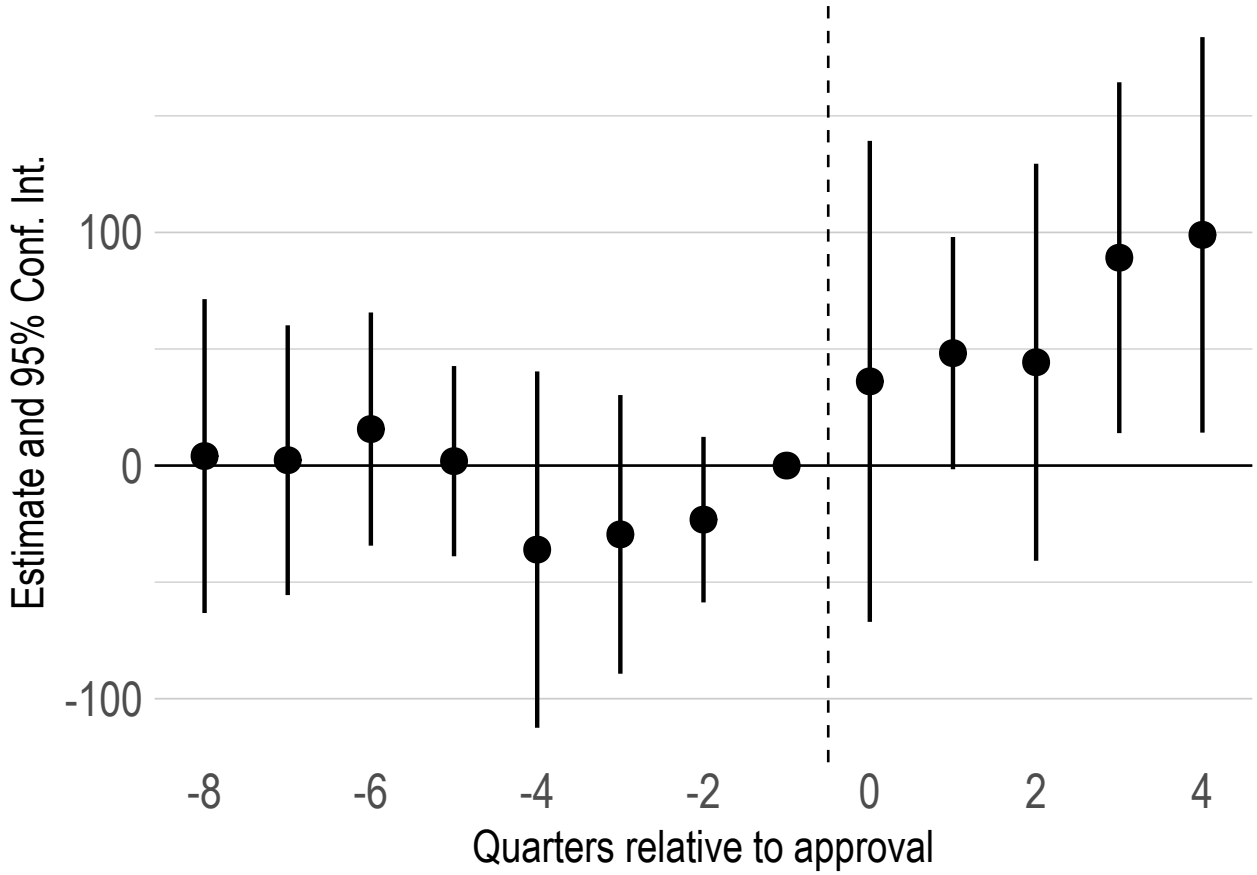


Figure 2: Event Study Estimates, First Fills

This figure depicts point estimates and 95% confidence intervals for the event study coefficients in Equation 1. We estimate this coefficient using the weighted least squares method described by Wing et al. (2024), which reweights control observations to match the distribution of event times among treated units. The dependent variable is the number of first fills of drug i for disease d in calendar quarter t. The 95% confidence intervals are calculated using standard errors clustered at the drug level.

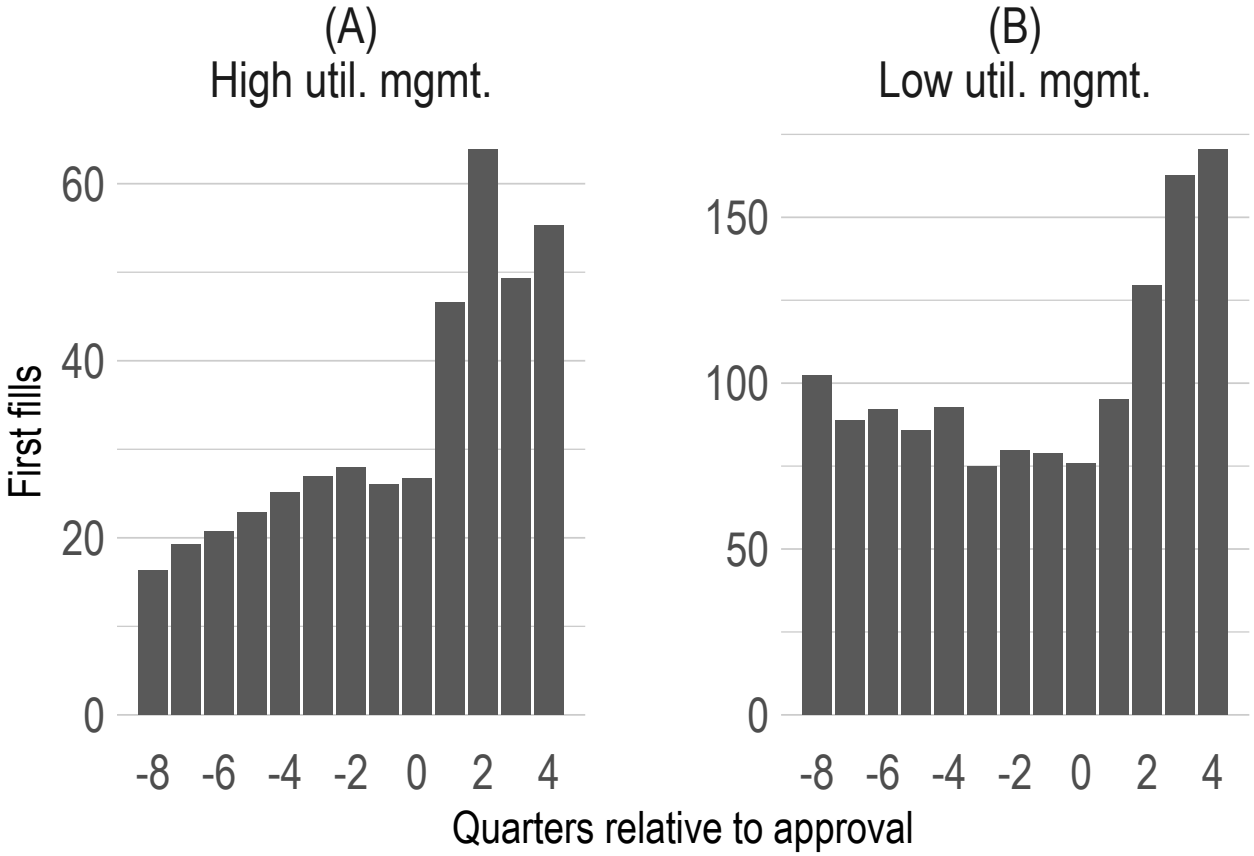


Figure 3: First Fills of Self-Administered Drugs, by Prevalence of Utilization Management

This figure depicts average first fills of 82 drug-disease pairs in windows around the approval of their first indications. Drug-disease pairs are included if the drug is a self-administered medication with known prevalence of utilization management, they are approved via a follow-on indication (i.e. not via a primary indication), and are approved in 2012 or later. The Panel A sample includes 38 drug-disease pairs for drugs with a utilization management share of 50% or greater. The Panel B sample includes 44 drug-disease pairs for drugs with a utilization management share less than 50%. First fills count only those drugs covered by the plan's pharmacy benefit, and thus do not include medications administered by physicians.

A Supplemental Tables and Figures

Table A.1: Difference-in-Differences Estimates, Alternate Censoring Assumptions

Assumed Value	4	1	10
Model:	(1)	(2)	(3)
Variables			
Treat × Post	88.1*	88.3*	87.9*
	(35.0)	(35.0)	(35.0)
Treat × Quarters relative to approval	-1.41	-1.40	-1.43
	(4.50)	(4.51)	(4.50)
Fixed-effects			
Event-Drug-Disease	Yes	Yes	Yes
Event-Year-Quarter	Yes	Yes	Yes
Fit statistics			
Observations	2,600	2,600	2,600
Indicated pairs	70	70	70
Control pairs	130	130	130
Unique pairs	121	121	121
Dependent variable mean	294.1	293.1	296.1
Counterfactual mean	355.9	355.2	357.3
R^2	0.996	0.996	0.996

Clustered (Drug) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, †: 0.1*

This table reports estimates of Equation 1 assuming different values for interval-valued cells (1–10 fills). Our main specification, reported in Column 1 and identical to Column 2 of Table 2, assumes all censored cells equal 4. Columns 2 and 3 assume 1 or 10 fills, respectively.

Table A.2: Difference-in-Differences Estimates, Nonlinear Transformations

Dependent Variables:	$\log$(First fills + 1)		$\sinh^{-1}$(First fills)	
Model:	(1)	(2)	(3)	(4)
Variables				
Treat × Post	0.313**	0.162	0.330*	0.159
	(0.115)	(0.121)	(0.127)	(0.142)
Quarters relative to approval × Treat		0.022		0.024
		(0.019)		(0.022)
Fixed-effects				
Event-Drug-Disease	Yes	Yes	Yes	Yes
Event-Year-Quarter	Yes	Yes	Yes	Yes
Fit statistics				
Observations	2,600	2,600	2,600	2,600
Indicated pairs	70	70	70	70
Control pairs	130	130	130	130
Unique pairs	121	121	121	121
Dependent variable mean	2.42	2.42	2.86	2.86
Counterfactual mean	3.27	3.43	3.82	3.99
R^2	0.967	0.967	0.961	0.961

Clustered (Drug) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, †: 0.1*

This table reports estimates of Equation 1 with nonlinear transformations of the dependent variable.

Table A.3: Fixed Effects Models of Utilization Management, 2013–2019

Dependent Variables:	Prior auth. (PA)		Step therapy (ST)		PA/ST	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Approved for multiple diseases	0.030		-0.053		0.023	
	(0.043)		(0.039)		(0.044)	
Approved diseases		0.048*		0.039*		0.049*
		(0.023)		(0.017)		(0.022)
Fixed-effects						
Drug	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,689	1,689	1,689	1,689	1,689	1,689
Drugs	706	706	706	706	706	706
Dependent variable mean	0.365	0.365	0.218	0.218	0.402	0.402
R^2	0.733	0.740	0.660	0.665	0.737	0.744

Clustered (Drug) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, †: 0.1*

This table reports estimates of the model $UM_{it} = \alpha_i + \lambda_t + \beta f(\text{Approved Diseases}_{it}) + u_{it}$, where UM_{it} is the utilization management share for drug i in year-and-quarter t and $\text{Approved Diseases}_{it}$ is the total number of diseases with at least one indication granted by t. Approved diseases either enter linearly or as an indicator for having more than one approved disease. Columns 1 and 2 measure the dependent variable as the share of covered lives exposed to prior authorization, Columns 3 and 4 use the share of covered lives exposed to step therapy, and Columns 5 and 6 use the share of covered lives exposed to either.

Table A.4: Example Indications, Diseases, and Diagnosis Codes

Drug	Indication	Disease	ICD-9	ICD-10
Carvedilol	Treatment of patients with left ventricular dysfunction following myocardial infarction	Myocardial Infarction	410	I21-I23
Certolizumab Pegol	Treatment of Psoriatic Arthritis	Psoriatic Arthritis	696.0	L40.5
Docetaxel	Treatment of patients with locally advanced or metastatic non-small cell lung cancer after failure of prior platinum-based chemotherapy	Lung Cancer	162.2-162.5; 162.8; 162.9	C34
Eculizumab	Treatment of neuromyelitis optica spectrum disorder (NMOSD) in adult patients who are antiaquaporin-4 (AQP4) antibody positive	Neuromyelitis Optica	341.0	G36.0
Lanreotide Acetate	Treatment of patients with unresectable, well- or moderately-differentiated, locally advanced or metastatic gastroenteropancreatic neuroendocrine tumors (GEP-NETs) to improve progression-free survival	Neuroendocrine Tumor	209	D3A; C7A; C7B

This table reports five example indications, their corresponding diseases, and the ICD-9 and ICD-10 codes assigned to those diseases. See Section B.1 for details on how diseases and diagnoses were assigned.

Table A.5: Difference-in-Differences Estimates, Alternate Diagnosis Windows

Diagnosis Window	30-day	90-day	180-day
Model:	(1)	(2)	(3)
Variables			
Treat × Post	69.6*	81.6*	88.1*
	(27.0)	(31.8)	(35.0)
Treat × Quarters relative to approval	-0.943	-1.28	-1.41
	(3.11)	(3.95)	(4.50)
Fixed-effects			
Event-Drug-Disease	Yes	Yes	Yes
Event-Year-Quarter	Yes	Yes	Yes
Fit statistics			
Observations	2,600	2,600	2,600
Indicated pairs	70	70	70
Control pairs	130	130	130
Unique pairs	121	121	121
Dependent variable mean	206.6	259.7	294.1
Counterfactual mean	249.6	314.3	355.9
R^2	0.997	0.997	0.996

Clustered (Drug) standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, †: 0.1*

This table reports estimates of Equation 1 using different windows prior to first fills of medications from which to source diagnoses. Columns 1 and 2 source diagnoses from 30- and 90-day windows, respectively. Column 3 re-reports the baseline specification, using a 180-day window.

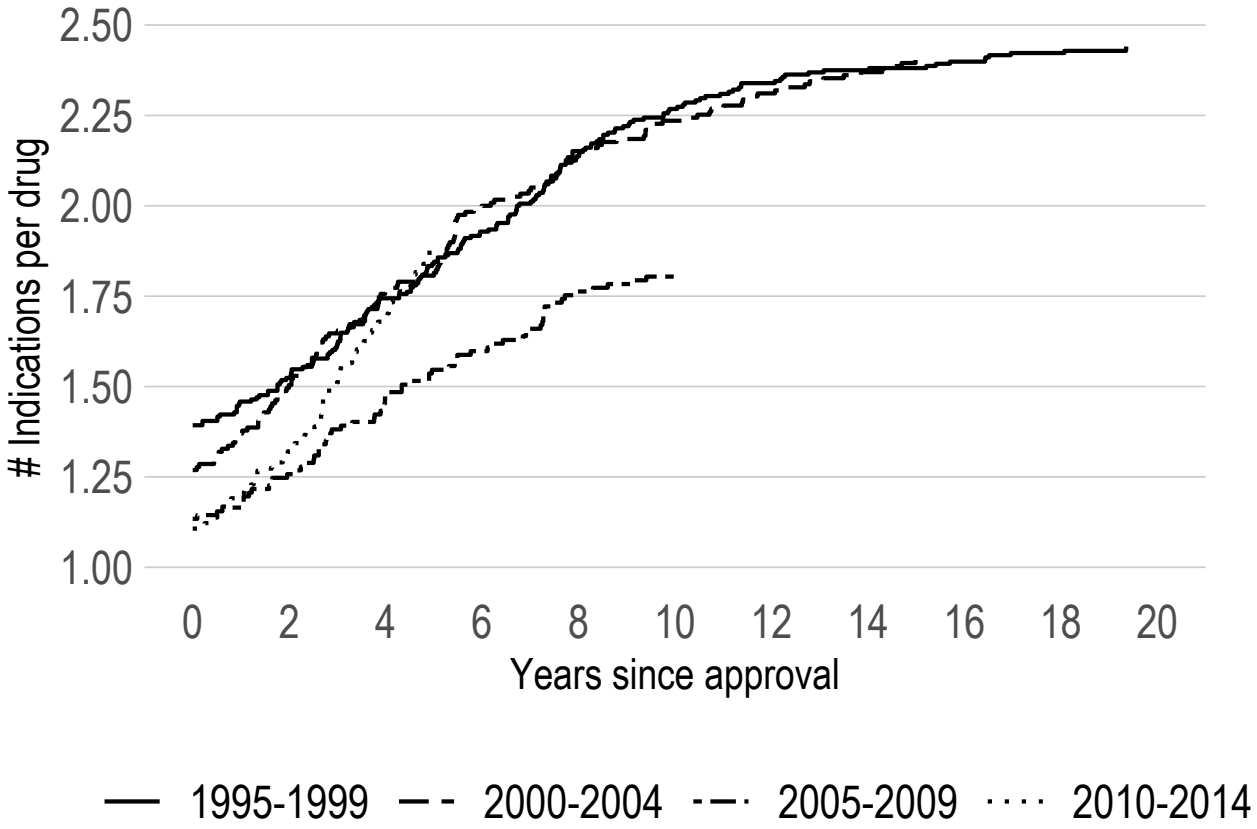


Figure A.1: Cumulative Approved Indications per Drug

This figure depicts the cumulative number of indications for drugs initially approved between 1995 and 2014. Drugs are grouped into 5-year cohorts according to the year in which they received primary approval. Indication approvals include both authorization of use in new diseases and expansion of use within previously approved diseases.

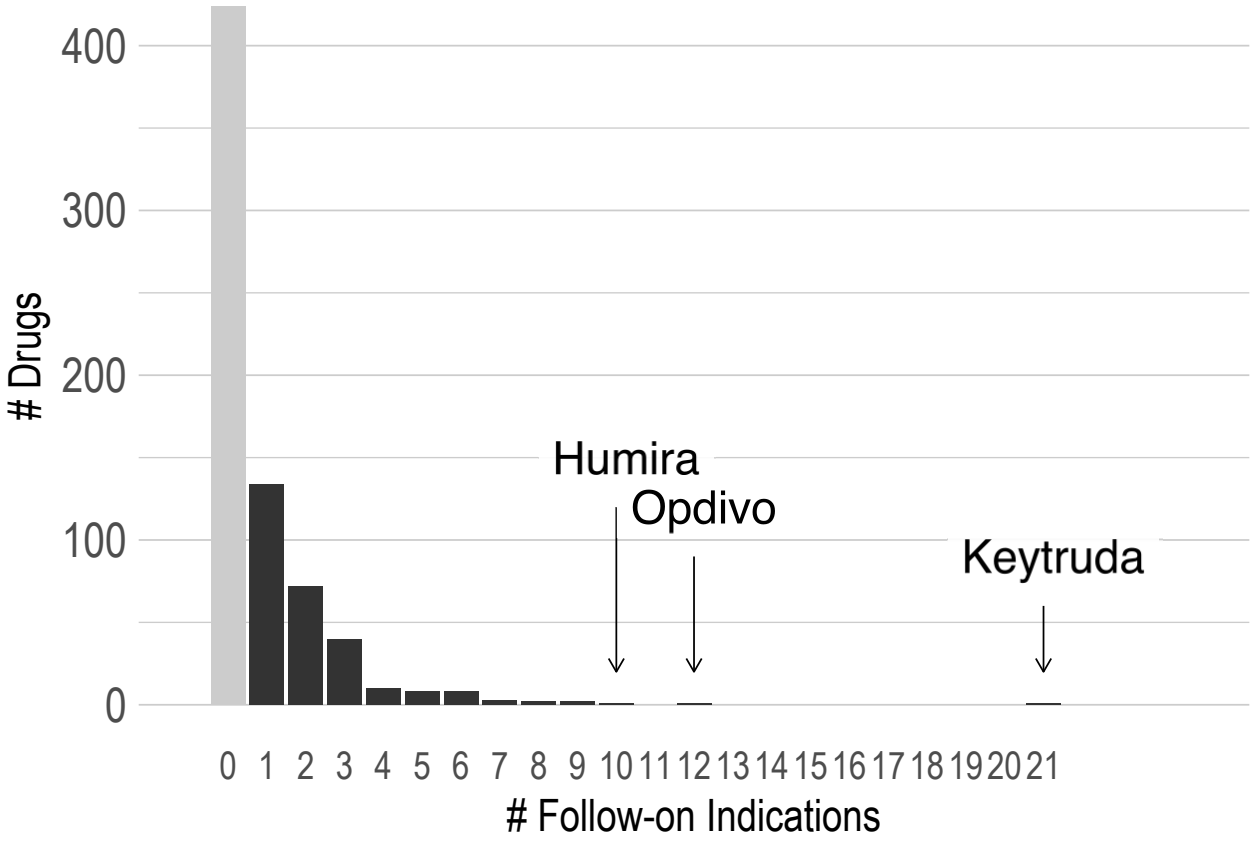


Figure A.2: Approved Follow-on Indications per Drug

This figure depicts the distribution of the number of follow-on indications per drug as of December 31, 2019. The sample includes all drugs approved initially approved between 1995 and 2019. Indication approvals include both authorization of use in new diseases and expansion of use within previously approved diseases but exclude indications approved on the date of primary marketing approval. Forty (40) percent of drugs have at least one follow-on indication. The three labeled drugs had the most follow-on indications as of the cutoff date.

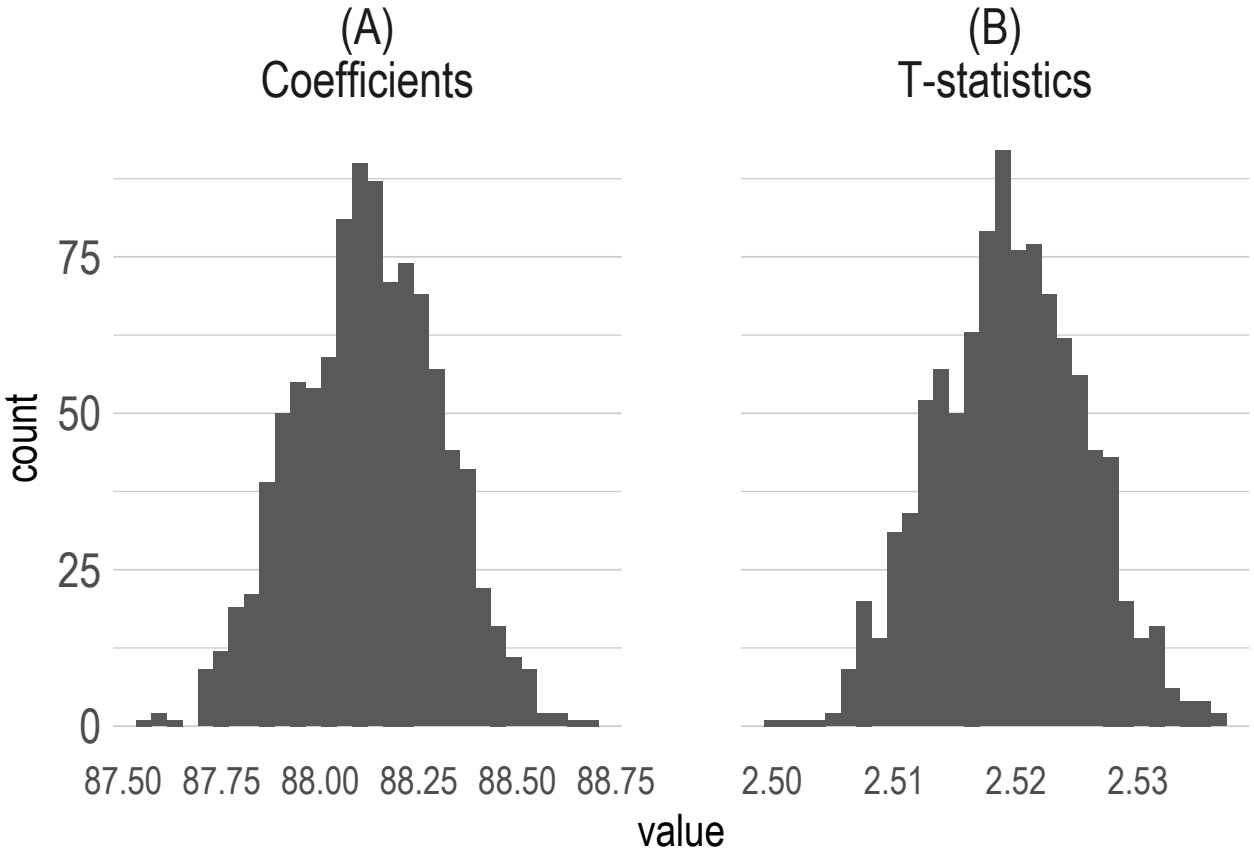


Figure A.3: Simulated Censored Value DiD Distribution

This figure depicts the distribution of 1,000 simulations of difference-in-differences, with each sample replacing interval-valued cells with randomly chosen values between 1 and 10. We assume the random values are drawn from a conditional Poisson distribution with parameter $\lambda = 3.95$, a value chosen so that the conditional expectation of the Poisson distribution matches the empirical mean of censored cells (4). Then, in each simulation, we simulate random values of first fills for all interval-valued cells, compute the difference-in-differences coefficient estimate, and calculate the t-statistic.

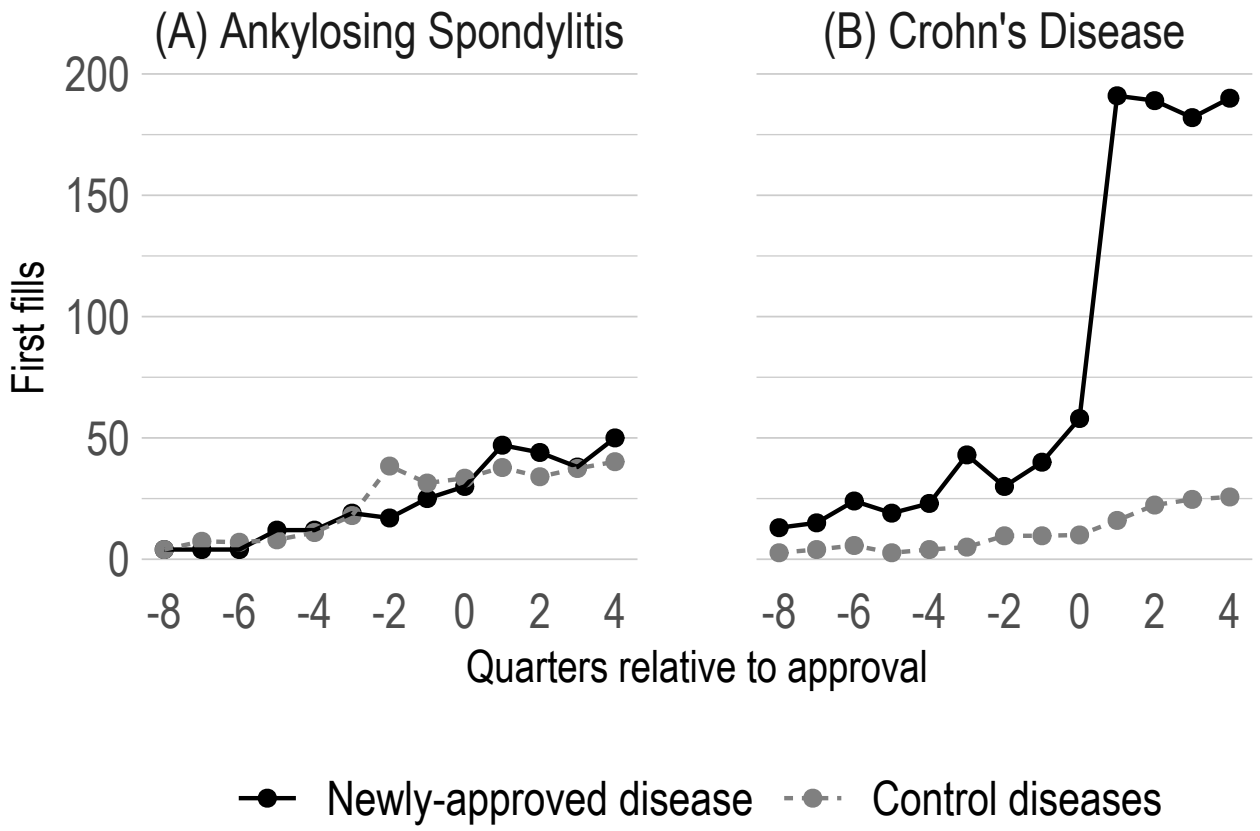


Figure A.4: Humira First Fill Trends

This figure depicts fills in windows around two approval events for the drug Humira (adalimumab). Panel A depicts quarterly fills for ankylosing spondylitis in a window around its approval in Q3 2006 and average quarterly fills for control diseases (plaque psoriasis, juvenile idiopathic arthritis, ulcerative colitis, hidradenitis suppurativa, and uveitis) over the same time. Panel B depicts quarterly fills for Crohn's disease in a window around its approval in Q1 2007 and average quarterly fills for control diseases (ulcerative colitis, hidradenitis suppurativa, and uveitis) over the same time.

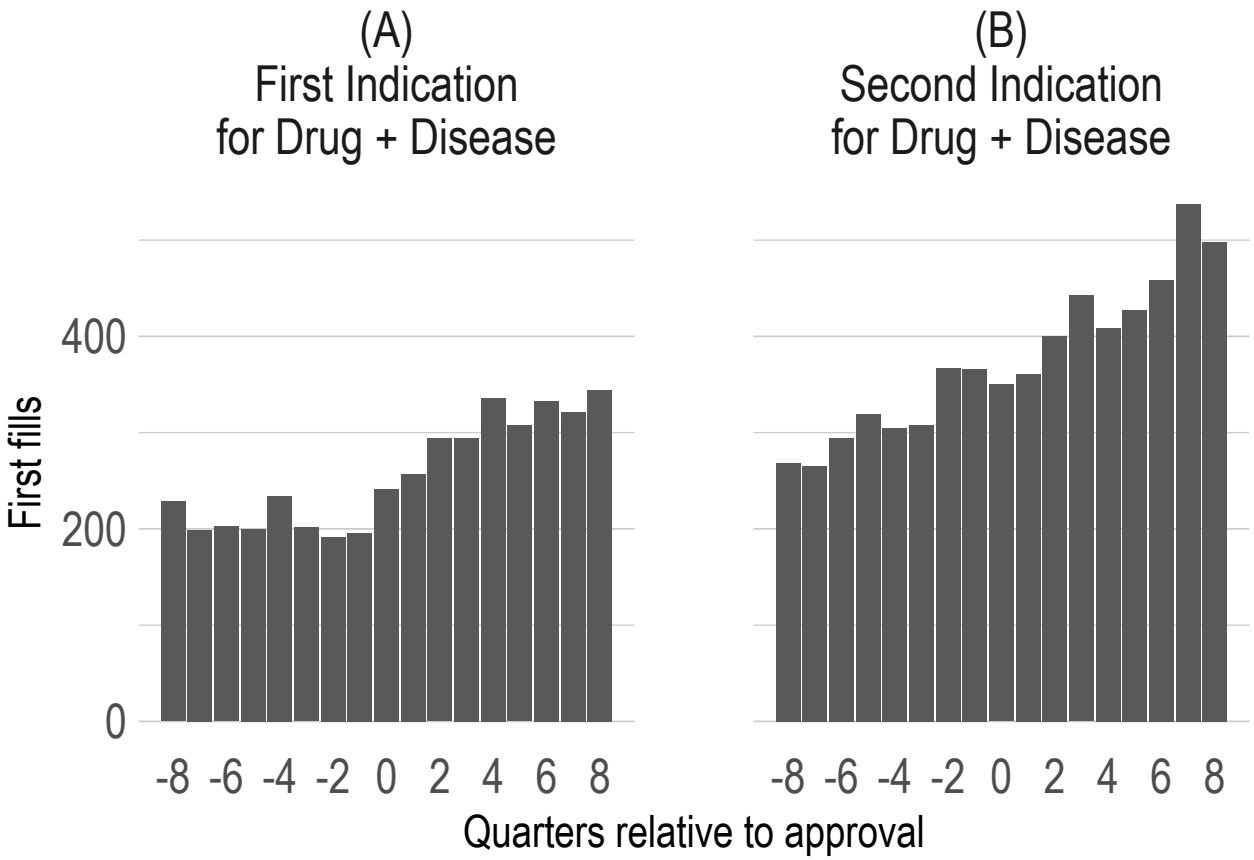


Figure A.5: Average Fills around Follow-on Indication Approval

This figure depicts the average number of first fills across drug-disease pairs in windows around the approval of their first and second indications. The Panel A sample includes all 316 drug-disease pairs that were first approved with a follow-on indication. First fills are averaged within each of the 8 quarters prior to and following approval of that first indication. The Panel B sample includes the subset of all drug-disease pairs (including those that received a primary indication) that received 2 or more indications. Additionally, Panel B excludes pairs if their second indication was approved in the same quarter as the first indication. Panel B depicts average fills within a window beginning 8 quarters prior to approval of the second indication and ending 8 quarters after.

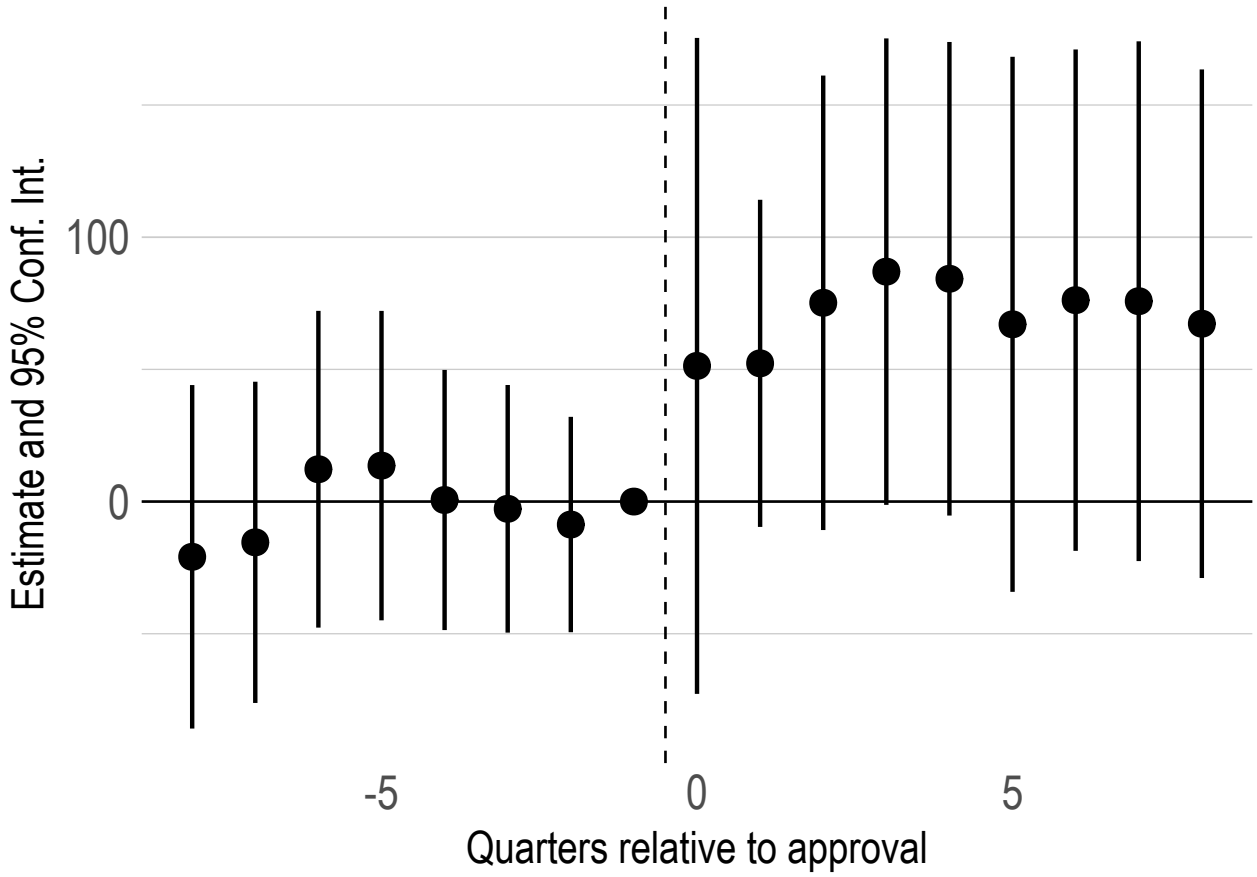


Figure A.6: Event Study Estimates, Extended Post Period

This figure depicts event study estimates for a stacked specification in which 8 post-approval quarters are included (as opposed to 4 post-approval quarters in the main results). Note that controls from the main DiD sample are dropped if they were approved between 5 and 8 quarter after approval of their corresponding indicated pairs.

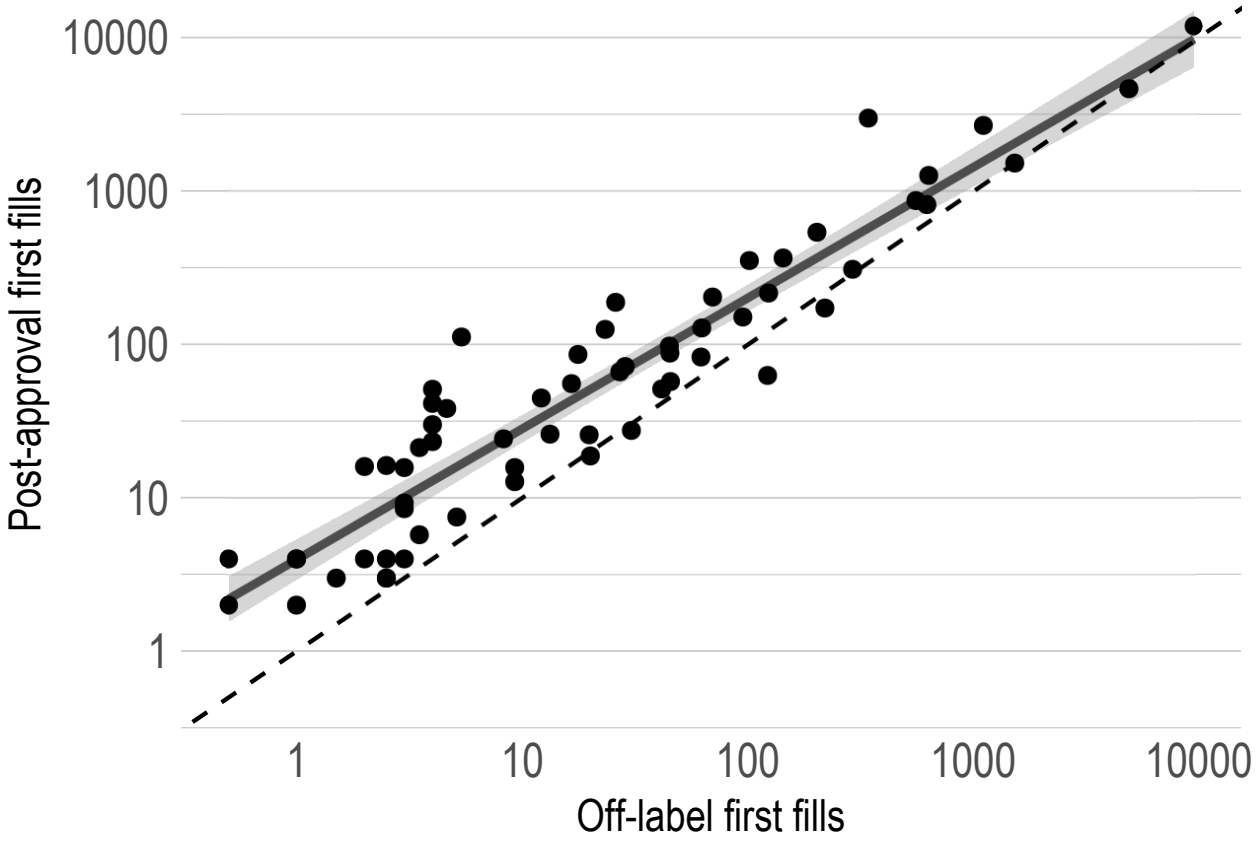


Figure A.7: First Fills before and after Indication Approval

This figure depicts average first fills of approved drug-disease pairs prior to and after they are granted indications. Included are the 70 drug disease pairs in the DiD sample, omitting pairs that had zero fills in all of the 8 quarters prior to approval. Off-label first fills average fills over the 8 quarters prior to approval and post-approval first fills average fills over the 4 quarters after approval. Axes are transformed by a logarithm base 10. The dashed line represents the 45 degree line of equality, and represents no growth in fills after indication approval. The gray line estimates the line of best fit: $\log(\text{Post Fills}_i) = \rho_0 + \rho_1 \log(\text{Pre Fills}_i)$. 95 percent confidence intervals for the predictions assume clustering at the drug level.

B Methodology Appendix

B.1 Indication Data

Our analysis relies on enumerating distinct drug-disease pairs, determining when they were granted indications, and assessing their prescribing patterns. This task is made challenging by the complex and overlapping nature of indications and the diagnosis codes necessary for tracking the prescribing patterns. Indications can be highly specific, including not only the approved disease but also numerous particularities. To give a few examples, a drug may be approved only for second-line therapy, only for patients with particular mutations, only in adults, only after progression to a certain stage, or only in combination with another drug.

Diagnosis codes tend to be defined more broadly than indications. They focus narrowly on the disease, rather than other circumstances warranting treatment, and tend to include little information on the disease's severity, progression, or genetic underpinnings. However, diagnosis codes are not always broader than indications: an example is breast cancer, which has different ICD-10 codes for each quadrant of each breast. Because breast cancer indication approvals may be relevant to patients with cancer of any part of the breast, this level of granularity in diagnosis codes is not needed. Researchers often define diseases at a fixed level of aggregation like a 3- or 4-digit ICD-9/ICD-10 codes, as opposed to the finest levels used for billing, but this too has its issues. Importantly for this paper, this type of aggregation can conflate closely related diseases, having the undesirable effect of reducing the number of distinct drug-disease pairs and therefore the number of valid control pairs. In summary, linkages between indications and ICD diagnosis codes are nontrivial and require a careful approach.

For these reasons, we chose to manually construct disease labels for each indication. This approach allowed us to choose a level of aggregation that maps cleanly to both indications and diagnosis codes. To do this, one author manually assigned disease names to

each indication by reviewing their text. The author then reviewed previous publications for diagnoses used to identify those diseases and created a crosswalk from each disease to one or more diagnoses. He then re-reviewed the distinct diseases and consolidated diseases that could not be differentiated by distinct diagnosis codes. Table A.4 reports diseases and diagnosis codes for five randomly selected indications.

Most indications, as listed in labels and letters on Drugs@FDA, correspond to a single disease, however, 31 indications had approved uses for multiple diseases. For example, a single indication for Repatha (evolucumab) approves treatment for both patients with primary hyperlipidemia and patients with atherosclerosis. In all such cases, we split the labeled indication such that Repatha for primary hyperlipidemia and Repatha for atherosclerosis are different indications.

For a very small sample of our data comprising indications for preventive uses, we ascribed codes to characterize the diagnosis that justifies use of the drug. For example, an indication for "reduction in risk of invasive breast cancer in postmenopausal women with osteoporosis" will be linked to osteoporosis, not breast cancer because diagnoses for prevented diseases do not appear in claims data, making it impossible to identify individuals based on that criterion. We did not code diagnoses for drugs with diagnostic uses (such as radiocontrast agents) and those primarily used in perioperative settings (such as anesthetics) due to the difficulty of identifying individuals based on diagnoses in these settings.

We grouped diseases into broader disease groups using the ICD-10 hierarchy. ICD-10 codes are organized into sub-chapters (corresponding to a diverse group of diseases), and chapters (corresponding to an even larger group of diseases). For example, ICD-10 code C91.1 "Chronic lymphocytic leukemia of B-Cell type", which in turn has several indications within it, falls in sub-chapter "Malignant Neoplasms Of Lymphoid, Hematopoietic And Related Tissue,", which is in the chapter "Neoplasms." Because a single disease may match a range of codes, we determined both a "primary" code, which we use to determine

the sub-chapter and chapter of the disease, and "secondary" codes that we only leverage to identify patients with the disease.

B.2 Medical and Pharmacy Claims Data

We restrict claims data to include only beneficiaries of commercial plans — which means excluding beneficiaries of Part D prescription drug plans — and only claims filed by beneficiaries enrolled in both medical and pharmacy coverage through the plan. Requiring patients to have both medical and pharmacy coverage significantly increases the probability that we observe both prescription fills (pharmacy coverage) and diagnoses (medical coverage).

For similar reasons, we source diagnoses from within 180-day windows prior the first fill of each drug. If we were to use shorter continuous enrollment windows (e.g. 30 days), then the prescriptions of beneficiaries who switched from another insurer's plan would be more likely to erroneously appear as first fills due to the lack of claims from the previous insurer. On the other hand, we expect that diagnoses sourced from shorter windows prior to first fills are more likely to reflect the diagnoses that warranted the prescription, rather than unrelated diagnoses which may lead our algorithm to misattribute diseases to those fills.

In Table A.5, we report how using alternate diagnosis windows impacts our estimates. Columns 1 and 2 report estimates of Equation 1 using 30- and 90-day diagnosis windows prior to first fill (180 days of continuous enrollment are still required). Column 3 re-reports data using the baseline diagnosis window of 180 days. The average number of fills shrinks when using the 30- and 90-day windows because fewer total diagnoses are reported in the shorter windows. The difference-in-differences point estimates are also smaller compared to the 180-day window. However, the percent effect is similar across specifications (30-day: 28%, 90-day: 26%, and 180-day: 25%). Thus, our estimates are qualitatively unchanged and quantitatively unchanged on a percent scale; sourcing diagnoses

from shorter windows does not impact any of our conclusions.

B.3 Empirical Methodology

B.3.1 Stacked Sample Construction

Drugs that are approved for only 1 disease have neither pre-period fills for that disease (because it is approved with a primary indication that simultaneously authorizes the drug to be sold) nor within-drug controls (because no other diseases are ever approved for that drug). Drugs that are approved for 2 diseases have pre-period fills for the second approved disease, but no within-drug controls because the only other approved disease has already been granted an indication by the time the second disease is granted an indication. Thus, a drug approved for n diseases may contribute $n - 2$ drug-disease pairs at most.

In principle, we could include drug-disease pairs that are never approved as controls, allowing us to construct controls for more treated pairs. However, we determined this was infeasbile for this study. The foremost reason is that it would be incredibly challenging to enumerate a representative sample of off-label uses due to the very fact that they are not labeled. While sources of information on off-label uses exist (compendia, medical literature, clinical trials, social media, etc.), the accuracy and comprehensiveness of that information varies across sources, and the magnitude of labor required to clean and standardize that data made it intractable. Also, while a larger control group would likely make our estimates more precise, it could also introduce bias because utilization trends for never-approved pairs likely differ systematically from trends in approved pairs.

B.3.2 Dependent Variable

We prefer to model first fills in levels rather than logarithms for four primary reasons. First, owing to low demand among many off-label uses, 23 percent of observations of the dependent variable equal zero, such that the logarithm is not defined. While it is com-

mon to use an alternative "log-like" transformation in this case (e.g. $\log(x + 1)$), using such a transformation makes it so that the model coefficients do not maintain the same convenient interpretation in terms of percent changes in the dependent variable. Second, 34 percent of observations in our sample fall in the range 1–10. In our main specification, we assume these cells equal 4. While our results in levels are robust to changes in this assumption, using log-like transformations of the dependent variable would lead our results to be highly sensitive to both the transformation used and the assumption on the censored range of values. Third, "log-like" transformations have recently been shown to have undesirable characteristics; in particular, treatment effect magnitude may be arbitrarily increased or decreased by rescaling the units of the dependent variable (Chen & Roth 2024). Fourth, our fixed effects models are intended to contrast marginal effects of the first and second indications within drug-disease pair. If both indications increased fills by the same amount on average, the first indication would tend lead to a larger log-point change in fills than the second even though the two indications impact demand by the same amount.

B.3.3 Estimation

When a treated pair is approved for multiple indications within the window, we leave the pair in the sample. This means that while the estimates are mostly driven by the effect of the first indication, it will also include some of the effect of later indications for that pair. However, only 6 percent of treated pairs (4 out of 70) have additional indications approved within the 4 quarter window. Additionally, two (2) of those indications are approved in the same quarter as the first indication, so our analysis does not distinguish their timing from that of the first indication; practically, they are the same as a single indication.

B.3.4 Calculating Corrective Weights

Weights for indicated pairs are given by $w_{idtk} = 1$. Weights for control pairs are given by $w_{idtk} = \frac{N_1^{E_k}/N_1}{N_0^{E_k}/N_0}$ where $N_1^{E_k}$ is the total number of indicated pairs in approval events occurring at time E_k, $N_0^{E_k}$ is the equivalent for control pairs, and N_1 and N_0 are the total number of treated and control pairs, respectively. Intuitively, these corrective weights increase the contribution of controls in event periods in which they are relatively less prevalent. For example, 2 out of 70 indicated pairs (2.9 percent) appear in approval events occurring in Q1 1998, and 2 out of 130 (1.5 percent) control pairs appear in events occurring in that period. Thus, the corrective weight for Q1 1998 controls is $\frac{2/70}{2/130} = 1.86$.

B.3.5 Calculating Counterfactuals and Percent Effects

Percent effects correspond to the ratio of β_{DD} to counterfactual average fills of treated pairs after approval. Using conventional notation for potential outcomes, the parallel trends assumption implies $\mathbb{E}[\text{First Fills}_{idtk}(0)|\text{Treat}_{idk} = 1, \text{Post}_{idk} = 1] = \mathbb{E}[\text{First Fills}_{idtk}(1)|\text{Treat}_{idk} = 1, \text{Post}_{idk} = 1] - \beta_{DD}$: this counterfactual mean is equal to the treated mean minus the treatment effect. Because we observe the treated outcome for treated pairs, the right hand side simplifies to $\mathbb{E}[\text{First Fills}_{idtk}|\text{Treat}_{idk} = 1, \text{Post}_{idk} = 1] - \beta_{DD}$, which can be estimated with our data. Thus we estimate percent effects with the sample analogue of $\frac{\beta_{DD}}{\mathbb{E}[\text{First Fills}_{idtk}|\text{Treat}_{idk} = 1, \text{Post}_{idk} = 1] - \beta_{DD}}$.

B.4 Humira Case Study

In Figure A.4, We illustrate our difference-in-differences approach with a case study of the two Humira approvals that we discussed earlier, for AS and Crohn's. The figure shows quarterly first fills of Humira by patients with AS and its control diseases or Crohn's Disease and its control diseases in windows around the approval of the first indications for those diseases. First fills by patients with AS changed little more than patients with control

diseases after approval of the AS indication (suggesting that market learning about this indication was relatively complete), while first fills by Crohn's disease patients increased by 159 after approval of the Crohn's disease indication (controls increased by only 16), suggesting that regulatory approvals gave meaningful information to market participants and that market learning would have taken many years to generate the same level of fills. Importantly, the panels of Figure A.4 each correspond to a single dataset for our stacked difference-in-differences model. In each, we include fills for a single approved disease that is impacted by the approval event and one or more controls diseases that are not.

B.5 Off-label Use

Our estimate of off-label use is larger than that reported by Radley et al. (2006) who find that off-label uses comprise 21 percent of prescriptions. There are several reasons why our results differ. Their data is based on survey evidence rather than administrative claims. Second, we report off-label use at the very beginning of the marketable life of drugs when drugs have the fewest approved indications and more off-label uses. Third, Radley et al. report off-label use in a cross section of drugs prescribed during a given year and thus observe both drugs that were recently approved and those that had been approved for decades (their sample includes generic drugs). Fourth, our sample differs dramatically in the types of drugs included. For instance, 24 percent of the drugs in our sample are cancer drugs while the Radley et al. sample includes none. Because cancer drugs are often used to treat unapproved cancers, this would increase the off-label share in our sample.