Incremental Innovation and Pharmaceutical Productivity[*]

Kristopher J. Hult[†]

Abstract

This paper investigates the role of novel and incremental innovation in biopharmaceutical markets. Previous research focusing only on novel innovation—FDA-approved new molecules—has led to the conclusion that the pharmaceutical industry is in a "productivity crisis," since R&D spending has increased exponentially while FDA-approved new molecules have remained flat over time. I find that incremental innovation—new drugs created by modifying existing FDA-approved molecules—accounts for 49% of the health impact of new innovations, and productivity of pharmaceutical innovation has increased 30% between 1980 and 2009 when considering the health impact of both novel and incremental innovations. I construct and estimate a model of how firms trade-off between novel and incremental innovation to predict future innovation trends, and I find that the productivity of new innovations will decline by 40% during the 2010s.

1 Introduction

The pharmaceutical industry is perceived to be in a "productivity crisis" because R&D spending has grown exponentially while innovation, measured as new FDA-approved molecules, has remained relatively flat over time.[1] These trends have persisted since the 1950s and have received attention in recent years as the number of new molecules fell by 60%, while R&D spending increased by more than 280% between 1996 and 2010. Figure 1 illustrates the stark contrast in perceived productivity between the top two R&D industries in the US: computers, on the left, have experienced exponential returns to R&D spending while pharmaceuticals, on the right, have experienced exponential declines in output per R&D dollar.
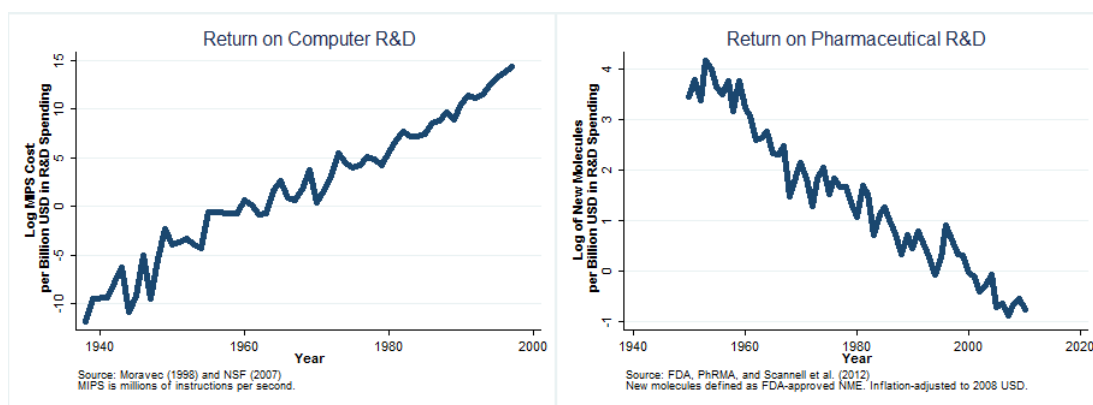

Figure 1: Pharmaceutical Productivity Crisis

I start with two simple, but important, observations. The first is that the productivity of pharmaceutical innovation should be measured as the health impact of new innovations, not the count of new molecules. Health impact incorporates three components: how widely the drug is used; adherence, or what fraction of users take the drug as prescribed; and efficacy, or the health impact per prescription conditional on adherence.

The second observation is that incremental innovation is an overlooked, but increasingly important, component of pharmaceutical innovation. There are two types of FDA-approved pharmaceutical innovation; novel innovations are new molecules and incremental innovations are new drugs created by modifying existing molecules. Incremental innovations represent a growing share of pharmaceutical innovation and utilization, and I estimate here that they account for more than half of all nongeneric prescriptions. Incremental innovations can generate value by: creating new drugs that use existing molecules to treat different diseases; changing the chemical formulation or active ingredient of a drug to increase the drug's efficacy and reduce

---

[1]Examples include Pammolli et al. (2011), Hu et al. (2007), Munos (2009), and Scannell et al. (2012).

side effects; creating combination drugs or reducing the number of pills or doses, which increases a patient's adherence to a drug regimen; and creating new delivery methods for certain patients, such as pediatric and elderly patients, who could not take the drug in its originally approved form.[2]

My main finding of this paper is, contrary to previous findings, the productivity of pharmaceutical innovation has actually increased by 30% between 1980 and 2009. This increase in productivity is driven by incremental innovation, which I find accounts for 49% of the total health impact of new innovations during the 2000s.

The response of pharmaceutical innovation to the spread of HIV/AIDS in the US illustrates the impact of incremental innovation.[3] From 1981 to 1992, the number of AIDS diagnoses in the US went from zero to over 75,000 per year and the number of deaths per year exceeded 40,000 by 1997 (CDC, 2013). From 1987 to 1996, there were 11 new molecules produced to treat HIV/AIDS. However, this period of HIV treatments was characterized by high mortality rates—the probability of surviving two years was 30%—and low adherence rates, which ranged from 30% to 50%.[4] Despite having treatments that increased life expectancy, the low adherence rates were mainly due to the numerous and serious side effects caused by the drugs and the difficult-to-follow drug regimen. In 1997, previously approved drugs were combined into a treatment known as highly active antiretroviral therapy (HAART), which drastically increased the probability of surviving two years after diagnosis to over 60%; however, in the early years of this treatment, adherence rates were only 55%–60% because many treatments featured a very difficult drug regimen with serious side effects, food interactions, pill refrigeration requirements, and five doses per day of 20 to 30 pills (Murphy et al., 2003). By 2000, despite featuring the same top-selling molecules, incremental innovations improved HAART treatments by combining existing molecules into drugs that reduced pill burden, creating new formulations to eliminate food interactions and the need for refrigeration, and producing new dosage formulations that allowed pediatric patients to take the drug. These innovations increased the probability of living two years after diagnosis to over 80% and adherence increased to 65%–70% (Rebick and Walmsley, 2012). By 2006, HAART treatments were replaced by combination treatments featuring a one-pill, once-a-day dosage due to the combination of new molecules and additional incremental innovation on existing molecules. As a result, treatment adherence for HIV increased to over 85%, and life expectancy continued to increase (Rebick and

---

[2]This definition does not include off-label prescriptions, which are drugs prescribed for an indication that differs from the indications they were approved by the FDA to treat. My definition also does not include the FDA's supplemental approvals, which are generally labeling or manufacturing revisions. I do not include supplemental approvals because they do not have a significant effect on the efficacy or adherence rate of the drug.

Incremental innovations are sometimes referred to as incrementally modified drugs or IMDs. Incremental innovations are not synonymous with me-too drugs, which are drugs that are chemically similar to existing drugs. Me-too drugs could be either novel or incremental innovations.

[3]Appendix A provides a more detailed summary of the role of innovation in HIV/AIDS.

[4]All survival probabilities in this section are from Couzigou et al. (2007). Adherence rates are from Wall et al. (1995).

Walmsley, 2012). In addition, new drugs have expanded the potential use for molecules created for HIV to treat hepatitis B and C, and are being used in research for infectious diseases such as malaria.

I do two things in this paper. First, I construct a novel dataset to estimate the health impact of every FDA-approved drug, where health impact incorporates the number of people taking the drug, the adherence rate of the drug, and the efficacy of the drug. I construct adherence rates and utilization measures for each drug using prescription survey data from the Medical Expenditure Panel Survey (MEPS), and I estimate efficacy using measures of quality-adjusted life years (QALYs) from the Cost-Effectiveness Analysis Registry (CEAR) at Tufts University. Using these data, I find incremental innovation makes up 49% of the health impact of new pharmaceutical innovation, and the share is growing over time. In addition, I find that the productivity of both novel and incremental innovation, measured as the health impact of new innovations divided by R&D spending, increased by 30% over the 1980s, 1990s, and 2000s. By contrast, productivity measures that only use the count or health impact of novel innovations decrease over this time period.

The second thing I do in this paper is I construct and estimate a model of the dynamic trade-off that firms face between investing in novel and incremental innovation. My model is based on the assumption that incremental innovation is easier with a larger stock of novel innovation on which to expand, and harder when more incremental innovation has taken place. I use this model to explain past trends and predict future trends in pharmaceutical innovation. The model finds that the increase in novel innovation during the 1990s, mainly due to changes in FDA procedures, produced a significant increase in the number of incremental innovations during the 2000s.[5] This increase, combined with a significant increase in the health impact per drug during this period, produced an increase in the productivity of pharmaceutical innovation. However, the substitution toward incremental innovation during the 2000s means that there is a relatively low stock of novel innovation, so incremental innovation will slow down during the 2010s. I predict that this slowdown in the number of innovations, combined with the expected doubling in R&D spending over the 2010s, will produce a 40% decline in pharmaceutical productivity in the 2010s.

This paper provides the first systematic analysis of the health impact of incremental innovation in pharmaceuticals and constructs a novel dataset to measure the health impact of new innovations, but relates to several previous strands of work. There is a wealth of literature measuring pharmaceutical innovation as the count of novel innovations, but it largely ignores the role of incremental innovation. This literature examines how market size or insurance affects innovation (Acemoglu and Linn, 2004; Cerda, 2003; Blume-Kohout and Sood, 2013), the cost of new innovations (DiMasi and Grabowski, 2007), and productivity in pharmaceuticals

---

[5]Berndt et al. (2005) find that the Prescription Drug User Fee Act (PDUFA) increased the number of novel innovations by over 3% each year, following its approval in 1992. The FDA also expanded expedited approval in reaction to the spread of HIV/AIDS and launched a fast-track program in 1997.

(Pammolli et al., 2011; Hu et al., 2007; Munos, 2009; Scannell et al., 2012).[6] Related papers also include findings that one type of incremental innovation (new indications) makes up 70%–80% of drug utilization in certain disease classes (Berndt et al., 2006), and argue that incremental innovation, in general, should be considered when measuring R&D output (Cockburn, 2007).[7]

2 Understanding Incremental Innovation

In this section, I describe incremental innovation and how it affects health. I show that the health impact of new innovations comes through three channels: number of users, adherence, and efficacy. I also show that incremental innovation is an important component of prescription utilization, accounting for over half of all nongeneric prescriptions, and of pharmaceutical innovation, accounting for roughly three times more innovation than novel innovation.

2.1 Defining Incremental Innovation

Pharmaceutical innovations are new FDA approvals, which take two forms: novel and incremental innovation. Novel innovations are new FDA-approved drugs whose ingredient is a previously unapproved molecule. Incremental innovations are either new FDA-approved drugs created from an already existing molecule or FDA-approved modifications to existing drugs. There are five types of incremental innovation: new dosage form, new formulation, new combination, new indication, and new active ingredient.[8]

New Dosage Form

Changes in dosages comprise 60% of FDA-approved incremental innovations and can affect any of the three parts of a dosage regimen: the dosage route (how the drug is administered, i.e., oral, injection, topical, etc.), the dosage form (whether the drug is a tablet, capsule, solution, cream, etc.), and the dose amount (how much of the active ingredient is in the drug, usually in milligrams). Route changes are the most common type with 41% of dosage changes, and only 5% are small dosage amount changes.[9]

Dosage changes have two main effects. First, they can increase adherence—whether a patient takes the medication as prescribed—by making drugs easier to take through reductions in pill burden, dosage

---

[6]Acemoglu and Linn (2004) may contain incremental innovation in their definition of nongeneric approvals, but since they use unpublished FDA data, it is unclear how they define nongeneric approvals.

[7]Cockburn (2007) discusses how incremental innovation should be considered when measuring R&D output. However, he uses the term differently than I do. He refers to incremental innovation as supplemental new drug approvals (NDAs), which are generally labeling revisions and changes in, or additions to, the manufacturing process. I only refer to incremental innovations as FDA-approved innovations using original NDAs, meaning that incremental approvals are newly approved drugs, not supplemental changes to existing drugs. He also refers to new indications, formulations, and dosage, which are subsets of my definition of incremental innovation.

[8]The classification is listed by the FDA under chemical type.

[9]See Appendix B for a breakdown of what types of dosage innovations are most common.

frequency, or side effects, or by changing the dosage route. Second, dosage changes can affect the number of people taking a drug within a disease class by changing the dosage form, such as creating a solution or injection form of the treatment instead of pills. This dosage form allows the treatment to be taken by pediatric, elderly, and pregnant patients who are unable to take the treatment in the original form.

New Formulation

Changes in formulation are the second most common type of incremental innovation at 18% of FDA incremental approvals. A drug's formulation is how the chemicals in the drug are combined to produce the drug and is nearly identical to the drug's dosage form. HIV/AIDS treatments, such as Norvir, featured changes in formulation that eliminated the need for refrigeration, reduced the number of drug and food interactions, and provided extended release for drugs. These innovations affected adherence rates, because they made drug regimens easier and also increased efficacy by reducing drug and food interactions.

New Combination

The creation of combination drugs from existing molecules, a process that accounts for 12% of FDA incremental approvals, played an important role in HIV/AIDS treatments. As combination drugs hit the market throughout the HAART and CART phases, adherence was increased due to reduced pill burden. In addition, combination drugs reduced potential drug interactions.

Some new combination approvals exist as separate drugs, which can be taken separately. However, utilization rates for two active ingredients increase ninefold after the FDA approves them in a combination drug.[10] Since a drug can't be promoted for uses for which it is not approved, and since getting FDA approval provides information on the efficacy and side effects of combining the two drugs, information could play a role in explaining why combinations are not widely used prior to FDA approval. Other new combinations, such as those that combine drugs with devices, do not exist in the market prior to the combination approval.

New Indication

The fourth type of incremental innovation, new indications, makes up 6% of incremental approvals; these take an existing drug and use it to treat a different condition. Each FDA-approved drug is approved exclusively to treat a specific condition. If a drug can be used to treat a different condition, then it can be approved as a new indication. Doctors are allowed to prescribe any FDA-approved drug for any condition they see fit, and off-label prescriptions—a prescription for an indication for which the drug is not FDA-approved—are a major component of drug sales. However, companies cannot promote drugs for a use that

---

[10]This estimate is based on the ten best-selling combinations in the MEPS dataset.

is not FDA-approved, so there is an incentive for firms to obtain new indication approval. Within a disease class, a new indication has the same health impact as a novel innovation that increases efficacy but does not require the producer to invest over $1 billion to create a new molecule (DiMasi et al., 2003).

Off-Label Prescriptions

Off-label prescriptions occur when a doctor prescribes a drug in a different manner from what was approved by the FDA. They are incremental innovations in that they are changes in a drug from its original approval, exactly like a new indication approval, except they do not have FDA approval. More than 20% of all prescriptions are off-label (Stafford, 2008). However, prescription companies cannot promote the use of a drug for an indication for which it was not approved. In certain diseases, like HIV/AIDS, where the disease spread quickly, the off-label utilization can be higher. Brosgart et al. (1996) report that 80% of HIV patients received at least one off-label prescription during treatment.

New Active Ingredient

New active ingredients make up 4% of incremental approvals. These are drugs that contain the same active moiety but include a different enantiomer, racemate, salt, ester, complex, chelate, or clathrate. A molecule may exist in two forms that are mirror images; these forms are known as enantiomers. Each enantiomer may have very different effects in a drug. A racemate is the combination of both enantiomers. For example, Fetzima, an SNRI drug used to treat major depressive disorder, is an enantiomer of a previously approved racemate milnacipran HCI with the brand name Savella, used to treat fibromyalgia. Both Fetzima and Savella use the same molecule, but in different orientations.

Active ingredients are also essentially the same in their impact as novel innovations, because they can produce entirely different drugs used to treat different diseases.

2.2   Measuring Health Impact

The health impact of a drug is how much it increases patients' length and quality of life. As illustrated in the previous section, innovations affect health through three channels: adherence, quantity measured as the number of users, and efficacy.

To derive the health impact of new innovations, I define the health impact of drug j, h_j, to be:

$$h_j = \sum_y q_{jy} a_j e_j$$

where q_{jy} is the quantity measured as the number of users in year y, a_j is the adherence rate of the drug, and e_j is the average efficacy of the drug with adherence. If 100 people take a drug with a 60% adherence rate that adds one QALY on average, then the health impact of the drug is 60 QALYs.[11]

Summing h_j over the set of all drugs (D) produces the total health impact of all drugs in year y (H_y):

$$H_y = \sum_{j \in D} \sum_y q_j a_j e_j$$

To measure the increase in health impact produced by new innovations approved in year y, which is how new innovations increase health impact relative to the standard of care (SOC) that existed before the innovation, I construct ΔH_y:

$$\Delta H_y = \frac{\partial H_y}{\partial q} \Delta q + \frac{\partial H_y}{\partial a} \Delta a + \frac{\partial H_y}{\partial h} \Delta h$$
$$= \sum_{j \in D_y} [\Delta q_j a_j e_j + \Delta a_j q_j e_j + \Delta e_j q_j a_j]$$

where q_j is the average quantity of drug j per year, Δq_j is how drug j changes the quantity relative to the SOC, Δa_j is how drug j changes the adherence rate relative to the SOC, Δe_j is how drug j changes efficacy relative to the SOC, and D_y is the set of all drugs approved in y. Hence, the health impact of new innovations is the sum of the effects of the change in the quantity, adherence, and efficacy relative to what was used before the innovation. For instance, if a drug innovation with 100 users and an efficacy of one QALY increases the adherence rate relative to the previous SOC by five percentage points, then the health impact of that innovation is $0.05 * 100 * 1 = 5$ QALYs. If that drug innovation had an adherence rate of 60% and also increased efficacy by 5%, then the health impact would be $5 + 0.05 * 100 * 0.6 = 5.3$ QALYs.

2.3 Overall Trends in Incremental Innovation in the US

Incremental innovation has several features. First, it requires much less R&D per innovation. Although it is hard to break down costs and R&D spending into novel and incremental innovation, the average novel innovation takes roughly 20 times more capitalized costs or 10 times more R&D spending than the average incremental innovation.[12]

---

[11]$100 * 0.6 * 1$ QALY $= 60$ QALYs.

[12]These estimates are based on dividing the count of incremental and novel innovations by cost estimates. DiMasi et al. (2003) find that post-approval R&D is roughly 10% of capitalized costs or 25% of R&D, and Frank (2003) estimates that 30% of R&D spending is used on new or modified uses for an existing product.

Second, while novel innovation remained pretty flat since 1980, except for an increase around 1996 that may be due to the Prescription Drug User Fee Act (Munos, 2009), incremental innovation trended upward since 1970 as shown in Figure 2.
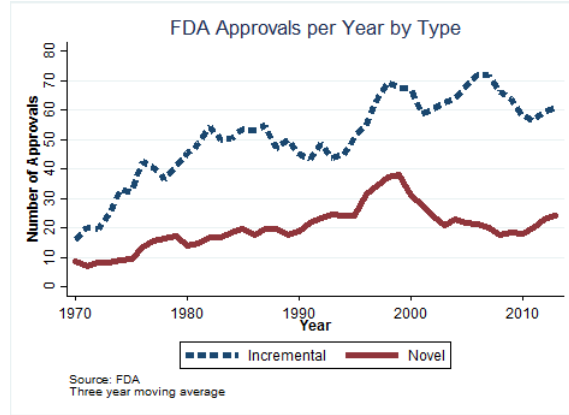
Figure 2: FDA Approvals by Type

Third, the majority of consumed nongeneric prescription drugs are from incremental innovation, and this fraction is rising quickly, as shown in Figure 3.
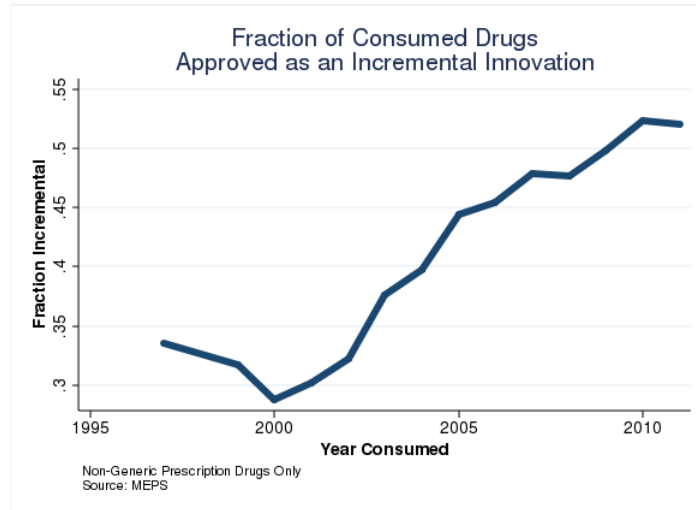
Figure 3: Incremental Innovation Fraction Share

Fourth, incremental innovation is adopted much more quickly than novel innovation. Using the MEPS data, I find that 17% of novel innovations were purchased within five years of being approved, while 51% of incremental innovations were purchased within the same period.

Fifth, patients with diseases experiencing increases in incremental innovation driven by Medicare Part D, which greatly increased prescription insurance coverage for the elderly, have been shown to display decreased physical difficulty, physical limitation, and mortality (Hult, 2014).

Sixth, incremental innovation requires less approval time (five to six years instead of around 10 to 12 years for novel innovation) and gets less market exclusivity (three years versus five years for novel innovation).[13] One important issue is whether pharmaceutical companies use incremental innovation simply as a method of extending market exclusivity. I address this concern by examining whether the number of incremental innovations changes around patent and exclusivity expiration dates. Figure 4 shows the fraction of incremental innovations by number of years before or after the patent expires. Negative numbers are the number of years before expiration and positive numbers are the number of years after expiration. I find that although the number of incremental innovations increases just after a patent expires, this spike is not a major component of incremental innovation. Exclusivity is a decreasing function of the expiration date.
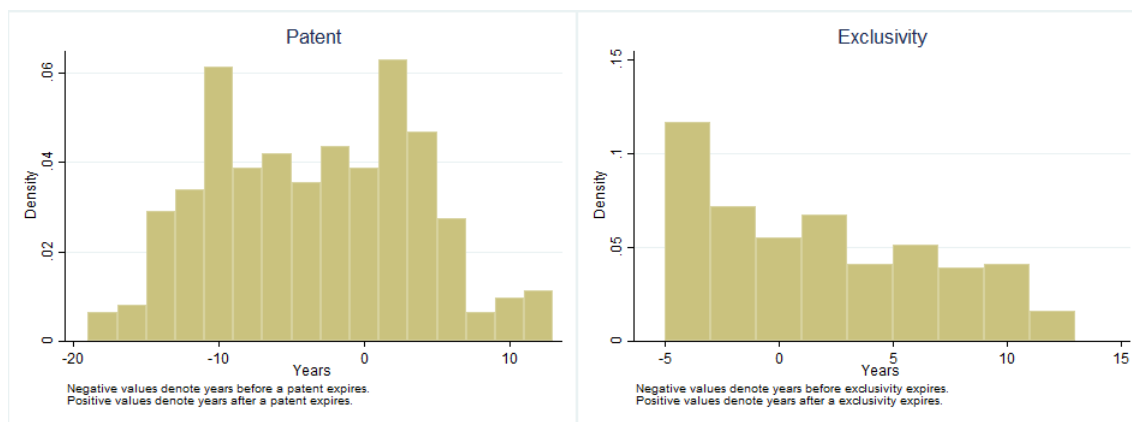
Figure 4: Number of Incremental Innovation Approvals by Year, Before and After Patent or Exclusivity Expires

3   Role of Incremental Innovation in Pharmaceutical Productivity

This section measures the health impact of new pharmaceutical innovations—how much health new innovations add relative to existing drug treatments—and the productivity of pharmaceutical innovation—

---

[13]See Appendix B for more detail on the approval process of incremental versus novel innovations.

the health impact per dollar of R&D spending. I find that the health impact of new innovations increased 3.5 times between 1980 and 2009 and incremental innovation is responsible for 49% of the health impact in the last decade. In addition, the productivity of pharmaceutical innovations has increased 30% between 1980 and 2009.

3.1 Data

To construct health impact measures for every drug, I use four main datasets: the FDA's Drugs@FDA website, the Medical Expenditure Panel Survey (MEPS), the National Ambulatory Medical Care Survey (NAMCS), and the Tufts Medical Center Cost-Effectiveness Analysis Registry (CEAR).

The FDA's website lists every FDA-approved drug by approval date, chemical type, and new drug application (NDA) number. The chemical type identifies whether an approval is a novel innovation (NME) or an incremental innovation (new active ingredient, new dosage form, new combination, new formulation or manufacturer, and new indication). Using the FDA's Orange Book, I map each FDA-approved drug (identified by an NDA) to National Drug Codes (NDCs), codes that uniquely identify drugs by manufacturer, product code, and packaging.

Using the NDCs, I match the FDA data with observations from the MEPS. MEPS is a large-scale collection of national representative surveys with detailed information on prescription utilization and demographics. The dataset is a panel with two years of data for each individual in the dataset. The prescription data contains around 144,000 individuals and 3.5 million prescriptions per year, from 1996 to 2012. It collects data on the individual taking the drug, drug name, NDC, ICD-9 disease category the drug is intended to treat, and prescription date. This data matches to individual data files in the MEPS, which contains information on every medical condition an individual has, listed by ICD-9 disease category. I construct 19 disease classes based on the ICD-9 classification and assign each drug to its primary disease class. The disease classes are listed in Appendix C.

The National Ambulatory Medical Care Survey (NAMCS) is an annual survey of randomly selected private-practice doctors that collects prescription data from 1980 to 2010.[14] Each doctor is surveyed about a random selection of their patient visits and lists all prescriptions, as well as the doctor's diagnosis listed as an ICD-9 code. I map the ICD-9 code to the 19 drug classes listed in Appendix C. In all, there are 3.2 million prescriptions in the NAMCS data.

For the efficacy measurement, I use the Tufts Medical Center Cost-Effectiveness Analysis Registry (CEAR). Hult and Philipson (2014) provide a detailed description of CEAR with an analysis of the trends

---

[14]There is no data from 1982 to 1984 or from 1986 to 1988.

in the dataset.[15] CEAR includes 1,377 pharmaceutical cost-utility analyses in the peer-reviewed medical literature. It is intended to be a comprehensive dataset of all cost-utility articles analyzed by trained professionals, who rate the quality of the study and provide information about the quality level and quality relative to the standard of care found in the study. The dataset lists the drug's name or active ingredient; the drug's disease class, which can be uniquely mapped into my 19 disease classes; and the year of the study. The dataset includes studies from 1977 to 2011, with 11% of the studies coming before 2000. The year is not necessarily the year the treatment variable became available, but since I aggregate all statistics within a decade, the timing is not a significant issue.

3.2 Measurement of Components of Incremental Innovation

I construct a dataset that measures the health impact of of each FDA-approved drug innovation from 1980 to 2009. Each drug innovation has six measurements that correspond to the six variables in the health impact produced by new innovations:

$$\Delta H_y = \sum_{j \in D_y} [\Delta q_j a_j e_j + \Delta a_j q_j e_j + \Delta e_j q_j a_j]$$

Adherence

Adherence (a_j) and quantity (q_j) are measured from the MEPS.[16] For chronic conditions—conditions that persist across survey years—adherence for drug j is measured by drug persistence, which is the fraction of patients who remain on drug j across both survey years. Individuals who switch to a different drug are omitted from the calculation.[17] In other words, if 100 individuals fill a prescription for drug j in their first year of the survey, 80 of those individuals still have the condition in the second year of the survey, and 60 of those individuals fill a prescription for drug j in the second year of the survey, then the adherence rate is $60/80 = 75\%$.

For conditions that do not persist across survey years, adherence for drug j is measured by a medical possession ratio (MPR). MPR is the average fraction of days that a patient had his prescription filled over

---

[15]I use the same data cleaning process as Hult and Philipson (2014).

[16]Appendix D has a more detailed description of the adherence calculation and a robustness check with different adherence specifications.

[17]The drug condition for adherence is measured by the ICD-9 condition. The denominator in the adherence calculation is all individuals who took drug j in the first year of the survey, had the same ICD-9 condition in the second year of the survey that drug j is intended to treat and did not switch from drug j to a different drug to treat the same ICD-9 condition. The numerator is the number of people who are in the denominator and also filled a prescription for drug j in the second year of the survey.

the course of the survey. In other words, if the average person had his prescription for drug j filled for 75 out of the 100 days the person was in the survey, then his adherence rate is 75%.

The average MPR is 70% and the average drug persistence rate is 74%, which is consistent with adherence estimates in the literature (Briesacher et al., 2008). The average innovation increases adherence by four percentage points.

Quantity

Quantity for drug j is measured as the number of users for that drug over the first 14 years the drug is on the market. The measure is constructed from three components. The appendix discusses the details of the quantity measurement and the data used. The first component is the total number of nongeneric, non-refilled prescriptions in year y, which is the number of people who take a prescription in year y multiplied by the number of different nongeneric drugs that person takes in year y. The second component is the share of this total prescription measure that comes from each disease class in year y. The third component is the share of each disease class that comes from drug j in year y.

Multiplying these three components together and summing them over the first 14 years after a drug's approval creates an estimate of the number of people who take drug j.[18] I measure quantity in this way because I have to match data across two drug level datasets. This method eliminates any difference in aggregate drug levels across the datasets that would influence health impact trends.

The first component, total prescriptions, mainly uses data from Census (2012). The second component, disease class shares, uses data from NAMCS. The third component uses NAMCS data from 1980 to 1995 and MEPS data from 1996 to 2012. If a drug has been on the market for fewer than 14 years, then I use a regression with drug fixed effects, fixed effects for the number of years since approval, and a year time trend to predict future quantity.

As a simple example, consider drug j in disease class c that was approved in 1990. If one million, non-refill prescriptions were filled in 1990, 1% of those prescriptions were filled for disease class c, and 10% of disease class c prescriptions were for drug j, then there are 1,000 users for drug j in 1990. Summing over this calculation for each year from 1990 to 2003 produces my measure of quantity.

Efficacy

CEAR contains a measure of the efficacy of a drug in QALYs, a frequently used measure of the efficacy of a medical treatment. A QALY measures the increase in the quality and quantity of life that treatments provide to patients. Since CEAR does not contain all FDA-approved drugs, I use the average quality measures within

---

[18]Table 6 in Appendix D shows why 14 years is a reasonable time frame and does a robustness check using 10 and 20 years.

a disease class, broken into either novel or incremental innovation, as the estimate for q_j.[19] The average drug adds 0.3 QALY over the SOC, which is a 4% increase in quality.

Changes in Adherence, Quantity, and Efficacy

For the measures of how a drug changes adherence (Δa_j), quantity (Δq_j), and efficacy (Δe_j), I compare the estimates for drug j relative to the treatment that would be prescribed if drug j had not been approved. This counterfactual treatment can be thought of as the SOC that existed before drug j.

For adherence, the SOC is the adherence rate of the other already-approved drugs that treat the same condition as drug j. Therefore, the change in adherence for drug j is the difference between the adherence rate for drug j and the average adherence rate for drugs approved before drug j but are in the same disease class as drug j.

The change in the quantity that results from drug j is the number of users for drug j minus the number of users that drug j crowds out from other drugs and a disease class time trend. Crowd-out is measured as the decrease in users by the other drugs in drug j's disease class after drug j hits the market.

For efficacy, the SOC is determined in the cost-utility analysis and the change in quality from drug j is directly measured in the CEAR data.

Approximately 30% of new innovations between 1980 and 2009 do not show up or do not have enough observations to measure adherence and quantity with the MEPS data. These observation get the average health impact within the disease class and innovation type.

3.3   Decomposition Results

With the adherence, quantity, and efficacy estimates, estimating the health impact of new innovations in year y (ΔH_y) is straightforward. I estimate this measure for all new innovations in each decade (1980s, 1990s, and 2000s) and break them down into novel and incremental innovations. Table 1 lists the results. Recall that the index measures the health impact added, not the level, so it tells us how much all new innovations in the average year of a decade increased quality relative to the average novel innovation in the 1980s. The average health impact of novel innovation in 1980s is normalized to 1. The second and fourth columns in Table 1 list the health impact of a novel innovation $\left(\frac{\Delta H_y^N}{\Delta H_{1980}^N} \right)$ and an incremental innovation $\left(\frac{\Delta H_y^I}{\Delta H_{1980}^N} \right)$ relative to the health impact of a novel innovation in the 1980s.

---

[19]To identify whether a study is for a novel or incremental innovation, I match the CEAR study to all FDA approvals with the same drug name or active ingredient. If a drug had an incremental innovation before the study year and the novel innovation occurred more than ten years before the study, then I treat the drug as an incremental innovation. Otherwise, the drug is treated as a novel innovation. See the Appendix for a robustness check on the ten-year time frame.

y	Average Health Impact per Novel Innovation	Novel Innovations Per Year	Average Health Impact per Incremental Innovation	Incremental Innovations Per Year	Share of Health Impact from Incremental
1980s	1.0	18.7	0.1	51.7	20%
1990s	1.5	30.6	0.3	53.4	23%
2000s	2.0	20.7	0.6	63.7	49%

Note: All health impacts are relative to a novel innovation in the 1980s. The health impact of a novel innovation is $\Delta H_y^N / \Delta H_{1980}^N$. The health impact of an incremental innovation is $\Delta H_y^I / \Delta H_{1980}^N$.

Table 1: Overall Gain Across Decades

By contrast, the health impact of a novel innovation in the 1980s is 1. In the 1990s, the 1.5 value in the second column means that the average novel innovation in the 1990s produced 1.5 times the value of the average novel innovation in the 2000s. The 0.1 value for an incremental innovation in 1980s means that each new incremental innovation is worth 10% of the health impact of a novel innovation in the same decade. Accounting for the number of new innovations of each type (there are 2.8 times more incremental than novel innovations), incremental innovation accounts for 20% of the health impact of all innovations during the 1980s. Incremental's share rose from 20% to 49% over the three decades.

The three main takeaways from Table 1 are that incremental innovation accounts for 49% of the health impact of new innovations, the health impact of innovations increased substantially over time, and incremental innovation's share rose faster than novel innovation's share. Omitting incremental innovation would not only underestimate the level, but also underestimate the trend in the health impact of innovation. This table also shows that each novel innovation is more valuable than each incremental innovation. However, there are two to three times more incremental innovations, which means that in recent years incremental innovations accounted for just under half of the health impact of all innovations.

	Share of Novel Innovation's Health Impact	Share of Incremental Innovation's Health Impact
Efficacy	27%	23%
Adherence	26%	29%
Quantity	47%	48%

Table 2: Overall Gain Broken into Components

Table 2 breaks down the health impact measure into how much is being driven by changes in the quantity, efficacy, and adherence. Although a new innovation's effect on efficacy often gets the most attention, most of the health impact from innovations comes from its ability to increase the quantity and the adherence. The quantity is the biggest driving force in the health impact of innovation and adherence plays a major role in the health impact of incremental innovations.

3.4 Measurement of R&D Expenditures

To estimate productivity, I need estimates of the level of R&D spending that went into new innovations. Pharmaceutical R&D comes from private companies and federal funding. Private measures are obtained from PhRMA, a trade group that represents almost every major company in the pharmaceutical industry.[20] They compile R&D measures from PhRMA companies from 1975 to 2012 provide a frequently used measure of pharmaceutical R&D.[21] Federal measures are obtained from the NIH budget appropriation estimates.

To adjust for inflation, I used the NIH's Biomedical R&D Price Index, which is designed to measure the change in the price of NIH biomedical research inputs, such as personnel services, supplies, and equipment.[22] This index rose nearly twice as much as the CPI inflation index since the 1960s due to increases in biomedical prices.

The second necessary adjustment is to match R&D spending with its expected year of output. R&D output and R&D spending do not occur in the same year; it often takes at least 14 years of investment before an innovation hits the market. Therefore, I match R&D spending to R&D output by using an estimate from Paul et al. (2010) of what fraction of a drug's total R&D is spent at each year of a drug's R&D life cycle. By knowing how many innovations are in each year of the pipeline and measuring how much each year in the R&D process accounts for total spending, I map pharmaceutical R&D into the year in which the drug hits the market. See Appendix D for a more detailed description. My timing-adjusted R&D measure says how much R&D went into the investments approved in a given year. If no investments were approved, then the timing–adjusted R&D would be zero. If three innovations hit the market, then it would consist of an estimate of what proportion of the total amount of R&D went into those three innovations over the past 14 years. This adjustement does not change the main findings of the paper.

3.5 Implied Productivity Trend

With estimates of R&D output measured in health impact and R&D spending, productivity is straightforward to measure since I define it as the ratio of R&D output to R&D spending. Table 3 shows that, for the 1980s, 1990s, and 2000s, the health impact of new innovations increased faster than R&D spending, which means that productivity increased. New innovations from the 2000s were three and a half times more valuable than the new innovations from the 1980s, and R&D spending increased nearly threefold over this period. While the health impact of novel innovation grew over this period, productivity using only the health

---

[20]Until 2009, PhRMA counted all top 15 pharmaceutical R&D companies as members, but Roche is no longer a member after merging with Genentech.
[21]See Scannell et al. (2012).
[22]Cockburn (2007) and Austin (2007) also use this index for adjusting pharmaceutical R&D.

impact of novel innovation declined. Productivity only increased when incremental innovation is included in the productivity measurement.

y	Health Impact of Innovations (Relative to 1980)	Adjusted R&D (Billions of USD)	Adjusted R&D (Relative to 1980)
1980s	1.0	18.8	1.0
1990s	2.6	34.9	1.9
2000s	3.5	50.1	2.7

Note: All health impact measures are relative to 1980: $\Delta H_y / \Delta H_{1980}$.

Table 3: Productivity of Pharmaceutical R&D

There are several forces that should make productivity gains difficult, even with increasing technology; these forces are discussed in the pharmaceutical literature. The first effect, the low-hanging fruit effect, says the disease classes with the lowest cost of innovation (the lowest δ_j in my model) get innovation first. This means that, in the future, companies will either have to compete in crowded low-cost disease classes or in high-cost disease classes. This effect should push up the cost per innovation. The second effect is related: since past innovation raised quality, future innovation has to have an increasingly higher quality level, which raises the cost of innovation. An industry like computers and electronics doesn't experience the same effects, because while the number of potential disease categories humans have is relatively constant over time, the number of categories available to computers and electronics grows rapidly with innovation. Innovation in this industry not only improves desktop computers, but creates new devices, from smartphones to tablets or wearable electronics, that did not even exist as potential categories decades ago.

4   Calibrating the Future Importance of Incremental Innovation

In this section, I outline a model of a firm's decision to invest in either novel or incremental innovation. The model is based on fact that novel innovation is a necessary precursor to incremental innovation, and firms equalize the return between R&D spending on novel and incremental innovation. The return to incremental innovation is proportional to the ratio of the stock of novel to incremental innovation. This ratio is a measure of how cheap incremental innovation is to produce, since a higher stock of novel innovation makes incremental innovation cheaper and a higher stock of incremental innovation makes incremental innovation more expensive to produce. The coefficient on this ratio, θ, determines the optimal share of novel and incremental R&D spending.

4.1 Model

The model I calibrate in this paper is based on the Schumpeterian growth model used by Acemoglu and Linn (2004) since consumers have no preference over whether a drug is novel or incremental. The key difference between their model and mine is that I add the interaction between novel and incremental innovation, as opposed to just considering novel innovation. I build this interaction into the model by requiring firms to initially create novel innovation within a disease class. Once a firm has a novel innovation in a disease class, that firm can innovate through incremental innovation.

The demand side of the model is exactly the same as in Acemoglu and Linn (2004). There are $i \in I$ individuals with $j \in J$ disease classes with Cobb-Douglas preferences:

$$\int_0^\infty \exp(-rt)[c_i(t)^{1-\gamma}(q_j(t)x_{ji}(t))^\gamma]dt$$

where r is the discount rate; x_{ji} is the consumption of drug j; $c_i(t)$ is a basic good that can be used for consumption, production of x_j, and R&D expenditures; and $\gamma \in (0,1)$.

The drug with the highest-quality q_j in each drug class captures the entire market. Prices of the basic good are 1, and prices of drugs are determined by the next-best drug, shown in Acemoglu and Linn (2004) to be:

$$p_j(t) = \lambda$$

since demand for drug j according to the Cobb-Douglas preferences is $X_j(t) = \frac{\gamma Y_j(t)}{p_j(t)}$.

A firm may choose to invest in novel R&D (z_j^N), which has a flow rate of innovation $(n_j(t))$ described below, or the firm may invest in incremental R&D (z_j^I), which has a flow rate of innovation $(i_j(t))$:

$$n_j(t) = \delta_j z_j^N(t)$$
$$i_j(t) = \zeta(N_j(t), I_j(t))\delta_j z_j^I(t)$$

The flow rate is a measure of how many new innovations are approved in a year. Each flow rate is a function of δ_j, which is the factor that converts R&D into innovation and varies by disease class. Incremental innovation has a second factor, $\zeta(N_j(t), I_j(t))$, which is a function of the stock of novel innovation, $N_j(t) = \sum_{s=0}^t n_j(s)$, and the stock of incremental innovation, $I_j(t) = \sum_{s=0}^t i_j(s)$, with $\frac{\partial \zeta(N_j(t), I_j(t))}{\partial N_j(t)} > 0$ and $\frac{\partial \zeta(N_j(t), I_j(t))}{\partial I_j(t)} < 0$. In other words, the return on investing in incremental innovation is higher with a higher stock of novel innovation, because there are more molecules from which to create incremental innovation, and it is negatively related to the stock of incremental innovation, since already-approved incremental innovation crowds out

future incremental innovation. The δ_j incorporates that it is easier to innovate in certain disease classes, and this effect is assumed to be the same for both types of innovation. Appendix E works out the details of solving the model.

When $\zeta(N_j(t), I_j(t)) = 1$, firms are indifferent to the choice between novel and incremental innovation and:

$$z_j^I + z_j^N = \frac{\delta_j(\lambda - 1)\gamma Y_j - r}{\delta_j}$$

If $\zeta(N_j(t), I_j(t)) < 1$, then firms only invest in novel innovation. R&D spending is $z_j^N = \frac{\delta_j(\lambda-1)\gamma Y_j - r}{\delta_j}$, which produces new innovations: $n_j = \delta_j(\lambda - 1)\gamma Y_j - r$. Similarly, if $\zeta(N_j(t), I_j(t)) > 1$, then firms only invest in incremental innovation. R&D spending under this scenario is $z_j^I = \zeta(N_j, I_j)^2 \frac{\delta_j(\lambda-1)\gamma Y_j - r}{\delta_j}$ which produces new innovations: $i_j = \zeta(N_j, I_j)(\delta_j(\lambda - 1)\gamma Y_j - r)$.

$$n_j + i_j = (\delta_j(\lambda - 1)\gamma Y_j - r)$$

If I assume a function form of ζ being proportional to the ratio of the stocks, $\zeta(N_j, I_j) = \theta \frac{N_j}{I_j}$, then I can solve for the steady-state. The steady-state R&D spending and innovation flows are:

$$
\begin{aligned}
z_j^N &= \frac{1}{1+\theta} \frac{\delta_j(\lambda - 1)\gamma Y_j - r}{\delta_j} \\
z_j^I &= \frac{\theta}{1+\theta} \frac{\delta_j(\lambda - 1)\gamma Y_j - r}{\delta_j}
\end{aligned}
$$

and

$$
\begin{aligned}
n_j &= \frac{1}{1+\theta}\delta_j(\lambda - 1)\gamma Y_j - r \\
i_j &= \frac{\theta}{1+\theta}\delta_j(\lambda - 1)\gamma Y_j - r
\end{aligned}
$$

In this steady-state, firms invest $\frac{\theta}{1+\theta}$ of their R&D on incremental innovation and $\frac{1}{1+\theta}$ of their R&D on novel innovation, which keeps the returns equalized between novel and incremental innovation and $\zeta(N_j, I_j) = 1$.

These equations reveal three observations. First, firms equalize the return between novel and incremental innovation. If the stock of novel innovation is large relative to the stock of incremental innovation,

$\zeta(N_j, I_j) > 1$, then firms get a higher return on incremental innovation and shift their resources to incremental investment. This investment will reduce $\zeta(N_j, I_j)$ until novel and incremental innovation have the same return. If the relative stock of novel innovation is low, then firms will invest in novel innovation, which raises $\zeta(N_j, I_j)$ and increases the return on incremental innovation in the future. The trade-off depends on the specific functional form of ζ.

Second, novel innovation is a bad proxy for innovation because novel and incremental innovation are production substitutes in the short term. If a firm has higher output in novel innovation in the previous period, then it will choose to invest more in incremental innovation. It would appear as if innovation slows when a firm switches to novel innovation, but the firm may actually have higher productivity. Consider PDUFA in 1992, which made it easier to get drugs through the FDA approval process and increased innovations, and especially novel innovations, for a short period in the mid-1990s. After the spike in novel innovations, there was a higher incentive for firms to invest in incremental innovations, which means that considering only novel innovation underestimates productivity during this period.

Third, firms do not invest when they have the highest-quality drug in the disease class. As a result, if there is only one firm investing in a particular disease class, then it will not crowd out its own highest-quality drug by producing more innovations. As a result, novel stocks can build up across disease classes when there isn't much competition. Orphan drugs, which treat disorders affecting fewer than 200,000 people, are an example of a situation where there may not be competition within the disease class, so incremental innovations are not funded.

4.2 Effects of Policy and Demographic Changes

The model has several implications for innovation in the pharmaceutical industry and related public policy. Demographic changes affect the incentive to create more aggregate innovation, but do not affect the share of innovation that comes from novel or incremental innovation. An increase in market size can be seen in the models as an increase in γY_j. When market size increases, both novel and incremental innovation increase. Since β does not increase, the relative share of novel and incremental innovation does not change. This result also holds for changes in insurance coverage, such as Medicare Part D, which greatly increased prescription insurance coverage for the elderly. Increases in prescription insurance coverage act exactly like a change in market size and do not affect β or the relative share of novel or incremental innovation.

For policy changes that affect incremental and novel innovation differently, consider a factor (ρ), that affects the flow rate of novel innovation:

$$n_j(t) = \rho\delta_j z_j^N(t)$$

$$i_j(t) = \zeta(N_j(t), I_j(t))\delta_j z_j^I(t)$$

where $\rho = 1$ with no policy change. As a result, the share of novel innovation is $\frac{\rho}{\rho+\theta}$ and the share of incremental innovation is $\frac{\theta}{\rho+\theta}$.

Some policy changes affect ρ, such as the Orphan Drug Act. Since these drugs are more likely to be first-in-class drugs and, as a result, have a significantly higher rate of being new molecules, the Act increases the rate of novel innovation more than incremental innovation ($\rho > 1$). If the lack of increase in novel innovations is due to inefficiencies or increased thresholds in the FDA approval or clinical trials program, then $\rho < 1$. As a result, firms will increase the relative share of incremental innovation. As incremental innovation increases, its stock increases, reducing the return on future incremental innovation. As a result, while novel innovation decreases initially, incremental innovation increases, diminishing any potential aggregate decrease in innovation.

In recent decades, the pharmaceutical industry has been characterized by a large number of mergers and acquisitions.[23] These mergers affect the stocks of innovation, because two merging firms combine their stocks of novel innovations. The effect of mergers is ambiguous in the model, but they could have important effects on innovation. Consider if large pharmaceutical firms merged with smaller firms after creating new molecules. This type of merger occurs because larger pharmaceutical firms have the resources and knowledge needed to get this molecule through the long and expensive FDA-approval process. In this scenario, larger firms would then produce the incremental innovation once the drugs are approved. The effects of this merger would be that larger firms would seem less productive if only novel innovation were measured, and if the health impact of future incremental innovation is not included, the health impact of the acquisition would be undervalued.

5   Empirical

This section estimates the model outlined in the previous section. I find empirical support for the main assumption of the model, which is that the flow of incremental innovation depends on how much incremental innovation has already been done on the stock of novel innovation. I find the optimal ratio of the stock of novel to incremental innovation is 55%. As a result, if the ratio is below 55%, then firms have a higher

---

[23]See Danzon et al. (2007).

return to invest in novel innovation, and if the ratio is below 55%, then firms have a higher return to invest in incremental innovation.

5.1 Empirical Strategy

The model assumes that the flow of novel and incremental innovation depends on the stocks of both types of innovation, $\zeta(N, I)$. To test this assumption, I use the same functional form, $\zeta(N, I) = \theta \frac{N}{I}$, that I use in the model. The theory assumes the flow of novel innovation would be negatively impacted by the ratio, and incremental innovation would be positively impacted by the ratio. The reasoning is that, if the stock of novel innovation is high relative to incremental innovation, then there are higher returns to capturing the lower-cost incremental innovation rather than investing in more novel innovation.

I test this assumption empirically using a model similar to the empirical strategy in Acemoglu and Linn (2004), but I add the ratio of the stocks. In the specification, innovation is measured as the number of FDA-approved novel innovations in year group t and drug class c (n_{ct}) as well as the number of FDA-approved incremental innovations (i_{ct}). Market size, which has been shown in several papers to affect innovation, is included as potential market size, M_{ct}.[24] Potential market size is a measure of the number of users multiplied by their marginal willingness to pay, but it allows for changes in population and income while maintaining a constant age profile of use and expenditures to deal with potential endogeneity issues.[25] It is measured by:

$$M_{ct} = \sum_a u_{ca} \cdot \iota_{at}$$

where u_{ca} is a time-invariant measure of the fraction of drug expenditures for individuals in age group a that come from drugs in disease class c, and ι_{at} is the income in age group a at time t.

The ratio of the stocks, $\left(\frac{N}{I}\right)$, is included as a determinant of the flow of innovation and as a test of the model. Year group controls are also included in μ_t, ϕ_c are class-fixed effects, and X'_{ct} represents any other controls included in the estimation.

I construct a Poisson model for the count of new innovations incorporating these variables:

$$E[n_{ct} | \phi_c, \bar{X}_c, \bar{n}_c] = \exp(\alpha_n \cdot \log M_{ct} + \beta_n \cdot \frac{N_{ct}}{I_{ct}} + X'_{ct} \cdot \theta_n + \mu_{n.t})$$

$$E[i_{ct} | \phi_c, \bar{X}_c, \bar{i}_c] = \exp(\alpha_i \cdot \log M_{ct} + \beta_i \cdot \frac{N_{ct}}{I_{ct}} + X'_{ct} \cdot \theta_i + \mu_{i,t}) \bar{i}_c$$

---

[24] Acemoglu and Linn (2004), Cerda (2003), Blume-Kohout and Sood (2013), and Hult (2014).

[25] Since the effect of market size is not of primary importance, see Acemoglu and Linn (2004) for a full description of why this measure is used to deal with the potential endogeneity issue that higher-quality products will have larger market size.

The estimates of interest are the coefficients on the stock ratios (β_n and β_i). I expect β_n to be negative and significant if the assumptions of the model are correct, and β_i to be positive and significant. However, to obtain an unbiased estimate of the β's, I use the Hausman et al. (1984) transformation also used in Acemoglu and Linn (2004). This transformation is used because the class-fixed effects cannot be estimated consistently.

$$E[n_{ct}|\phi_c, \bar{X}_c, \bar{n}_c] = \frac{\exp(\alpha_n \cdot \log M_{ct} + \beta_n \cdot \frac{N_{ct}}{I_{ct}} + X'_{ct} \cdot \theta_n + \mu_{n.t})}{\sum_{\tau=1}^{T} \exp(\alpha_n \cdot \log M_{c\tau} + \beta_n \cdot \frac{N_{c\tau}}{I_{c\tau}} + X'_{c\tau} \cdot \theta_n + \mu_{n,\tau})}\bar{n}_c \tag{1}$$

$$E[i_{ct}|\phi_c, \bar{X}_c, \bar{i}_c] = \frac{\exp(\alpha_i \cdot \log M_{ct} + \beta_i \cdot \frac{N_{ct}}{I_{ct}} + X'_{ct} \cdot \theta_i + \mu_{i,t})}{\sum_{N=1}^{T} \exp(\alpha_i \cdot \log M_{c\tau} + \beta_i \cdot \frac{N_{c\tau}}{I_{c\tau}} + X'_{c\tau} \cdot \theta_i + \mu_{i,\tau})}\bar{i}_c \tag{2}$$

Equations (1) and (2) are estimated by quasi-maximum likelihood (QML) with an observation being a five-year-group for each disease class.[26]

5.2 Data

The data used to estimate the QML is similar to that used in Acemoglu and Linn (2004). The dataset consists of six variables: income measures (ι_{at}), the flow of novel innovation (n_{ct}), the flow of incremental innovation (i_{ct}), the stock of novel innovation (N_{ct}), the stock of incremental innovation (I_{ct}), and drug expenditure shares (u_{ca}).

Income data is measured by age group using the Annual Demographic File from the CPS for each year from 1970 to 2010.[27] The five age groups used are 0–4, 5–30, 31–60, 61–90, and 90-plus with family income equally divided between family members. The flow and stocks of innovation are from the FDA's Drugs@FDA website. The FDA lists every FDA-approved drug by approval data, chemical type, and new drug application (NDA) number. The chemical type identifies whether an approval is a novel innovation (NME) or an incremental innovation (new active ingredient, new dosage form, new combination, new formulation or manufacturer, and new indication). Drug expenditures are measured using the MEPS and are matched to the FDA data as discussed in Section 3. With this mapping between the FDA and MEPS dataset, I categorize each drug into one of 19 drug classes (listed in Appendix D) that correspond to ICD-9 disease categories.[28] Drug expenditure shares are computed as the fraction of drug expenditures for individuals in each of the five age groups, a, that come from drugs in each of the 19 disease classes.

---

[26]See Acemoglu and Linn (2004) for testing the effects of different models, such as a linear model, and including lags and leads using a similar model with similar data.

[27]The Annual Demographic File is also known as the March CPS file.

[28]Drugs are classified by the drug class that they most frequently treat.

	(1) Novel	(2) Novel	(3) Incremental	(4) Incremental
Ln Market Size	2.204*	2.294*	2.421*	2.322*
	(0.355)	(0.352)	(0.208)	(0.202)
Ratio of Novel to Incremental Stock	-0.696*		0.377*	
	(0.349)		(0.169)	
Year Group 80-84	0.499*	0.544*	0.0809	0.0554
	(0.155)	(0.153)	(0.0975)	(0.0969)
Year Group 85-89	0.425*	0.486*	0.316*	0.288*
	(0.139)	(0.135)	(0.0808)	(0.0798)
Year Group 90-94	0.415*	0.498*	0.0152	-0.0297
	(0.130)	(0.124)	(0.0831)	(0.0805)
Year Group 95-99	0.491*	0.555*	-0.0649	-0.0962
	(0.113)	(0.109)	(0.0757)	(0.0743)
Year Group 00-04	0.670*	0.696*	0.183*	0.171*
	(0.0971)	(0.0963)	(0.0628)	(0.0625)
Year Group 05-09	-0.468*	-0.449*	-0.206*	-0.210*
	(0.139)	(0.139)	(0.0752)	(0.0751)
Observations	152	152	152	152

Estimated by quasi-maximum likelihood (QML) with Hausman et al. (1984) transformation. Huber-White robust standard errors are reported in parenthesis. The dependent variable in columns 1 and 2 is the count of novel innovation approvals, and in columns 3 and 4 is the count of incremental innovation approvals. Approval counts are from the FDA, market size is from the MEPS and CPS. Estimates are weighted by total expenditure for the category in the MEPS. Year group 70-74 is the omitted group and the coefficients for Year Group 75-79 are not listed. Marginal Effect listed.
* $p < 0.05$

Table 4: Effect of Stock Ratio on Innovation

5.3 Results

The results are shown in the regression table below, with and without the ratio included. There are two main results. First, the results provide empirical support for the assumption of the model that the flow of novel and incremental innovation depends on the stocks of each type of innovation. Second, the results calibrate that the ratio of the stock of novel to incremental innovation should be 0.55.

Table 4 provides results for the estimation of the QML. The dependent variable is the number of novel approvals in the first two columns and the number of incremental approvals in the last two columns. Columns 1 and 3 include the ratio of the stock of novel to incremental innovation, and columns 2 and 4 exclude it. Errors for all QML estimates are corrected for heteroskedasticity with Huber-White.

Columns 1 and 3 show empirical support for the assumption of the model that the flow of novel and incremental innovation depends on the stocks of each type of innovation. The positive coefficient on the ratio says there are higher returns to incremental innovations when there are more novel innovations from

which to innovate. The negative coefficient on the ratio, when the dependent variable is the flow of novel innovations, says that there are lower returns to investing in novel innovation when there is more potential to create incremental innovation.

The interpretation of the magnitude of these coefficients is: if the ratio of novel to incremental stock went up one unit from one more novel innovation, then novel innovation in the next five-year group would be roughly 50% lower. Similarly, if the ratio went up one unit, the amount of incremental innovation in the next year group would be 50% higher. These coefficients produce a $\theta = 1.81$, which means that, in the steady-state, 65% of innovation is incremental and 35% is novel. These results produce a ratio of the stock of novel innovation to incremental innovation of 0.55. As a result, this estimation presents evidence of the assumptions of the model that the flows of innovation depend upon the relative size of the stocks of each type of innovation.

The coefficients on market size are lower but similar to the coefficients of 3.54 for novel innovation found by Acemoglu and Linn (2004). They differ because my estimation covers different years of data. The coefficient on the year group 2005–2009 shows that innovation may be slowing down in recent years, although the other year groups seem to feature a general increase in innovation over this period.

Table 5 includes a nonlinear version of the stock ratio, and a version including the level of the stocks separately. Columns 1 and 2 of Table 5 show novel innovation exhibits concavity, while incremental innovation does not. This result means that as the relative stock of novel innovation gets large, it has less effect on the flow of novel innovation. Firms will still engage in some novel innovation, even if they already have a relatively large stock of it. Columns 3 and 4 in Table 5 show that the ratio of the stocks is a more important determinant of innovation than the stocks are separately. The reason this ratio matters is that it measures how much potential exists for novel innovations. If the stock of novel innovation is high, but most of the novel innovation has already been captured—that is, the stock of incremental innovation is high—then the returns to incremental innovation are low.

6 Predictions

In the health impact estimation, I found, contrary to the consensus in the literature, the productivity of new pharmaceutical innovation increased by 30% from the 1980s through the 2000s. In this section, I use my model calibrations to predict future trends in innovation. I find that the productivity of new innovations will decline by 40% over the 2010s as a result of the relative decrease in incremental innovation and the decline in the health impact per novel innovation.

	(1) Novel	(2) Incremental	(3) Novel	(4) Incremental
Ratio of Novel to Incremental Stock	-1.413*	0.463		
	(0.493)	(0.297)		
Ratio Squared	0.234*	-0.0274		
	(0.102)	(0.0781)		
Novel Stock			0.00113	-0.00123
			(0.00136)	(0.000931)
Incremental Stock			-0.00486*	-0.00485*
			(0.00172)	(0.000933)
Ln Market Size	2.158*	2.435*	3.649*	4.007*
	(0.355)	(0.211)	(0.537)	(0.304)
Year Group 80-84	0.478*	0.0841	0.542*	0.0406
	(0.156)	(0.0979)	(0.154)	(0.0971)
Year Group 85-89	0.400*	0.319*	0.483*	0.253*
	(0.139)	(0.0813)	(0.135)	(0.0800)
Year Group 90-94	0.375*	0.0211	0.545*	-0.0110
	(0.130)	(0.0848)	(0.125)	(0.0807)
Year Group 95-99	0.463*	-0.0608	0.597*	-0.0776
	(0.113)	(0.0765)	(0.110)	(0.0745)
Year Group 00-04	0.657*	0.186*	0.687*	0.152*
	(0.0971)	(0.0632)	(0.0964)	(0.0625)
Year Group 05-09	-0.483*	-0.205*	-0.448*	-0.217*
	(0.139)	(0.0752)	(0.140)	(0.0758)
Observations	152	152	152	152

Estimated by quasi-maximum likelihood (QML) with Hausman et al. (1984) transformation. Huber-White robust standard errors are reported in parenthesis. The dependent variable in columns 1 and 2 is the count of novel innovation approvals, and in columns 3 and 4 is the count of incremental innovation approvals. Approval counts are from the FDA, market size is from the MEPS and CPS. Estimates are weighted by total expenditure for the category in the MEPS. Year group 70-74 is the omitted group and the coefficients for Year Group 75-79 are not listed. Marginal Effect listed.
* $p < 0.05$

Table 5: Effect of Stocks on Innovation

6.1 Innovation Prediction

My model is based on the assumption that incremental innovation is easier to create with a higher stock of novel innovation, and that incremental innovation is harder to create the more incremental innovation has taken place. The estimates in the empirical section predict that 64% of innovation will be incremental and 36% will be novel during the 2010s. This result produces a ratio of the stock of novel to incremental of 0.55. In 2005, the ratio of the stocks of novel to incremental innovation was 0.48 and, due to the high fraction of incremental innovation during this time—partially due to PDUFA—the ratio dropped to 0.45 by 2010. Since these stock ratios are below the steady-state estimate, the share of future novel innovation will likely increase.

I predict the flow of future novel and incremental innovation using the QML estimates. To obtain these predictions, I start by projecting current trends in market size and total spending within a disease class. Plugging these estimates into the empirical model with the estimated coefficients predicts that the flow of novel innovation will increase from 20% of all innovation to 32% across the 2010s. I see evidence of this increase in novel innovation from 2010 to 2013, since novel innovation increased by 17% while incremental innovation decreased by 16% from the 2000s to the 2010s. I expect this trend to continue through the rest of the 2010s.

These estimates assume no other factors affect innovation, such as changes in FDA procedures. As evidenced by the year group coefficients in the QML estimation, other factors can have a significant effect on innovation flows.

6.2 R&D Prediction

R&D data can be projected using recent R&D spending measures and projecting current trends. I predict timing-adjusted R&D spending will nearly double from the 2000s to the 2010s, as shown in Table 6, because R&D spending has more than doubled from the late 1990s to the early 2010s and much of that R&D is going to future innovations.

6.3 Health Impact per Drug Prediction

I predict the health impact per drug using the trends from 1980 to 2009. These trends predict a 45% increase in the health impact per novel innovation between the 2000s and the 2010s and a 26% increase for novel innovations. Assuming the same health impact per drug between the 2000s and 2010s would further exacerbate the predicted decline in productivity.

6.4   Health Impact Prediction

Putting these estimates together produces a projection that continues the estimates of Table 3 to include predictions for the 2010s. This table shows that, barring significant increases in the health impact per innovation, there will be a significant decline of 60% in the productivity of pharmaceutical innovation during the 2010s. This decline is mainly driven by the enormous increase in R&D spending growth, but is also due to the decrease in the number of innovations as a result of fewer incremental innovations. I predict fewer incremental innovations because of the relatively low number of novel innovations during the 2000s.

y	Health Impact per Drug (Relative to 1980)	Adjusted R&D (Relative to 1980)	Incremental's Share of Health Impact
1980s	1.0	1.0	20%
1990s	2.5	1.9	23%
2000s	3.5	2.7	49%
2010s	4.0	5.1	36%

Note: The estimate from the 2010s is a prediction.

Table 6: Projecting Future Trends

7   Conclusion

This paper investigates the role of incremental innovation in pharmaceuticals, which has largely been ignored by previous literature. I find that 49% of the health gains from pharmaceutical innovation in the last decade is due to incremental innovations. I also find that the pharmaceutical industry is not in a productivity crisis when innovation is measured as the health impact of incremental and novel innovations, but rather productivity increased 30% between 1980 and 2009.

Understanding the health impact of incremental innovations is important for several reasons. First, my findings contradict the view that me-too drugs—drugs that are chemically similar to existing drugs—produce little health impact and should be subject to less generous reimbursement. Although the literature generally considers the novel innovation form of me-too drugs, I show that chemically similar drugs affect health impact through adherence, number of users, and efficacy.

Second, my findings present a more accurate picture of the trends in pharmaceutical innovation. These trends show that the wealth of literature on the productivity crisis in pharmaceuticals undervalues innovation by excluding incremental innovation. In addition, my model shows that novel and incremental innovation act as production substitutes in the short term, so periods of low novel innovation, such as the 2000s, can be periods of very high productivity in innovation overall.

Third, my findings have direct health care policy implications. Changes in FDA policies that encourage novel innovation, including PDUFA, the Orphan Drug Act, and fast-track programs, have the added benefit of encouraging incremental innovation in the future. Since the health impact of a novel innovation is roughly equal to the health impact of the subsequent incremental innovations based on that molecule, measuring the effect of these policy changes with novel innovation captures only half of their health impact. My model also predicts that if there is a slowdown in novel innovation during the 2010s due to rising costs or regulatory issues, then the impact on innovation will be nearly twice as costly because less novel innovation today makes it harder to produce incremental innovation in the future.

Appendix

A. Case Study of HIV/AIDS

I illustrate the role of novel and incremental innovation with a case study of the pharmaceutical industry's response to HIV/AIDS.

HIV is a retrovirus that replicates in a host cell by utilizing the reverse transcriptase enzyme, which converts its RNA genome into DNA in a process known as reverse transcription. HIV/AIDS treatments consist of different types of inhibitors that work by interfering with different points of the HIV replication process. Reverse transcriptase inhibitors (RTIs), which can take nucleoside (NRTIs) and non-nucleoside (NNRTIs) forms, make up 68% of the market for antiretroviral medications. Protease inhibitors (PIs), which make up 31% of the market for antitretroviral medications, work by interfering with the essential protease enzyme that cuts long HIV proteins into shorter proteins. By inhibiting the protease enzyme, HIV cannot successfully replicate and infect additional host cells. AIDS is the final stage of the HIV infection, and successful treatments prevent HIV from reaching the AIDS stage.

The first case of AIDS was reported in 1981 and it was only in 1984 that the medical profession realized HIV caused AIDS.[29] From the mid-1980s until 1995, HIV/AIDS-related deaths skyrocketed until it was the leading cause of death in 25- to 44-year-old men in the United States from 1992 to 1995.[30] There are four phases of treatments for HIV/AIDS that correspond to the introduction of highly active antiretroviral therapy (HAART), which produced a major breakthrough in the treatment of HIV/AIDS (see Table 7). These are the pre-HAART phase from 1987 to 1996, the early HAART phase from 1997 to 1999 just after HAART became available, the late HAART phase from 2000 to 2005 which featured significant improvements to HAART treatments, and the CART (or combined antiretroviral therapy) phase from 2006 to 2014 which features drugs that combine the different HAART treatments into one combination pill.

In 1987, the FDA approved the first new molecule, or novel innovation, for the treatment of AIDS, Retrovir, commonly known by its generic name, AZT. Initially, Retrovir was approved for high dosage amounts and required high-dosage frequency; it had to be administered every four hours throughout the day and night. The high-dosage level of Retrovir produced small increases in life expectancy for patients, between several months to 1.6 years.[31] The drug also had serious drawbacks. It produced serious side effects, including serious blood problems (anemia and neutropenia), liver damage (hepatotoxicity), heart disease (cardiomyopathy), and muscle weakness (myopathy). Drug resistance was very high, which meant that the efficacy of the drug decreased with usage, and having to take a dose every four hours made adherence very

---

[29]See FDA's HIV/AIDS Historical Time Line.
[30]CDC (2014).
[31]See Becker et al. (2007).

difficult. The increase in life expectancy was enough for many patients to tolerate the side effects, although the adherence rate was incredibly low, with only 30%–50% of patients properly adhering. In addition, the original approval of Retrovir was only available in capsule form, so pediatric patients and some elderly patients could not take the drug.

Although it produced increases in life expectancy, the health impact of the novel innovation of Retrovir in 1987 reached a limited share of the HIV/AIDS population due to low adherence and restricted dosage forms. Three subsequent incremental innovations expanded the population of the drug to include pediatric patients and other patients who were unable to take capsules by approving an oral syrup form and an injectable form, and provided a higher-dose pill, which reduced pill burden and dosage frequency.

Between 1991 and 1997, there was significant novel innovation in HIV/AIDS treatments. Six additional molecules were approved as RTIs and five molecules were approved as PIs. However, these new molecules had little effect on patient outcomes because researchers did not know how to use them in effective treatments.

	Pre-HAART	Early HAART	Late HAART	CART
Years	1987-1996	1997-1999	2000-2005	2006-2014
Medications	Retrovir (AZT)[*]	Epivir and Zerit[*], Viracept[†], Norvir[‡]	Epivir and Zerit[*], Kaletra[†]	Altripa[§]
Adherence	30%–50%	55%–60%	65%–70%	over 85%
Probability of Living 25 Years	under 4%	50%	75%	over 75%
Doses	6 per day, 12 pills	5 per day, 20–30 pills	2 per day, 8 pills	1 per day, 1 pill
Drug Toxicity Issues	√	√	√	√
High Drug Resistance	√			
Serious Side Effects	√	√		
Drug Interactions	√	√	√	
Food Interactions		√		
Refrigeration		√		

Sources: Survival Probabilities are from Lohse et al. (2007) and CART rates are estimated by combining Lohse et al. (2007) and statistics from CDC (2013) that show HIV related deaths have fallen by over 65% between 2006 and 2010. Adherence rates are from Wall et al. (1995) (pre-HAART), Murphy et al. 2001 (early HAART), and Rebick and Walmsley (2012) (late HAART and CART). Early HAART pill counts are estimated from Ickovics and Meade (2002). Other pill counts are calculated from FDA dosage information.

[*] NTRI
[†] PI
[‡] PI Booster
[§] Combination Drug

Table 7: HIV/AIDS Treatment Phases

By 1997, researchers realized that they could use the already-approved FDA molecules much more effectively with HAART, a drug therapy that generally combined two NRTIs, a PI, and a PI booster. The PI booster (Ritonavir) prevented other enzymes from metabolizing the PI. As a result of the new treatment, the probability of living for 25 years if diagnosed with HIV at age 25 increased from less than 5% to 50%.[32] However, the new treatment was very difficult to take and continued to have significant side effects. For instance, early-period HAART treatments often required five doses per day with 20–30 pills per day. Some HAART pills, such as Norvir, required refrigeration and had significant food interactions so food could not be consumed an hour before or two hours after taking medication. In addition, individuals could develop drug resistance. As a result, adherence to early HAART treatments was only 55%–60%, even though adherence was demonstrated to have enormous effects on longevity.

From 2000 to 2005, incremental innovations continued to improve HAART treatments. Although more novel innovations occurred during this period, the most commonly used molecules (listed in Table 7) were approved by 1997, but these molecules were being used in different ways. PI drugs, such as Kaletra, had boosters incorporated in them by 2000, so instead of taking two PI drugs, patients would only take one. In addition, Kaletra had a new formulation that did not require refrigeration, had a higher absorption rate which reduced side effects and pill burden, and did not have food interactions. The most common NRTI drugs during this time were still Epivir and Zerit, but new NTRI combinations such as Combivir gained market share and reduced pill burden.
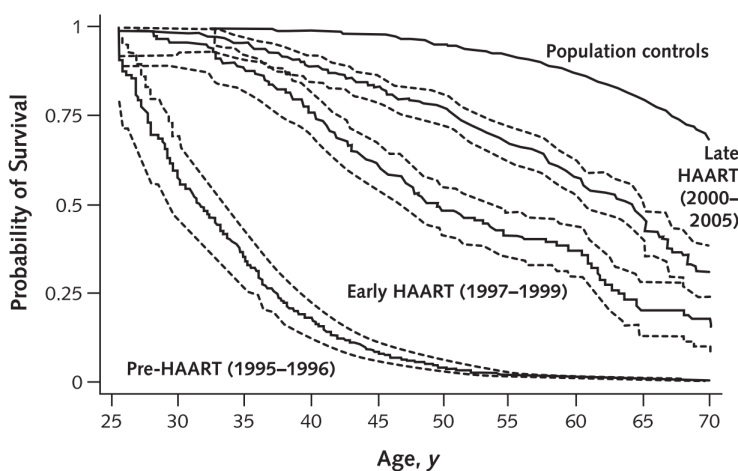
From: Lohse et al. (2007)

Figure 5: HIV Probability of Survival If Diagnosed at 25

---

[32]Lohse et al. (2007).

While patients were taking more complicated drug regimens during the early HAART phase, which included up to five doses and up to 30 pills per day, with pills refrigeration and food interactions, later HAART treatments were much easier. Incremental innovations reduced the dosage regimen to only two doses and around eight pills per day without food interactions or refrigeration, as well as fewer side effects. These innovations increased adherence to 65%–70% and greatly increased life expectancy. The probability of living 25 years after an HIV diagnosis increased from 50% during the early HAART phase to 75%.

These significant changes in mortality and morbidity were mainly due to incremental innovations, since the molecules used in the most common treatment had been around since before HAART treatments. It was changes in drug dosage, combinations of existing molecules, and new drug formulations that produced most of the change during the late HAART phase. Dosage changes increased adherence through reduced pill burden, number of doses, and side effects. Dosage changes also expanded the potential number of users by making drugs available to subsets of the population, such as pediatric patients, who could not take the drug as it was originally approved. Combining existing molecules increased efficacy through more effective treatments and increased adherence through easier dosage regimens. New formulation increased adherence and efficacy by eliminating the need for refrigeration and reducing drug and food interactions.

Incremental innovation continued through the introduction of CART, or combination antiretroviral therapy, during the 2006–2014 period. During this time, new drug treatments combined previously approved molecules into a one-pill, once-a-day HIV treatment. Atripla, for instance, combined three RTIs (a molecule originally approved in 1998 for Sustiva, a molecule approved in 2003 for Emtriva, and a molecule originally approved in 2001 for Viread). These new combination pills increased adherence to over 85% and decreased HIV-related mortality rates and morbidity.

The HIV/AIDS example shows that new molecules are a very important part of pharmaceutical innovation. Without new molecules, it would be impossible to develop effective HIV treatments. However, the creation of the molecules only provided part of the health impact of the innovation. When most of the most important drugs were originally approved, they did not affect the outcomes of patients. Instead, incremental innovation on those molecules unlocked a large amount of the health impact from innovation. Many of the most commonly used NTRI and PI drugs in the late HAART phase hit the market at a time when an HIV diagnosis carried a short life expectancy before HAART treatments. It wasn't until incremental innovation modified drugs and treatment regimens that HIV treatments saw a massive increase in quality and quantity of life.

B. Incremental Innovation Breakdown, Examples, and Approval Process

Dosage Change Breakdown

Table 8 breaks down dosage form innovations into four main classes: dosage route change (41% of dosage form approvals), dosage form change (35%), dosage amount change (13%), and duplicates (11%).[33] Duplicates are changes in dosage form that are identical to a drug sold by another company. For instance, albuterol sulfate, used to treat asthma, was approved in the same dosage amount, route, and form by Teva and 3M.

I break down each category into whether the FDA approval was granted priority status, and I further classify dosage form changes and dosage amount changes. I categorize dosage form changes into whether the change was between two forms that have a similar category or a different category. Dosage form categories include pills (tablets and capsules), inhalation (aerosols and sprays), topical (creams, lotions, gels, and ointments), and liquids (solutions and concentrates). Fourteen percent of dosage form changes are within a similar category (i.e., change from a tablet to a capsule, or cream to a lotion) while the rest are across different categories.

I define dosage amount changes as either small changes (less than a 40% change from any previously approved dosage amount) or large changes (more than a 40% change from a previously approved dosage amount). Only 5% of all dosage form changes are small changes in the dosage amount. Instead, most of these innovations are changes in the way the drug is taken or delivered, which can have a significant effect on the efficacy and adherence of the drug.

Main Category	Sub Category	Share of Approvals	Adherence Effect	Quality Effect	Population Effect
Route (41%)	Priority	9%	Possible	Yes	Possible
	Non-Priority	32%	Yes	Possible	Possible
Dosage Form (35%)	Similar Category	5%	No	No	No
	Different Category	26%	Yes	Possible	Possible
	Priority	4%	Possible	Yes	Possible
Dosage Amount (13%)	Small Change	5%	No	No	No
	Large Change	7%	Yes	Possible	Possible
	Priority	2%	Possible	Yes	Possible
Duplicate (11%)		11%	No	No	No

Note: Share of approval for main categories sum to 100%. Share of approvals in sub categories sum to share of approval in the main category.

Table 8: Dosage Form Breakdown

---

[33]Route change includes any innovation that had a route change, including those that also had dosage form changes or dosage amount changes.

Approval Process for Incremental Innovation

The FDA approval process differs between novel and incremental innovation. For a novel innovation to be available for sale in the United States it must complete a process of clinical studies and FDA approval which can take between eight and nine years on average.[34] When a drug is developed, it starts by having to pass a round of preclinical animal testing, followed by an Investigational New Drug application to the FDA. If it passes these rounds, it then goes through three phases of clinical testing. The first phase tests the drug's more frequent side effects, using 20 to 80 healthy volunteers, and takes 20 months on average. The second phase uses hundreds of patients to test the drug's effectiveness and takes 2.5 years on average. The third phase, which takes four years on average, uses thousands of patients to gather more thorough information on the drug's safety and effectiveness. If the drug passes all three of these phases, then it goes to the FDA for approval, which generally takes from six to ten months. Prior to approval, the FDA conducts meetings, reviews the application, and reviews the drug's labeling and production facilities. If the FDA approves the drug, it can be sold on the market, where it is subject to a final fourth phase of monitoring consisting of safety checks once the drug is used by the wider market.

Unlike NMEs, which have a standard FDA approval process, the innovation process for incremental innovation is not as standardized. It depends on what the innovation is and how it differs from what has already been approved. It can take anywhere from months to the same timeline as an NME if full approval is required. Hult (2014) finds that it takes five to six years for incremental innovations to hit the market.

Patent and exclusivity rights also differ between novel and incremental innovation. Novel innovations are usually patented during clinical trials and the patent lasts for 20 years. Since patenting occurs before drugs are approved, drugs generally have patent lengths of around 12 years. A patent prevents another company from using what is patented, such as a molecule, but does not prevent a competing company from creating a different drug that competes with the patented drug. Exclusivity rights, which are generally awarded to both novel and incremental innovation, prevent a competing company from introducing any drug with a similar indication. Exclusivity rights are generally five years for a novel innovation, seven years for an orphan drug (which treats a disorder affecting fewer than 200,000 people).

---

[34] All drug approval time averages are from Blume-Kohout and Sood (2013), which compares estimates across different sources.

C. Disease Classes

1	Infectious and Parasitic Diseases (ICD: 001–139)	11	Diseases of the Genitourinary System (ICD: 580–629)
2	Neoplasms (ICD: 140–239)	12	Complications of Pregnancy, Childbirth, and the Puerperium (ICD: 630–679)
3	Endocrine, Nutritional and Metabolic Diseases, and Immunity Disorders (ICD: 240–279)	13	Diseases of the Skin and Subcutaneous Tissue (ICD: 680–709)
4	Diseases of the Blood and Blood-Forming Organs (ICD: 280–289)	14	Diseases of the Musculoskeletal System and Connective Tissue (ICD: 710–739)
5	Mental Disorders (ICD: 290–319)	15	Congenital Anomalies (ICD: 740–759)
6	Diseases of the Nervous System (ICD: 320–359)	16	Certain Conditions Originating in the Perinatal Period (ICD: 760–779)
7	Diseases of the Dense Organs (ICD: 360–389)	17	Symptoms, Signs, and Ill-Defined Conditions (ICD: 780–799)
8	Diseases of the Circulatory System (ICD: 390–459)	18	Injury and Poisoning (ICD: 800–999)
9	Diseases of the Respiratory System (ICD: 460–519)	19	External Causes of Injury and Supplemental Classification (ICD: E and V)
10	Diseases of the Digestive System (ICD: 520–579)		

Table 9: Disease Classes

D. Data Appendix

Matching between Datasets

This section discusses merging between the different datasets. I start with all original new FDA drug approvals listed on the Drugs@FDA dataset. This definition does not include abbreviated new drug approvals (generic drug approvals) and supplements to approvals (changes in labeling or manufacturing). Within original approvals, I define novel innovations as NMEs and incremental innovations as new active ingredients, new dosage form, new combination, new formulation or manufacturer, or new indication. I do not include drugs already marketed without an approved NDA or over-the-counter switches. I also do not include drugs that are discontinued, which are drugs not listed in the FDA Orange Book. The FDA and FDA Orange Book, therefore, match all FDA approvals from 1980 to 2009 that are not discontinued.

	Fraction that Match at Drug Level	Fraction that Match at Disease Class
FDA	100%	.
FDA Orange Book	100%	.
MEPS	71%	93%
NAMCS	59%	93%
CEAR	.	93%

Match fractions are the fraction of FDA-approvals from 1980 to 2009 that match to drug observations in the different datasets. Drugs are matched to the MEPS by NDC and are matched to the NAMCS by name. Periods indicate that the dataset is not matched at that level.

Table 10: Fraction of 1980–2009 FDA Approvals that are Matched Across Datasets

There are 3,859 new innovations in the FDA dataset, 2,506 of which are from the 1980 to 2009 period. Of the drug innovations in that period, 1,792 (71%) can be matched to drugs in the MEPS by NDC and 1,469 (59%) can be matched to drugs in the NAMCS data by name. The drug's disease class is determined by the most common drug class listed in the MEPS and NAMCS data for which the drug is intended to treat. I calculate this measure for the first five years after a drug is approved to avoid off-label usage changing the disease class. I use drugs.com—a comprehensive website with detailed information on over 4,000 prescription drugs, over-the-counter medicines, and natural products—to determine the disease class for 271 drugs not listed in the MEPS or NAMCS datasets. The remaining 181 drugs, or 7% of the FDA approvals from 1980 to 2009, cannot be matched to a disease class and are omitted. I assume that if the drug is not common enough to show up in any of these datasets that its quantity is zero, which is equivalent to excluding it from the dataset. The 93% of drugs that are matched all have a disease class.

I only match CEAR at the disease class level, where 93% of drugs are matched, corresponding to the 93% of drugs that have a disease class.

Quantity and Adherence Measurement

This section describes the quantity and adherence measurement in more detail. Quantity is a measure of the number of prescription drug users over the first 14 years a drug is on the market. The quantity of drug j in disease category c is constructed from three components:

$$q_{j,c} = \sum_{y \in [0,14]} N_y \frac{N_y^c}{N_y} \frac{N_y^j}{N_y^c} = \sum_{y \in [0,14]} N_y s_y^c s_y^d$$

N_y is a measure of the number of people in year y multiplied by the number of different nongeneric drugs that person takes in year y (i.e. the total number of nongeneric, non-refilled prescriptions in year y), $s_y^c \equiv \frac{N_y^c}{N_y}$

is the share of those prescriptions that come from disease class c in year y, and $s_y^d \equiv \frac{N_y^j}{N_y^c}$ is the share of disease class c that come from drug j in year y.

I measure quantity in this way because I have to match data across two drug level datasets. This method eliminates any difference in aggregate drug levels across the datasets that would influence health impact trends.

N_y is measured from Farnsworth and Soejarto (1985) and Census (2012), with missing years linearly interpolated in between. Aggregate prescriptions increased by roughly 280% from 1980 to 2009. The disease class shares, s_y^c, are measured from NAMCS data. The drug level shares from 1996 to 2012 are measured from MEPS data. From 1980 to 1995, measuring drug shares is more complicated because the NAMCS data only lists drug name, which is not a unique identifier of a drug innovation. Therefore, from 1980 to 1995, I estimate the drug share by drug name in each year in the NAMCS. Then I divide this drug share across the different innovations that existed in that year and under that drug name, according to the relative share of those innovations in the MEPS dataset.

The data sources are listed in Table 11.

Years	N_y	s_y^c	s_y^d
1980–1995	Census (2012) and Farnsworth and Soejarto (1985)	NAMCS	NAMCS and MEPS
1996–2012	Census (2012)	NAMCS	MEPS

Table 11: Data Sources for Quantity Measurement

For instance, consider Retrovir, which had four innovations: a novel innovation in 1987 and incremental innovations in 1989, 1990, and 1995. To measure the drug shares in 1991—$s_y^d \equiv \frac{N_y^j}{N_y^c}$—I start by measuring the number of people prescribed Retrovir in 1991 in NAMCS. Assume for illustrative purposes that 10% of the disease class was prescribed Retrovir in 1991. Then I estimate what share of Retrovir went to each different innovation in the MEPS dataset. If an equal share of 25% of Retrovir prescriptions went to each of the four innovations, then the drug shares of the four Retrovir innovations would be 3.3%, 3.3% 3.3%, and 0%.[35] The 0% represents the drug innovation from 1995, which is not on the market in 1991, and the 3.3% is the share of each of the three innovations on the market in 1991.

---

[35] $.10 * \frac{0.25}{0.25 + 0.25 + 0.25} = 0.033 = 3.3\%$

	Quantity
Year	-0.208*
	(0.00757)
Year Squared	0.0000519*
	(0.00000189)
Drug Innovation FE	Yes
Years Since Approval FE	Yes
Observations	16213
R^2	0.888

OLS with dependent variable of quantity measurement. Observation is a drug-year.
* $p < 0.05$

Table 12: Quantity Prediction Regression

I measure quantity over the first 14 years a drug is on the market to control for the different amount of time different drugs have data available. I use 14 years as the cutoff because as Figure 6 shows, quantity drops off after 14 years on the market. Table 13 provides a robustness check using 10 and 20 years, instead of 14 years. For drugs that have been on the market less than 14 years, I project future spending with the regression in Table 12. This regression uses year trends, disease class fixed effects, and years since approval fixed effects to predict the future trend in quantity.

Coefficients plotted are from a regression with dependent variable of normalized number of users and controls for year, year squared, and disease class fixed effect. This regression is listed in Table 12. An observation is a new drug innovation from 1980 to 2009 and year. Year 0, or the year of approval, is the baseline. 95% confidence interval is bar and coefficient is dot.
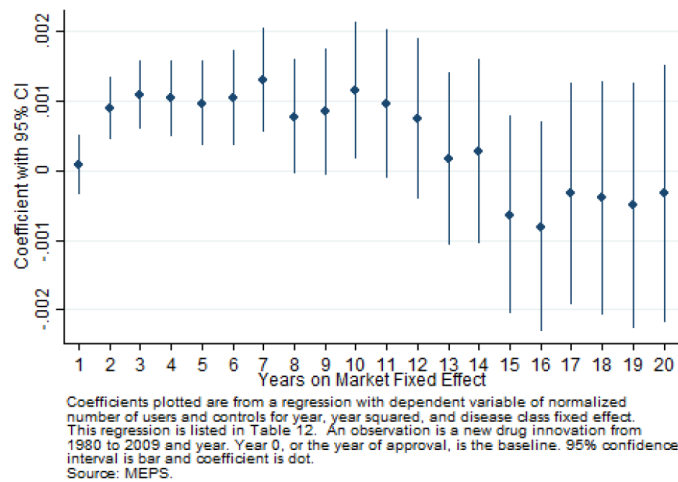Source: MEPS.

Figure 6: Number of Users as a Function of Years on the Market

Adherence is measured in two ways depending on whether the drug is used to treat chronic or non-chronic conditions.[36] Two-thirds of drugs in the dataset treat chronic conditions, defined as conditions in which at

---

[36]See Kaplan (2011) and Graden (2003) for discussions of measuring adherence with MEPS data.

least 10% of users take the drug across multiple survey years. If the condition is chronic, then adherence is measured as drug persistence, which is the fraction of patients who remain on drug j across both survey years. This fraction is calculated for individuals who had the same ICD-9 condition in both years of the survey that drug j was intended to treat for that individual.

If the condition is non-chronic, then adherence is measured as the medical possession ratio, or MPR. MPR is the ratio of days supplied of a drug to the number of days in the survey. From 2010 to 2012, days supplied was measured by the MEPS. Using the days supplied variable, I calculated the average days supplied for each drug by the drug's quantity, form, and strength. If the drug does not have observations in the MEPS between 2010 and 2012, then I use the average MPR for drugs with the same ICD-9, quantity, form, and strength.

Drugs that are not matched to the MEPS or that are taken as needed use the average adherence rate for drugs within their disease class. I do two robustness checks on the adherence measure in Table 13. I do a specification where adherence is only measured as drug persistence and one where adherence is only measured as MPR.

Role of Adherence in CEAR Efficacy Measure

One issue with using cost-effectiveness analyses is that adherence is that adherence is not always separated from the QALYs measure (see Rosen et al. (2009)). I do a robustness check to determine how much adherence measures in QALY could affect the health impact.

For the health impact measures I use:

$$\frac{\Delta H_y}{\Delta H_{1980}^N} = \sum_{j \in D_y} \left[\frac{\Delta q_j a_j e_j + \Delta a_j q_j e_j + \Delta e_j q_j a_j}{\Delta q_{N,80} a_{N,80} e_{N,80} + \Delta a_{N,80} q_{N,80} e_{N,80} + \Delta e_{N,80} q_{N,80} a_{N,80}} \right]$$

where the N superscript and 80 subscript denotes the average health impact for a novel innovation from the 1980s.

For this exercise, I consider measurement error of the type $\tilde{e}_i = e_i \psi(a_i)$, where e is the true measure and $\tilde{e}$ is the measure observed in the CEAR data. In the extreme case, $\psi(a_i) = a_i$ and adherence is measured though both e and a.

$$\frac{\Delta H_y}{\Delta H_{1980}^N} = \sum_{j \in D_y} \left[\frac{\Delta q_j a_j e_j \psi(a_j) + \Delta a_j q_j e_j \psi(a_j) + \Delta(e_j \psi(a_j)) q_j a_j}{\Delta q_{80}^N a_{80}^N e_{80}^N \psi(a_{80}^N) + \Delta a_{80}^N q_{80}^N e_{80}^N \psi(a_{80}^N) + \Delta(e_{80}^N \psi(a_{80}^N)) q_{80}^N a_{80}^N} \right]$$

Notice that this effect shows up in both the numerator and the denominator, dampening the effect. If all adherence rates were the same, then this effect would cancel out.

Table 13 provides a robustness check on the health impact measures considering the specification where $\psi(a_i) = a_i$.

Alternative Specification Robustness

This section provides robustness checks on several alternative specifications. It provides four measures over five alternative specifications. The four measures are how this specification changes the measure of health impact per novel innovation, health impact per incremental innovation, and total health impact relative to the main specification used in Table 1 and incremental innovation's share of health impact in the 2000s.

The five alternative specifications are measuring quantity over the first 10 years a drug is on the market, measuring quantity over the first 20 years a drug is on the market, measuring adherence only by persistence, measuring adherence only by MPR, and adjusting the calculations to allow for efficacy to include the adherence rate.

The measure of 1.5 in the second row of Table 13 says that under the alternative specification of measuring quantity over the first 10 years a drug is on the market, the change in the health impact per novel innovation from the 1980s to the 2000s increased by 150% (or a ratio of 1.5). Under the main specification, this measure increased by 200%.

Specification	Change in Health Impact per Novel	Change in Health Impact per Incremental	Change in Total Health Impact	Incremental's Health Impact Share in 2000s
Main Specification	2.0	7.2	3.5	49%
10 Year Quantity	1.5	5.1	2.5	49%
20 Year Quantity	3.3	12.2	5.9	50%
Adhere. Only Persis.	1.8	7.4	3.3	50%
Adhere. Only MPR	1.4	6.6	2.6	49%
Efficacy includes Adh.	1.7	7.2	3.1	51%

Changes are measured as the ratio of the health impact from the 2000s to the health impact from the 1980s. A value of 2.0 for health impact per novel innovation means that the health impact of a novel innovation in the 2000s is twice as large as the health impact of a novel innovation in the 1980s.

Table 13: Alternative Specifications

There are two main takeaways from Table 13. The first is that the only specification that has a significant effect is how many years are considered for the quantity measurement. The reason this specification matters is that, for drugs approved later in the sample, up to 85% of the years in the quantity calculation are

projections. Since the projection regression has a time trend in it, most of the increase in quantity is driven by projection trends.

The second is that the share of the health impact that comes from incremental innovation is very consistent across specifications. This measurement is so consistent because the different changes affect the measurement of both novel and incremental innovation, so they do not have a significant effect on the share.

R&D Matching

This section outlines the method for matching R&D to the innovation year. The goal is to match the spending with the health impact of the innovations produced from that spending. To accomplish this, I use an estimate from Paul et al. (2010) of the fraction of R&D spending at each year of the innovation process for an NME. Table 7 presents this R&D breakdown over time. Matching the number of innovations in each year with their drug share over time and the fraction of R&D spending in each year with aggregate R&D, produces a measure of how much R&D spending goes into the outputs in each year.
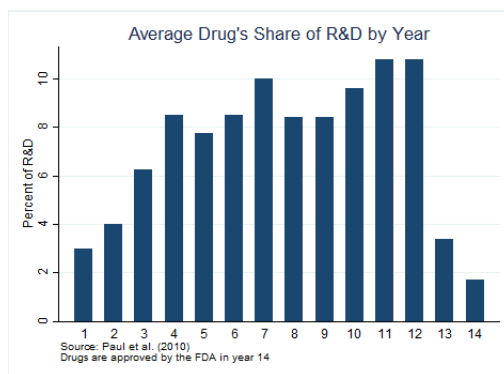

Figure 7: R&D Spending Per Year

E. Model Extras

The flow rates of innovation:

$$n_j(t) \quad = \quad \delta_j z_j^N(t)$$
$$i_j(t) \quad = \quad \zeta(N_j(t), I_j(t)) \delta_j z_j^I(t)$$

create value functions for the highest-quality firm for novel ($V_j(t|q_j)$) and incremental innovation ($W_j(t|q_j)$), which are a function of the per-period profits ($\pi_j(q_j)$) and the probability that another firm innovates to a higher quality level, reducing profits to zero.

$$rV_j(t|q_j) - \dot{V}_j(t|q_j) = \pi_j(q_j) - \delta_j \left(z_j^N(t) + \zeta(N_j(t), I_j(t)) z_j^I(t) \right) V_j(t|q_j) \tag{3}$$

$$rW_j(t|q_j) - \dot{W}_j(t|q_j) = \pi_j(q_j) - \delta_j \left(z_j^N(t) + \zeta(N_j(t), I_j(t)) z_j^I(t) \right) W_j(t|q_j) \tag{4}$$

The zero-profit condition from free entry implies that:

$$\delta_j V_j(t|q_j) = 1 \tag{5}$$

$$\zeta(N_j(t), I_j(t)) \delta_j W_j(t|q_j) = 1 \tag{6}$$

if a firm invests in either incremental or novel innovation. In addition, differentiating (5) and (6) with respect to time tells us that:

$$\dot{V}_j(t|q_j) = 0 \tag{7}$$

$$\dot{W}_j(t|q_j) = 0 \tag{8}$$

Plugging (5), (6), (7), and (8) into the value functions in (3) and (4) yields:

$$\frac{r}{\delta_j} = \pi_j(q_j) - z_j^N(t) - \zeta(N_j(t), I_j(t)) z_j^I(t)$$

$$\frac{r}{\zeta(N_j(t), I_j(t)) \delta_j} = \pi(q_j) - \frac{z_j^N(t)}{\zeta(N_j(t), I_j(t))} - z_j^I(t)$$

If $\zeta(N_j(t), I_j(t)) < 1$, solving these equations gives us (omitting time and quality notation for simplicity):

$$z_j^N = \frac{\delta_j(\lambda - 1)\gamma Y_j - r}{\delta_j}$$

$$z_j^I = 0$$

which means:

$$n_j = \delta_j(\lambda - 1)\gamma Y_j - r$$

$$i_j = 0$$

43

If $\zeta(N_j(t), I_j(t)) > 1$, solving these equations gives us (omitting time and quality notation for simplicity):

$$
\begin{aligned}
z_j^N &= 0 \\
z_j^I &= \zeta(N_j, I_j)^2 \frac{\delta_j(\lambda - 1)\gamma Y_j - r}{\delta_j}
\end{aligned}
$$

which means:

$$
\begin{aligned}
n_j &= 0 \\
i_j &= \zeta(N_j, I_j)(\delta_j(\lambda - 1)\gamma Y_j - r)
\end{aligned}
$$

F. Variables

h_j	Health Impact of drug j	N	Stock of Novel Innovation
H_y	Health Impact of Drugs in y	I	Stock of Incremental Innovation
ΔH_y	Health Impact of New Innovations in y	ζ	Incremental Innovation Flow Rate Adj.
q_j	Number of Users for drug j	V	Value Function of Novel Innovation
a_j	Adherence rate of drug j	W	Value Function of Incremental Inn.
e_j	Efficacy of drug j	π	Profits
Δq_j	New Inn. Change in Quantity	z^N	Novel Inn. R&D
Δa_j	New Inn. Change in Adh. rate	z^I	Incremental Inn R&D
Δe_j	New Inn. Change in Efficacy	δ	Innovation Flow Rate Adjuster
y	Year	M	Expected Market size
n	Flow of Novel Innovation	X	Regression Controls
i	Flow of Incremental Innovation	ρ	Policy Counterfactual Adjuster

Table 14: Variables

References

Acemoglu, Daron and Joshua Linn, "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics*, 2004, (3), 1049–1090.

Austin, David H, *Research and Development in the Pharmaceutical Industry*, Congressional Budget Office, 2007.

Becker, Gary, Kevin Murphy, and Tomas Philipson, "The Value of Life near Its End and Terminal Care," Technical Report, National Bureau of Economic Research 2007.

Berndt, Ernst R, Adrian HB Gottschalk, Tomas J Philipson, and Matthew W Strobeck, "Industry Funding of the FDA: Effects of PDUFA on Approval Times and Withdrawal Rates," *Nature Reviews Drug Discovery*, 2005, *4* (7), 545–554.

_ , **Iain M Cockburn, and Karen A Grepin**, "The Impact of Incremental Innovation in Biopharmaceuticals," *Pharmacoeconomics*, 2006, *24* (2), 69–86.

Blume-Kohout, Margaret E and Neeraj Sood, "Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development," *Journal of Public Economics*, 2013, *97*, 327–336.

Briesacher, Becky A, Susan E Andrade, Hassan Fouayzi, and K Arnold Chan, "Comparison of Drug Adherence Rates Among Patients with Seven Different Medical Conditions," *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 2008, *28* (4), 437–443.

Brosgart, Carol L, Thomas Mitchell, Edwin Charlebois, Rebecca Coleman, Steven Mehalko, Jamie Young, and Donald I Abrams, "Off-Label Drug Use in Human Immunodeficiency Virus Disease," *Journal of Acquired Immune Deficiency Syndromes*, 1996, *12* (1), 56–62.

CDC, "Mortality Slide Series," 2013.

Census, "Statistical Abstract of the United States," *Government Printing Office*, 2012.

Cerda, Rodrigo Andres, "Drugs, Population and Market Size." PhD dissertation, University of Chicago, Department of Economics 2003.

Cockburn, Iain M, "Is the Pharmaceutical Industry in a Productivity Crisis?," in "Innovation Policy and the Economy, Volume 7," MIT Press, 2007, pp. 1–32.

Couzigou, Carine, Caroline Semaille, Yann Le Strat, Roselyne Pinget, Josiane Pillonel, Florence Lot, Francoise Cazein, Daniel Vittecoq, J-C Desenclos, Aids Survival Study Group et al., "Differential improvement in survival among patients with AIDS after the introduction of HAART," *AIDS care*, 2007, *19* (4), 523–531.

Danzon, Patricia M, Andrew Epstein, and Sean Nicholson, "Mergers and Acquisitions in the Pharmaceutical and Biotech Industries," *Managerial and Decision Economics*, 2007, *28* (4-5), 307–328.

DiMasi, Joseph A and Henry G Grabowski, "The Cost of Biopharmaceutical R&D: Is Biotech Different?," *Managerial and Decision Economics*, 2007, *28* (4-5), 469–479.

_ , **Ronald W Hansen, and Henry G Grabowski**, "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics*, 2003, *22* (2), 151–185.

Farnsworth, Norman R and Djaja Doel Soejarto, "Potential Consequence of Plant Extinction in the United States on the Current and Future Availability of Prescription Drugs," *Economic Botany*, 1985, *39* (3), 231–240.

Frank, Richard G, "New Estimates of Drug Development Costs," *Journal of Health Economics*, 2003, *22* (2), 325–330.

Graden, Suzanne, "National Estimate of Cost of Illness for Hypertension and Non-Persistence with Drug Therapy Using the Medical Expenditure Panel Survey." PhD dissertation, The Ohio State University 2003.

Hausman, Jerry, Bronwyn H Hall, and Zvi Griliches, "Econometric Models for Count Data with an Application to the Patents-R & D Relationship," *Econometrica: Journal of the Econometric Society*, 1984, pp. 909–938.

Hu, Michael, Karl Schultz, Jack Sheu, and Daniel Tschopp, "The Innovation Gap in Pharmaceutical Drug Discovery & New Models for R&D Success," *Kellogg School of Management*, 2007.

Hult, Kristopher, "Measuring the Value of Incremental Innovation Using Evidence from Medicare Part D," *Working Paper*, 2014.

_ **and Tomas Philipson**, "How Does Technological Change in Health Care Affect Quality-Adjusted Prices? A Systematic Analysis," *Working Paper*, 2014.

Ickovics, Jeannette R and Christina S Meade, "Adherence to Antiretroviral Therapy among Patients with HIV: A Critical Link between Behavioral and Biomedical Sciences," *Journal of Acquired Immune Deficiency Syndromes (1999)*, 2002, *31*, S98–102.

Kaplan, Cameron, "Compliance and Copayments: The Case of Statins," 2011.

Lohse, Nicolai, Ann-Brit Eg Hansen, Gitte Pedersen, Gitte Kronborg, Jan Gerstoft, Henrik Toft Sørensen, Michael Vaeth, and Niels Obel, "Survival of Persons with and without HIV Infection in Denmark, 1995–2005," *Annals of Internal Medicine*, 2007, *146* (2), 87–95.

Munos, Bernard, "Lessons from 60 Years of Pharmaceutical Innovation," *Nature Reviews Drug Discovery*, 2009, *8* (12), 959–968.

Murphy, Debra A, Moussa Sarr, Stephen J Durako, Anna-Barbara Moscicki, Craig M Wilson, and Larry R Muenz, "Barriers to HAART Adherence among Human Immunodeficiency Virus–Infected Adolescents," *Archives of Pediatrics & Adolescent Medicine*, 2003, *157* (3), 249–255.

Pammolli, Fabio, Laura Magazzini, and Massimo Riccaboni, "The Productivity Crisis in Pharmaceutical R&D," *Nature Reviews Drug Discovery*, 2011, *10* (6), 428–438.

Paul, Steven M, Daniel S Mytelka, Christopher T Dunwiddie, Charles C Persinger, Bernard H Munos, Stacy R Lindborg, and Aaron L Schacht, "How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge," *Nature Reviews Drug Discovery*, 2010, *9* (3), 203–214.

Rebick, Gabriel and SL Walmsley, "Issues in Resistance, Adherence, and Comparative Efficacy of the Single-Tablet Regimen Combination of Tenofovir, Emtricitabine, and Efavirenz in the Management of HIV-1 Infection," *Virus Adaptation and Treatment*, 2012, *4*, 51–63.

Rosen, Allison B, Alicen B Spaulding, Dan Greenberg, Jennifer A Palmer, and Peter J Neumann, "Patient Adherence: A Blind Spot in Cost-Effectiveness Analyses," *American Journal of Managed Care*, 2009, *15* (9), 626–32.

Scannell, Jack W, Alex Blanckley, Helen Boldon, and Brian Warrington, "Diagnosing the Decline in Pharmaceutical R&D Efficiency," *Nature Reviews Drug Discovery*, 2012, *11* (3), 191–200.

Stafford, Randall S, "Regulating Off-Label Drug Use: Rethinking the Role of the FDA," *New England Journal of Medicine*, 2008, *358* (14), 1427–1429.

Wall, Tamara L, James L Sorensen, Steven L Batki, Kevin L Delucchi, Julie A London, and Margaret A Chesney, "Adherence to Zidovudine (AZT) among HIV-Infected Methadone Patients: A Pilot Study of Supervised Therapy and Dispensing Compared to Usual Care," *Drug and Alcohol Dependence*, 1995, *37* (3), 261–269.