5-2015

Essays on the Health Economics of Pharmaceuticals

Anna Chorniy
Clemson University

Essays on the Health Economics of Pharmaceuticals

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Economics

---

by
Anna Chorniy
May 2015

Accepted by:
Dr. Daniel Miller, Committee Chair
Dr. Thomas Mroz
Dr. Michael Maloney
Dr. Raymond Sauer
Dr. Chungsang Lam

Abstract

The effects of pharmaceutical treatment on patient health, pricing of pharmaceuticals and their regulation are the backbone of my research. My work reflects two current trends used to advance our knowledge in the field: the use of dynamic structural models that is supplemented by detailed administrative individual-level data.

This thesis consists of three chapters that address a number of policy-relevant questions in health economics using both individual- and market-level outcomes. In the first chapter I take a market-level approach to look at the effect of mergers between insurance companies on Medicare Part D plan premiums and generosity of coverage. In the following two chapters I study the effects of ADHD treatment on children's health and behavioral outcomes.

The first chapter focuses on the insurance design and pricing of insurance plans that cover prescription drugs. We examine horizontal mergers amongst Medicare Part D insurers with the aim of decomposing market power, cost efficiency, and bargaining power merger effects. We apply a differences-in-differences identification strategy to panel data on plans offered between 2006 and 2012 to document the effects of mergers on plan premiums and drug coverage characteristics. The results indicate substantial market power as plans affected by a merger increase premiums on average. But, premiums fall and drug coverage improves for merging insurers that restructure plans and renegotiate contracts with drug suppliers by consolidating existing plans. We attribute these effects to improved cost efficiencies and increased bargaining power.

In the second and third chapters I look at the individual-level outcomes following medical treatment of Attention-deficit/hyperactivity disorder (ADHD). In the U.S. the incidence of ADHD diagnosis among children increased significantly over the past decade. The most recent National Survey of Children's Health 2011/12 reports that over 5 million children aged 2-17 (7.9%) have been diagnosed with ADHD. Over 68% of these children are taking medications for the disorder. However, little is known about the existing prescribing practices, physician learning process, and

relative efficacies of various ADHD treatment strategies. In Chapter 2, I build on the literature on investment in human capital (see for example Heckman et al. (2006)) to model the timing of the first diagnosis, treatment, and adverse outcomes over time. ADHD is a common chronic mental condition that impairs noncognitive skills. A child who has ADHD has a relatively low stock of abilities at birth. Once a child is diagnosed, her family can invest in medical treatment to reduce the gap in abilities of a child with ADHD compared to her non-ADHD peers. In the model, ADHD treatments are the only available type of investment. While on treatment, the child is able to improve her outcomes in the short-run, accumulate cognitive and noncognitive skills and possibly improve her long-run outcomes.

Using a 10-year panel of South Carolina Medicaid claims, I model the probability of the initial diagnosis of ADHD, dynamic treatment choice decisions and subsequent adverse events later in life. Controlling for endogeneity, I find that there is a strong persistence in treatment choices across time periods. The results also suggest that pharmacological treatment has only short-term positive impact on the probability of such adverse events as injuries, teenage pregnancy, and STDs, and no impact on substance abuse disorders. Behavioral therapy alone is not as effective as it is in combination with ADHD drugs, but for STDs and substance abuse disorders it seems to show relatively long-lasting effects in contrast to drugs alone.

In Chapter 3, I extend Crawford and Shum's (2005) model to explore the effect of treatment interruptions (drug holidays) in addition to the effects of various drug therapies. The evidence suggests that children diagnosed with ADHD face significant uncertainty regarding efficacy and severity of adverse effects of ADHD medications. Almost half of these children switch therapies during the first six months of treatment. This suggests a considerable amount of experimentation by doctors. Using South Carolina Medicaid claims data for 2003-2012, I estimate a dynamic model of demand for ADHD drugs under uncertainty. In the model, highly heterogeneous patients learn about the efficacy of available treatments through experimenting. I find that patients are heterogeneous in the underlying illness severity. The probability that the patient will be able to function successfully in their everyday life without ADHD treatment varies from 1.8% to 76.7% in the baseline model specification. Although merely suggestive, it might point at the presence of overdiagnosis and overprescription practices. I also find that there is a lot of uncertainty regarding patient-drug match by both symptomatic and curative properties. Although some drugs are better than others for each of the patient types by severity of their condition, their match value distributions overlap significantly.

In other words, knowing patient type, does not resolve patient-drug match uncertainty.

Although the model with drug holidays yields overall similar results to the baseline model, in their current formulation they cannot be directly compared because of the differences in the choice set. Notably, drug holidays rank first for the healthy type by symptomatic relief properties.

Both dynamic models allow for policy-relevant simulations, for example one could evaluate the effect of interruptions in treatment on the overall treatment cost and disease duration, accounting for patient heterogeneity in response to treatment for ADHD and potentially develop better guidelines that can improve the quality of drug-patient matches and patients outcomes. This is left for future work.

Dedication

Сей труд посвящается моим родителям, брату и мистеру Дабл-ю.

To my parents, brother, and Mr. W.

Acknowledgments

I owe special thanks to my advisors Tom Mroz, Daniel Miller, and Michael Maloney for their guidance and continuous and unconditional support. I would also like to thank my dissertation committee members Raymond Sauer and Tom Lam. I am indebted to Donna Gilleskie for her comments on previous drafts and for sharing her Fortran code with me; Matthew Shum and Tanja Saxel for their help on the estimation of dynamic learning models and for sharing programming codes for their respective research papers; participants of the Clemson Industrial Organization seminar and Clemson Labor Economics seminar for their helpful suggestions; Art Carden, James Bailey, Bruce and Dot Yandle for helpful suggestions and support. I am also grateful for the financial support for my research from the J. Walker Department of Economics at Clemson University. Finally, I would like to thank Clemson IT Department and Mark Harouff in particular; SC Revenue and Fiscal Affairs Office, in particular Heather Kirby and Joe Magagnoli, who helped me through a long and tedious process of obtaining access to SC Medicaid data.

Contents

List of Tables

List of Figures

Chapter 1

Mergers in Medicare Part D: Decomposing Market Power, Cost Efficiencies, and Bargaining Power (with Daniel Miller and Tilan Tang)

1.1 Introduction

The landscape of competition in the health insurance industry has experienced many changes in the past several years, starting with the introduction of managed care plans in the 1980s, privatized Medicare plans, expanded prescription drug coverage, and most recently the reforms in the 2010 Patient Protection and Affordable Care Act. Throughout this period there have been waves of merger and acquisition (M&A) activity as insurers adapted to the evolving marketplace (Town and Park (2011)).

In this paper, we examine the effect that horizontal M&A activity amongst health insurers has on prices and coverage characteristics of prescription drug plans offered in the Medicare Part D market. Part D is a recently created program that established a regulated and subsidized insurance exchange for senior citizens to purchase prescription drug coverage from competing private insurers. The program lifetime overlapped with a dozen large scale horizontal M&A deals involving the parent companies of insurers offering Part D plans. Each year, an average of 17% of all plans is directly affected by an M&A deal. More, even larger deals are on the docket. If they all proceed, 22 of the top 25 Part D insurers will have gone through a merger.

Theory suggests three major channels through which mergers affect markets. First, horizontal mergers may be beneficial if they result in increased productive efficiency. In health insurance, efficiency gains can be achieved through scale economies that appear as firms consolidate their ad-

ministrative and marketing activities. Second, horizontal mergers alter bargaining dynamics with upstream suppliers as the combined firm gains monopsony power over suppliers. For health insurers the upstream suppliers are the providers of healthcare goods and services (doctors, hospitals, drug manufacturers, and pharmacies). With greater bargaining power, an insurer may be able to negotiate more favorable terms with providers. This merger effect is of particular importance in Part D. The program designers relied heavily on the ability of private insurers to bargain with drug suppliers and explicitly prohibited the government from participating in negotiations (Duggan and Scott-Morton (2010); Frank and Newhouse (2008)). Mergers could have a positive effect if the improved bargaining position allows insurers to increase the scope of covered drugs or negotiate lower drug acquisition costs, which can be passed to enrollees either directly through reduced cost sharing on drug copays or indirectly through lower insurance premiums. Finally, horizontal mergers give firms more market power as markets become more concentrated. Reduced competition can lead to higher prices for consumers or lower product quality if firms compete on quality dimensions.

Anti-trust authorities care about whether the beneficial effects of mergers (cost efficiencies and monopsony power) in fact exist, and if so, whether they outweigh negative market power effects. Stylized facts about Medicare Part D give reason for concern. Since the program's inception in 2006, premiums increased by more than 26% in real terms. Coverage has declined. The number of drug offerings on plans' formularies has fallen by 29% and out-of-pocket costs paid by enrollees for the most popular drugs has nearly doubled. While the typical consumer still has many choices—an average of 30 plans available in each market—there has been a drastic 31% decrease in the number of plan offerings coinciding with this period of rising premiums and declining coverage.

Much of the decrease in the number of plan offerings can be attributed to merging insurers consolidating their plan offerings; even more is due to non-merging insurers consolidating their plans. By consolidation we mean that an insurer takes two or more plans offered in the previous year and consolidates them in a single plan for the upcoming year. In any given year, about 20% of plans are consolidated. To distinguish terminology, mergers can be thought of as *inter*-firm combinations; plan consolidation, as *intra*-firm combinations. The distinction is important for anti-trust purposes. If an insurer can realize the beneficial effects of mergers (cost efficiencies and monopsony power) organically by consolidating its own plans, without engaging in a merger with an outside firm, then there is a weaker case to be made in favor of mergers. Our empirical methodology explicitly distinguishes mergers from consolidation to test whether merger effects only appear through external

mergers or can be achieved internally.

Plan consolidation is a particularly important policy topic in Medicare Part D. In 2011, Medicare began publishing regulations encouraging insurers to consolidate their plans. It recommended that insurers consolidate low enrollment and "meaningfully" similar plans. Many insurers complied, however there is no evidence of this rule being enforced. As of 2014, significantly more stringent rules have been proposed that not only restrict incumbent insurers, but also limit entry of new Part D providers.

In our application to Medicare Part D, we analyze the effects that horizontal mergers have on market outcomes with the aim of separately identifying the three channels through which M&A activity affects plans: cost efficiencies, monopsony power with upstream drug suppliers, and market power. We use panel data on all plan offerings between 2006 and 2012 (over 9,000 plan-year observations) and consider two types of outcome variables: plan premiums and measures of plan coverage, specifically the number of drugs covered on insurers' formularies and an index of the out-of-pocket cost sharing an enrollee pays in drug copays.

To identify the treatment effect that M&A deals have on plans we use a differences-in-differences approach. In our first specification, we examine how plans affected by a merger change in the year following a merger as compared to the control group of plans unaffected by mergers. This approach measures the combined effect of all three channels, which is useful to run a horse race gauging whether the beneficial effects outweigh the adverse effects for insurers. However, simply comparing outcomes of merged and non-merged plans is not informative about the magnitudes of the three competing effects and indicates nothing about whether the benefits of mergers can be achieved internally through plan consolidation.

In our second specification, we sort out the three competing theories of mergers. To do so, we modify the differences-in-differences treatments to distinguish mergers that involved plan consolidation from mergers that did not. Our hypothesis is that merging on its own—without consolidating plans—does not allow a firm to realize cost efficiencies and implies it is not exercising its increased monopsony power to renegotiate contracts with drug suppliers. Thus, only market power effects appear as the merging insurers coordinate pricing decisions. By merging and restructuring plan offerings through consolidation, merging insurers can realize all three merger effects. In other words, we can separate market power from cost efficiency and monopsony power effects by contrasting mergers with and without plan consolidation. Finally, we examine cases where non-merging firms

consolidate plans. Our hypothesis is that non-merging insurers only improve cost efficiencies by consolidating plans; they gain no additional market power, nor monopsony power.

To further gauge outcomes, we examine coverage characteristics. The effects of mergers on coverage are important as both prices and the terms of coverage are jointly determined in insurance contracts. Under Part D regulations, coverage is heavily determined by the bargaining process between insurers and drug suppliers. These results provide more robust evidence about the monopsony power effects than can be gleaned from evidence on insurance premiums and constitute an important contribution to the merger literature which often lacks detailed analysis of product characteristics.

In summary, our results show that all three channels are at play. When insurers merge and do not consolidate plans, premiums increase by an average of 9%. We attribute the rise to a strong market power effect. For insurers that merge and consolidate plans, the net effect on premiums is an average decrease of 4%, outweighing market power effects. Breaking down the results based on our comparisons of non-merging insurers that consolidate plans, about two-thirds of the premium decrease is due to cost efficiencies that even non-merging firms can realize, and the remaining one-third comes from the increased monopsony power gained by merging.

The results for coverage characteristics corroborate the findings on premiums and highlight the significance of the bargaining process between insurers and drug suppliers. For insurers that merge and consolidate plans, there are large improvements in coverage. These plans increase the number of drug offerings on their formulary by an average of 14%, and decrease enrollee out-of-pocket copay costs by 4%. Merging without consolidating plans has a near zero effect on drug coverage. Likewise, there is little effect for non-merging firms that consolidate. The evidence supports our hypothesis that bargaining gains cannot be achieved internally, only for merged insurers that consolidate plans.

The remainder of the paper is organized as follows. In section 2 we discuss related literature. In section 3 we provide the background for our application to Medicare Part D. In section 4 we discuss the data. In section 5 we present the econometric method, and in 6, the results. Section 7 concludes.

1.2 Healthcare Competition Literature

Economists have long been concerned about whether healthcare markets are competitive and, if so, whether unfettered competition ensures the first best. Ellis (2012) cites evidence of high

levels of concentration and raises concerns about market power in both provider markets (hospitals, physician networks, pharmaceuticals) and insurance markets. Apart from market power, two other channels—cost efficiencies and the balance of bargaining power in the vertical relationship between insurers and healthcare providers—determine the performance of markets. This paper contributes to the literature by decomposing these three channels as they apply to health insurance markets. Merger studies provide an excellent avenue for analyzing competition as mergers events change the structure of the industry.

The literature on health insurance claims an insurer's scale as measured by enrollment, which we associate with cost efficiencies, is an important determinant of its cost structure. There is a strong correlation between scale and insurance loads: the difference between what is collected in premiums and paid out in benefits. For employer sponsored health insurance plans Karaca-Mandic et al. (2011) document loads ranging from 4% for the largest insurance plans with over 10,000 enrollees to over 40% for the smallest with under 50. In Part D, the size of plans spans this same range. A leading cause is that large insurance plans economize on administrative costs. Part D administrative costs may be particularly high due to Medicare's stringent compliance and reporting standards and the added complexities of real-time pharmacy claims processing at the point of sale. In the Medigap market, insurers have high loads because of marketing costs (Starc (2012)). Insurers use the same marketing tools for their Part D plans. Horizontal mergers may have tremendous benefits if the increased scale of merging insurers reduces administrative and marketing costs. Legislation in the PPACA aims to reduce loads by imposing minimum loss ratios (MLR) on insurers. Starting in 2014, MLRs will be implemented in Medicare Part D. Mergers may be one of the most effective ways for insurers to reduce costs so that they can meet the new MLR requirements.

The next channel we consider is the vertical market relationship between insurers and providers. The industry has shifted towards a model where insurers selectively contract with providers through a bargaining process. Insurers decide which providers to include in their network, providers decide which networks to join, and the two parties negotiate over reimbursement rates and the terms of enrollee cost sharing. There is a large literature on bargaining from the perspective of hospitals, (Ho (2009); Ho and Lee (2013); Gowrisankaran et al. (2013); Lewis and Pflum (2011)), but less is known from the insurance side, particularly for prescription drugs. In Part D, bargaining is quite important and has been credited with reducing drug prices for the Medicare population (Duggan and Scott-Morton (2010)).

Our merger study allows us to gain a greater understanding of how competition impacts the bargaining process. Mergers alter bargaining positions. The threat point in the Nash-bargaining models applied to the industry is determined by the number of people enrolled by the insurer. Insurers can expand their base of enrollees through merger to gain greater bargaining power. That can translate into some combination of lower premiums, expanded network coverage, and reduced cost sharing for its enrollees. We also provide evidence on whether internal plan consolidation, which makes plans larger but doesn't change the size of the insurer, affects bargaining power.

Much less is known about the effects of M&A deals in health insurance markets. Two of the most comprehensive studies are Dafny (2010) and Dafny et al. (2012). Dafny (2010) uses a large panel of insurers offering plans in the employer sponsored health insurance market to investigate whether health insurers have market power. The authors find non-trivial market power as evident in their ability to price discriminate by charging higher premiums to more profitable employers, especially so in highly concentrated markets. A similar conclusion is reached by Bates et al. (2012) that finds higher prices and lower rates of health insurance enrollment in more concentrated markets.

Dafny et al. (2012) employs the same data set as Dafny (2010) to study the effect of concentration on premiums and payments to physicians and nurses. They focus on the 1999 merger of Aetna and Prudential, two of the largest insurers in their sample. The deal between them resulted in a sharp change in the Herfindahl-Hirschman concentration Index (HHI). Their estimates show that the average market-level changes in HHI between 1998 and 2006 caused a 7 percentage points increase in premiums. They also find evidence of increased bargaining power with health care providers. They estimate that payments to physicians and nurses decreased by 2% to 3% over the same time period.

We build on Dafny et al. (2012) in two important ways. First regards the data. Whereas they examine just 1 merger case, we use panel data that includes all merger activity between 2006 and 2012. The high churn rate of mergers yields a large sample of both treated (merged plans) and a control group of plans (unmerged plans) to identify merger treatment effects. We also have detailed plan-level data on coverage characteristics, not just premiums, that we consider as merger outcomes. This is important as both prices and the terms of coverage are jointly determined in insurance contracts. Our second contribution is to disentangle the three merger effects. Their results show market power dominates, but are not informative of the extent to which the merger created cost efficiencies or altered bargaining power.

The effect of mergers on market performance is also an important topic in the finance literature. While we address the question using product-level data, most papers in finance use event studies on a set of multiple M&A deals. Most closely related is Fee and Thomas (2004) that specifically aims to identify how mergers affect market power, cost efficiencies, and vertical bargaining power. They use a large cross-industry sample of deals from 1980 to 1997 and examine stock price movements for the merging firms, horizontal rivals, and upstream suppliers. Maksimovic et al. (2011) examines post-merger plant closures and restructuring of supplier contracts as means of improving efficiency. The analog to plant closures and restructuring in our paper is plan consolidation.

Finally, our paper contributes to a growing literature on Medicare Part D. Several papers (Lucarelli et al. (2012); Miller and Yeo (2013); Ericson (2014); Decarolis (2012)) examine firm conduct and competition, include important institutional details related to subsidies and market regulations. Another strand of the literature (Abaluck and Gruber (2011); Heiss et al. (2013); Ketcham et al. (2011); Kling et al. (2012)) uses individual level data on consumer choice and finds evidence that enrollees make poor plan choices. These studies have been influential in guiding policy decisions. The consumers' choice problem could be eased by reducing the number of available plan offerings. The question becomes a matter of how to implement policy to reduce choice without compromising competition or the breadth of offerings. There are two standing proposals involving plan consolidation; forced consolidation of low enrollment plans and forced consolidation of meaningfully similar plans. The most recent 2014 proposals extend these criteria to forbid new entry. Alternatively, anti-trust authorities could adopt a tolerant stance towards merger cases. This study sheds light on the policy debate by showing the effect that mergers and consolidation have on prices and coverage.

1.3 Medicare Part D Background

Medicare Part D introduced a prescription drug benefit to the Medicare program. It was authorized under the 2003 the "Medicare Prescription Drug, Improvement, and Modernization Act" and fully enacted in 2006. The legislation created a coverage mandate requiring beneficiaries to obtain prescription drug coverage when they first become eligible for Medicare or face penalties for late enrollment. The act established a regulated and subsidized health insurance exchange where beneficiaries can choose amongst plans offered by competing private insurers. The prescription drug plans offered in this exchange are the focus of our study. About 60% of the Medicare population is

covered by a Part D plan; the remainder either lack coverage or obtain prescription coverage through other means such as employer/retiree benefits or another government program.

The Part D exchange was designed to rely on free market principles to provide competitive drug plans. The benefit is offered by private insurers who may freely enter and exit the market, choose the number of plans to offer, and set monthly premiums. Insurers are also largely responsible for the benefit design. Each insurer selectively chooses which drugs to cover on its formulary and sets cost sharing copay/coinsurance rates on a drug-by-drug basis. Drug prices are determined through a bargaining process between and drug manufacturers, wholesalers, and pharmacies. Per regulation, negotiated prices must be passed on to enrollees. This has been seen as a controversial feature of the program because the legislation explicitly prohibits the government from being involved in price negotiations with the pharmaceutical industry (Frank and Newhouse (2008)) as is the case for other government drug benefits such as Medicaid.

The regulations establish a number of coverage standards. All providers are required to offer at least one basic plan that meets (or is actuarially equivalent to) a minimum coverage level with respect to the deductible, coinsurance and copay rates, and the scope of drugs covered on the formulary. In addition to a basic plan, insurers may offer enhanced plans that have more generous coverage through a combination of lower deductibles, lower copay/coinsurance rates, and drug coverage for a larger set of medical conditions.

Plans have a large toolbox of "formulary management" techniques that they can use as bargaining levers with drug suppliers and as a means to steer enrollees' usage of drugs. With the exception of six therapeutic classes, they are allowed to selectively choose which drugs to include on their formularies, place drugs on pricing tiers such as "preferred," "non-preferred," and "specialty," as well as impose usage restrictions in the form of quantity limits, step therapy routines, and prior authorization requirements. These techniques are thought to be important tools for negotiating favorable drugs prices, which will ultimately be reflected in the generosity of plans coverage and premiums (Duggan and Scott-Morton (2010)).

Nearly all major health insurance companies and many regional insurers entered the Part D market in the first two years of the program. There has been almost no entry in later years. Geographically, the market is separated into 39 markets drawn around state boundaries. Insurers offer and price plans individually for each market. In the typical market, enrollees can choose from about 40 plans offered by 20 insurers.

1.4 Data

1.4.1 Plan-Level Data

We utilize detailed longitudinal data on plans that includes an average of 1,500 stand-alone, Part D plans (PDPs) per year. We exclude Medicare Advantage plans that bundle Part D coverage with other Medicare coverage components. The data span 7 years from 2006 when Medicare Part D was introduced to the most recently available data in 2012 and cover all 39 geographical markets. The sample is constructed using both publicly available and restricted use data obtained from the Centers for Medicare and Medicaid Services (CMS).

Enrollment in stand-alone Part D plans has grown from about 17 million in 2006 to over 20 million by 2012. The average plan has 11,347 individuals enrolled per year. However, the plans differ significantly on this margin. There are plans that have fewer than 10 insured, while others insure more than 300,000 individuals. About 40% of the enrollees receive additional premium and copay subsidies through the low income subsidy (LIS) program. Table 1.1 presents information on market level trends. In the first year of the program, there were only 1,446 plan offerings, which rose to 1,908 in the second year. But following 2007, the number of plan offerings has steadily decreased down to 995 by 2012. Much of this decrease can be attributed to merger activity and plan consolidation. During the sample period average premiums increased by 26% in real terms (by 43% in nominal terms), and the average plan's market share increased 37%.

We collect information on each plan's premium, deductible, gap coverage, and drug formulary. Table B.1 reports summary statistics on the plan-level data for 2006-2012. A plan's *premium* is set up once a year, when private insurance companies submit their bids for contract with Medicare. The deadline for the plan sponsors to submit their bid is the first Monday in June each year. The open enrollment runs from October through December, and the contract year begins January 1st. Premiums are paid monthly by the insured. Qualified individuals are provided with the "Extra Help", or low-income subsidy (LIS) by Medicare. This LIS program covers in full or partially the monthly premium amount, deductible, copayments and coinsurance, and eliminates the coverage gaps.

The *deductible*, followed by the initial coverage zone, is the amount the insured must pay out-of-pocket before the drug plan cost-sharing kicks in. The yearly deductible for what Medicare determines as the standard Part D benefit was set to $250 in 2006. Updated using annual percentage

Table 1.1: Trends in Medicare Part D market, 2006-2012.

	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Monthly premium	42.55	40.63	42.99	49.03	48.61	54.73	53.41
	(14.60)	(16.70)	(21.35)	(22.15)	(20.14)	(25.79)	(26.72)
Plan market share	0.009	0.007	0.007	0.008	0.008	0.012	0.013
	(0.018)	(0.016)	(0.015)	(0.015)	(0.016)	(0.024)	(0.023)
N plans offered	37.08	48.92	45.54	41.69	38.28	26.51	25.51
	(13.82)	(16.47)	(14.54)	(13.10)	(12.29)	(8.65)	(8.74)
Plan enrollment	10,730	8,473	8,573	9,415	10,594	16,201	17,297
	(25,159)	(23,066)	(21,155)	(21,912)	(24,187)	(37,194)	(36,155)
LIS enrollment	5,588	4,196	4,051	4,377	5,042	7,699	8,069
	(13,368)	(13,820)	(11,104)	(12,387)	(14,401)	(20,340)	(20,431)
Eligible population, in'000	1,275	1,279	1,305	1,329	1,364	1,396	1,480
	(951)	(963)	(986)	(1,010)	(1,029)	(1,049)	(1,104)
Insurer regional presence	26.33	31.14	29.76	31.30	30.10	31.23	28.85
	(12.04)	(9.25)	(11.15)	(7.96)	(10.68)	(8.99)	(12.12)
N plans affected by merger	293	4	541	173	129	272	245
N plans offered	**1,446**	**1,908**	**1,778**	**1,626**	**1,493**	**1,034**	**995**

Notes: All plans: renewed, consolidated, new and terminated in the next calendar year are included. Premiums are given in 2012 dollars. Number of plans offered and eligible population are calculated per Part D region. Standard deviations are in parentheses.

increase, it was raised to $320 by 2012. Most enhanced PDPs eliminate the deductible so that the enrollee receives first dollar coverage.

The gap in coverage or "*donut hole*" begins when the insured reaches the limit on the expenses covered by the initial coverage zone ($2250 in 2006). Prescription costs beyond the limit and below the "catastrophic" level ($5100 in 2006) are paid by the insured out-of-pocket. Many enhanced PDPs provide full or partial coverage in the donut hole. The ACA legislation eliminated the donut hole effective 2014.

The *formulary* is a comprehensive list of the medicines covered by the plan, identified by the National Drug Code (NDC).[1] The formulary files contains data on the drug's tier, usage restrictions, and copay/coinsurance provisions that determine the cost to a beneficiary. The formulary file is complemented with drug pricing data that was first published in 2009. The pricing data contain information on the average drug prices for every drug and plan. Specifically, the reported price is the average transaction price, net of all rebates for a 30-day supply filled at the plan's preferred pharmacies in the third fiscal quarter of each year.

To measure the comprehensiveness of formulary coverage, we count the number of drugs

[1] NDC is an 11-digit classification issued by the Food and Drug Administration (FDA) for all the approved drugs. Under this system, different package and dosage sizes of the same drug molecule have separate NDCs.

listed on the plan's formulary. The first measure counts the number of top 100 drugs. In early years, the average plan covered more than 90 of the top 100 and fell to 75 by 2012. The second measure counts the total number of NDCs on a formulary which plans select from a set of 5300 unique drugs that qualify for coverage under Part D.[2] Like the top 100 drug, the total number of covered NDCs fell throughout the sample period.

Part D formularies typically have three *pricing tiers* that separate preferred drugs with relatively more favorable coverage from non-preferred ones. Lower tiers indicate better coverage. For example, a three-tier plan that has 1/3 of its drugs on tier 1, 1/3 on tier 2, 1/3 on tier 3 has an average pricing tier of 2. Since the plans differ in the number of tiers (up to 7 tiers), for the purposes of comparison we normalize a 2 on a scale of 1 to 3, to 0.5 on a 0 to 1 scale. The formularies also might have up to three types of *restrictions* placed on drug consumption: step therapies, prior authorization, and quantity limits. We sum up the restrictions and calculate the average number of restrictions on a formulary using a 0 to 3 scale.

We use drug prices and cost sharing rates to construct a price index to compare out-of-pocket copay prices across plans. This is our most refined measure of the generosity of plan coverage. It is constructed by using actual copay/coinsurance rates and pharmacy prices to calculate the out-of-pocket price an enrollee pays for a basket of the top 100 drugs ranked by the number of prescriptions filled. These hundred drug prices are combined into a price index, where each drug is weighted equally. If a drug is not covered by a particular plan, we assume that enrollees will have to pay the full retail price out-of-pocket. We construct separate price indexes for the initial coverage zone and donut hole. Three sources of variation affect the out-of-pocket price index: number of covered drugs, drug pricing tiers, and a plan's negotiated price with the pharmacy and drug manufacturer. More comprehensive formularies, lower pricing tiers, and lower pharmacy prices all contribute to a lower value of the out-of-pocket price index.

The other measures of plan design are distinguishing characteristics of basic and enhanced plans. Recall basic plans meet or are actuarially equivalent to minimum coverage standards set by the Part D regulations, enhanced plans offer some form of additional coverage. Slightly more than half of the plans are basic. Benchmark plans are a subset of basic plans that are priced below the market average of basic plans. Benchmark plans qualify for the full subsidy amount of the low

[2]The method for counting NDCs changed after 2006. In 2006, identical drugs made by different manufacturers were "double-counted" as distinct drugs. 2007 onward, identical drugs were only counted once.

income subsidy (LIS). They also qualify to receive Medicare/Medicaid dual eligible beneficiaries. Dual eligibles—who account for about 20% of the Medicare and 40% of Part D enrollment—are randomly and uniformly assigned to the LIS eligible plans if they don't otherwise actively select a plan. Given the large number of dual eligibles, LIS eligible plans receive a big boost in enrollment from random assignment, which can be thought of as a characteristic making those plans more desirable. The theoretical foundations for this interpretation are explained in companion work by Miller and Yeo (2012). We include these other plan characteristics as control variables to ensure that our differences-in-differences results attribute price changes to merger effects, and not pricing responses to changes in coverage characteristics.

1.4.2 Data on M&A Deals

We collect data on M&A activity from the Securities Data Company (SDC) merger and acquisition module which contains detailed information on all deals involving public and private companies. In the time frame suitable for our analysis, from 2006 to 2011, we identified a total of 11 completed horizontal M&A deals amongst companies that offer Medicare Part D policies. Table 1.2 lists the details on each of the selected deals. All of the deals involve major Part D insurers that offer plans across the entire nation with the exception of the Medical Mutual of Ohio/ Carolina Care Plan acquisition. Note that some of the major plan providers were involved in multiple deals during the sample period.

Table 1.2: M&A Deals' Details

N	Acquiror	Target	Value	Date	Form
1	United HealthCare Services	PacifiCare Health Systems	7,511	12.21.05	M
2	MemberHealth	AmeriHealth Ins Co-Medicare	N/A	11.16.06	AA
3	Medical Mutual of Ohio	Carolina Care Plan	N/A	05.18.07	AA
4	Universal Holding Corp	MemberHealth	780	09.21.07	AA
5	UnitedHealth Group	Sierra Health Services	2,425	02.25.08	M
6	CVS Caremark Corp	Longs Drug Stores Corp	2,637	10.30.08	M
7	CVS Caremark Corp	Universal American Corp	N/A	12.31.08	DJV
8	United HealthCare Services	Health Net-US Northeast	630	12.11.09	AA
9	HealthSpring	Bravo Health	545	11.30.10	M
10	Munich Health North America	Windsor Health Group	131	01.04.11	M
11	CVS Caremark Corp	Universal American Corp	1,059	04.29.11	M

Notes: We list the acquiror and target names as they are recorded in the SDC data. For example, in deal #6 the acquiror is UnitedHealth Group Inc. It is a parent of the United HealthCare Services Inc, a company that was the acquiror in deals #1 and #8. Merger value is given in millions of dollars. The date is merger completion date. "AA" stands for acquisition of assets; "M" for merger; "DJV" for dissolution of joint venture. AA is the purchase of a company by acquisition of its assets rather than its stock.

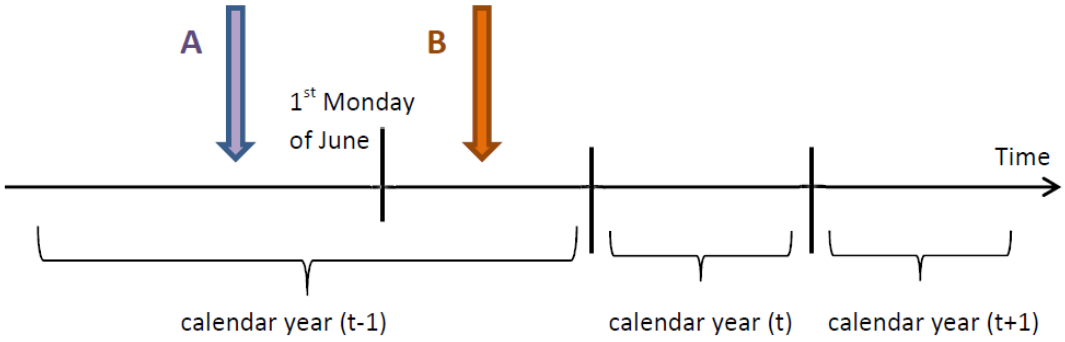

Figure 1.1: M&A deals timing with repect to the bid deadline date

We restrict attention to horizontal mergers and acquisitions of assets where either participants or their immediate subsidiary offered a Part D plan at least in the year prior to the merger completion date. We exclude all the deals where one or both companies belong to a non-Part D line of insurance (such as life insurance), joint ventures of Part D insurers into related lines of business (such as pharmacy management) and vertical mergers with pharmacies. It is worth noting that we exclude a few large deals that took place in the second half of 2011 and in 2012 due to our assumption on the relative timing of the deal and its effects. The bids for each successive calendar year are submitted before the first Monday in June of the previous calendar year. Thus, for the deals completed prior to the deadline we measure the "before" period as the current calendar year and "after" as the following calendar year assuming that their bid will reflect the effects of merger. For example, case A in Figure 1.1 demonstrates a merger that was completed prior to first Monday in June of year (t-1). In this case, year (t-1) will represent the "before" period and year (t) - the "after" period. The merger from case B was completed after the bid date. It means that its "before" period is year (t) and "after" period is year (t+1). We also go through the news reports and companies' press releases for each of the 11 deals to obtain factual support to our assumption. The mergers that were completed after June 2011 when all the bids for 2012 calendar year had been submitted would require data from 2013. The latest CMS data available at the time of study are for 2012. Including these later deals, 22 of the top 25 Part D insurers have been involved in an M&A deal with the notable exception being the number 2 insurer, Humana.

We match the SDC data on deals to the plan-level data by company name. There are about 100 unique parent companies whose subsidiaries offer Part D plans during the sample period. Some parents control more than one insurance company. As multi-product firms, insurers offer between

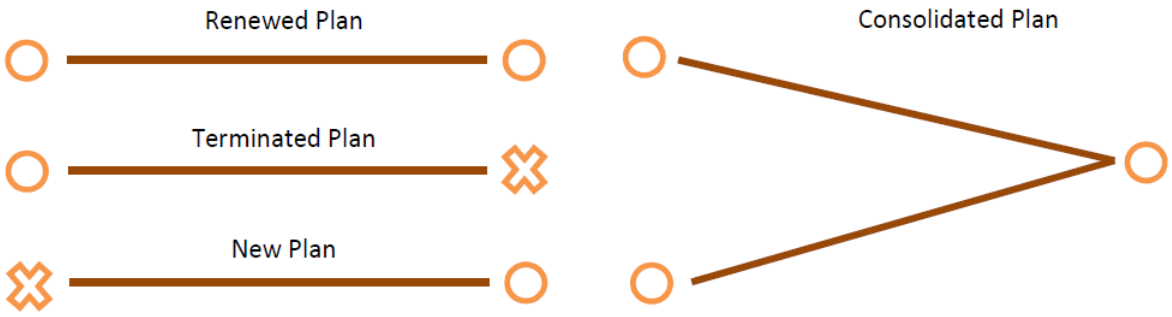

Figure 1.2: Plan transitions from year-to-year

one and three plans per region with the requirement that at least one plan qualifies as a basic plan.

We look at the short-term merger effects by comparing plans prices and coverage characteristics before and after the deal was completed. From year-to-year, plans can evolve in one of four ways as depicted in figure 1.2. Plans can be renewed, terminated, consolidated, or new plans can be introduced. To determine each plan's transition status we use the CMS "crosswalk" file that links plans across years. Renewed plans carry-over enrollees from the previous year and typically maintain the same product segment: basic or enhanced status. However, plan characteristics such as the monthly premium, formulary list, and copay/coinsurance tiers, and drug prices can change across years. Terminated plans simply stop being offered for the new calendar year, and previously enrolled individuals have to actively select another plan. New plans are introduced to the market for the first time and they have no enrollees from the previous calendar year. Consolidated plans combine two or more plans from the previous year into one plan. Enrollees from the previous year's plans carry over into the new plan. Like renewed plans, the product characteristics can differ from the previous year's plan characteristics. Most consolidations combine two or more basic plans or two or more enhanced plans, but there are examples of cross segment, basic-enhanced consolidation.

Consolidation of plans is undertaken by merging firms as well as by firms that did not participate in a deal. We posit that the main reasons behind plan consolidation are to achieve cost efficiency gains and, for merging insurers, as a means to renegotiate contracts with drug suppliers. A similar idea is presented by Maksimovic et al. (2011). They find evidence of extensive restructuring in a short period following an M&A deal. In the sample of U.S. manufacturing firms, acquirors were likely to sell or close down targets' plants. It resulted in a boost in productivity in the retained plants comparing to the industry. Health insurance is fundamentally different from manufacturing

in that terminating plans is highly undesirable because enrollees are lost. Part D insurers are better off consolidating plans when they want to restructure plans offerings so as to retain enrollees.

Table 1.1 shows the total number of plans offered during the sample period in each year and the number of plans directly affected by an M&A deal. In each year, an average of 17% of all plans are affected by a merger. Table B.2 shows how all plans and M&A affected plans evolve. There is no systematic tendency for the plans of merged firms to evolve differently from non-merger affected plans. Most plans are renewed or consolidated, few plans exit or newly enter the market. The only difference between the two groups of plans is that firms that were not affected by a merger were more likely to create a new plan. For our analysis we restrict attention to renewed and consolidated plans because our empirical method requires a plan to be observed for at least two consecutive years. By definition, terminated and new plans do not meet this criteria. Excluding them from the sample is unlikely to bias results because they compose such a small fraction of the market.

Table B.2 also reports comparative summary statistics for the control group, plans unaffected by merger, and treatment group, plans offered by companies involved in a merger deal. The pre-merger plan characteristics of merger affected plans are generally similar to all other plans.

1.5 Estimation Strategy: Differences-in-Differences

To estimate the effect of mergers and plan consolidation, we use a differences-in-differences (DD) identification strategy. Differences-in-differences is a popular method for identifying effects of policy "treatments" most often applied to household-level data in labor, health, and development economics fields (Bertrand et al. (2004)). DD and treatment effect approaches are used less often for studies of the firm and in particular merger outcome studies. However, there are notable applications — Hastings (2004) (retail gas stations) and Dafny et al. (2012) (health insurance). The detailed panel of product-level data and large sample of merger-"treated" plans make such a DD approach feasible and provide an attractive alternative to structural-based modeling and estimation of merger outcomes (Angrist and Pischke (2010)).

1.5.1 Merger Treatment Effects

We run several specifications of DD regressions to estimate the treatment effect of an M&A deal on plan outcomes. Specification (1) considers the effect of deals on our first outcome of interest

— the monthly premium, p.

$$p_{it} - p_{it-1} = \alpha + \beta D_{it-1} + (\mathbf{X}_{it} - \mathbf{X}_{it-1})'\beta + \varphi_t + \varphi_{market} + \varphi_{insurer} + \epsilon_{it-1} \tag{1.1}$$

where i indexes the plan, and t the year. The deal treatment $D_{it-1} = 1$ if plan i was involved in an M&A deal that was completed in year $t-1$, such that the effect of the deal could be expected to appear in year t. Note that the dating of deals is determined by the time line in figure 1.1 and does not necessarily match the calendar year in which the deal was officially announced. The controls for plan characteristics $\mathbf{X}_{it}$ include various measures of plan design and drug coverage. We also include fixed effects for years (φ_t), markets (φ_{market}), and also insurer fixed effects ($\varphi_{insurer}$) in our most heavily controlled specification. The term ϵ_{it-1} is a plan-year specific error term. To estimate the effect of mergers on plan characteristics, we apply the DD approach to drug formulary counts, f, and the out-of-pocket drug price index, *copay*. The dependent variables in these regressions are the first differences in outcome measures, $f_{it} - f_{it-1}$ and $copay_{it} - copay_{it-1}$ respectively.

To identify the merger effect, we take advantage of the two dimensions present in the data: time and merger status. First, we look at the across time variation in outcomes, i.e. plan premiums immediately before the deal to premiums immediately after. This comparison is possible if a plan is observed in the data for at least two consecutive years. For this reason, our sample includes renewed and consolidated plans, excluding new and terminated plans (see figure 1.2). The unit of observation is indexed to year $t-1$ in equation (1.1). This timing issue matters for consolidated plans. For example if plans A and B sold in year $t-1$ are consolidated into plan C for year t, there are two observations in the data for plans A and B in year $t-1$. Observations of A and B may have different p_{it-1} and X_{it-1} values in year $t-1$, but will have the same p_{it} and X_{it} values in year t because of consolidation.[3]

On the merger status dimension, we compare merger-affected plans to a control group of plans unaffected by an M&A deal. Combining both sources of variation in the DD estimator provides a very robust means of identifying average treatment effects.

To understand the intuition behind the DD approach, it is useful to break down the components of the estimator. Applying only one of the differences could result in confounded estimates of

[3]Note that there is no "splitting" of plans. That is, plan A in year $t-1$ cannot be split into plans B and C for year t.

the treatment effect. In the raw data, a before and after comparison across time of average premiums for merger-treated plan shows a (44.81-40.27=)$4.54 *increase* in premiums caused by a merger (see table B.2). A comparison of average premiums for merger (treatment group) and non-merger (control group) plans shows a (44.81-45.16=)$0.36 *decrease* in premiums caused by a merger.

Neither of these results necessarily measures the causal treatment effect. The increase indicated by time differencing could simply reflect an increasing trend in premiums over time that affects all plans. Such a trend is plausible given plans not affected by a merger experience average premium increases of (45.16-42.54=)$2.62. The decrease indicated by differencing the treated and untreated group could be attributed to differences in unobserved plan characteristics of the two groups of plans. The DD estimate of (44.81-40.27)-(45.16-42.54=)$1.92 controls for both confounding time trend effects and unobserved plan characteristics. The estimate of $1.92 is the causal average treatment effect if firms' decisions about merging are orthogonal to plan, market, and time period characteristics. To control for selection on observables, we include first differences in plan characteristics $\mathbf{X}_{it} - \mathbf{X}_{it-1}$. For example, if merger-affected plans are more likely to lower the deductible between years than non-merger plans, the $1.92 could simply reflect the fact that lower deductible plans are more costly. The year and market fixed effects control for their respective correlation with mergers. Year fixed effects are needed because mergers do not all occur in the same year. From the data (table 1.1), mergers happened more intensively in the years following the 2010 health reform legislation, which itself may have altered trends in health insurance premiums. Market fixed effects control for market characteristics, such as the number of competing plans in the market and its size. Note, unlike Dafny et al. (2012), we do not include measures of market competition such as Herfindahl-Hirshman Index (HHI) as it is controlled for by the fixed effects.

The DD estimate of the merger effect is the causal treatment effect if the decision to merge is exogenous or random, conditional on the control variables and fixed effects. Two features of the insurance industry during this time period support the plausibility of the merger exogeneity assumption. First, the mergers in our sample involve large diversified insurance companies. Part D is a relatively small component of the firms' business activities, which suggests merger decisions are likely exogenous to the Part D market. Second, nearly every major firm offering a Part D plan has been involved in a merger since 2006. Including recent mergers announced after our sample period, 22 of the top 25 Part D insurers have merged with another Part D insurer. This high intensity of merger activity suggests merger decisions are not a matter of "if" a firm will merge, but rather

a question of "when" it will merge. Matters of "if" firms merge raise concerns about whether the DD estimator measures causal treatment effects; matters of "when" to merge are controlled for by the year fixed effects. These two justifications aside, we cannot rule out the possibility that there are other unobserved insurer characteristics correlated with the specific year, when a particular insurer merges. To purge such correlation our most heavily controlled specifications include insurer fixed effects. The DD estimator becomes a triple differences-in-differences-in-differences (DDD) with insurer fixed effects (Bertrand et al. (2004)). Identification is a comparison of year-to-year differences in premiums within an insurer in the year(s) it merges compared to year-to-year differences in premiums in the year(s) it does not merge. Insurer fixed effects change the control group from being all other Part D plans that don't merge, to plans of the same insurer in years that the insurer does not merge. We should note that for these specifications it is necessary to compute insurer heteroskedasticity-robust standard errors, which given the limited variation in the data results in large standard errors. Nonetheless our results are economically significant and in many specifications statistically distinguishable from the null hypothesis of zero merger effect.

Interpreting the DD estimates requires care because of equilibrium effects and the possibility of multiple merger events occurring simultaneously in the same time period. In the product and upstream supplier market, equilibrium effects can cause a merger event to have an effect on all plans in a market, not just plans sold by the parties to the merger. In the product market, Bertrand pricing models of differentiated products predict that all firms, including rivals to merging parties, gain market power when a merger increases market concentration. Likewise, mergers can increase monopsony power with upstream suppliers for all firms in a market. The analysis in Dafny et al. (2012) estimates the market-wide effects of concentration induced by the Aetna-Prudential merger on product market pricing and payments to the upstream market for doctors and nurses. Lucarelli et al. (2012) estimate a structural discrete choice model of the Part D market under Bertrand pricing and simulates the effect on premiums from the 2006 merger of United Healthcare and Pacificare. The average premium increases 4.7% for the plans of the merged firms, and just 0.9% for all other plans. Our DD results measure the merger effect on a treated plan over and above the equilibrium effects of mergers on the untreated group. For example, if the data matched that in the simulated model in Lucarelli et al. (2012), the DD estimator on premium would show a (4.7-0.9=)3.8% increase in premiums. When there are multiple merger events occurring at the same time, the estimator measures the marginal effect of a merger on a particular plan, not the total effect of all simultaneously

occurring mergers. Market and year fixed effects control for the intensity of merger activity in a given year and market. For example, there was a lot of merger activity in 2008 when prices increased by a very large amount of $6 on average. The 2008 fixed effect would be higher than other years.

The last consideration for the DD estimator is sample selection. In Part D, plans are allowed to freely enter and exit the market. The DD estimator requires observation of a plan across two consecutive years. As such, new and terminated plans must be dropped from the sample. The DD estimate is potentially biased by sample selection if factors that influence decisions to terminate or introduce a new plan are also related to merger decisions. The issue of plans selecting into or out of the market is analogous to the issue of program participation decisions in the typical DD estimator used for household studies. In our case, selection is not a major concern because there is very little churn in plans entering and exiting the market, and the little churn that exists does not appear to be related to merger decisions.[4] In particular, plans of merged firms are not more or less likely to introduce new plans or terminate plans than non-merging firms (see table B.2). There are good reasons to expect little churn in Part D. First, lock-in effects stemming from switching costs give strong incentives for plans to renew plans from year-to-year and make it difficult for new plans to attract enrollees (Miller and Yeo (2012); Ericson (2014)). Second, subsidy amounts are calculated based on the previous year's enrollment figures which discourages plan entry and exit (Miller and Yeo (2013)). For these reasons new insurers that want to enter the Part D market do so by acquiring the plans of incumbent insurers, not by organically creating new plans. The leading example is the 2012 acquisition of Medco by Express Scripts.

1.5.2 Plan Consolidation Treatment Effects

The next set of DD specifications includes plan consolidation as an additional treatment effect. In contrast to a merger that is a combination of two distinct insurance companies offering Part D plans into a joint company, plan consolidation is a combination of two or more plans offered by an insurance company into a single plan for the upcoming year. In this sense, our classification of a merger event can be though of as an *inter*-firm combination, and plan consolidation is an *intra*-firm combination. Note that a non-merging insurer can consolidate its own plans; in periods that an insurer merges it can consolidate its own plans or consolidate with plans offered by its merger

[4] The exceptions where a lot of entry is observed are 2006, when all plans were new plans by definition, and 2007 when the market was still in its nascency.

partner. Insurers cannot consolidate plans with a rival company.

We specify the following DD estimator for consolidation:

$$\begin{aligned} p_{it} - p_{it-1} = \alpha + \beta_1 D_{it-1}^{merge} + \beta_2 D_{it-1}^{cons} + \beta_3 D_{it-1}^{cons} * D_{it-1}^{merge} + (\mathbf{X}_{it} - \mathbf{X}_{it-1})'\beta \\ + \varphi_t + \varphi_{market} + \varphi_{insurer} + \epsilon_{it-1} \end{aligned} \tag{1.2}$$

The treatment dummy for plan consolidation $D_{it-1}^{cons} = 1$ if plan i is consolidated with another plan between years $t-1$ and t, and the M&A treatment dummy $D_{it-1}^{merge} = 1$ follows the same definition as that described in equation (1.1). The additional term $D_{it-1}^{cons} * D_{it-1}^{merge}$ measures the interaction effect of a plan being affected by both a merger and consolidation event. We also consider the treatment effect on formulary counts $f_{it} - f_{it-1}$ and the copay price index $copay_{it} - copay_{it-1}$.

The same identification issues discussed above for mergers apply for plan consolidation treatment effects. The exogeneity assumption is perhaps more tenuous. A major concern is that insurers consolidate under-performing plans as a way to remove them from the market. In addition to the many product characteristic control variables, we control for under-performance by including measures of prior year enrollment and markets shares. There is also strong evidence that institutional features of the Part D program are primary drivers of plan consolidation. The rules for determining the LIS threshold and subsidies are pegged to enrollment figures, giving insurers a strong incentive boost enrollment by consolidating plans. This is evident in the data. The normal frequency of consolidation is 20%, but for plans that switch status to becoming LIS eligible benchmark plans, the frequency rises to 42%. The other marked increase in consolidation came in 2011 when Medicare first announced guidelines directing insurers to consolidate low enrollment and "meaningfully" similar plans. Year fixed effects and covariates for LIS status capture both of these institutional features. The interaction term of mergers and consolidation is plausibly exogenous given the data indicate a similar fraction of plans are consolidated by merging firms as non-merging firms (see table B.2).

1.5.3 Testing the Three Theories of Mergers

One our of main objectives is to distinguish the three channels through which mergers affect markets: market power, cost efficiencies, and upstream monopsony power. In many industries, all three channels likely impact merger outcomes. Retrospective merger studies that examine product

market prices are interesting in that they show the net effect of the three channels, but do not distinguish how much each factor contributes to the outcome. Prospective merger simulation studies have difficulty forecasting cost efficiency and monopsony power effects and instead are often based on modeling approaches that assume there are only market power effects (Weinberg and Hosken (2013)). Our contribution is to show that all three are important.

We use two extra pieces of information—over and above price data—to test the theories. First, we exploit the distinction between inter-firm mergers and intra-firm plan consolidation. Second, we test how mergers affect product characteristics: in our application coverage characteristics. Throughout the results section, we discuss a series of assumptions about the market to test the theories. The basic idea of our hypothesis can be summarized as follows. Only merging firms gain market power and monopsony power. Firms realize cost efficiencies and monopsony power by consolidating plans. Taken together, the hypothesis implies the merger dummy in specification (2) measures the market power effect on prices, the consolidation dummy measures cost efficiencies, and the interaction term measures monopsony power. The same logic applies to the product characteristic measures; however given the design of the Part D program we expect upstream monopsony power to be a more important determinant of coverage characteristics than market power. There is little reason to believe that administrative and marketing cost efficiencies would translate into changes in coverage characteristics.

1.6 Results

In this section we report results of the differences-in-differences estimates for plan premiums and the three coverage characteristics: the total number of drugs covered on formularies, the number of top 100 drugs on formularies, and the out-of-pocket cost for a basket of the top 100 drugs.

The results for each outcome variable are presented using three panels. Our main findings are shown in the panel labeled C. They are estimates from specification (2) that includes the merger treatment D^{merge}, consolidation treatment D^{cons}, and their interaction $D^{merge} \times D^{cons}$. Panel A shows results from specification (1) that includes only the merger treatment D^{merge}; panel B reports for the specification that only includes the consolidation treatment D^{cons}. These two specifications are reported for comparison purposes. We also show estimates with and without insurer fixed effects. The standard errors are large in specifications with insurer fixed effects because there is less within-

insurer variation in the covariates. However, the point estimates generally have the same signs and magnitudes as the specifications without insurer fixed effects. We focus our interpretation on the results that include insurer fixed effects.

1.6.1 Mergers and Plan Premiums

Table 1.3 reports the results for the effect on premiums. The tables suppress coefficients on the control variables; full results are in the appendix. Panel A shows the merger treatment effect in isolation, without regard to consolidation. The results indicate that when insurers merge, the premiums on their plans go up by \$3.61 relative to the premiums for insurers that do not merge. Given the average premium of \$45 across years, the rise corresponds to an 8% increase. Theory suggests the higher premium for merged firms is due to a strong market power effect dominating cost efficiency and upstream-monopsony power effects.

Panel B reports the consolidation treatment effect in isolation. The results show how premiums for plans that were consolidated (treatment group) change with respect to the premiums for plans that were renewed (control group). Premiums for consolidated plans are \$3.86 (8.7%) lower relative to the control group of plans that are renewed across years. This result suggests that insurers are either achieving cost efficiencies or gaining monopsony power over drug suppliers by consolidating their plans.

Panel C reports estimates from the specification that jointly estimates merger and consolidation effects. This specification measures three treatment effects relative to the omitted category of not-merging/not-consolidating. The coefficient on the merger dummy, D^{merge}, indicates premiums are \$3.84 (8.5%) higher for the plans of merged insurers that are renewed but not consolidated. This result supports a strong market power effect of mergers. The coefficient on the consolidation dummy, D^{cons}, shows consolidated plans of non-merging insurers are \$3.42 (7.6%) lower than renewed plans of non-merging insurers. This drop could either be caused by a cost efficiency or upstream-monopsony power effect. This result is not influenced by market power effects because the comparison is between plans of non-merged insurers. The difference in premiums between consolidated plans of merged insurers and renewed plans of non-merged insurers is given the by sum of the merger, consolidate, and interaction term coefficients, $D^{merge} + D^{cons} + D^{merge} \times D^{cons}$. The premiums are \$1.69 (3.8%) lower, suggesting cost efficiencies and/or monopsony power effects

dominate market power effects when merging insurers consolidate their plans.[5] This result stands in stark contrast to the finding that renewed plans of merged insurers are priced higher.

The results for plan premiums provide the first set of evidence that we use to disentangle the three competing effects in the merger theory. The effects are separately identified under two assumptions. First, if the act of renewing plans by merging insurers implies that the insurers do nothing to restructure the management of plans or renegotiate contracts with drug suppliers, then there is no cost efficiency or upstream-monopsony effect. Under this assumption the coefficient on the merger dummy measures the market power effect stemming from the ability of merging insurers to coordinate pricing decisions. Second, the cost efficiency and monopsony power effects can be separated by further assuming that monopsony power over drug suppliers is solely determined at the insurer level, not the plan level. To the extent that enrollment determines bargaining positions with drug suppliers, this assumption can be interpreted to mean that insurer-wide enrollment (in both Part D and non-Part D plans) matters for monopsony power, not how an insurer's enrollees are allocated across individual plans. Under this assumption the coefficient on the consolidation dummy measures the cost efficiencies achieved from restructuring the management and marketing of its plans. This coefficient does not measure a market power effect because no merger takes place, and, under our assumptions, it does not represent a monopsony power effect because there are no overall gains in enrollment at the insurer level for a non-merging insurer consolidating its plans. The monopsony power effect is given by the coefficient on the interaction of the merger and consolidate dummy: $D^{merge} \times D^{cons}$. If insurers renegotiate contracts with drug suppliers when they consolidate plans, a merged insurer with a larger base of enrollees will have stronger monopsony power.

In summary, the disentangled results indicate the market power effect of mergers raises premiums \$3.84, cost efficiencies reduce premiums \$3.42, and the extra monopsony power effect reduces premiums \$2.11. The net effect for merging insurers that consolidate plans is the sum of the three effects: a decrease in premiums of \$1.69.

1.6.2 Mergers and Drug Coverage: Formularies

Our next set of results investigates how mergers and plan consolidation affect coverage characteristics. First, we look at the composition of drug formularies to gouge the generosity of

[5]When insurer fixed effects are excluded and the estimates are less noisy, the combined effect of merging and consolidating remains negative and passes an F-test of joint significance differing from zero. However it fails at reasonable significance levels in the specification with insurer fixed effects.

Table 1.3: Difference-in-Difference Estimates: Premiums.

	A		B		C	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Merger-affected plan	1.703	3.607			2.241	3.840
	(0.363)	(2.219)			(0.400)	(2.494)
Consolidated plan			-4.221	-3.861	-3.911	-3.422
			(0.320)	(1.339)	(0.343)	(1.547)
Consolidated x Merger plan					-2.199	-2.105
					(0.827)	(2.127)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	8,839		F-test		29.7	0.6
N of M&A affected plans	1,375					
N of consolidated plans	1,994					
N of M&A consolidated plans	296					

Notes: Panel A shows estimates for the plans involved in a merger; this specification does not distinguish between mergers that consolidated plans and mergers that didn't. Panel B shows estimates for the plan consolidation effect on premiums. Panel C includes the merger-consolidated plan interaction term. The F-test null hypothesis is that the sum of the coefficients on merger dummy, consolidation dummy and their interaction term is zero. Standard errors are in parentheses, clustered by pre-merger insurer for specification with pre-merger insurer fixed effects. Coefficients on the suppressed controls are presented in Table B.4 of the Appendix.

drug coverage offered by a plan. We use two measures: the number of top 100 drugs covered on a plan's formulary in table 1.4 and the total number of all NDCs in table 1.5. The top 100 captures how generous coverage is for a general Medicare population that is likely to take some of the most popular drugs. The all NDCs list reflects how well the plan serves a diverse population, with some individuals requiring special treatments outside of the most common medicines list. Note that these measures are not necessarily closely correlated. At the extreme, one plan may cover all drugs from the top 100 and a minimal number of drugs outside the top 100. Another plan may have a limited selection of the most common drugs but have a variety of other options on its formulary.

For the top 100 drugs, panel A and panel B show that mergers and plan consolidation when taken in isolation have a near zero effect on drug formularies. For the all NDCs list, the effects are also near-zero, however there may be some evidence in the specification with insurer fixed effects in panel A that mergers lead to less formulary coverage. Although these results don't reveal any meaningful effect on formulary coverage, we find large effects in the specification that includes the interaction of merging and plan consolidation in panel C. The coefficient on the merger dummy, D^{merge}, indicates renewed plans of merged insurers delist 1 of the top 100 drugs and 320 from the all NDCs list. Given that the average plan lists 90 out of top 100 drugs and 2,700 NDCs, these

changes represent decreases in percentage terms of 1.2% and 11.9% respectively. The top 100 figure may seem small, but, stated equivalently, one fewer listed drug corresponds to a 12% increase in the number of top 100 excluded from formularies. The coefficient on the consolidation dummy D^{cons} indicates a decrease in coverage, slightly less than the merger effect for the top 100 drugs (-0.9), and much smaller in magnitude for all NDCs (-62). The largest effect is for merged insurers that consolidate plans. The interaction term $D^{merge} \times D^{cons}$, is an increase in the top 100 of 4.5 top 100 drugs and 550 NDCs, which in percentage terms represent increases of 5% and 20% respectively. The combined effect of merging insurers consolidating plans nets a very large increase in drug coverage relative to the more modest effects for merging insurers that renew plans and non-merged insurers consolidating plans.

These results provide further evidence on the three theories of mergers. The large increase in coverage for consolidated plans of merged insurers indicates a strong monopsony power effect. By consolidating and renegotiating contracts with drug suppliers, merged insurers with a larger base of enrollees have greater bargaining power to extract better terms from drug suppliers. The results suggests greater bargaining power allows insurers to offer substantially broader drug coverage for both top 100 drugs and across the full spectrum of all NDCs. Apart from greater bargaining power, it is also plausible that the merging insurers are able to combine their pre-merger formularies into a single more extensive formulary.

The near zero effects (or modest effects) on formularies found for merged/non-consolidated and non-merged/consolidated plans are also of interest. That consolidation by non-merging insurers does not increase coverage (or somewhat decreases for top 100 coverage) supports the hypothesis that bargaining power is not determined at the plan level. Returning to interpretation of the premium results, these formulary results indicate the large drop in price from consolidation are attributed to efficiency factors, not monopsony power. The modest negative effect on coverage for merged insurers that renew plans could be indicative of a market power effect, whereby the larger firm exercises market power by reducing the quality of their plan offerings. That the negative effect is larger for the NDCs measure than the top 100, could indicate insurers exercise monopoly power by horizontally differentiating their formularies. That is, after the merger, drugs for some specialized classes of medical conditions are retained for one of their plans, yet dropped on another plan to make the plans appeal to different sets of consumers.

Table 1.4: Difference-in-Difference Estimates: Formulary, Top 100 Drugs.

	A		B		C	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Merger-affected plan	0.391	-0.146			-0.492	-1.081
	(0.172)	(1.872)			(0.189)	(2.025)
Consolidated plan			-0.196	-0.176	-0.866	-0.880
			(0.155)	(0.922)	(0.165)	(0.940)
Consolidated x Merger plan					4.357	4.459
					(0.396)	(2.244)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	8,839		F-test		77.4	1.48
N of M&A affected plans	1,375					
N of consolidated plans	1,994					
N of M&A consolidated plans	296					

Notes: Dependent variable is the change in the number of drugs ranked in top100 by prescriptions filled, in the formulary. Standard errors are in parentheses, clustered by insurer for specification with insurer fixed effects. Coefficients on the suppressed controls are presented in Table B.5 of the Appendix.

1.6.3 Mergers and Drug Coverage: Out-of-pocket Drug Cost

For a complete picture of the effect on drug coverage, we consider out-of-pocket drug costs. The outcome of interest is the out-of-pocket cost in copays/coinsurance that an enrollee pays for a basket of top 100 drugs in the initial coverage zone after deductibles have been met. Three components influence out-of-pocket costs: the number of drugs out of top 100 list covered by a plan's formulary, copay and coinsurance rates, and the list price for each drug negotiated with drug manufacturers. The negotiated price matters for out-of-pocket costs for drugs covered by a coinsurance scheme (percentage of drug price) as opposed to copayment which is a fixed dollar amount. If a drug is covered by the plan, it enters the basket with its respective copay rate or its coinsurance rate times negotiated price. For drugs not listed on the formulary, we assume that an enrollee pays the full retail price which we set to the average regional (if available) or national drug price. The out-pocket-cost complements the formulary count outcome as it measures not just the number of covered drugs, but also the cost of covered drugs. With negotiated prices and copay/coinsurance rates included, it encompasses the most direct measure of the bargaining power insurers have with drug manufacturers and as such may be a better indicator of monopsony power effects.

The results for the out-of-pocket cost measure are generally consistent with those found

Table 1.5: Difference-in-Difference Estimates: Formulary, All Drugs.

	A		B		C	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Merger-affected plan	43.56	-182.80			-47.08	-320.23
	(25.83)	(338.65)			(29.15)	(354.33)
Consolidated plan			16.57	30.60	-45.12	-62.34
			(22.58)	(109.96)	(24.29)	(123.18)
Consolidated x Merger plan					373.07	552.93
					(56.41)	(221.75)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	7,396		F-test		34.9	0.2
N of M&A affected plans	1,082					
N of consolidated plans	1,746					
N of M&A consolidated plans	276					

Notes: Dependent variable is the change in the number of drugs included into the formulary. 2006-2007 year-plan pairs are excluded. Standard errors are in parentheses, clustered by pre-merger insurer for specification with insurer fixed effects. Coefficients on the suppressed controls are presented in Table B.6 of the Appendix.

for the drug formulary measures but are noisier. The most stark result in panel C of table 1.6 is the large negative coefficient (-$3) on the interaction term of merging and consolidating. Given an average cost for the basket of top 100 drugs of $63, the result represents a decrease in cost of 4.8%. Following our interpretation of the theories, the decrease indicates a strong monopsony power effect that merging insurers can achieve by consolidating plans. For non-merging insurers, consolidation has the opposite effect; out-of-pocket costs increase $1.40. This supports the notion that insurers cannot increase their monopsony power by consolidating plans, and further supports the hypothesis that premium reductions for consolidated plans are due to cost efficiency effects. The estimate on the merger dummy D^{merge} indicates a monopoly power effect for merging insurers that renew plans. Although the regulations require insurers to pass on all negotiated drug prices to enrollees, they can exercise monopoly power over out-of-pocket drug costs by raising copay and coinsurance rates. This appears to be happening for renewed plans of merged insurers, in which out-of-pocket costs increase by $2.41. However, the result is not robust to the exclusion of insurer fixed effects.

Comparing the results on formulary coverage to out-of-pocket costs for the interaction term, $D^{merge} * D^{cons}$ leads to the same conclusion that merging and consolidating plans improves coverage through increased monopsony power. But the combined effects $D_{merge} + D^{cons} + D_{merge} * D^{cons}$, which is the ultimate outcomes for consumers, leads to divergent conclusions. Drug coverage in-

creases in terms of the number of drugs on the formulary (+2.5 top 100 drugs and +170 NDCs), yet decreases in terms of out-of-pocket costs (a rise of $0.80 for the top 100 drugs). Whether coverage improves depends on what is more important: drug costs or the scope of covered drugs. The bargaining process between insurers and drug manufacturers is certainly very complicated, involving many decisions about the inclusion of drugs, copay/coinsurance rates, and drug prices. The relatively stronger effect on the interaction term for formulary counts relative to that for the out-of-pocket cost, suggests that the decision about what drugs to include on formularies matters more in the bargaining process than the costs of those drugs.

Table 1.6: Difference-in-Difference Estimates: Price Index.

	A		**B**		**C**	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Merger-affected plan	-0.424	1.755			0.076	2.441
	(0.311)	(2.240)			(0.344)	(2.033)
Consolidated plan			1.706	0.908	2.132	1.440
			(0.280)	(1.152)	(0.300)	(1.299)
Consolidated x Merger plan					-2.723	-3.070
					(0.722)	(3.311)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	8,839		F-test		0.7	0.98
N of M&A affected plans	1,375					
N of consolidated plans	1,994					
N of M&A consolidated plans	296					

Notes: Dependent variable is the change in the weighted price of the basket of top100 drugs under each plan. Standard errors are in parentheses, clustered by pre-merger insurer for specification with insurer fixed effects. Coefficients on the suppressed controls are presented in Table B.7 of the Appendix.

1.7 Conclusion

This paper examines the effects of horizontal mergers amongst Part D insurers on prices and coverage characteristics. Our method applies a differences-in-differences identification strategy to a large panel of all Part D plans sold between 2006 and 2012. We make a distinction between mergers—inter-firm combinations—and plan consolidation—intra-firm combinations—to decompose the three channels through which mergers affect markets: market power, cost efficiencies, and upstream monopsony power.

We draw two main conclusions. First, we find evidence that mergers cause premiums to rise,

indicative of a strong market power effect. However, market power is offset when merging insurers consolidate plans. These cost savings stem from two sources: economizing on administrative expenses and market activities (cost efficiencies) and improving bargaining positions with drug suppliers (monopsony power). As further evidence on bargaining power, we find merging and consolidating plans leads to greatly improved drug coverage, yet merging on its own has a near zero effect on coverage. Our second conclusion is that plan consolidation by non-merging firms results in lower premiums, but does not improve drug coverage. These results suggest insurers can organically achieve cost efficiencies through plan consolidation, but only mergers alter market power and monopsony power.

Given the rapid pace of M&A activity in the industry, there is keen interest amongst anti-trust authorities and healthcare policy makers to scrutinize these deals. Our results offer a few lessons. Merger deals create considerable market power. However, there can be benefits in the form of lower premiums and improved coverage if the merging insurers restructure their plans to streamline costs and exercise monopsony power. Yet, cost efficiency alone is not a sufficient justification as non-merging insurers can also realize cost efficiencies. Balancing bargaining power and market power and weighing the importance of coverage versus price become the keys to an anti-trust investigation. There are also specific ramifications for Part D. Current policy aims to reduce the number of plans. Our results suggest policies should favor plan consolidation, as opposed to the elimination of insurers and restrictions on new entrants. Consolidation has the added benefit of creating cost efficiencies, and the further benefit of improved drug coverage if consolidation involves merging insurers.

There are several avenues for extending this work. A similar analysis could be conducted for vertical mergers. There are two types: mergers with pharmacies, such as the CVS Caremark deal, and M&A deals with pharmacy benefits managers (PBMs). PBMs historically acted as third party administrators who process claims and consult on formulary construction. Recently PBMs have been entering the market by acquiring the Part D assets of health insurers; at the same time, health insurers have been bringing PBM functions in-house through acquisition. Much of the current merger activity impacts broader health insurance markets outside Part D. A key difference is that bargaining with providers (hospitals, doctors) occurs at a local level, whereas it is at a national level for prescription drugs. Finally, new individual level administrative claims data is becoming available for Part D. Future work could examine how mergers and plan consolidation affect enrollment decisions and prescription drug usage.

Chapter 2

The Effects of Investment in Child Well-Being over Time: Children with ADHD

2.1 Introduction

The most recent National Survey of Children's Health 2011/12 reports that over 5 million children aged 2–17 (7.9%) have been diagnosed with attention-deficit/hyperactivity disorder (ADHD) in the U.S. Over 68% of these children are taking medications for the disorder.[1] However, very little is known about the relative effectiveness of available treatments and their effects on health, behavioral, and school outcomes, especially in the long-run.

In the Medicaid population, these statistics are even more pronounced. For example, in a cohort of children born in 1996 and ever eligible for SC Medicaid between 2003–2012, over 23% of children have been diagnosed with ADHD during the sample period. About 80% of those diagnosed with the condition were prescribed pharmacological or non-pharmacological treatment.

In 2012, SC Medicaid spent $62.1 million on ADHD prescription medications. That is three-fold the amount the program spent in 2003, despite a 200% increase in share of generic prescription claims over this time period.[2] The increase comes from the rise in the number of enrollees diagnosed with ADHD (68%), prescriptions per patient (18%), and cost of medications for Medicaid (98%). In contrast, SC Medicaid total spending on physician visits by patients with ADHD increased by just 3% between 2003 and 2012, which translates into a 38% decline in per patient spending.

In recent years the media launched an attack on the rapidly rising trend of ADHD diag-

[1] National Survey of Children's Health. NSCH 2011/12. Data query from the Child and Adolescent Health Measurement Initiative, Data Resource Center for Child and Adolescent Health website. Retrieved on 09/25/2014 from www.childhealthdata.org.

[2] Author's calculations from the SC Medicaid 2003-2012 claims data set.

noses and prescriptions.[3] In these articles, mental health professionals argue that ADHD drugs are overprescribed. The drugs are said to be overused by young people due to their immediate effect on the ability to concentrate. They warn the reader that the life-long consequences of taking these medicines are unknown. Moreover, there is no evidence of any long-term positive effects on educational or behavioral outcomes, but there are a number of worrisome side effects such as slowdown in growth and addiction.

This stance on ADHD medications may come from the fact that clinical studies last just a few weeks; they are unable to study long-term effects of treatment. Excluding the extreme cases when ADHD drugs are taken solely to improve performance on a particular test, the argument misses the fact that the stock of human capital is inherently dynamic in nature. Even if the contemporaneous effects of treatment are short-lived, the child has an opportunity to learn how to manage her condition while on treatment and accumulate social and cognitive skills that will improve her outcomes later in life.

This paper focuses on the effects of human capital investments on health and socio-economic outcomes of children with ADHD. It is a common chronic mental condition that impairs children's noncognitive skills. Patients of the hyperactive type lack self-control and patience; they demonstrate immature behavior that is inconsistent with their age group. Inattentive type patients have a poor ability to concentrate and complete tasks; they are forgetful.

Once a child is diagnosed, her family can invest in medical treatment to reduce the gap in abilities of a child with ADHD compared to her non-ADHD peers. While on treatment, the child is able to improve her outcomes in the short-run, accumulate cognitive and noncognitive skills and possibly improve her long-run outcomes.

Using a large 10-year panel of SC Medicaid claims, I evaluate available ADHD treatment strategies in the framework of investment in child development. The length of the panel allows me to take advantage of an empirical approach that is commonly used for dynamic processes that have potential unobserved heterogeneity problem (see Mroz and Savage (2006) and Yang et al. (2009)).

I model and simultaneously estimate the event of the initial diagnosis, treatment choice, and the probability of adverse events later in life. The discrete factor random effects estimator

[3]See for example, "Ritalin Gone Wrong." by Sroufe, L. Alan. The New York Times, January 28, 2012; "Risky Rise of the Good-Grade Pill." Schwarz, Alan. The New York Times, June 9, 2012; "Drowned in a Stream of Prescriptions." Schwarz, Alan. The New York Times, February 2, 2013; "A Nation of Kids on Speed." Cohen, Pieter and Rasmussen, Nicholas. The Wall Street Journal, June 16, 2013; and "The Truth About Smart Drugs" by Marek Kohn, BBC, July 29, 2014.

is beneficial in this setting because it can be used to control for endogeneity biases in nonlinear models where fixed effect estimators would be inconsistent. For comparison purposes, I estimate single-equation discrete choice models for all of the events and outcomes.

I find that there is a strong persistence in treatment choices across time periods. The results also suggest that pharmacological treatment has only short-term positive impact on the probability of such adverse events as injuries, teenage pregnancy, and STDs, and no impact on substance abuse disorders. Behavioral therapy alone is not as effective as it is in combination with ADHD drugs, but for STDs and substance abuse disorders it seems to show relatively long-lasting effects in contrast to drugs alone.

The rest of the paper is organized as follows. In the next section I review the relevant literature and outline my contribution. In section 2.3 I discuss the background of ADHD and develop a model using human capital accumulation framework; section 2.4 outlines the empirical model, followed by the data section 2.5. I conclude with results and discussion.

2.2 Literature Review

The literature on child development indicates that gaps in abilities that form early in life persist into adulthood and can explain a large array of differentials in adult outcomes. Conti and Heckman (2014), for example, provide an extensive review of the empirical evidence on the effects of the two dimensions of child well-being, cognitive and noncognitive skills, on educational attainment, asocial and risky behaviors, and health. They emphasize the importance of modeling multidimensional capabilities as opposed to the earlier literature on human capital development that concentrated on cognitive abilities of a child, often measured by IQ, to explain the outcomes later in life.

One of the earliest studies to account for the latent noncognitive skills is Heckman et al. (2006). They find that both cognitive and noncognitive abilities affect wages, schooling, work experience, occupational choice, and participation in a range of adolescent risky behaviors. These results have important policy implications, but most interventions target children's cognitive rather than noncognitive abilities. The Perry Preschool experiment, for example, did not result in the IQ improvements. Nonetheless, the program had a beneficial impact on many child outcomes. Heckman et al. (2006) argue that these beneficial impacts were achieved by altering social skills.

There is an emerging health economics literature on the effects of ADHD treatment on short- and long-term outcomes. For example, Currie et al. (2014) use a quasi-natural policy experiment that lowered prices on all prescription drugs in Quebec, Canada but not in other provinces. They find little evidence of positive effects on academic outcomes and even some evidence of negative impact of treatment on grade repetition, math scores, and emotional stability of girls. Dalsgaard et al. (2014) look at health services utilization (hospital and ER visits) and behavioral outcomes (crime), using the variation in the doctor propensity to prescribe pharmacological treatment as an IV. They find a positive effect of treatment on patient health and behavior. Treated children had fewer hospital visits, due to fewer injuries, and they also had fewer encounters with the police. Using the same IV applied to a sample of children and young adults enrolled in SC Medicaid in 2003–2012, Chorniy and Kitashima (2014a) find that ADHD treatment reduces the probability of teenage pregnancy, contraction of an STD, and substance abuse for ADHD population – teenage pregnancy.

I contribute to this literature by explicitly capturing the dynamic nature of the process of human capital accumulation. I concentrate on the long-run effects of ADHD treatment on health outcomes (injuries), consequences of the risky sexual behavior (teen pregnancy and STDs) and other risky behaviors (substance abuse). I also distinguish between meaningfully different pharmacological treatments and include behavioral therapy into the choice set. This approach allows me to compare effectiveness of particular treatment sequences, accounting for treatment interruptions.

The problem of treatment selection is also addressed in the recent literature on choice under uncertainty (Crawford and Shum (2005), Dickstein (2014a), and Saxell (2013)). They use learning with a Bayesian updating framework to model the process of patient search for most suitable and cost-efficient drug. This approach is limited in that the only relevant information in the current period is the choice made in the period immediately prior to the current period.

In contrast, my model allows the entire past sequence of treatments to affect the current decision. Thus, I can directly test a hypothesis that some treatments are more valuable in the beginning period of treatment and others are more suitable for an established patient. To my knowledge, this approach has not yet been applied to the problem of the demand for treatment.

Another limitation of the literature on demand for medical treatment under uncertainty is the lack of data on patient outcomes. They rely on the assumption that a patient is cured when she exits treatment. ADHD is not curable. My data allow me to introduce a more realistic measure of treatment effectiveness – a number of behavioral and health outcomes that I identified from the

medical literature (see Barkley (2006) for a detailed review) and the literature on child well-being (e.g. Heckman et al. (2006)). The three adverse outcomes identifiable in my data are outcomes associated with risky sexual behavior, teenage pregnancy and STDs, health outcomes associated with poor attention and hyperactivity, injuries.

2.3 Conceptual framework

2.3.1 ADHD and noncognitive ability

Every child is born with a multidimensional endowment of abilities. They include cognitive (e.g. IQ, memory) and noncognitive skills (e.g. self-control, patience, time preference)(Conti and Heckman (2014)). Most recent medical research suggests that genetic and neurological factors are the greatest contributors to the ADHD (see Barkley (2006) for an extensive review). Due to their genetic condition, children who suffer from ADHD have a relatively low initial stock of noncognitive skills.

Poor noncongitive abilities may lead to a number of negative health and social outcomes, such as teen pregnancy, contraction of STDs, injuries, and substance abuse (as described in Section 2.4.3). According to the medical literature, ADHD can seldom be cured. It persists into adolescence in up to 70% of cases and into adulthood in up to 66% of childhood cases[4] (Barkley (2006)). In order to relieve symptoms of the condition and augment the stock of noncognitive skills, patients take ADHD drugs and/or attend psychotherapy sessions. The pharmacology of ADHD medicines is such that the contemporaneous effect of treatment goes away as soon as the patient stops taking them. However while on treatment, pharmacological or behavioral, patients are able to accumulate human capital. They can learn planning and self-control skills in order to better manage their ADHD symptoms in the future. Accumulation of ADHD "management" skills reduces the probability of adverse events in the future. Furthermore, treated children are more likely to do better in school and accumulate cognitive abilities. This makes up a link between previous ADHD treatment, current stock of noncognitive and cognitive skills, and future health and social outcomes, as well as the future treatment choices.

ADHD treatments can only be prescribed after the initial diagnosis. According to the Amer-

[4]The same professional opinion was expressed in an interview that I conducted with a developmental pediatrician who specializes in developmental pediatrics in December 2013.

ican Academy of Pediatrics guidelines[5], primary care clinicians should evaluate children between 4 and 18 years old for ADHD if they show some of the symptoms.[6] Since ADHD is a hereditary rather than an acquired condition, the timing of diagnosis depends on the severity of symptoms and ADHD type. Hyperactive and impulsive types are more likely to be diagnosed earlier than inattentive types simply because inattentiveness might be confused with poor cognitive skills. But if a child is acting up, parents and teachers are more likely to suggest medical diagnosis and treatment.

One of the important findings of the literature on child development is that investments in human capital are more productive earlier in life (Cunha and Heckman (2007)). I directly test this prediction by comparing the outcomes of early to later diagnosis and treatment. To formalize the model, I use the general theoretical framework of the technology of skill formation and investment in human capital laid out in Cunha and Heckman (2007) and Cunha et al. (2010).

2.3.2 Notation

In what follows I introduce the main variables and vectors of variables that I use in the model.

1. Initial endowment (latent variable), θ_{i0}. Child i is born with an initial endowment of cognitive and noncognitive abilities.

2. Adverse outcomes, $Y_{it}^{outcome}$. ADHD-related adverse outcomes are injuries, $Y_{it}^{inj,k}$, risky-sexual behavior outcomes, $Y_{it}^{sex,m}$, and substance abuse, $Y_{it}^{s.abuse}$ at time period t. For injuries, k indexes the three most prevalent kinds of injuries for ADHD population: superficial injuries, open-wound injuries, and internal injuries; and all other injuries as defined by the ICD9 categories, $k = 1, ..., 4$. For the outcomes of risky sexual behavior, $m = 1, 2$ and corresponds to teenage pregnancies and contraction of STDs. More detailed description of the outcomes is outlined in Section 2.4.3.

3. The event of the first ADHD diagnosis, D_{it}.

[5]Subcommittee on Attention-deficit/hyperactivity disorder, steering committee on quality improvement and management, "ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents", *Pediatrics*, 2011.

[6]Since a number of ADHD prescription drugs are approved by the FDA for use in children as young as 3 years old (e.g. Adderall, Adderall XR), I use age 3 as the first time period when a diagnosis can be made and treatment initiated.

4. Medical treatment, M_{ijt}. Children with ADHD can be prescribed pharmacological, behavioral, or a combination treatment, M_{ijt}, where j indexes the type of treatment. The full choice set of pharmacological treatments is presented in Table 3.1. In this paper I estimate a simplified empirical model with $m = 1, 2, 3$ corresponding to medicines-only, behavioral therapy-only, and a combination of treatments choice options.

5. Medical treatment characteristics, Z_{jt}. A number of pharmacological treatment characteristics affects treatment decision. They include drug branded status, side effects, dosing frequency, preferred-drug list status, and drug prices paid by the patient and by Medicaid.[7] For behavioral therapy treatment, this vector includes session duration, type (individual, group, and/or with parent present), and its cost.

6. Child characteristics, X_{it}^{child}. This vector includes constant (race and gender) and period-specific (age and county of residence) variables.

7. Mother characteristics at the child's birth, X_{i0}^{mother}. They include mother's age, race, education level, and history of mental disorders.

8. Medical provider characteristics, $X_{i}^{provider}$. They include provider office location (county) and specialty (e.g. pediatrician, psychiatrist, etc.).

9. Family and home environment characteristics, $X_{it}^{environment}$. I follow previous literature and use the child's foster status in year t and mother's current mental health status to approximate for the home environment. In addition, I use yearly information on the number of children and adults in the family.

2.3.3 Timing

The model timeline can be divided into three parts. First, when a child i is born she receives an initial endowment of noncognitive skills, θ_{i0}, that depends on genetic and environmental factors. Second, once the child reaches the age of 3, she can be tested for and diagnosed with ADHD in year t (D_{it}) if she has some of the symptoms. Finally, once the patient who has the condition is diagnosed, she can be prescribed a medical treatment (M_{it}) to relieve and treat symptoms of ADHD.

[7] Earlier literature finds that insurers exert pressure on doctors to prescribe cheaper options, e.g. see Dickstein (2014b).

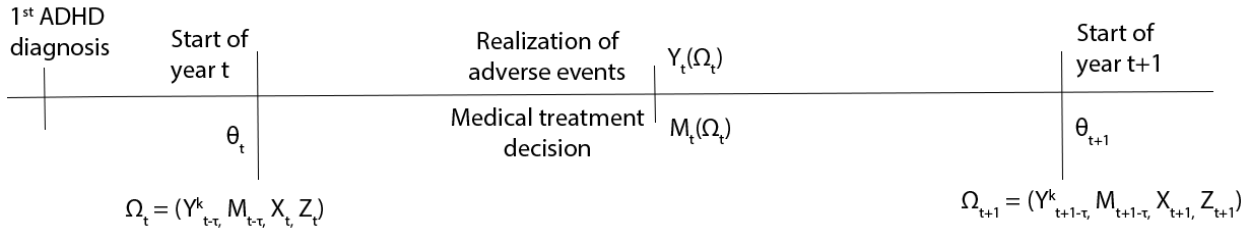

Figure 2.1: Decision Timeline

The treatment augments the stock of noncognitive skills that feeds into the next time period. It also affects the probability of the adverse outcomes that could be realized in the next period ($Y_{it+1}^{outcome}$).

Figure 2.1 depicts the dynamics of the stock of child's abilities and the probability of adverse events linked to the low level of human capital. It shows a representative year of the time period after the patient was diagnosed with ADHD. At the start of the year t the patient has information (Ω_t) on her stock of skills, adverse events that were realized in the past years, and past treatment (if any). During year t she will be making decisions on treatment for her condition. The adverse events will or will not be realized. Then, by year $t+1$ the individual will have an updated stock of skills due to medical treatment and she makes decisions based on the updated information set, Ω_{t+1}.

More precisely, at the beginning of each time period (year) an individual and her parents have the following information that influences the treatment choice in that time period: current stock of skills (θ_{it}), occurrence of adverse events in the past years ($Y_{i,t-\tau}^{outcome}, ..., Y_{i,t-1}^{outcome}$), ADHD diagnosis status ($D_{i,t-1}$) and if diagnosed, what treatment have they undergone in the previous periods ($M_{ij,t-\tau}, ..., M_{ij,t-1}$). The observed information also includes patient characteristics (X_{it}^{child}), physician characteristics ($X_i^{physician}$), and drug characteristics (Z_{jt}). Finally, they have information on current and lagged variables unobservable to the researcher that feed into the optimization problem.

Prior to the initial diagnosis of ADHD no treatment can be prescribed. The timeline for a representative year before the initial diagnosis is similar to the one shown on Figure 2.1, except for there is no treatment decision to make since no medical treatment can be prescribed before without the ADHD diagnosis. Individuals transition into the next period with an unaffected stock of noncognitive skills ($\theta_{i,t+1} = \theta_{it}$). In what follows I describe an empirical specification for every component of this dynamic system.

2.4 Empirical specification

The focus of this paper is investments in child noncognitive skills made in the form of ADHD treatments. The model has three main components: the event of the initial ADHD diagnosis, per-period treatment decisions, and the incidence of adverse events (injuries, teen pregnancy, STDs, and substance abuse).[8] Below I specify an equation for each of these components and an initial condition for the stock of the child's abilities at birth.

2.4.1 Initial Condition

Children are born with a multidimensional initial endowment of abilities that include cognitive (θ_0^C) and noncognitive skills (θ_0^N).

$$\theta_{i0} = (\theta_{i0}^C, \theta_{i0}^N) \tag{2.1}$$

ADHD is a chronic medical condition that impairs noncognitive skills. Some children, albeit being capable learners of high intellectual ability, have difficulty concentrating and controlling their behavior.

The process I am modeling includes two effects of ADHD treatment: direct effect on noncognitive skills and indirect effect on cognitive skills. When a child is able to manage her behaviors (noncognitive abilities), she is able to boost her IQ as well (cognitive abilities). Previous literature found that both cognitive and noncognitive abilities explain behavioral outcomes later in life (e.g. Heckman et al. (2006)). Since both kinds of skills are latent, my model is not distinguishing between the effect of treatment of cognitive and noncognitive skills. Nor does it distinguishes between the effect of cognitive and noncognitive stock of skills on the probability of adverse outcomes. I model the stock of skills as a unidimensional vector. It means that I am estimating the effect of investment in noncognitive skills that might be potentially multiplied via a cognitive skills channel, or an overall effect of treatment on child skills accumulation and thus the incidence of adverse outcomes later in life. It is left for the future research to disentangle these potentially important multidimensional effects.

Genetic factors are major inputs into the formation of the initial endowment of skills (Olds (2002), Levitt (2003)). They are captured by a set of mother's characteristics at birth (X_{0i}^{mother}):

[8] In the current version of the draft I do not estimate the effect of treatment on STDs and onset of substance abuse.

mother's age, education level, and history of mental disorders prior to the child's birth.[9]

$$\theta_{i0} = \alpha X_{i0}^{mother} + \rho_{i0}^{I} \mu_i, \tag{2.2}$$

where i indexes individuals with ADHD and μ_i denotes permanent unobserved heterogeneity, or mother, child, family environment characteristics at birth that are not observed by the econometrician. I do not directly estimate the Equation 2.2. Instead, I substitute it into the equations on the timing of the first diagnosis, treatment decision, and adverse outcomes equations.

2.4.2 Treatment

The initial stock of noncognitive skills can be altered with investments in child's development. Higher noncognitive abilities reduce the likelihood of adverse health and social outcomes that children and adolescents with ADHD are prone to. One way to improve the abilities of children with ADHD is to treat their disabling mental condition. In each period when a treatment is administered, ADHD symptoms subside and the child's ability to concentrate and control her impulses improves. Moreover, the child gets a chance to perform better at school and learn more cognitive and noncognitive skills. In other words, ADHD treatment alters the stock of noncognitive skills in the current period which is passed on to the next period.

In the model, pharmacological and behavioral ADHD treatments are assumed to be the only investments available to the parents to improve their child's noncognitive skills. I assume that both physician and parents are perfect agents of the child.[10]

At the beginning of year t patient i diagnosed with ADHD has the stock of abilities θ_{it}. The current stock of skills depends on the stock of skills and investments in skills in the previous period, $t-1$ as shown in Equation 2.3. Investments in the form of a variety of ADHD medical treatments, M_j, are available, where j denotes a particular treatment regimen: a medication from the choice set or behavioral therapy, or a combination of the two. Acquired skills do not depreciate with time. If a child does not receive any treatment in the current period, her stock of skills remains unchanged

[9] In the current version of the draft the empirical estimation does not include mother's mental health.

[10] Admittedly, this assumption is a significant simplification. In the literature, there is evidence on the importance of parent preferences and overall family environment on child skills formation (see Conti and Heckman (2014) for a review). Furthermore, physician preferences, incentives and information available to them were shown to be important in the treatment decision process (see for example, Dickstein (2014b), Saxell (2013)). However, it is left for the future work to relax this assumption.

from the current period to the next.

$$\theta_{it} = f_t(\theta_{it-1}, M_{it-1}) \tag{2.3}$$

Recursively, it can be written as follows.

$$\theta_{it} = g(\theta_{i0}, M_{it-1}, M_{it-2}, \ldots, M_{it-9}) \tag{2.4}$$

where t is a year in survey, $t = 1, \ldots, 10$; i is a patient, and j is a medical treatment.

The indirect utility from each treatment alternative ($j = \{0, ..., 11\}$) depends on the severity of ADHD, θ_{it}, adverse events realized in the past years ($Y^{outcome}_{i,t-\tau}, ..., Y^{outcome}_{it}$), provider characteristics, $X^{provider}_{i}$, patient characteristics X^{child}_{it}, treatment characteristics, Z_j, and family environment, $X^{environment}_{it}$, where $\tau = \{1, ..., 9\}$ since there are at most 9 years of history available for each patient i at time t. For detailed description of these vectors see Section 2.3.2.

At the beginning of period t, the patient, her parents and doctor observe $\Omega_{it} = (Y^{outcome}_{it}, ..., Y^{outcome}_{i,t-\tau}, M_{ijt}, ..., M_{ij,t-\tau}, X_{it}, Z_{jt})$, where $X_{it} = (X^{child}_{it}, X^{provider}_{it}, X^{mother}_{i0}, X^{environment}_{it})$. Then, the expected indirect utility of treatment j for child i is

$$V^M_{ijt} = v(\theta_{it}, Y^{outcome}_{i,t-\tau}, ..., Y^{outcome}_{it}, X_{it}, Z_{jt}, M_{ij,t-\tau}, ..., M_{ij,t-1}, M_{it} = j) + u^M_{ijt} \tag{2.5}$$

where $\tau = \{1, ..., 9\}$ and u^M_{ijt} is unobserved individual heterogeneity that influences treatment choice decisions.

There are unobserved individual characteristics in this model that may affect the choice of treatment and the efficacy of treatment. For example, child's cognitive skills may have a positive effect on treatment choice if the child is relieved from ADHD symptoms and improves her academic performance significantly when treated. To account for the unobserved characteristics, I follow Mroz and Savage (2006) and Yang et al. (2009) among others and decompose the error term, u^M_{ijt}, into the three components. The first component, μ_i, captures permanent unobserved heterogeneity (e.g. preference for medical treatment). The second component captures time-varying unobserved heterogeneity, ν_{it} (e.g. cognitive abilities). The third part is a serially uncorrelated error term, ϵ^M_{ijt}, that expresses an individual's random preferences for medical treatment of ADHD.

Equation 2.6 details the error term structure.

$$u_{ijt}^{M} = \rho_{ij}^{M}\mu + \omega_{ij}^{M}\nu_t + \epsilon_{ijt}^{M} \tag{2.6}$$

where ρ^M, μ, and ω^M, and ν_t are estimated parameters of the empirical model. The discrete mass points of the permanent and time-varying heterogeneity distributions are denoted $\mu \in \{\mu_1, \mu_2, ..., \mu_G\}$ and $\nu_t \in \{\nu_{1t}, \nu_{2t}, ..., \nu_{Lt}\}$, respectively, where G and L are the number of mass points in the discrete approximations to the distributions.

Substituting Equation 2.6 into Equation 2.5 and assuming an Extreme Value distribution of the additive idiosyncratic error term, ϵ_{ijt}^{M}, in the alternative-specific value function for treatment, the individual's decision rule is to choose a treatment regimen with the highest indirect utility. This assumption yields a multinomial logit distribution of current treatment choices as a function of the theoretically relevant variables known to the individual at the start of the period t including treatment choice in the previous periods.

$$\begin{aligned} ln\Big[\frac{Pr(M_{it}=j)}{Pr(M_{it}=0)}\Big] = \beta_{0j} + \beta_{1j}\theta_{it} + \sum_{\tau=1}^{9}\lambda_{\tau j}M_{ijt-\tau} + \sum_{\kappa=1}^{4}\sum_{\tau=1}^{9}\nu_{\kappa\tau j}Y_{t-\tau}^{inj_\kappa} + \beta_3 X_{it} + \beta_{4j}Z_{jt} \\ + \rho_j^M\mu_i + \omega_{ijt}^M\nu_{it} \end{aligned} \tag{2.7}$$

where $j = 1, 2, ..., 11$.

Substituting θ_{t-1} for θ_t from the equation 2.3, we obtain

$$\begin{aligned} ln\Big[\frac{Pr(M_t=j)}{Pr(M_t=0)}\Big] = \beta_{0j} + \beta_{1j}[\theta_{t-1} + M_{t-1}] + \sum_{\tau=1}^{9}\lambda_{\tau j}M_{jt-\tau} + \sum_{\kappa=1}^{4}\sum_{\tau=1}^{9}\nu_{\kappa\tau j}Y_{t-\tau}^{inj_\kappa} + \beta_3 X_{it} + \beta_{4j}Z_t \\ + \rho_j^M\mu_i + \omega_{ij}^M\nu_{it} \end{aligned}$$

Continued substitution results in the following expression:

$$ln\Big[\frac{Pr(M_{it}=j)}{Pr(M_{it}=0)}\Big] = \beta_{0j} + \alpha\beta_{1j}X_{i0}^{mother} + \sum_{\tau=1}^{9}\delta_{\tau j}M_{ijt-\tau} + \sum_{\kappa=1}^{4}\sum_{\tau=1}^{9}\phi_{\kappa\tau j}Y_{i,t-\tau}^{inj_\kappa} + \beta_{3j}X_{it} + \beta_{4j}Z_t + \rho_j^M\mu_i + \omega_{ijt}^M\nu_{it} \tag{2.8}$$

where $j = 1, 2, ..., 11$.

2.4.3 Adverse events

A low level of noncognitive skills is an important determinant of poor educational, labor market, and social outcomes. Following the medical literature (a detailed review is provided by Barkley (2006)) and the literature on child development and well-being (Dalsgaard et al. (2014), Heckman et al. (2006), Carneiro et al. (2007), Conti and Heckman (2010), Goodman et al. (2011)), I concentrate on the three adverse events that are common among children and young adults with ADHD: teen pregnancy, contraction of STDs, and injuries. Chorniy and Kitashima (2014b) extend this paper to include educational attainment and schooling outcomes.

2.4.3.1 Risky sexual behavior

Adolescents with untreated ADHD have difficulty controlling their impulses and planning ahead. These teens also tend to struggle with low self-esteem and for that reason, teenage girls often seek affirmation through the sexual attentions of boys (Arnold (1996)).[11] Their condition makes them more likely to become sexually active earlier than their peers, to have more partners on average, and to use inconsistently use birth control (Kessler et al. (1997), Payne (2014)). I focus on the two adverse events associated with risky sexual behavior: teen pregnancy and contraction of a sexually transmitted disease (STD).

In 2013, in the U.S. 274,641 babies were born to mothers aged 15–19 years and 3,108 babies to mothers under 15 years old, a live birth rate of 26.6 and 0.3 per 1,000 women in these age groups

[11] Adolescent girls' symptoms of ADHD often worsen due to the hormonal changes at puberty (Resnick (2005)).

respectively (Hamilton et al. (2014)).[12] About 80% of teenage births are unplanned or unwanted (Mosher et al. (2012)) and only 59% of them ended with a live birth in 2008 (Finer and Zolna (2011)).

Teenage pregnancy is a negative social and health outcome. Adolescent mothers are more likely to be single, to be on welfare and to have a hard time getting off welfare. Teenage pregnancy is also associated with negative consequences for the mother later in life (low educational attainment, poor employment outcomes, and marital instability) and poor child outcomes (low birth weight, delay in cognitive development, school problems, and behavioral disorders, see Kessler et al. (1997) for a review.

While the incidence of teen pregnancy in the U.S. is declining, the trend for cases of STDs has been increasing since the early 2000s. In 2012, there were 49,903 cases of STDs (16.0 per 100,000 population). Adolescents ages 15–24 account for nearly a half of the new cases of STDs each year (STD Fact Sheet (2013)).

To my knowledge, there is no research that looks at the effects of ADHD treatment on teenage pregnancies and incidence of STDs directly. However, there is empirical evidence of the importance of both latent cognitive and noncognitive skills for teenage pregnancy, among other adverse behavioral outcomes (Carneiro et al. (2007) and Heckman et al. (2006)).

2.4.3.2 Injuries

Inattentiveness, difficulty in assessing potential outcomes, and motor incoordination are a frequent cause of accidental injuries (e.g. fractures) for patients with ADHD. Besides having more frequent injuries, these children also tend to have more severe injuries than their peers (Barkley (2006), Swensen et al. (2004)). In particular, among the stronger findings in the medical literature is that ADHD adolescents are more likely to have a car crash and they are more often at fault in such accidents (Barkley (2006), Weiss and Hechtman (1993)).

In their work on the long-term consequences of ADHD treatment, Dalsgaard et al. (2014) find that pharmacological treatment of ADHD results in fewer hospital and emergency room visits. They argue that this result is driven by the reduction in injuries.

[12] CDC classifies births as teenage births if mother is between 10 and 19 years old. The subclassification by age splits teen moms into three groups: ages 10-14, 15-17, and 18-19.

2.4.3.3 Substance use and abuse

According to the 2011 Youth Risk Behavior Survey, 20.5% of the teens drank alcohol and 8.1% tried marijuana for the first time before 13 years old; 6.8% ever used cocaine, 11.4% ever used inhalants, and 2.9% ever used heroin in the U.S. The estimated economic cost of substance abuse is non-trivial. Miller and Hendrie (2009) combine condition-specific studies published between 2000 and 2004 and report that alcohol abuse cost the nation $191.6, tobacco use was responsible for $167.8 billion, and drug abuse accounted for $151.4 billion making up a total of $510.8 billion.[13] These costs include the costs of medical treatment and productivity costs.

Medical literature documents conflicting evidence on the association of ADHD and substance use and abuse, including alcohol, tobacco, and drugs. Looby (2008) provides a thorough review of major studies that address this question. Some of them find that teens with ADHD are more likely on average than individuals without ADHD to smoke, use and abuse alcohol and drugs, and develop health problems related to these activities. However, others conclude that there are additional related conditions that contribute to the likelihood of these adverse events, e.g. conduct disorder symptoms and association with deviant peers.

Despite a disagreement on the relationship between ADHD and substance use, Looby (2008) review suggests that ADHD treatment reduces the risk of substance use disorders in children with ADHD. Using a meta-analysis, Wilens et al. (2003) also find that stimulant medications reduce the risk for subsequent drug and alcohol use disorders.

Following the methodology described in Bouchery et al. (2012), I was able to identify cases of substance abuse from the insurance claims data using ICD-9 diagnosis codes: 291, 292, 303, 304, and 305.

2.4.3.4 Empirical specification

In order to estimate the effect of treatment on the adverse outcomes of interest I specify an equation for an occurrence of each outcome. Each of these events is modeled as a discrete outcome. Adverse outcomes depend on the current stock of skills the child i has accumulated by period t.

$$Y_{it}^{*outcome} = \gamma_0 \theta_{it} + \gamma_1 X_{it} + \rho_{it}^{Y} \mu + \omega_{it}^{Y} \nu_t + \epsilon_{it}^{Y} \tag{2.9}$$

[13] Estimates are given in 1999 dollars. For alcohol, Harwood (2000), trend-adjusted from 1998 to 1999; for tobacco, Fellows et al., (2002) except illness earnings loss from Harwood & Bouchery (2001); for other drugs, Harwood & Bouchery (2004).

where $Y_{it}^{outcome} = 1$ if $Y_{it}^{*outcome} > 0$ and 0 otherwise, t indexes years, $t = 1, 2, ..., 10$ that correspond to the survey period 2003-2012.

Similarly to above, by continuously substituting θ_{t-1} for θ_t from the equation 2.3, I obtain the following expression:

$$Y_{it}^{*outcome} = \gamma_0 \alpha X_{i0}^{mother} + \sum_{\tau=1}^{9} \zeta_1 M_{ij,t-\tau} + \gamma_1 X_{it} + \rho_{it}^{Y} \mu_i + \omega_{it}^{Y} \nu_{it} + \epsilon_{it}^{Y} \tag{2.10}$$

The outcomes of risky sexual behavior are age-specific. As discussed earlier, teenage pregnancies are defined as a pregnancy-related medical treatment for female patients older than 11 and younger than 19 years of age. STDs are an adverse outcome for the same age group of the entire ADHD population.

2.4.4 First diagnosis of ADHD

An eligible child-enrollee can be tested for and diagnosed with ADHD at a medical provider office. Any doctor is able to diagnose and prescribe treatments (except for psychologists). In order to be diagnosed ($D_{it} = 1$), the test should reveal at least six of the inattention symptoms and/or at least six of hyperactivity-impulsivity symptoms that "have persisted for a least 6 months to a degree that is maladaptive and inconsistent with developmental level."[14] It is extremely rare for a child to be diagnosed before age 3 because the symptoms are not apparent at this age.

Whether ADHD is diagnosed in any given year depends on the contemporaneous stock of noncognitive skills (θ_{it}, $t = 3, 4, ...19$ and $\theta_{i3} = \theta_{i0}$ by assumption) and on the history of adverse outcomes ($Y_{it-\tau}^{outcome}$). My default specification uses five lags leaving it to future research to refine time period and have a less restrictive specification.

Besides noncognitive skills level and history of adverse events associated with ADHD, the probability of being diagnosed depends on individual, X_i^{child}, family, $X_i^{environment}$, and medical

[14] The American Psychiatric Association publishes the Diagnostic and Statistical Manual of Mental Disorders (DSM), where it sets criteria for the classification of mental disorders. It is the standard classification of mental disorders used by mental health professionals in the United States. The DSM consists of three major components: the diagnostic classification, the diagnostic criteria sets, and the descriptive text. The most current version is DSM-5 published in May 2013, a revision of DSM-IV-TR that came out in 2000.

provider, $X_i^{provider}$ characteristics.

Diagnosis is specified as a latent variable and can be written as follows.

$$D_{it}^* = \delta_0 \theta_{it} + \sum_{\kappa=1}^{4} \sum_{\tau=1}^{5} \chi_{\kappa\tau j} Y_{t-\tau}^{inj_\kappa} + \delta_2 X_{it} + \rho_{it}^Y \mu_i + \omega_{it}^Y \nu_{it} + \epsilon_{it}^D \tag{2.11}$$

where $D_{it} = 1$ if $D_{it}^* > 0$ and 0 otherwise; t indexes the year of potential diagnosis, $t = 1, 2, ..., 10$ that correspond to the survey period 2003-2012.

Substituting θ_{i0} from the equation 2.4, I get the following expression:

$$D_{it}^* = \delta_0 \alpha X^{mother} + \sum_{\kappa=1}^{4} \sum_{\tau=1}^{9} \chi_{\kappa\tau j} Y_{t-\tau}^{inj_\kappa} + \delta_2 X_{it} + \rho_{it}^Y \mu_i + \omega_{it}^Y \nu_{it} + \epsilon_{it}^D \tag{2.12}$$

This is a discrete time hazard model of the age at which ADHD is first diagnosed.

2.4.5 Likelihood function

Following Mroz and Savage (2006), Yang et al. (2009), and Fout and Gilleskie (2014) I use the discrete factor maximum likelihood (DFML) method to control for heterogeneity and endogeneity by integrating out the unobserved factors μ_i and η_{it}. the contribution to the likelihood of the individual i in year t is:

$$\begin{aligned} L_{it}(\Omega|\mu_i, \nu_{it}) = &\Big[\Pr\{D_{it} = 1|\mu_i, \nu_{it}\} \cdot \Pr\{M_{ijt} = j|\mu_i, \nu_{it}\}^{M_{ijt}} \cdot \Pr\{M_{ijt} = 0|\mu_i, \nu_{it}\}^{(1-M_{ijt})}\Big]^{D_{it}} \cdot \\ &[\Pr\{D_{it} = 0|\mu_i, \nu_{it}\}]^{(1-D_{it})} \cdot \\ &\big[\Pr\{Y_{it}^{outcome} = 1|\mu_i, \nu_{it}\}\big]^{Y_{it}^{outcome}} \cdot \big[\Pr\{Y_{it}^{outcome} = 0|\mu_i, \nu_{it}\}\big]^{(1-Y_{it}^{outcome})} \end{aligned} \tag{2.13}$$

where Ω is a vector of parameters to be estimated. I use Fortran programs to obtain maximum likelihood estimates.

For comparison purposes, I first estimate single-equation specifications for every outcome of interest which are the incidence of injuries and teenage pregnancy, as well as for the event of the first ADHD diagnosis and decision on medical treatment. All my dependent variables are specified

as discrete and I use binary logit regressions for adverse outcomes and a three-choice model for the treatment decision. I compare the results from these single-equation regressions to the results obtained using the discrete factor maximum likelihood approach.

2.4.6 Identification

In my model, I simultaneously estimate a system of dynamic equations that requires exogeneity of some of the explanatory variables conditional on the unobserved individual-level heterogeneity for identification. The theoretical argument for identification in these models is outlined in Bhargava (1991). Empirical applications include Mroz and Savage (2006), Yang et al. (2009), and Fout and Gilleskie (2014) among others. I follow these studies in relying on time-varying exogenous variables to identify the model. In this study they are treatment prices to individual and Medicaid, children's age, number of adults and children in the family, and county-level variables. The latter include yearly unemployment rate, average income, population density, number of physicians who accept Medicaid patients, and the share of patients with ADHD that receive ADHD medication. Additionally, the equations' functional form is nonlinear, and this specification reduces the number of exogenous variables needed for identification.

2.5 Data

2.5.1 Medicaid Claims

I use a large panel data set of SC Medicaid claims that spans 10 years from 2003 to 2012. In South Carolina, Medicaid is one of the major health insurance providers with about 20% of the state population being active enrollees[15] and over $5 billion in spending in FY2012[16]. The state administers a separate Children's Health Insurance Program (CHIP) program as a part of its Medicaid program.

Medicaid eligibility is based on income determined on the state level. Families with income below 150% of the Federal Poverty Level (FPL) are eligible for Medicaid and children from families with income below 200% of the FPL are eligible for Medicaid coverage through CHIP.[17] My sample

[15] State Medicaid Fast Facts SFYTD 2014 (July-December). https://www.scdhhs.gov/historic/countyleveldata.html. Accessed on 03.04.2015.

[16] The estimate includes CHIP program expenditures. CMS-64 Quarterly Expense Report, www.medicaid.gov. Accessed on 03.04.2015.

[17] Changed slightly over time. I requested these stats.

represents a large part of children population of the state. Over half of the Medicaid insured are children[18] and 87.5% of eligible children are estimated to be enrolled.[19]

My data set includes Medicaid eligibility information, hospital, outpatient, and pharmacy claims for individuals who were diagnosed with ADHD between 3 and 21 years old in 2003-2012. These data are supplemented by several variables from the enrollees' birth certificates. They are mother's de-identified ID, age, race, and education level. Given that children who were diagnosed with ADHD arguably do not constitute a random sample, I face a potential sample selection problem. In the future work, this data limitation can be addressed by including children who were never diagnosed with the condition.

Eligibility file contains information on the individual monthly eligibility status, Medicaid qualifying category, demographic characteristics (date of birth, gender, and race), living arrangement and family characteristics (number of children and adults in the family and income). Once eligibility for Medicaid is established, the health insurance coverage is available for an enrollee for a 12-month period (unless the enrollee becomes ineligible during this time), after which the eligibility needs to be reconfirmed. An eligible individual who received services prior to the actual enrollment, can be covered retroactively for up to three months before the month when eligibility was established.

Eligibility status file also specifies the type of plan an individual is enrolled in. Medicaid has two components: traditional fee-for-service (FFS) and services provided through managed care organizations (MCO). Due to the differences in reporting requirements, the complete information on all services provided to a patient are only available for those enrolled in the FFS plan. However, mental health is one of the "carved-out" conditions that are covered by the FFS component even if an individual is enrolled into a managed care plan. I use all available claims and when possible, perform robustness checks by excluding MCO enrollees.

Medicaid hospital and outpatient claims files have a similar structure. Each claim includes at least one ICD-9 diagnosis code.[20] ICD-9 codes for ADHD are 314.00 (Attention deficit, without hyperactivity) and 314.01 (Attention deficit, with hyperactivity). Every patient in our sample has at least one claim with ADHD diagnosis. Claims also have details on the provider (location, specialty, and a unique identifier), services provided (timing, CPT procedure code(s)), and the amount that

[18] CMS, Medicaid & CHIP Monthly Application, Eligibility Determination, and Enrollment Reports and Updated Data, July - December, 2014, as of February 23, 2015.

[19] http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-State/south-carolina.html. Accessed on 03.04.2015

[20] A hospital claim may have up to 9 diagnosis codes and an outpatient claim may have up to 3 codes.

was paid by Medicaid. In 2013, most eligible individuals faced a small copay per doctor visit ($3.30), per prescription ($3.40 for adults over 19 years old and zero otherwise) and per hospital stay ($25).

ICD-9 diagnosis codes and CPT procedure codes are also useful in identifying adverse events in the data. In order to find cases of pregnancy and STDs among medical records, I compile a list of ICD-9 diagnosis codes that correspond to each of these events respectively. In addition, I distinguish cases of testing for STDs using CPT procedure codes. The ICD-9 codes for injuries were borrowed from Marcus et al. (2008) and we followed the methodology described in Bouchery et al. (2012) to find cases of substance abuse among patient claims. For the outcomes of risky sexual behavior and substance abuse disorders I focus on the first occurrence of the adverse event, or the earliest date when a respective diagnosis code appears in the data. Claims data are not as detailed as medical history records making it hard to distinguish two different instances of an STD or even pregnancy.

In addition to adverse events, hospital and outpatient claims allow me to identify instances of behavioral therapy treatment. Patients with ADHD may benefit from pharmacological treatment and/or from behavioral therapy. Therapy usually consists in assisting children in managing their condition and in educating parents and teachers on how to provide positive feedback on desired behaviors and how to discourage unwanted behaviors. Behavioral therapy alone was found to be less effective than pharmacological treatment alone, but no consensus exists on whether medications are inferior to the combination treatment (Barkley, 2006). A combination of behavioral therapy and pharmacological treatment constitutes yet another choice in the set.

Pharmacy claims contain records of all prescriptions filled by a patient. Each records has a dispense date, National Drug code (NDC)[21], quantity purchased, pharmacy ID, dispense fee and the amount paid by Medicaid. Note that pharmacy claims do not have diagnosis record. I use medical literature to identify drugs that were approved by the FDA to be prescribed for patients with ADHD. Table 3.1 lists these medications with their respective in-sample market shares calculated for the entire period between 2003 and 2012. The last category, "Others" includes medicines that had an in-sample market share lower than 5%. The market is dominated by the extended-release formulations of relatively old drugs: together amphetamine salts and methylphenidate comprise almost a half of the market for ADHD pharmacological treatments. Stimulants are often recommended as the first step in treatment. In this paper, I combine all pharmacological treatment into a single choice set

[21] NDC is an 11-digit classification issued by the Food and Drug Administration (FDA) for all the approved drugs. Under this system, different package and dosage sizes of the same drug molecule have separate NDCs.

option leaving it to the future research to refine the selection. I also ignore off-label medications that were not approved for the treatment of ADHD.[22]

In SC, children and young adults (under age 19) face zero copayment for the prescription drugs and are only responsible for the pharmacy dispense charge (about $5). The state maintains preferred drug lists for medicines that include the most commonly used ADHD medications. These drugs do not require prior authorization; all other drugs can be prescribed and covered by Medicaid if a doctor-filed authorization request is approved. The quantity restrictions are also common with a typical prescription capped at a 30-day supply.

My original sample contains 131,008 Medicaid enrollees who had at least one ADHD claim between 2003 and 2012. For a cohort born in 1991, my sample represents 23% of all Medicaid enrollees. We estimate the model on a sub-sample of all identified children with ADHD.

First, I exclude individuals with missing family and demographic records leaving me with a sample of 118,655 patients. We also exclude 129 children (0.1%) who died between 2003 and 2012. Second, we only select Medicaid enrollees who are continuously eligible for the program because there is no information on individual outcomes when they are ineligible. In the data, about 30% of the individuals have lapses in eligibility that are on average 9 months long. For lapses in eligibility that last under three months, we assume that patients received no medical treatment. For inconsistent eligibility periods that have longer lapses in coverage, I only keep the medical history to the point prior to the lapse. Furthermore, I exclude 6,836 (5.2%) patients who had no single eligibility spell that lasted 365 days or the spell they were diagnosed in or took a prescription lasted less than that between 2003 and 2012. If a patient is observed for less than a year, I can not observe a potential effect of treatment that is realized in the next time period (year).

Third, my identification strategy hinges on the outcome of the first doctor visit, so I only keep patients for whom I observe the event of the first diagnosis. Based on earlier literature, I exclude patients who fill a prescription prior to their first doctor visit ADHD claim (30,700, or 23.4%) and patients who had their ADHD visit within the first 180 days from their first date in the sample. In addition, only patients who have at least one full year of eligibility after their first ADHD diagnosis and who were diagnosed between 3 and 18 years old are selected. This left-censoring

[22] Although not approved for the treatment of ADHD, certain antidepressants and sleep-disorder medications are prescribed to patients off-label. For example, Provigil (sleep disorders); Wellbutrin (antidepressant); tricyclic antidepressants; Catapres and Tenex (short-acting forms of high blood pressure medicines); Abilify, Zyprexa, Seroquel, Risperdal, and Geodon (antipsychotics).

problem eliminates a significant part of our sample.

Furthermore, I drop patients for whom I was unable to calculate provider propensity to prescribe. My final sample has 64,031 individuals. Table 2.2 shows summary statistics on individual, mother's and home environment characteristics. Boys comprise over 66% of the sample of children diagnosed with ADHD, whites and blacks are represented nearly equally. On average, children are first diagnosed with ADHD at 8 years old.

Data on mother's characteristics are pulled from in-state birth certificates. Only 40% of the original sample IDs of children with ADHD were matched to their mother's ID. Most mothers in the sample have at least some high-school education or a high school diploma. About 7% of children in the sample were in foster care at least for some time between 2003 and 2012. The families predominantly consists of a single adult and two children. Their reported net monthly income is about $620 on average.

Table 2.3 reports summary statistics on medical treatment of ADHD and adverse outcomes that I observe in the sample. The majority of all ADHD-diagnosed children have hyperactive symptoms rather than inattentiveness. Nearly 65% of the patients filled at least one prescription, 13% had a behavioral therapy session, and 27% had a combination of the two after they were diagnosed. To account for potential data issues, I define pharmacological treatment as a prescription filled within a year from the first diagnosis.

On average, I observe every Medicaid enrollee for seven years. During this time, 1,244 of them become pregnant before age 19; 4,301 contract an STD and/or are tested for an STD condition. I also observe 4,602 teens having at least one claim that indicates one of the substance abuse disorders. The most frequent outcome that I observe yearly are injuries. About 74% of children and teens had at least one injury while in sample. In order to take into account injury severity, I calculate the total Medicaid spending on injury-related injuries. These expenditures vary widely; on average SC Medicaid spent $4,291 per patient during their period of eligibility in 2003-2012.

2.6 Discussion of results

This section presents the results obtained using DFML estimator on 5 points of support for the permanent and 2 for the time-varying unobserved heterogeneity. For comparison, I estimate single-equation logit models for every outcome: the event of the first diagnosis, decision on medical

treatment, probability of teen pregnancy, STDs, substance abuse, and injuries. There are three treatment modalities: medications only, behavioral therapy only, and a combination of the two. Time periods are years. It is left for future research to refine both the choice set and time periods.

2.6.1 Initial diagnosis

For the event of the initial diagnosis, the effects estimated by the model with and without unobserved heterogeneity point in the same direction, but the effect is larger in the model with unobserved heterogeneity. The effect of injuries on the probability of being diagnosed fades with time. Given the severity of ADHD, the effect of injuries on the probability of the initial diagnosis with ADHD is the highest in the time period $t-1$ and smallest in $t-3$. The estimated coefficients are presented in Table 2.11 and marginal effects for single-equation logit are shown in Table 2.5.

Boys are more likely to be diagnosed and the probability increases with age since everybody in the sample has the condition. Similarly, children who show hyperactivity-related symptoms rather than inattentiveness are more likely to receive the ADHD diagnosis. Consistent with the work by Doyle (2013), I find that patients who were eligible for Medicaid as foster children are more likely to receive the diagnosis since they have to go through a checkup procedure. Family income is negatively related to the probability of the diagnosis and is supportive of the argument that relatively better off parents resort to other than pharmacological treatment options as their first choice.

2.6.2 Choice of treatment

Treatment choice estimates mostly coincide in signs across DFML and simple multinomial logit models (see Tables 2.6). Medical choice in this specification is binary. Treatment decisions are persistent over time. In particular, if a patient was on particular type of treatment in previous period, she is extremely likely to continue with the same choice of treatment, although the impact fades with time.

Having a history of injuries positively affects the probability of treatment. This suggests that ADHD children are in fact more prone to injuries and treatment is seen as a way to reduce the probability of these adverse events. Children diagnosed with a hyperactive type of ADHD are more likely to receive treatment.

2.6.3 Adverse outcomes

2.6.3.1 Injuries

The model with unobserved heterogeneity shows that individuals who were on either pharmacological, behavioral, or combination treatment were less likely to experience injuries in the following year. The effect of treatment from further back in time is less precise. Only combination treatment seems to have effects that hold over time and from which patients could benefit by following the regimen consistently, year after year. These results differ from the single-equation logit model that is affected by the endogeneity problem. The results are presented in Tables 2.8 and 2.14.

Consistent with the medical literature, behavioral therapy is the least efficient treatment and its combination with pharmacological treatment has the largest effect. Notably, behavioral therapy does not create lasting effects. If the patient was on treatment two years ago, he or she is no less likely to experience injuries in the current year than the patient who did not go through therapy sessions.

Among other results, boys are more injury-prone than girls as are relatively hyperactive kids when compared to inattentive types.

2.6.3.2 Teen pregnancies & STDs

The estimates from the model with unobserved heterogeneity also show that ADHD treatment is successful in reducing risky sexual behavior activity, measured by the probability of teen pregnancy (Tables 2.7 and 2.12) and STDs (Tables 2.9 and 2.13). As with injuries, combination of behavioral therapy and medicines has the highest impact on reducing the probability of adverse events. However, for both outcomes, STDs and pregnancy, behavioral therapy is preferred to drugs alone. Moreover, for STDs there seem to be a relatively longer lasting effect of behavioral therapy, suggesting that these sessions might be effective in teaching children and their parents on how to manage their condition effectively.

Teens with hyperactive type of ADHD are more likely to experience teen pregnancy and STDs than predominantly inattentive types. Higher family income and being in foster care are negatively related to the probability of pregnancy.

2.6.3.3 Substance Abuse

Finally, I look at another potential outcome of risky behavior – substance abuse. In this case, the most effective treatment is behavioral therapy that has not only immediate effect from $t-1$ period, but also a smaller long-run effect from earlier treatment. Medicines have no beneficial impact on the probability of substance abuse but the combination of behavioral therapy and drugs has positive effect (Tables 2.10 and 2.15). This result may follow from the fact that some of the ADHD treatments may cause addiction and individuals who attend behavioral therapy sessions are likelier to seek treatment for their substance abuse disorders as well.

2.7 Conclusions

Attention-deficit/hyperactivity disorder (ADHD) is a common mental chronic condition that negatively affects noncognitive abilities. Over 5 million children aged 2–17 (7.9%) have been diagnosed with ADHD in the U.S. and the majority of them are taking medications for this condition.

This paper sheds light on the effectiveness of ADHD treatment in reducing the probability of adverse health and behavioral outcomes. I use a 10-year panel of SC Medicaid claims to model the probability of the initial diagnosis of ADHD, dynamic treatment choice decisions and subsequent adverse events later in life. Controlling for endogeneity, I find that there is a strong persistence in treatment choices across time periods. The results also suggest that pharmacological treatment has only short-term positive impact on the probability of such adverse events as injuries, teenage pregnancy, and STDs, and no impact on substance abuse disorders. Behavioral therapy alone is not as effective as it is in combination with ADHD drugs, but for STDs and substance abuse disorders it seems to show relatively long-lasting effects in contrast to drugs alone.

In general, these results are consistent with medical literature and the theory of human capital accumulation. It is left for the future research to simulate the cost of treatment to Medicaid under various treatment sequences, including the cost of poor adherence and late diagnosis.

Table 2.1: Summary of Empirical Model Specification

Outcome	**Estimator**	**Explanatory Variables**		
		Endogenous	Exogenous	Unobserved Heterogeneity
First diagnosis, D_{it}	logit	Y_{it-1}	X_i^{mother}, X_{it}^{child}, $X_{it}^{h.env.}$, $X_{it}^{provider}$	μ_i^D, ν_{it}^D, ϵ_{it}^D
Treatment choice, M_t	mlogit	Y_{it-1}, $(T_{it-\tau}^D, ..., T_{it-1}^D)$	X_i^{mother}, X_{it}^{child}, $X_{it}^{h.env.}$, X_{it}^{prov}, Z_{jt}	μ_i^M, ν_{it}^M, ϵ_{it}^M
Adverse outcomes, Y_t:				
Teen Pregnancy, Y^{pregn}	logit	$(M_{it-\tau}, ..., M_{it-1})$	X_i^{mother}, X_{it}^{child}, $X_{it}^{h.env.}$	μ_i^Y, ν_{it}^Y, ϵ_{it}^Y
STD, Y^{std}	logit			
Subst. abuse, Y^{sa}	logit			
Injuries, Y^{injury}	logit			
Stock of skills at birth, θ_{i0}	latent	–	X_i^{mother}	μ_i^I

Notes: The table decomposes the main components of the equations comprising the empirical model. The notation is defined in Section 2.3.2

Table 2.2: Summary statistics: Individual and family characteristics.

	N obs	**Mean**	**Median**	**Std**	**Min**	**Max**
Individual characteristics						
Age 1st in sample	64,031	5.56	5.00	4.23	0	18
Age at 1st ADHD diagnosis	64,031	8.54	8.00	3.50	3	19
Male	64,031	0.66			0	1
Race: White	64,031	0.53			0	1
Black	64,031	0.44			0	1
Family & home environment						
Family net monthly income	64,031	619.95	438.28	641.60	0	6,352
N adults	64,031	1.07	1.00	0.62	0	6.00
N children	64,031	2.09	2.00	1.02	0	10.00
Ever in foster care	64,031	0.07			0	1
Mother's characteristics						
Age when gave birth	39,576	23.05	22.00	5.30	11	46
Educ: Less than HS	39,576	0.05			0	1
Some HS	39,576	0.38			0	1
HS diploma	39,576	0.40			0	1
Some college	39,576	0.13			0	1
College diploma/grad school	39,576	0.04			0	1

Notes: The sample includes every SC Medicaid enrollee, who was eligible for at least one year after their first diagnosis and who was diagnosed with ADHD for the first time at any age between 3 and 19 years old in 2003-2012.

Table 2.3: Summary statistics: Medical Treatment and Adverse Outcomes.

	N obs	Mean	Median	Std	Min	Max
Medical diagnosis & treatment						
1st diagnosis: hyperactive type	64,031	0.73			0	1
inattentive type	64,031	0.25			0	1
mixed type	64,031	0.02			0	1
Ever filled 1+ Rx	64,031	0.65			0	1
Behavioral therapy	64,031	0.13			0	1
Combo treatment	64,031	0.27			0	1
Years of data	64,031	6.93	7.00	2.68	1	10
Outcome: Risky sexual behavior						
1. Teen Pregnancy						
Age at 1st pregnancy	1,244	16.71	17.00	1.70	11	19
Race: White	1,244	0.53			0	1
Black	1,244	0.46			0	1
2. STD						
Age at 1st STD (incl. testing)	4,301	15.02	15.00	2.34	11	19
Male	4,301	0.40			0	1
Race: White	4,301	0.52			0	1
Black	4,301	0.46			0	1
Outcome: Substance Abuse						
Age at 1st substance abuse	4,602	15.25	15.00	2.06	11	19
Male	4,602	0.64			0	1
Race: White	4,602	0.54			0	1
Black	4,602	0.44			0	1
Outcome: Injuries						
Ever injured	64,031	0.74			0	1
Age at injury	47,293	8.71	8.00	3.93	0	21
Male	47,293	0.67			0	1
Race: White	47,293	0.54			0	1
Black	47,293	0.43			0	1

Notes: The sample includes every SC Medicaid enrollee, who was eligible for at least one year after their first diagnosis and who was diagnosed with ADHD for the first time at any age between 3 and 19 years old in 2003-2012.

Table 2.4: Choice Set in the U.S. ADHD Drugs Market, 2003-2012.

Active Ingredient	**Speed**	**Mkt share**	**Major Brands**	**G**	**Entry**	**Avg. Price**
Amphetamine salts	E	25.16	Adderall XR	Y	2001	150.67
Methylphenidate	E	20.26	Concerta	Y	2000	131.00
Methylphenidate	N	11.13	Ritalin LA, Metadate CD, Methylin ER	Y	2002	127.35
Lisdexamfetamine	E	11.04	Vyvanse	N	2007	141.11
Amphetamine salts	M	8.15	Adderall	Y	1996	37.27
Dexmethylphenidate	E	7.19	Focalin XR	N	2005	144.85
Atomoxetine	n/a	6.37	Strattera	N	2002	130.27
Methylphenidate	M	5.82	Methylin, Ritalin	Y	2002	30.16
Others	–	4.89	Various	–		81.15

Notes: "Speed" stands for the drug release speed, where "E" means extended release, "N" - intermediate and "M" – immediate release speed. Extended release drugs are superior than immediate release drugs in that their active ingredient is released over a longer period of time, often allowing for once-a-day dosing. In-sample market share is based on the number of prescriptions filled in 2003-2012. "G" stands for generic drugs availability. Average price is calculated by averaging SC Medicaid reimbursement payments to pharmacies.

Table 2.5: First ADHD Diagnosis. Single-equation logit estimation.

Selected variables	**(A)**			**(B)**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of injuries	0.017^a	0.015^a	0.003	0.019^a	0.012^b	0.002
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)
Individual characteristics						
Male	0.041^a	(0.002)	0.038^a	(0.003)		
Race (white omitted):						
Black	-0.036^a	(0.002)			-0.040^a	(0.010)
Other	-0.048^a	(0.006)			-0.041^a	(0.010)
ADHD symptoms (inattentive omitted)						
hyperactive	0.037^a	(0.002)			0.035^a	(0.002)
mixed	0.135^a	(0.090)			-0.146^a	(0.011)
Age	0.100^a	(0.001)			0.149^a	(0.002)
Age sq.	-0.004^b	(0.001)			-0.007^a	(0.0001)
Mother's characteristics						
Age when gave birth					-0.001^b	(0.0003)
Education (high school degree omitted):						
Less than HS					-0.010	(0.006)
Some HS					-0.0005	(0.003)
Some college					0.0005	(0.004)
College degree or higher					0.003	(0.006)
Family & home environment						
Family net income	-0.003^b	(0.001)			-0.004^b	(0.002)
Foster care	0.059^a	(0.005)			0.081^a	(0.007)
Number of adults	-0.001	(0.002)			-0.008^a	(0.002)
Number of children	-0.016^a	(0.001)			-0.015^a	(0.001)
Time trend (4th degree polynomial)			YES			YES
County time-varying characteristics			YES			YES
LF		-109,491			-67,904	
N person/year obs		**215,428**			**141,841**	

Notes: The coefficients are marginal effects at means and standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively. Panel (A) excludes mother's characteristics and uses the entire sample of 64,031 individuals in the original sample. Panel (B) includes mother's characteristics as controls. Many children were not matched to their moms because of the Medicaid data constraints in usage of birth certificate data and the fact not all mothers are on Medicaid when their children are eligible.

Table 2.6: Summary results on the probabilities of treatment choices. Single-equation mlogit.

Selected variables	**Time period**					
	t-1		**t-2**		**t-3**	
Medicines only						
History of treatment						
Medicines	0.740^a	(0.003)	0.531^a	(0.004)	0.501^a	(0.005)
Btherapy	0.112^a	(0.007)	0.390^a	(0.012)	0.439^a	(0.013)
Combo	0.272^a	(0.005)	0.485^a	(0.006)	0.467^a	(0.007)
No treatment	0.266^a	(0.002)	0.428^a	(0.003)	0.449^a	(0.002)
Behavioral therapy						
History of treatment						
Medicines	0.006^a	(0.001)	0.028^a	(0.002)	0.035^a	(0.003)
Btherapy	0.264^a	(0.009)	0.053^a	(0.004)	0.053^a	(0.005)
Combo	0.051^a	(0.003)	0.031^a	(0.002)	0.041^a	(0.004)
No treatment	0.031^a	(0.001)	0.038^a	(0.001)	0.036^a	(0.001)
Combination						
History of treatment						
Medicines	0.098^a	(0.002)	0.163^a	(0.003)	0.165^a	(0.004)
Btherapy	0.226^a	(0.008)	0.173^a	(0.007)	0.175^a	(0.009)
Combo	0.576^a	(0.005)	0.197^a	(0.004)	0.196^a	(0.006)
No treatment	0.071^a	(0.001)	0.157^a	(0.002)	0.159^a	(0.002)

Notes: The coefficients are marginal effects at means and standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively. This sample includes mother's characteristics as controls.

Table 2.7: Adverse Outcome: Teen Pregnancy. Single-equation logit estimates.

Selected variables	**(A)**			**(B)**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.002[a]	0.0003	-0.0001	-0.001	0.001	-0.002[b]
	(0.001)	(0.0009)	(0.0009)	(0.001)	(0.001)	(0.001)
Behavioral therapy only	-0.0001	0.002	-0.0004	-0.0002	0.004[b]	-0.001
	(0.001)	(0.0013)	(0.0012)	(0.0011)	(0.002)	(0.001)
Combination	-0.001	0.001	0.0002	0.0001	0.001	-0.0001
	(0.001)	(0.0013)	(0.0012)	(0.0011)	(0.001)	(0.001)
Individual characteristics						
Age	0.019[a]	(0.001)			0.011[a]	(0.002)
Age squared	-0.0004[a]	(0.00004)			-0.0003[a]	(0.0001)
Race (white omitted):						
Black	-0.001[c]	(0.001)			-0.001	(0.0005)
Other	-0.008[a]	(0.003)			–	–
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive	0.002[a]	(0.001)			0.001[c]	(0.0005)
mixed	-0.001	(0.002)			-0.003[a]	(0.0008)
Mother's characteristics						
Age when gave birth					-0.0001	(0.00005)
Education (HS diploma omitted):						
Less than HS					0.002[c]	(0.001)
Some HS					0.0004	(0.001)
Some college					-0.002[c]	(0.001)
College or higher					-0.001	(0.002)
Family & home environment						
Family net income	-0.002	(0.0004)			-0.001[a]	(0.0004)
Foster care	-0.002	(0.0008)			-0.002[b]	(0.001)
Number of adults	-0.001	(0.0003)			0.0001	(0.0003)
Number of children	0.0002	(0.0002)			-0.0002	(0.0002)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characterics		YES			YES	
LF		-4,014			-1,174	
N person/year obs		**39,125**			**18,631**	

Notes: The coefficients are marginal effects at means and standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with *a*, *b*, and *c* respectively. Panel (A) excludes mother's characteristics and uses the entire sample of 21,770 girls in the original sample. Panel (B) includes mother's characteristics as controls. Many children were not matched to their moms because of the Medicaid data constraints in usage of birth certificate data and the fact not all mothers are on Medicaid when their children are eligible.

Table 2.8: Adverse Outcome: Injuries. Single-equation logit estimates.

Selected variables	**(A)**			**(B)**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.0005	0.005	-0.001	-0.003	0.009^{c}	0.001
	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Behavioral therapy only	0.017^{a}	0.016	0.011	0.014^{b}	0.014	0.009
	(0.005)	(0.006)	(0.007)	(0.007)	(0.009)	(0.010)
Combination	0.017^{a}	0.006	0.004	0.014	0.009	0.007
	(0.004)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)
Individual characteristics						
Male	0.024^{a}	(0.002)			0.023^{a}	(0.003)
Age	0.014^{a}	(0.001)			0.005^{b}	(0.002)
Age squared	-0.001^{a}	(0.0001)			-0.0003^{a}	(0.0001)
Race (white omitted):						
Black	-0.071^{a}	(0.002)			-0.076^{a}	(0.003)
Other	-0.052^{a}	(0.006)			-0.054^{a}	(0.010)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive	0.012^{a}	(0.002)			0.011^{a}	(0.003)
mixed	0.001	(0.006)			-0.0004	(0.008)
Mother's characteristics						
Age when gave birth					-0.0005^{b}	(0.0002)
Education (HS dimploma omitted):						
Less than HS					0.018^{a}	(0.005)
Some HS					0.009^{a}	(0.003)
Some college					-0.011^{a}	(0.004)
College or higher					-0.018^{a}	(0.006)
Family & home environment						
Family net income	-0.008^{a}	(0.001)			-0.006^{a}	(0.002)
Foster care	0.057^{a}	(0.004)			0.056^{a}	(0.006)
Number of adults	-0.001	(0.001)			0.001	(0.002)
Number of children	-0.0004	(0.001)			-0.005^{a}	(0.001)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characteristics		YES			YES	
LF		-148,213			-91,638	
N person/year obs		**252,421**			**153,282**	

Notes: The coefficients are marginal effects at means and standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively. Panel (A) excludes mother's characteristics and uses the entire sample of 64,031 individuals in the original sample. Panel (B) includes mother's characteristics as controls. Many children were not matched to their moms because of the Medicaid data constraints in usage of birth certificate data and the fact not all mothers are on Medicaid when their children are eligible.

Table 2.9: Adverse Outcome: STDs. Single-equation logit estimates.

Selected variables	**(A)**			**(B)**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.001	-0.001	0.002	-0.002	0.002	0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Behavioral therapy only	0.003	0.0004	0.006^{c}	-0.005^{c}	0.0004	0.007
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
Combination	0.001	0.002	0.003	-0.002	0.007	-0.001
	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)
Individual characteristics						
Male	-0.028^{a}	(0.001)			-0.020^{a}	(0.001)
Age	0.017^{a}	(0.003)			0.009^{b}	(0.004)
Age squared	-0.0003^{a}	(0.0001)			-0.0001	(0.0001)
Race (white omitted):						
Black	-0.003	(0.001)			-0.006	(0.001)
Other	0.001	(0.003)			-0.003	(0.007)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive	0.001	(0.001)			-0.002	(0.001)
mixed	-0.001	(0.003)			-0.009	(0.004)
Mother's characteristics						
Age when gave birth					-0.00004	(0.0001)
Education (HS diploma omitted):						
Less than HS					0.002	(0.003)
Some HS					0.000	(0.001)
Some college					0.001	(0.002)
College or higher					0.002	(0.003)
Family & home environment						
Family net income	-0.001	(0.001)			0.0005	(0.001)
Foster care	0.009^{a}	(0.002)			0.003	(0.003)
Number of adults	-0.001	(0.001)			-0.001	(0.001)
Number of children	-0.0002	(0.0004)			-0.002	(0.001)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characteristics		YES			YES	
LF		-14,573			-6,175	
N person/year obs		**101,145**			**48,604**	

Notes: The coefficients are marginal effects at means and standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with *a*, *b*, and *c* respectively. Panel (A) excludes mother's characteristics and uses the entire sample of 64,031 individuals in the original sample. Panel (B) includes mother's characteristics as controls. Many children were not matched to their moms because of the Medicaid data constraints in usage of birth certificate data and the fact not all mothers are on Medicaid when their children are eligible.

Table 2.10: Adverse Outcome: Substance abuse. Single-equation logit estimates.

Selected variables	**(A)**			**(B)**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.0003	-0.001	0.002	0.0004	0.001	0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Behavioral therapy only	0.007^a	0.004^c	0.007^b	0.005^c	0.002	0.004
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Combination	0.015^a	-0.002	0.005^b	0.011^a	-0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Individual characteristics						
Male	0.004^a	(0.001)			0.005^a	(0.001)
Age	0.045^a	(0.002)			0.030^a	(0.003)
Age squared	-0.001^a	(0.0001)			-0.001^a	(0.0001)
Race (white omitted):						
Black	-0.008^a	(0.001)			-0.008^a	(0.001)
Other	-0.010^a	(0.003)			-0.009	(0.006)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive	0.005^a	(0.001)			0.005^a	(0.001)
mixed	0.007^b	(0.003)			0.001	(0.003)
Mother's characteristics						
Age when gave birth					0.00004	(0.0001)
Education (HS diploma omitted):						
Less than HS					0.007	(0.002)
Some HS					0.002	(0.001)
Some college					-0.004	(0.002)
College or higher					-0.007	(0.003)
Family & home environment						
Family net income	-0.003^a	(0.001)			-0.001	(0.001)
Foster care	0.008^a	(0.001)			0.007	(0.002)
Number of adults	-0.002^a	(0.001)			-0.002	(0.001)
Number of children	-0.001	(0.0004)			-0.001	(0.001)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characteristics		YES			YES	
LF		102,805			-5,841	
N person/year obs		**-14,925**			**50,234**	

Notes: The coefficients are marginal effects at means and standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively. Panel (A) excludes mother's characteristics and uses the entire sample of 64,031 individuals in the original sample. Panel (B) includes mother's characteristics as controls. Many children were not matched to their moms because of the Medicaid data constraints in usage of birth certificate data and the fact not all mothers are on Medicaid when their children are eligible.

Table 2.11: First Diagnosis: DFML estimates.

Selected variables	**LOGIT**			**DFML**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of injuries	0.089^a	0.077^a	0.013	0.213^a	0.175^a	0.109^a
	(0.013)	(0.016)	(0.020)	(0.020)	(0.021)	(0.023)
Individual characteristics						
Male		0.206^a	(0.011)		0.351^a	(0.017)
Age		0.505^a	(0.007)		1.005^a	(0.013)
Age squared		-0.019^a	(0.0003)		-0.038^a	(0.001)
Race (white omitted):						
Black		-0.183^a	(0.011)		-0.400^a	(0.017)
Other		-0.243^a	(0.031)		-0.426^a	(0.025)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive		0.195^a	(0.012)		0.393^a	(0.018)
mixed		0.642^a	(0.039)		1.397^a	(0.026)
Family & home environment						
Family net income		-0.017^a	(0.007)		-0.042^a	(0.013)
Foster care		0.298^a	(0.028)		0.703^a	(0.025)
Number of adults		-0.006	(0.009)		-0.064^a	(0.014)
Number of children		-0.081^a	(0.005)		-0.142^a	(0.009)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characteristics		YES			YES	
N person/year obs		**215,428**			**215,428**	

Notes: The table shows coefficient estimates from single-logit equation and DFML model; both exclude mother's characteristics. Standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively.

Table 2.12: Adverse outcomes: Teenage pregnancy. DFML estimates.

Selected variables	**LOGIT**			**DFML**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.317^a	0.047	-0.025	-0.207^a	0.096^a	-0.026
	(0.103)	(0.123)	(0.125)	(0.026)	(0.026)	(0.026)
Behavioral therapy only	-0.023	0.236	0.051	-0.314^a	0.116	-0.226^a
	(0.131)	(0.145)	(0.163)	(0.027)	(0.027)	(0.027)
Combination	-0.143	0.083	0.033	-0.405^a	-0.019	-0.176^a
	(0.14)	(0.167)	(0.166)	(0.027)	(0.027)	(0.027)
Individual characteristics						
Age		2.617^a	(0.294)		1.895^a	(0.023)
Age squared		-0.063^a	(0.009)		-0.039^a	(0.001)
Race (white omitted):						
Black		-0.116^c	(0.068)		-0.067^a	(0.025)
Other		-1.164^a	(0.388)		-1.283^a	(0.027)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive		0.267^a	(0.070)		0.307^a	(0.026)
mixed		-0.086	(0.294)		-0.103^a	(0.027)
Family & home environment						
Family net income		-0.247^a	(0.054)		-0.249^a	(0.024)
Foster care		-0.304^a	(0.111)		-0.340^a	(0.027)
Number of adults		-0.110^b	(0.044)		-0.160^a	(0.023)
Number of children		0.034	(0.034)		0.051^a	(0.021)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characterics		YES			YES	
N person/year obs		**39,125**			**39,125**	

Notes: The table shows coefficient estimates from single-logit equation and DFML model; both exclude mother's characteristics. Standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively.

Table 2.13: Adverse outcomes: STDs. DFML estimates.

Selected variables	**LOGIT**			**DFML**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.040	-0.032	0.068	-0.003	-0.013	0.059^a
	(0.055)	(0.068)	(0.067)	(0.023)	(0.024)	(0.024)
Behavioral therapy only	0.111	0.014	0.205	-0.082^a	-0.098^a	0.025
	(0.08)	(0.094)	(0.099)	(0.026)	(0.026)	(0.026)
Combination	0.026	0.064	0.103	-0.157^a	-0.012	-0.043
	(0.074)	(0.089)	(0.087)	(0.025)	(0.025)	(0.025)
Individual characteristics						
Male		-1.068	(0.035)		-1.191^a	(0.022)
Age		0.656	(0.099)		0.509^a	(0.019)
Age squared		-0.013	(0.003)		-0.008^a	(0.001)
Race (white omitted):						
Black		-0.133	(0.037)		-0.171^a	(0.022)
Other		0.022	(0.121)		-0.043	(0.027)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive		0.051	(0.038)		0.058^a	(0.022)
mixed		-0.056	(0.132)		-0.025	(0.027)
Family & home environment						
Family net income		-0.036	(0.024)		-0.033	(0.018)
Foster care		0.354	(0.063)		0.396^a	(0.025)
Number of adults		-0.038	(0.024)		-0.064^a	(0.018)
Number of children		-0.009	(0.017)		-0.009	(0.015)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characterics		YES			YES	
N person/year obs		**101,145**			**101,145**	

Notes: The table shows coefficient estimates from single-logit equation and DFML model; both exclude mother's characteristics. Standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively.

Table 2.14: Adverse outcomes: Injuries. DFML estimates.

Selected variables	**LOGIT**			**DFML**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.002	0.025	-0.005	-0.068^a	0.003	-0.028
	(0.015)	(0.019)	(0.021)	(0.013)	(0.014)	(0.015)
Behavioral therapy only	0.086^a	0.079^b	0.054	-0.047^a	0.006	-0.016
	(0.026)	(0.031)	(0.035)	(0.019)	(0.021)	(0.022)
Combination	0.084^a	0.030	0.022	-0.071^a	-0.035^a	-0.069^a
	(0.019)	(0.025)	(0.026)	(0.016)	(0.017)	(0.018)
Individual characteristics						
Male		0.118^a	(0.01)		0.134^a	(0.009)
Age		0.068^a	(0.007)		0.146^a	(0.005)
Age squared		-0.004^a	(0.0003)		-0.007^a	(0.0002)
Race (white omitted):						
Black		-0.354^a	(0.010)		-0.376^a	(0.009)
Other		-0.258^a	(0.029)		-0.318^a	(0.020)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive		0.063^a	(0.011)		0.094^a	(0.010)
mixed		0.008^a	(0.03)		0.118^a	(0.021)
Family & home environment						
Family net income		-0.040^a	(0.006)		-0.038^a	(0.006)
Foster care		0.286^a	(0.021)		0.416^a	(0.016)
Number of adults		-0.004	(0.007)		-0.001	(0.006)
Number of children		-0.0002	(0.005)		-0.013^a	(0.004)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characterics		YES			YES	
N person/year obs		**252,421**			**252,421**	

Notes: The table shows coefficient estimates from single-logit equation and DFML model; both exclude mother's characteristics. Standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively.

Table 2.15: Adverse outcomes: Substance Abuse. DFML estimates.

Selected variables	**LOGIT**			**DFML**		
	t-1	**t-2**	**t-3**	**t-1**	**t-2**	**t-3**
History of treatment						
Medicines only	-0.012	-0.050	0.095	0.004	-0.033	0.098^a
	(0.057)	(0.070)	(0.070)	(0.024)	(0.024)	(0.024)
Behavioral therapy only	0.295	0.171	0.283	-0.090^a	-0.052^b	-0.043
	(0.079)	(0.093)	(0.102)	(0.026)	(0.026)	(0.026)
Combination	0.541	-0.077	0.202	0.196^a	-0.253^a	-0.037
	(0.069)	(0.09)	(0.088)	(0.025)	(0.025)	(0.025)
Individual characteristics						
Male		0.152	(0.035)		0.183^a	(0.023)
Age		1.942	(0.109)		1.788^a	(0.019)
Age squared		-0.053	(0.004)		-0.046^a	(0.001)
Race (white omitted):						
Black		-0.360	(0.036)		-0.412^a	(0.022)
Other		-0.429	(0.136)		-0.554^a	(0.027)
ADHD symptoms at first diagnosis (inattentive omitted)						
hyperactive		0.240	(0.038)		0.317^a	(0.023)
mixed		0.314	(0.116)		0.407^a	(0.027)
Family & home environment						
Family net income		-0.108	(0.025)		-0.136^a	(0.019)
Foster care		0.322	(0.062)		0.465^a	(0.025)
Number of adults		-0.103	(0.024)		-0.163^a	(0.018)
Number of children		-0.022	(0.017)		-0.023	(0.015)
Time trend (4th degree polynomial)		YES			YES	
County time-varying characterics		YES			YES	
N person/year obs		**102,805**			**102,805**	

Notes: The table shows coefficient estimates from single-logit equation and DFML model; both exclude mother's characteristics. Standard errors are in parentheses. Coefficient estimates that are significant at 1%, 5%, and 10% level are denoted with a, b, and c respectively.

Chapter 3

Dynamic Sequencing of Drug Treatments for ADHD Patients with Medicaid Coverage

3.1 Introduction

The most recent 2011/12 National Survey of Children's Health reported that about 5 million children were reported as having an ADHD diagnosis and almost 70% of them were on medical treatment in the U.S. Although there is a strong belief that ADHD medication is overprescribed, very little is known about the existing prescribing practices and physician learning process in ADHD treatment. Based on the Medical Expenditure Panel Survey (MEPS) 1996-2010 data, ADHD is one of the top-25 conditions by the number of prescriptions filled. The evidence suggests that children and teenagers diagnosed with ADHD face significant uncertainty regarding efficacy and severity of adverse effects of ADHD medications. The typical patient is prescribed between one and two different drugs before they find a suitable treatment. The switch from the initial choice occurs approximately within half a year with over 33% of patients switching after the first prescription.

Using South Carolina Medicaid claims data for 2003-2012, I estimate a dynamic model of demand for ADHD drugs with learning under uncertainty. Uncertainty comes from two sources: little evidence on newly introduced ADHD treatments and uncertainty about the response to treatment of a particular patient. In the model, heterogeneous patients learn about the efficacy of available treatments through experimenting. Their preferences are embedded into the preferences of their physician (decision-maker).

This paper is an extension of Crawford and Shum (2005). Their analysis of demand for anti-ulcer drugs suggests that treatments' rankings differ in their curative and symptomatic effects and

although there is substantial heterogeneity in these effects across patients, learning enables them to reduce the cost of uncertainty. In a more recent paper, Dickstein (2014a) analyses depression drugs and suggests that insurer copayment policies and drug promotions for the most efficient treatments can improve patient outcomes while minimizing insurer cost.

In this paper, I explore a dimension that has not been addressed in the literature – drug holidays. I will evaluate the effect of interruption in treatment on its cost and duration, accounting for patient heterogeneity in response to treatment for ADHD. I will explore the potential to develop better guidelines that can improve the quality of drug-patient matches and patient outcomes.

3.2 Related Literature

This paper contributes to the literature on prescription drug demand under uncertainty and more broadly, on demand for experience goods. The quality of these so-called experience goods is imperfectly observed so, in order to establish product qualities, a consumer has to try the good(s). Erdem and Keane (1996) and Ackerberg (2003) look at the effects of advertisement on consumer demand in markets where there is uncertainty regarding product quality, which can be resolved with experience and outside information. The first paper looks at scanner data on sales of laundry detergent to estimate how changes in marketing strategy affect brand choice both in the short run and long run. They find that the intensity of promotion has small short-run effects but is significant in the long run.

Ackerberg (2003) examines sales of yogurt, where consumers also learn from experience and advertisement. However, they distinguish between the two potential effects of promotion: "informative" and "prestige" effects, where the latter affect both experienced and inexperienced consumers, while the former affect the demand of inexperienced consumers only. Ackerberg finds that consumers learn from their experience and informational component of advertisement.

Early work on the demand for prescription drugs (e.g. Ellison et al. (1997), Berndt et al. (1996), and Hellerstein (1998)) does not feature consumer learning. Crawford and Shum (2005) build a model of demand with learning based on the premise that patients and their doctors face uncertainty regarding the efficacy and severity of the side effects of a drug in a particular patient before she tries it. Prescription drug consumption is modeled as a bivariate matching problem, where the existing uncertainty is resolved over time through experimenting. With the agency problem

assumed away, physicians maximize their patients' expected utility by selecting a sequence of drug treatments.

The authors measure the effects of uncertainty and learning in the demand for drugs. They find that learning reduces the costs of uncertainty in the anti-ulcer pharmaceutical market. Learning in this class of drugs occurs very quickly. Over two-thirds of patients resolve initial uncertainty and remain on their choice drug after the first prescription. To determine the costs of uncertainty, two counterfactuals are estimated. First, patients make the "best-case" scenario choice under complete information about their matching values. Second, the learning is eliminated by forcing patients to stay on their first choice of treatment. Complete information results in about 9% higher average discounted utility over the baseline case with learning. If experimenting is not allowed, the average utility level drops 6% below the baseline case.

Two recent working papers extend Crawford and Shum (2005). Dickstein (2014a) develops a dynamic model of demand under uncertainty for antidepressant medications using employer-based commercial claims from 2003 until 2005. He relaxes the perfect agency assumption and adopts a different computational approach to accommodate a large set of available drug treatments. Patients are enrolled in a variety of health insurance plans with differential copayment rates. This variation allows Dickstein to study whether insurance cost-sharing policies and drug promotions can improve the efficiency of drug choice, measured by better patient outcome and lower long-run insurer costs. To answer the question, he estimates two sets of counterfactuals. In the first counterfactual, a series of copayment schemes is evaluated: uniform pricing applied across the border, uniform pricing applied to generic and brand drugs separately, and a "value-based" insurance design that channels consumption into the most cost-effective drug classes. In the second counterfactual, the effects of advertisement are simulated. The author selects two antidepressants and adjusts their product-level fixed effects that proxy for promotion. Based on the estimation results, Dickstein argues that value-based policies that are built from observed quit rates can improve patient health, and promotion of cost-effective treatments is beneficial.

Another recent extension of Crawford and Shum (2005) is by Saxell (2013), who takes into account physician private and social learning to study the importance of long-term continuous physician-patient relationship. Using data on cholesterol drug prescriptions in Finland in 2003-2006, Saxell analyzes the physician's attempt to learn patient-drug match value from her own experience and the past choices of other doctors. She finds that treatment strategy is highly responsive to

changes in the length of the doctor-patient relationship; changing physicians slows down learning and increases the total cost of treatment.

This paper extends Crawford and Shum (2005) to account for treatment interruptions, or drug holidays that are taken by most patients in my data. Also, I plan to adjust my model to account for behavioral treatment therapy that is common for patients with ADHD. It is especially beneficial to estimate such a model, because there is no consensus in the medical literature on the effects of behavioral treatment alone and in combination with pharmacological treatment.

3.3 Background

3.3.1 ADHD

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common mental conditions affecting children. It is often first diagnosed in school-aged children. The average age for children to be identified as having the condition is seven years old. The National Comorbidity Survey Adolescent Supplement of 2001-2004 showed that 8.7% from a nationally representative sample of youth aged 13–18 years had ADHD, with males being three times more likely to be diagnosed than females. Approximately half of these cases were classified as severe ADHD (Merikangas et al. (2010)). ADHD can also affect adults. The American Psychiatric Association (APA) estimates that 4.1% of adults in the U.S. have this disorder, with 1.7% of adults affected severely.[1]

More than half of these children are reported as receiving medication treatment for the disorder. Although there is a strong belief that ADHD drugs are overprescribed, very little is known about the existing prescribing practices and physician learning process in ADHD treatment (e.g. see Goldman et al. (1998)). In the Medical Expenditure Panel Survey (MEPS) 1996-2010 dataset, ADHD is one of the top-25 conditions by the number of prescriptions filled. The IMS Institute for Healthcare Informatics (IMS) ranks ADHD 11th among therapeutic classes by U.S. spending in 2011.[2] Since 2007, expenditures on ADHD drugs increased from $4 billion to $7.9 billion in 2011. The market is likely to continue to grow as more adults are being recognized as having ADHD as well.

The description of the syndrome first appeared in 1902. Since then, its definition, catego-

[1] American Psychiatric Association website, http://www.psychiatry.org/adhd. Accessed on April 14th, 2013.
[2] IMS National Sales Perspectives, February 23, 2012.

rization, and treatment practices have changed. The APA defines ADHD as a "brain condition" that is said to be present if either six or more of the inattention symptoms or six or more hyperactivity-impulsivity symptoms "have persisted for a least 6 months to a degree that is maladaptive and inconsistent with developmental level."[3]

Within the past several decades, ADHD has become one of the most studied childhood behavior disorders (Barkley, 2006). In part, it can be explained by a high potential for fruitful medical intervention. In the 1960s, it was shown that stimulant drugs have beneficial effects on both behavioral and cognitive aspects of the condition. More recently, new drugs were introduced, with the most recent being Kapvay and Intuniv in 2009. Development of both medications and therapies broaden the set of choices for managing a wide variety of mental issues united under the umbrella of ADHD. Most stimulant medications are now offered in a number of different strengths, forms and dosages.

Specific causes of ADHD are not fully understood. It was found that although ADHD runs in families, other factors like environment, biological proneness to the condition, and brain injury may play a role in the onset of the condition. Some studies also suggest that children whose mothers smoked, drank alcohol, or were exposed to extreme stress during pregnancy have an increased risk of ADHD.

ADHD is a behavioral disorder that adversely impacts many major life activities from childhood to adulthood. The condition is severe enough to be distressing for children, their families, and teachers. On average, children with ADHD display lower levels of intellectual and academic performance than non-disabled children. They are also more likely to develop a learning disability, to have delays in speech development, and to have lower working memory capacity. Individuals affected by the syndrome were also found to discount the future more heavily than unaffected individuals, to have problems with self-control and self-regulation, and to display riskier behavior, including more dangerous driving. Although there are no studies of the impact of ADHD on life expectancy, issues like more frequent accidents in childhood, auto accidents in adolescence and adulthood, higher crime rates, and use and abuse of substances all can be associated with reduced life expectancy.

If they are left untreated, ADHD sufferers are at a greater risk for potentially serious conse-

[3]The American Psychiatric Association publishes the Diagnostic and Statistical Manual of Mental Disorders (DSM), where it sets criteria for the classification of mental disorders. It is the standard classification of mental disorders used by mental health professionals in the United States. The DSM consists of three major components: the diagnostic classification, the diagnostic criteria sets, and the descriptive text. The most current version is DSM-5 published in May 2013, a revision of DSM-IV-TR that came out in 2000.

quences. Individuals with ADHD, when compared to their unaffected peers, are found to be 32-40% more likely to drop out of school, to rarely complete college (5-10%), to have fewer or no friends (50-70%), to underperform at work (70-80%), to engage in antisocial activities (40-50%), and to use tobacco or illicit drugs more than normal. Furthermore, children growing up with ADHD are more likely to experience teen pregnancy (40%), STDs (16%), depression (20-30%), and personality disorders (18-25%) as adults (Barkley, 2006).

Although there is a consensus that ADHD is a disabling condition, there is little evidence and, thus, agreement on diagnosing and treatment practices. Diagnosing ADHD is a subjective evaluation. In addition to the direct child examination, it involves parents and teachers filling out questionnaires describing the patient's behavior in different settings. The process is complicated by the diversity of ADHD manifestations and by frequently present comorbid conditions.

Children with ADHD are a heterogeneous group who are believed to have in common the characteristics of developmentally inappropriate levels of inattention, and in most cases hyperactivity-impulsivity. With time, the definition of ADHD subgroups will become more refined but, as of today, there are only two diagnosing subcategories to describe different ADHD subpopulations.[4] ADHD has a number of serious comorbid disorders, and it shares symptoms with most of them. They include anxiety disorder, depressive disorder, bipolar I disorder, oppositional defiant and conduct disorders, and others.

3.3.2 SC Medicaid

Medicaid is a means-tested health insurance program. Its target population is low-income families, disabled, aged, and blind individuals, and pregnant women. Under broad federal guidelines, each state manages its own Medicaid program. The provision of prescription drugs' coverage is an optional benefit that is currently offered by all states.

Among other program parameters, the states decide on whether to charge premiums for enrollment and whether to have cost-sharing provisions for the enrollees, but their amount is capped by federal regulation. In most states, certain population groups, including children are exempt from out-of-pocket spending provisions. In SC, children and young adults (under age 19) face zero copayment for the prescription drugs. The state maintains preferred drug lists for medicines that

[4]ICD-9 codes for the Attention deficit disorder are 314.00 (Attention deficit, without hyperactivity) and 314.01 (Attention deficit, with hyperactivity). The American Psychiatric Association maps their classification into the ICD codes.

do not require prior authorization, all other drugs may be covered if a doctor-filed authorization request is approved. The quantity restrictions are also common with a typical prescription capped at a 30-day supply.

A child may be covered by the SC Medicaid program if she is a U.S. citizen or permanent resident. She may be eligible for Medicaid regardless of the eligibility status of her parents or guardians. In SC, the "traditional" Medicaid program is combined with the Children's Health Insurance Program (CHIP) that has looser eligibility criteria than Medicaid itself. Individuals whose income is below 150% of the Federal Poverty Level (FPL) are eligible for Medicaid in SC and those whose income is below 200% of the FPL are eligible for CHIP. All eligible enrollees have access to the same health benefits.

Once eligibility is established, Medicaid coverage is available for an enrollee for a 12-month period (unless the enrollee becomes ineligible during this time), after which the eligibility needs to be reconfirmed. If an eligible individual received medical services and applied for Medicaid after that, her coverage may be activated retroactively for up to three months prior to the month of application, if the individual would have been eligible during the retroactive period had she applied then.

Medicaid is similar to Medicare in that in addition to a traditional "fee-for-service" plan, there are also managed care plans available. Most states expanded their managed care programs in the past decade. However, the only complete data available are for the "traditional" Medicaid enrollees. For this reason, I concentrate on the "fee-for-service" SC Medicaid population.

In SC, 892,583 individuals, about 20% of the state population, were enrolled in Medicaid in FY2009. The majority of enrollees (62%) are female and half of the enrollees are children (52%). The overall program spending in SC was $5.2 billion, 4% of which was spent on prescription drugs.

3.4 Data Description

3.4.1 Treatment Choices

As for most mental disorders, there are three major treatment strategies: medications, behavioral therapy, or a combination of the two. Therapy usually consists of teaching parents and teachers how to provide positive feedback for desired behaviors and consequences for negative ones. Behavioral therapy alone was found to be less effective than pharmacological treatment alone, but no consensus exists on whether medications are inferior to the combination treatment (Barkley, 2006).

There are two major classes of ADHD medications: stimulant and, more recently, non-stimulant drugs. Central nervous system stimulant medications have been used since the 1930s in treatment of behavioral disorders. Today they are the most commonly prescribed drugs to ADHD patients. Stimulants were found to improve the core symptoms of ADHD and to enhance behavioral, academic, and social functioning in about 50-95% of children treated.[5] Stimulants are likely to be recommended as the first step in treatment. If one stimulant does not work, another one may be tried. Children can respond differently to the stimulants, as well as to the other drugs less often used to treat ADHD. The drugs are sometimes, but not often, used in combination. There is little evidence that some stimulants are more efficient than others. There is also uncertainty about whether these benefits last longer than two years.

Table 3.1: Choice Set in the U.S. ADHD Drugs Market, 2003-2012.

Active Ingredient	**Speed**	**Mkt share**	**Major Brands**	**G**	**Entry**	**Avg. Price**
Amphetamine salts	E	25.16	Adderall XR	Y	2001	150.67
Methylphenidate	E	20.26	Concerta	Y	2000	131.00
Methylphenidate	N	11.13	Ritalin LA, Metadate CD, Methylin ER	Y	2002	127.35
Lisdexamfetamine	E	11.04	Vyvanse	N	2007	141.11
Amphetamine salts	M	8.15	Adderall	Y	1996	37.27
Dexmethylphenidate	E	7.19	Focalin XR	N	2005	144.85
Atomoxetine	n/a	6.37	Strattera	N	2002	130.27
Methylphenidate	M	5.82	Methylin, Ritalin	Y	2002	30.16
Others	–	4.89	Various	–		81.15

Notes: "Speed" stands for the drug release speed, where "E" means extended release, "N" - intermediate and "M" – immediate release speed. Extended release drugs are superior than immediate release drugs in that their active ingredient is released over a longer period of time, often allowing for once-a-day dosing. In-sample market share is based on the number of prescriptions filled in 2003-2012. "G" stands for generic drugs availability. Average price is calculated by averaging SC Medicaid reimbursement payments to pharmacies.

To form the choice set, I group all drugs that were approved by the U.S. Food and Drug Administration (FDA) for treatment of ADHD in children, into nine classes by active ingredient and release speed. Table 3.1 lists these classes, indicating their in-sample market share over the entire sample period between 2003 and 2012. The last category, "Others" includes medicines that had an in-sample market share lower than 5%. Note that most stimulant drugs had seen their patent expire, and there are generic substitutes available in the market. The market is dominated by the extended-release formulations of relatively old drugs: together amphetamine salts and

[5]Connor, Daniel F., et al. "Proactive and reactive aggression in referred children and adolescents" American Journal of Orthopsychiatry 74.2 (2004): 129-136. Spencer, Thomas, et al. "Pharmacotherapy of attention-deficit hyperactivity disorder across the life cycle." Journal of the American Academy of Child & Adolescent Psychiatry 35.4 (1996): 409-432.

methylphenidate comprise almost a half of the market for ADHD pharmacological treatments.

Although not approved for the treatment of ADHD, certain antidepressants and sleep-disorder medications are prescribed to patients off-label. For example, Provigil (sleep disorders); Wellbutrin (antidepressant); tricyclic antidepressants; Catapres and Tenex (short-acting forms of high blood pressure medicines); Abilify, Zyprexa, Seroquel, Risperdal, and Geodon (antipsychotics). My data allow for identifying off-label prescription practices and I plan to examine them in the future.

3.4.2 Evidence of Experimenting

In this section I analyze raw data for the evidence of uncertainty and experimentation in the market for ADHD prescription drugs. I use panel data from SC Medicaid for children between ages 3 and 19 years old, who were covered by Medicaid and diagnosed with ADHD between 2003 and 2012. The sample only includes patients who had consistent Medicaid coverage: I exclude individuals who were covered for fewer than 10 months a year. For individuals who had periods of consistent coverage from the beginning of the sample that were followed by inconsistent coverage, I only keep the medical history to the point when the coverage becomes inconsistent and assign a right-censoring flag to the observations that I keep. From this sample of 66,748 individuals I exclude patients for whom I cannot determine the date of the onset of the ADHD condition. To avoid difficulties associated with left-censoring, I drop another 17,399 patients. Excluding patients who were prescribed several ADHD drugs at a time (polytherapy) and patients who had at least one Managed Care claim for the purpose of consistency, I am left with a sample of 12,338 children who take at least one ADHD prescription during the sample period. For the purposes of my baseline model, I also drop patients who were diagnosed with ADHD but did not take any prescriptions while being eligible for Medicaid. These 5,331 patients took an outside option that may stand for behavioral therapy treatment, an off-label prescription for ADHD, or simply no treatment.

Table 3.2 presents summary statistics of prescription patterns observed in my data. Patients take on average 12 prescriptions (drug purchase events). Most prescriptions filled that I observe in the data are for the 30-day supply. Over the course of the treatment, patients try about 1.73 different drugs (as presented in Table 3.1). Since the patients can switch to a drug they have tried before from a current treatment, the number of "spells" is higher – about 2 drugs.[6]

[6] A spell is a period of time when one particular drug is being prescribed.

On average, patient treatment lasts for a year and 9 months and for some patients – for the entire sample period of 10 years.

Table 3.2: Sample Summary Statistics

Variable	N	Mean	Median	St.d.	Min	Max
N of prescriptions	12,338	11.99	7.00	14.42	1	111
N of drugs, per person	12,338	1.73	1.00	1.00	1	8
Number of spells	12,338	2.00	1.00	1.79	1	39
Treatment length (in years)	12,338	1.77	1.05	1.94	0.1	9.69
N patients, who take ADHD drug	12,338					
N right-censored patients	4,303					
N patients, who took outside option	5,331					

Notes: SC Medicaid, 2003–2012. The sample only includes continuously insured individuals.

The variables presented in Table 3.2 are very skewed in the data. Figure 3.1 illustrates the extent of patient heterogeneity in the number of prescriptions filled, number of different drugs tried, length of treatment and number of spells. This heterogeneity speaks in favor of the assumption that patients differ in their underlying illness severity. Also, as Crawford and Shum (2005) suggest, the difference in the length of treatment might also be a result of patient choices. If more effective drugs are also more expensive, some patients and their doctors may favor a cheaper drug that is slower in its curative properties.

Prescription drugs are a good example of an experience good. Although certain qualities of most drugs have been established, there is no medical consensus on their relative efficacy. Even more importantly, it is hard to say how a specific drug would affect any given patient. The human body is very complex and it is nearly impossible to predict its reaction to a chemical component with certainty. This uncertainty means that patients are likely to try more than a single drug in the course of their treatment while they experiment with what is best for them. However, patients also tend to persist in their drug choices, which might look like a lack of experimenting. This persistence is usually explained with risk aversion and switching costs.

In the data, I see evidence of both: persistence in drug choices and experimenting. As Table 3.3 shows, 90.2% of individuals who took drug 2 in the previous month will continue with the same treatment in the next month, while only 79.2% of those who tried drug 9 will continue with it. Table 3.4 presents the transition probabilities between the first and second periods. Patients are much more likely to experiment when they are early in their treatment. The share of patients who stayed on drug 2 after they first tried it is 8 percentage points lower than the average share across all time

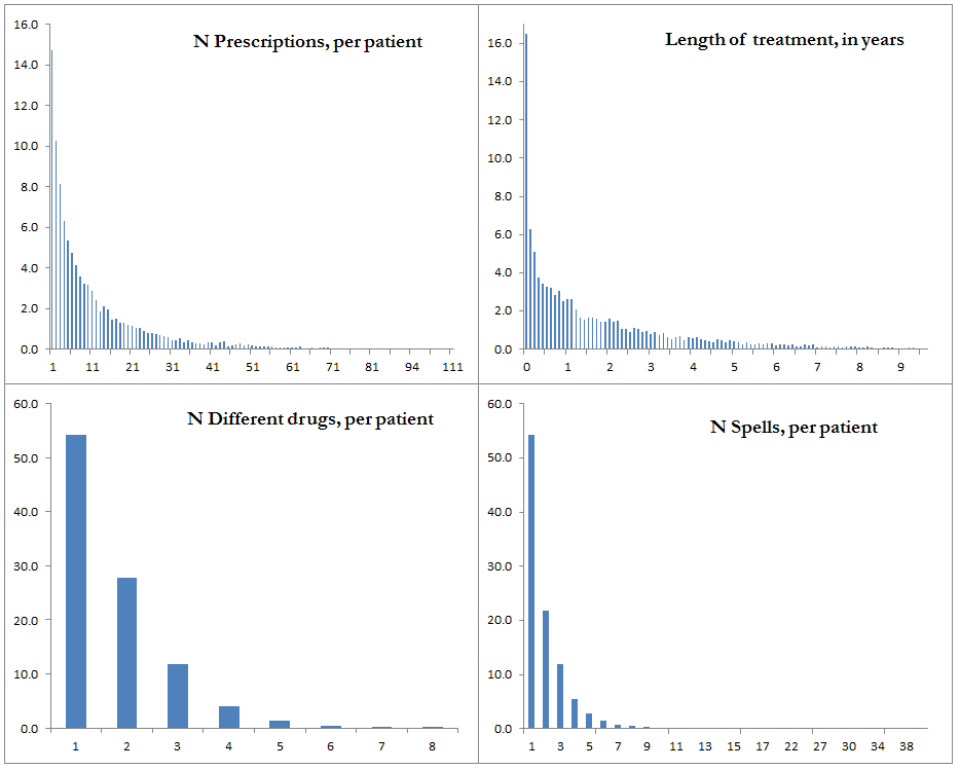

Figure 3.1: Histograms for Prescription Patterns in the Data.

periods (presented in table 3.3) and for drug 8 it is 14 percentage points lower.

Table 3.3: Transition probabilities between periods (t-1) and t.

(t-1)\t	1	2	3	4	5	6	7	8	9
1	**92.8%**	1.4%	1.0%	1.5%	1.5%	0.6%	0.5%	0.3%	0.3%
2	1.9%	**90.2%**	4.3%	1.3%	0.3%	0.7%	0.5%	0.4%	0.3%
3	2.0%	3.5%	**89.9%**	1.5%	0.3%	0.8%	0.5%	1.1%	0.4%
4	1.0%	1.1%	0.9%	**94.7%**	0.4%	0.8%	0.3%	0.2%	0.7%
5	9.5%	1.0%	1.0%	1.6%	**84.4%**	0.4%	0.6%	0.9%	0.6%
6	1.5%	1.4%	0.9%	2.3%	0.2%	**91.2%**	0.5%	0.5%	1.6%
7	2.7%	2.9%	1.4%	1.2%	0.7%	0.7%	**89.4%**	0.5%	0.6%
8	2.7%	3.8%	5.2%	1.4%	1.5%	1.4%	0.6%	**82.6%**	0.9%
9	3.0%	2.1%	2.6%	4.2%	1.1%	5.8%	0.8%	1.1%	**79.2%**

Notes: Previous period $(t-1)$ indexes rows and current period t indexes columns. For example, of patients who took drug #7 in period $(t-1)$ only 2.9% switched to drug #2 in the next period.

Table 3.4: Transition probabilities between first and second periods.

t=1\t=2	1	2	3	4	5	6	7	8	9
1	**84.3%**	3.7%	2.8%	2.1%	2.4%	1.6%	1.4%	1.0%	0.7%
2	6.0%	**82.2%**	4.6%	1.9%	0.9%	1.3%	1.4%	1.3%	0.5%
3	7.1%	5.1%	**78.9%**	2.0%	1.0%	1.9%	1.2%	2.1%	0.7%
4	2.5%	3.7%	2.6%	**84.8%**	0.9%	2.9%	0.4%	1.1%	1.1%
5	16.7%	2.6%	2.4%	2.5%	**69.2%**	0.6%	1.3%	3.0%	1.6%
6	3.8%	2.5%	2.3%	3.6%	0.6%	**82.9%**	0.5%	1.4%	2.5%
7	5.6%	6.0%	2.8%	0.7%	1.1%	0.9%	**80.5%**	1.3%	0.9%
8	6.1%	7.6%	7.4%	2.1%	3.7%	1.9%	1.4%	**68.5%**	1.3%
9	3.8%	2.8%	3.4%	2.0%	3.2%	7.3%	1.3%	2.8%	**73.5%**

Notes: Previous (first) period indexes rows and current (second) period indexes columns. For example, of patients who took drug #7 in the first period 6% switched to drug #2 in the second period.

3.4.3 Drug Holidays

One of the interesting data features is the ability to observe time intervals during which individuals did not consume any drugs. Earlier studies did not account for it. Formally, temporary suspension of pharmacological treatment is called "structured treatment interruption" or "drug holiday". In the case of ADHD, such treatment interruptions are generally not recommended by doctors, but are considered acceptable for patients on stimulant drugs. Non-stimulant medicines have a different mechanism of action. They take longer to produce effect and to leave the body, so they effectively cannot be suspended for a short period of time. In the data, 74% of ADHD patients have taken at least one month off their medication, and for a majority of them, it was no longer

than four months. On average, I see patients taking about two to three holidays over the course of treatment.

Drug holidays are thought to be initiated for one of the following reasons. Most commonly, side effects are bothersome. They include symptoms such as loss of appetite and anorexia, slow growth, and insomnia. Second, some parents may feel that they are "overmedicating" their children and would like to suspend treatment for the periods of low demand for attention, e.g. school holidays. Finally, taking a break from treatment could be an attempt for a patient, her parents and doctor to test her ability to cope with the condition without pharmacological intervention, i.e., to make sure that the drug is needed or not needed at all. There is significant heterogeneity in the need for drug holidays. It might depend on individual-specific side effects' manifestations: school attendance, presence of low versus high demand for attention periods, and parents' perception of drug treatment. Also note that predominantly hyperactive types are less likely to suspend treatment than predominantly inattentive types due to their condition symptoms' manifestation and their impact on other people. Doctors typically recommend a daily regiment for both types.

It is important to distinguish drug holidays from gaps in treatment that are due to parent forgetfulness to refill a prescription or/and adaptation of non-standard treatment regimen (i.e., one pill every two days). The latter is believed to be systematic, while drug holidays are spikes in the data.

Treatment suspension during school holidays can be considered "strategic" and can be identified in the data. These are periods of relatively low demand for attention, when side effects outweigh benefits from treatment. I do observe patterns of such behavior in my data. Table 3.5 shows that if there was treatment suspension of three months or more, patients are more likely to resume the treatment in the Fall, when school starts. Similarly, the last time I would see them filling their prescription is in April or May before the treatment suspension, which is the end of a school year.

Another important dimension of drug holidays is their outcome in terms of switches. When a patient takes a break from her medicine, does she then restart on the same treatment or a different treatment? In about 9% of cases, patients switch medication after a holiday. This switching behavior might suggest that if a medicine is not working, a patient gets off of it, but then as the disease's manifestations intensify they go to the doctor and obtain a prescription for a different drug. When the patient restarts on the same drug, it probably means that side effects were tolerable and they suspended the treatment for some other reason. Supportive of this assumption, average length of a

Table 3.5: Seasonality of Drug Holidays.

Month	First month after holiday		Last month before holiday	
	90+ days	60-120 days	90+ days	60-120 days
Jan	7.34	12.27	5.64	6.82
Feb	7.93	10.34	5.79	6.52
Mar	8.04	7.15	8.64	6.30
Apr	6.45	6.85	13.15	8.61
May	5.42	7.12	14.78	14.69
Jun	3.88	4.64	5.54	8.15
Jul	5.20	6.99	3.80	5.46
Aug	16.60	16.34	6.27	7.23
Sep	14.15	8.24	6.09	6.59
Oct	10.17	6.81	7.37	8.47
Nov	8.27	6.46	9.28	12.13
Dec	6.54	6.79	13.65	9.02
N obs	**22,235**	**16,987**	**22,235**	**16,987**

Notes: A drug holiday of 90 days means that 120 days pass since last prescription was filled until the treatment is restarted, given that the last prescription was written for a 30-day therapy.

holiday diminishes over the treatment length. As a result of experimenting early in the treatment, patients find a drug they are comfortable with and take it regularly as prescribed. The median length of drug holidays does not vary significantly with the choice of drug. In this paper I will examine how treatment interruption affects treatment length and cost.

3.4.4 Behavioral Therapy

Non-medicine treatments of ADHD include parental behavior training, psychosocial therapy and school-based programs, all of which can be combined with one another and/or with pharmacological treatment. Parental training includes teaching parents how to understand and correct their children's behavior. Psychosocial therapy sessions are designed to help a child acquire or improve social skills and to control her behavior and emotions. School-based programs include special education services for children who qualify. Parental training, psychosocial therapy sessions and a number of school- and community-based services are covered by SC Medicaid. I use Current Procedural Terminology codes (CPTs) listed on every doctor visit claim to identify behavioral therapy sessions.[7]

[7] In doing that, I follow the ADHD Coding Fact Sheet for Primary Care Pediatricians from the "Caring for Children With ADHD: A Resource Toolkit for Clinicians", 2^{nd} Edition published by the American Academy of Pediatrics in 2012 and "Evaluation/Management and Psychotherapy Coding Algorithm developed by the American Academy of Child and Adolescent Psychiatry, 2012, `www.aacap.org/App_Themes/.../Code_Selection_Algorithm_v3.pdf`. Accessed on December 1^{st}, 2013.

In the data, about 33% of identified behavioral therapy sessions are individual psychotherapy, (e.g. CPT 90807 "Individual psychotherapy, insight oriented, behavior modifying and/or supportive, in an office or outpatient facility, approximately 45 to 50 minutes face-to-face with the patient; with medical evaluation and management services"). The majority (54%) of therapy sessions are either group therapy or parent sessions or therapy with a parent present (e.g. CPT 96153 "Health & Behavior Intervention – Group"). I also include other psychological services like case management and school-based mental health services for children diagnosed with ADHD. Note that most school-based special education programs are not billed to Medicaid and are not included in my sample.

Table 3.6 presents summary statistics on behavioral treatment claims paid by SC Medicaid between 2003 and 2012. On average, a patient attends two to three sessions in a month. This number is slightly below a typically recommended regimen of weekly sessions. It suggests that such potential complications of non-medicine treatment as the need to adjust parent schedule and the availability of providers in the area are likely to be real. Another difficulty is that behavioral therapy treatment often requires long-term commitment: on average, patients attend sessions for 15 months. However, almost 15% of patients in my sample attend a single therapy session over the course of their treatment.

Therapy can also be costly. In SC, average monthly cost per patient with ADHD paid by Medicaid is about $668. It is consistent with an estimate provided in AHRQ (2012) of "between $300 and $2,000, depending on the individual therapist or program and the amount of time needed". SC Medicaid patients are responsible for a copay of $3.30 per visit and are limited to a certain number of visits per day, month and/or year depending on the procedure performed. There are also non-pecuniary costs of treatment – time spent to get to the physician's office and the session itself.

American Academy of Pediatrics (2001) and American Academy of Child and Adolescent Psychiatry (2007) guidelines state that behavioral treatment is not as effective as pharmacological intervention, and the latter guideline does not recommend non-medicine treatment unless drugs are found ineffective. In contrast, Fabiano et al. (2009) analyze 114 published research papers and conclude that there is strong and consistent evidence that behavioral interventions are effective in treating ADHD. However, they also note that there is not enough evidence of long-lasting positive effects of behavioral treatment. In my sample, only 1% of patients are treated with behavioral therapy alone, while most children (70%) rely completely on ADHD drugs for treatment.

Even though long-term effects of behavioral treatment are not established, it has a benefit of not causing side effects as drugs do. Also, while the ADHD drugs are known to decrease ADHD symptoms, they do not change behavioral habits. As soon as a patient discontinues her drug treatment, its symptom-relief effects go away due to the nature of ADHD medicines.

Table 3.6: Behavioral Therapy Summary Statistics, 2003-2012.

Variable	N	Mean	Median	St.d.	Min	Max
N sessions, per patient/month	4,879	2.52	1.67	2.97	1.00	43
Therapy cost, per patient/month	4,879	667.89	259.00	1,247.92	14.00	15,493
Months of therapy, per patient	4,879	14.88	6.90	19.63	0.00	113
% patients on therapy only	1.02%					
% patients on drugs only	70.33%					
N patients	16,444					
N observations	209,259					

Notes: The sample contains fee-for-service SC Medicaid enrollees only and excludes patients with left-censored observations, patients with noncontinuous eligibility status and patients, who take multiple ADHD prescription drugs in a month. Zero months of therapy duration is assigned to patients who attend a single session during their treatment.

3.5 Model

3.5.1 Initial Diagnosis

I adopt a dynamic model of demand for pharmaceuticals with learning under uncertainty developed by Crawford and Shum (2005) to study a phenomenon of drug holidays, or planned treatment interruptions. It is a discrete choice model, where patients choose between available drugs and occasionally choose to interrupt their treatment and take no medicines for a relatively short period of time.

When a patient comes to a doctor's office for the first time, she receives an initial diagnosis. In the context of ADHD, it is the type of the condition – with or without hyperactivity, and its severity. Although ADHD type is known to the econometrician, condition severity is not. To address the problem of unobserved heterogeneity, it is assumed that doctors classify patients into K discrete types based on the initial diagnosis. These patient groups differ in two dimensions. First, they have different condition severity, and second, they may respond differently to medical treatments. The model handles the former by assuming that the initial probability of being healed for individual j

(without any treatment at all) varies with patient type:

$$h_{0j} = \theta_k \tag{3.1}$$

with probability p_k, $k = 1, \ldots, K$, where j indexes patients, k indexes patient types by the underlying illness severity, and h_{0j} is the initial probability of being healed without taking drugs and $0 < p_k < 1$, $0 \leq \theta_1, \ldots, \theta_K \leq 1$. The more severe is the condition, the lower is the probability of recovery without receiving any treatment.

To account for the fact that patients of different types are likely to respond differently to ADHD medicines, doctors' prior beliefs regarding symptomatic and curative match values are allowed to vary by patient type.

3.5.2 Patient Preferences

Patient utility is a function of the symptomatic signal x_{jnt} of treatment, i.e. of how successful a drug is in treating the symptoms of ADHD for this patient and its side effects' profile, and the cost of treatment n, p_n. Side effects that are typically associated with ADHD drugs are decreased appetite, sleep problems, headache, irritability, jitteriness, and stomach pain. The side effects may be mild or in some cases severe enough to cause patients to switch to another treatment. Side effects also vary by age of the patient, with preschoolers often having stronger side effects.

During a drug holiday, a patient continues to receive symptomatic signals that are associated with the no-drug option and the cost of treatment becomes zero. Patients are modeled as risk-averse individuals in order to accommodate significant persistence in choices that are observed in the data. Imposing a Constant Absolute Risk Aversion specification yields the following utility function:

$$u(x_{jnt}, p_n, \epsilon_{jnt}) = -exp(-r \times x_{jnt}) - \alpha \times p_n + \epsilon_{jnt} \tag{3.2}$$

where ϵ_{jnt} is an additive idiosyncratic error that measures idiosyncratic tastes for drug n by patient j in period t, and r measures the level of risk aversion. Note that the utility is negative up to the error term and the utility in the period just before the patient is cured is normalized to zero. This is done to avoid a situation, when patients delay recovery in order to continue receiving positive utility.

3.5.3 Recovery probabilities

After patient j takes drug n in period t, she can recover at the end of period t and need no more prescriptions with probability h_{jt}. This posterior probability evolves through time according to the following expression:

$$h_{jt}(h_{jt-1}, y_{jnt}) = \frac{(\frac{h_{jt-1}}{1-h_{jt-1}}) + d_{jnt}y_{jnt}}{1 + ((\frac{h_{jt-1}}{1-h_{jt-1}}) + d_{jnt}y_{jnt})} \tag{3.3}$$

where $h_{0j} = \theta_k$. Note that since h_{jt} depends on the curative signal y_{jnt}, h_{jt} is random when patient j makes choice in period t.

3.5.4 Learning process

Upon initial diagnosis, doctors determine initial treatment choice according to their prior beliefs on how a particular drug n affects patients of type K. Person-specific match values are unknown, but they can be described by a multivariate normal distribution:

$$\begin{pmatrix} \mu_{jn} \\ \upsilon_{jn} \end{pmatrix} \sim N\left(\begin{bmatrix} \underline{\mu}_{nk} \\ \underline{\upsilon}_{nk} \end{bmatrix}, \begin{bmatrix} \underline{\sigma}_n^2 & 0 \\ 0 & \underline{\tau}_n^2 \end{bmatrix} \right) \tag{3.4}$$

where $\underline{\mu}_{nk}$ and $\underline{\nu}_{nk}$ denote prior mean symptomatic and curative matching values for each drug and patient type, and $\underline{\sigma}_n$ and $\underline{\tau}_n$ are standard deviations for these means respectively. Note that they are assumed not to vary across patient types. Patient-specific matching values μ_{jn} and ν_{jn} are not known at the start of the treatment, but every period patients receive two noisy normally distributed signals that are centered around the true match values.

$$\begin{pmatrix} x_{jnt} \\ y_{jnt} \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_{jn} \\ \upsilon_{jn} \end{bmatrix}, \begin{bmatrix} \sigma_n^2 & 0 \\ 0 & \tau_n^2 \end{bmatrix} \right) \tag{3.5}$$

Signals are drawn every period t. Doctors are assumed to have rational expectations in a sense that their prior beliefs about drug-specific match values distribution corresponds to their actual distribution. Patients and their doctors know values $\underline{\mu}_{nk}$, $\underline{\nu}_{nk}$, and their respective standard deviations

$\underline{\sigma}_n$, $\underline{\tau}_n$. I estimate these parameters.

Following Crawford and Shum (2005), I accommodate new-to-market drugs by allowing prior beliefs about match values to vary for the first two years since the drug entry. The idea behind this set up is that when a new drug enters the market, its actual symptomatic and curative match values are not clearly established. With time, doctors learn about these drugs through the experience of their patients and from medical studies and symposiums.

In my sample, several new brand drugs enter the market between 2003 and 2012. The two newest drugs are non-stimulants Kapvay and Intuniv that were launched in 2009. Their in-sample market share has not reached 5% by 2013, so I bundle them into a composite drug with other low market share drugs. Another relatively recent market entrant is Vyvanse, a stimulant that was launched in 2007 and soon gained significant market share. I allow prior beliefs for it to vary every half a year for two years, estimating four prior mean match value parameters for each drug/type, instead of just one as it is for older drugs with established qualities in the population.

Patient j's posterior beliefs regarding her symptomatic match value μ_{jn} and curative match value ν_{jn} are given by the sequence of normal distributions with the following mean and variance:

$$\mu_{jn}^{t+1} = \begin{cases} \frac{\frac{\mu_{jn}^t}{V_{jn}^t} + \frac{x_{jnt+1}}{\sigma_n^2}}{\frac{1}{V_{jn}^t} + \frac{1}{\sigma_n^2}} = \frac{\sigma_n^2 \mu_{jn}^t + V_{jn}^t x_{jnt+1}}{\sigma_n^2 + V_{jn}^t}, & \text{if drug } n \text{ is taken in } t+1, \\ \mu_{jn}^t, & \text{otherwise.} \end{cases}$$

$$V_{jn}^{t+1} = \begin{cases} \frac{1}{\frac{1}{\underline{\sigma}_n^2} + \frac{l_{jn}^{t+1}}{\sigma_n^2}} = \frac{\sigma_n^2 \underline{\sigma}_n^2}{\sigma_n^2 + l_{jt}^{t+1} \underline{\sigma}_n^2}, & \text{if drug } n \text{ is taken in } t+1, \\ V_{jn}^t, & \text{otherwise.} \end{cases} \tag{3.6}$$

where $\mu_{jn}^0 = \underline{\mu}_{nk}$, $V_{jn}^0 = \underline{\sigma}_n$ and l_{jn}^t is a count of number of times that patient j has taken drug n by time t, including t.

$$\upsilon_{jn}^{t+1} = \begin{cases} \dfrac{\tau_n^2 \upsilon_{jn}^t + R_{jn}^t y_{jnt+1}}{\tau_n^2 + R_{jn}^t}, & \text{if drug } n \text{ is taken in } t+1, \\ \upsilon_{jn}^t, & \text{otherwise.} \end{cases}$$

$$R_{jn}^{t+1} = \begin{cases} \dfrac{\tau_n^2 \underline{\tau}_n^2}{\tau_n^2 + l_{jn}^{t+1} \underline{\tau}_n^2}, & \text{if drug } n \text{ is taken in } t+1, \\ R_{jn}^t, & \text{otherwise.} \end{cases} \tag{3.7}$$

where initial conditions are $\nu_{jn}^0 = \underline{\nu}_{nk}$, $R_{jn}^0 = \underline{\tau}_n$.

3.5.5 Dynamic Drug Choice

Every period a diagnosed patient and her doctor choose a medication for treatment of ADHD. It can be one of the drugs from the choice set, or a no-drug option (i.e., patient takes a drug holiday). Entering period t, the state variables are: the patient j's posterior mean match values μ_{jn}^t and ν_{jn}^t; counts of the number of times patient j tried each drug n, l_{jn}^t; the recovery probability for patient j at the end of time period t, h_{jt}; and the idiosyncratic error terms υ_{jnt}. All of them are collected into a vector of state variables S_t for period t: $S_t = (\mu_{j1}^t, ..., \mu_{j10}^t, \upsilon_{j1}^t, ..., \upsilon_{j10}^t, l_{j1}^t, ..., l_{j10}^t, h_{jt}, \epsilon_{j1t}, ..., \epsilon_{j10t})$.

The transition rules for mean match values are defined by the form of their posterior distribution (see top rows of equations (3.6) and (3.7)). The recovery probability transition rule is defined in equation (3.3). Finally, the transition rule for the count of times a patient had experience with a particular drug n is:

$$l_{jn}^{t+1} = \begin{cases} l_{jn}^t + 1 & \text{if drug } n \text{ is taken in } t, \\ l_{jn}^t & \text{otherwise.} \end{cases} \tag{3.8}$$

3.5.6 Dynamic Optimization Problem

Patients and their physicians maximize utility $W(S_t)$ that is obtained as a solution to an infinite horizon problem defined in the equation (3.2). It can be defined recursively using Bellman's

equation:

$$
\begin{aligned}
&W(S_t) = \max_n \mathbb{E}[u(x_{jnt}, p_n, \epsilon_{jnt}) + \beta(1 - w_{jt})\mathbb{E}[W(S_{t+1})|x_{jnt}, y_{jnt}, n]|S_t] \\
&= \max_n \mathbb{E}[u(x_{jnt}, p_n, \epsilon_{jnt}) + \beta(1 - \mathbb{E}[w_{jt}|y_{jnt}])\mathbb{E}[W(S_{t+1})|x_{jnt}, y_{jnt}, n]|S_t] \\
&= \max_n \mathbb{E}[u(x_{jnt}, p_n, \epsilon_{jnt}) + \beta\mathbb{E}(1 - h_{jt}(h_{jt-1}, y_{jnt}))\mathbb{E}[W(S_{t+1})|x_{jnt}, y_{jnt}, n]|S_t] \\
&= \max_n (-exp(-r\mu_{jn}^t + \frac{1}{2}r^2(\sigma_n^2 + V_{jn}^t)) - \alpha p_n + \epsilon_{jnt} \\
&+\beta\mathbb{E}(1 - h_{jt}(h_{jt-1}, y_{jnt}))\mathbb{E}[W(S_{t+1})|x_{jnt}, y_{jnt}, n]|S_t] \\
&\equiv \max_n\{W_n(S_t)\}
\end{aligned}
$$

Optimal policy for the patient j in every time period t is to choose the drug n with the highest value function $W_n(S_t)$. Assuming stationarity of the optimal policy, the Bellman equation can be rewritten as follows:

$$
\begin{aligned}
W(S) = &\max_n \mathbb{E}[u(x_{jnt}, p_n, \epsilon_{jnt}) \\
&+ \beta\mathbb{E}(1 - h'(h_j, y_{jn}))\mathbb{E}[W(S')|\vec{x}_{jn}, n]|S]
\end{aligned} \tag{3.9}
$$

where $\vec{x}_{jn}$ is a vector of symptomatic and curative signals from drug n, $\vec{x}_{jn} \equiv (x_{jn}, y_{jn})'$. The value function is computed using Keane and Wolpin (1994) approximation method.

3.6 Econometric Model

Each period patient j selects one of n drugs from the choice set or a drug holiday option. Denote the sequence of choices as $d_{j1t}, ..., d_{j10t}$; treatment length as T_j and a censoring indicator as I_j. Recall that if I do not observe the entire treatment sequence for a patient (I do not see the period in which she is cured), I use the information on decision choices up to the point when censoring occurs.[8]

Then, the likelihood function for patient j in period t can be written as follows:

[8] See Appendix A for a detailed discussion on the assumption of cure.

$$\prod_n \mathbb{E}_{\epsilon_{j1l},...,\epsilon_{j10l}}(\mathbb{1}(W_{jn1,k} > W_{jn'1,k}, n' \neq n))^{d_{jn1}}, \text{ for } t{=}1,$$

$$\mathbb{E}_{x_{jnt-1,k},h_{0j,k}}[\big((1-h_{jt-1,k})\times$$

$$\prod_n \mathbb{E}_{\epsilon_{j1l},...,\epsilon_{j5l}}(\mathbb{1}(W_{jnt,k} > W_{jn't,k}, n' \neq n)^{d_{jn1}})\big)], \ 1 < t < T_j,$$

$$\mathbb{E}_{\vec{x}_{jnt-1,k},h_{0j,k}}[(1-I_j)\times h_{jT_j,k}] \text{ for } t = T_j, \tag{3.10}$$

where $\vec{x}_{jnt,k} \equiv (x_{jn1,k}, ..., x_{jnt,k}, y_{jn1,k}, ..., y_{jnt,k})$ is a vector of per-period symptomatic and curative signals from drug n, patient type j. The signals and healing probability depend on the severity of each patient's condition, group k. Note that the likelihood function is different for censored and uncensored individuals in the last period, because for the censored patients the actual last period in treatment is unobserved.

Assume that the per-period additive idiosyncratic shocks to the patient's utility are i.i.d. Type I Extreme Value, choice probabilities can be re-written in multinomial logit expression:

$$\mathbb{E}(\mathbb{1}(W_{jnt,k} > W_{jn't,k}, n' \neq n))$$
$$= exp(W_{jnt,k})/[\sum_{n'=1}^{10} exp(W_{jn't,k})] \equiv \lambda_{jnt,k}$$

Recall also that patient initial diagnosis is observed by doctors and patients, but not by an econometrician. The probability of being type k is denoted by p_k. Then the likelihood function becomes:

$$\sum_{k=1}^{K} p_k \cdot \mathbb{E}_{\vec{x}_{jnT_j,k}|h_{0j,k}}[\prod_{t=1}^{T_j-1}((1-h_{jt,k})\prod_n \lambda_{jnt,k}^{d_{jnt}})] \cdot h_{jT_j,k}\prod_n \lambda_{jnT_j,k}^{d_{jnT_j}} \tag{3.11}$$

for uncensored patients, and

$$\sum_{k=1}^{K} p_k \cdot \mathbb{E}_{\vec{x}_{jnT_j,k}|h_{0j,k}}[\prod_{t=1}^{T_j-1}((1-h_{jt,k})\prod_{n}\lambda_{jnt,k}^{d_{jnt}})] \cdot \prod_{n}\lambda_{jnT_j,k}^{d_{jnT_j}} \tag{3.12}$$

for censored patients.

3.7 Results

Table 3.7 and Table 3.8 report the estimates of two variants of a dynamic learning model of demand for ADHD treatment.[9] First, I estimate a baseline model that does not take into account drug holidays or behavioral therapy sessions. It assumes that prescriptions are taken regularly every month. Next, I estimate a model that includes drug holidays. I assume that drug holidays are a part of the choice set, similar to the actual medicines. Patient treatment is initiated with a drug holiday, when I observe a lag of at least two weeks between the first diagnosis of ADHD and first filled ADHD prescription.

In this draft, I estimate both models with two discrete patient types ($K = 2$). I also reduce my original choice set (see Table 3.1) and group drugs into 4 categories by their stimulant status and release speed. Drug 1 is the newest stimulant drug in the sample – Vyvanse, which was introduced in 2007 and is only available as a brand name. For Vyvanse, I estimate a two-period evolution in physician beliefs. When the drug entered the market, there was little information available, so the prior beliefs are allowed to change with time (for simplicity, only once after the first six months in the current variant of the model). Drug 2 has older extended-release stimulants like Adderall XR and Concerta. Drug 3 combines immediate-release stimulants like Adderall and Ritalin. Finally, drug 4 is reserved for nonstimulant medicines: Strattera, Intunive and Kapvay.

The two columns of Table 3.7 present estimates of the model parameters for each of the two unobserved patient types. The first panel indicates that the relatively "healthy" type (highest baseline recovery probability, h_o) comprises 33.1% of the population while the less healthy type makes up 66.9% of the sample. For the two patient types, the model suggests substantial heterogeneity in the underlying illness severity, measured as baseline recovery probability. ADHD is a chronic

[9] In both empirical models, I follow Crawford and Shum (2005) and restrict the variances of the curative signals, τ_n^2, prior variances of symptomatic ($\underline{\sigma}_n^2$) and curative ($\underline{\tau}_n^2$) distributions to be identical across drugs to reduce parameter space. In the future, I plan to estimate a less restrictive model.

condition and here by "recovery" I mean a situation, when a patient is able to function successfully in life without taking medicines.

The next panel in the table presents the prior means of the symptomatic match values for each drug and each patient type (see $\underline{\mu}_{nk}$ in Equation 3.4). The third panel presents the prior means for the curative match values for each drug and patient type (see $\underline{\nu}_{nk}$ in Equation 3.4). The fourth panel in Table 3.7 presents the estimates of the standard deviations of the symptomatic match values ($\underline{\sigma}_n^2$ in Equation 3.4) and the fifth panel presents the standard deviations for the symptomatic signals (σ_n^2 in Equation 3.5). For simplicity I assume the standard deviations of the match values ($\underline{\sigma}_n^2$ and $\underline{\tau}_n^2$) and signals (σ_n^2 and τ_n^2) do not vary across patient types. I also restrict standard deviation of the curative match values ($\underline{\tau}_n^2$) and curative signals (τ_n^2) to not vary across choices in the current specification. Finally, the table presents estimates of the price coefficient (α) and risk-aversion (r) parameters that enter the utility function (see Equation 3.2).

The parameter estimates of mean match values suggest that the two patient types differ not only in the severity of their condition, but also in their response to treatment. The drug that is the best for "sick" patients by symptomatic match (extended-release stimulants) ranks second for "healthy" patients. By symptomatic match, relatively healthy patients are better matched to Vyvanse, while for relatively sick patients it is their last choice. By curative mean match values, extended-release stimulants are again the best option for the "sick" type, while for the "healthy" type it is immediate-release stimulants.

Distributions of symptomatic and curative match values for both patient types are presented in Figure 3.2. Standard deviations for symptomatic match values are relatively large in magnitude in comparison to their respective means. Significant overlap between the four drugs' distributions suggests that in terms of symptomatic match ADHD drugs are horizontally differentiated, so even patients of the same type are likely to have very different match values by drug symptomatic properties. It means that within a group of patients of the same type, a significant uncertainty regarding a specific patient-drug match is present.

This phenomenon is even more pronounced in the curative match value distributions. For relatively sick type, extended-release stimulants stand-out, but all other treatments overlap completely. For healthy type, all drugs' distributions overlap with each other. Again, it means that there is uncertainty of how good would be a match between a specific patient and a drug even if the initial diagnosis (that determined baseline recovery probability) is definitive.

Table 3.7: Baseline Dynamic Model: Parameter Estimates.

Parameter	**Est.**	**Std. err.**	**Est.**	**Std. err.**
Illness heterogeneity distribution	**Type 1 ("Sick")**		**Type 2 ("Healthy")**	
Recovery probability	0.040		0.758	
Type probability	0.669		0.331	
Means, symptom match values				
$\underline{\mu}_{\text{vyvanse, Period 1}}$	1.331		1.001	
$\underline{\mu}_{\text{vyvanse, Period 2}}$	-0.114		1.145	
$\underline{\mu}_{\text{extended-release stimulants}}$	1.572		1.141	
$\underline{\mu}_{\text{immediate-release stimulants}}$	0.770		1.133	
$\underline{\mu}_{\text{nonstimulants}}$	0.869		1.127	
Means, curative match values				
$\underline{\nu}_{\text{vyvanse, Period 1}}$	1.085		0.976	
$\underline{\nu}_{\text{vyvanse, Period 2}}$	1.156		0.971	
$\underline{\nu}_{\text{extended-release stimulants}}$	1.717		0.708	
$\underline{\nu}_{\text{immediate-release stimulants}}$	1.071		1.034	
$\underline{\nu}_{\text{nonstimulants}}$	1.025		0.873	
Std. dev., symptom match values				
$\underline{\sigma}_{\text{vyvanse}}$	1.196			
$\underline{\sigma}_{\text{extended-release stimulants}}$	1.329			
$\underline{\sigma}_{\text{immediate-release stimulants}}$	1.262			
$\underline{\sigma}_{\text{nonstimulants}}$	1.134			
Std. dev., symptom signals				
σ_{vyvanse}	1.368			
$\sigma_{\text{extended-release stimulants}}$	1.220			
$\sigma_{\text{immediate-release stimulants}}$	1.193			
$\sigma_{\text{nonstimulants}}$	1.276			
Std. dev., curative match values				
$\underline{\tau}$	1.307			
Std. dev., curative signals				
τ	1.196			
Price coefficient, α	1.125			
Risk-aversion parameter, r	1.058			
Discount rate, β	0.950	(fixed)		
N patients	1,000			
N time periods	36			
N observations				
N draws	10			

Notes: Drug prices are averaged across time and patients. Vyvanse is a new drug and its mean match values vary in time. Time periods are months.

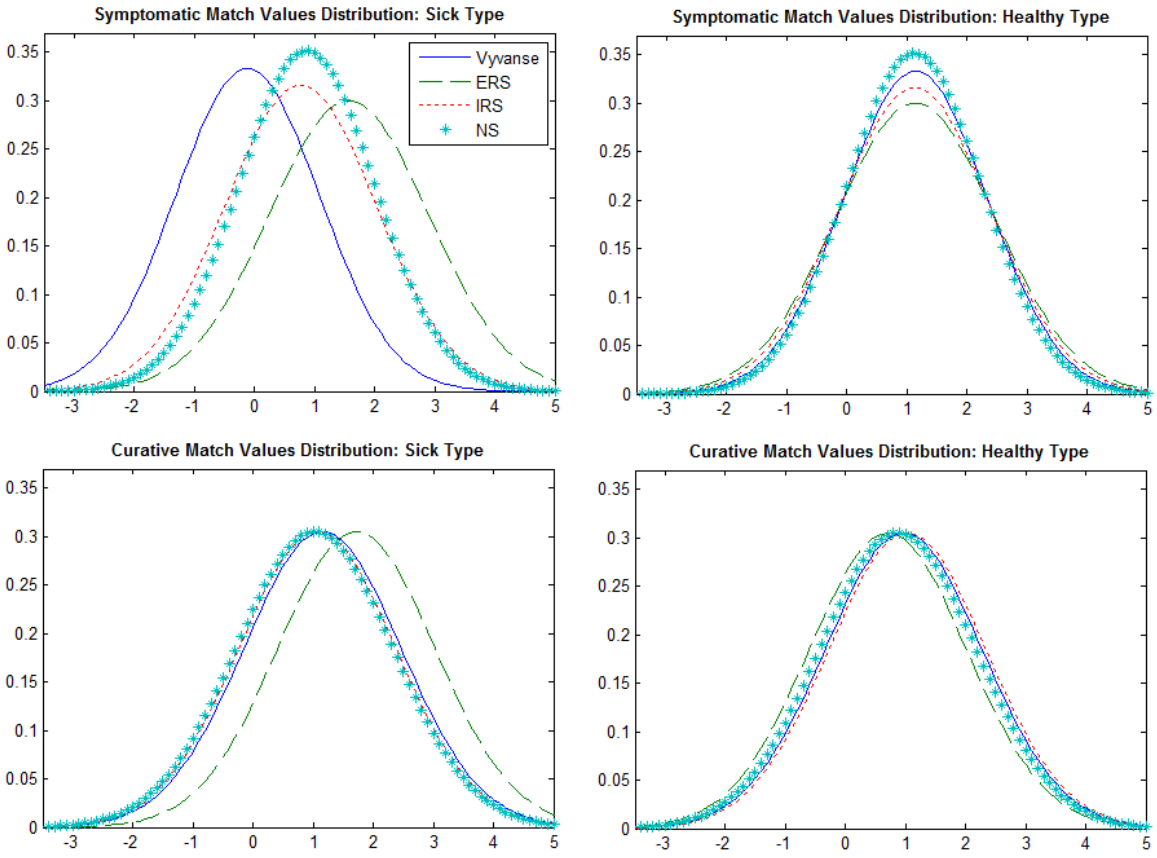

Figure 3.2: Estimated Symptomatic and Curative Match Values Distributions. Baseline Model.

Table 3.8: Dynamic Model with Drug Holidays: Parameter Estimates.

Parameter	**Est.**	**Std. err.**	**Est.**	**Std. err.**
Illness heterogeneity distribution	**Type 1 ("Sick")**		**Type 2 ("Healthy")**	
Recovery probability	0.000		0.959	
Type probability	0.932		0.068	
Means, symptom match values				
$\underline{\mu}_{\text{vyvanse, Period 1}}$	0.512		0.692	
$\underline{\mu}_{\text{vyvanse, Period 2}}$	-0.778		0.489	
$\underline{\mu}_{\text{extended-release stimulants \& nonstimulants}}$	4.823		1.067	
$\underline{\mu}_{\text{immediate-release stimulants}}$	0.279		0.452	
$\underline{\mu}_{\text{drug holiday}}$	1.551		3.454	
Means, curative match values				
$\underline{\nu}_{\text{vyvanse, Period 1}}$	1.477		0.772	
$\underline{\nu}_{\text{vyvanse, Period 2}}$	-0.178		3.153	
$\underline{\nu}_{\text{extended-release stimulants \& nonstimulants}}$	-4.919		1.737	
$\underline{\nu}_{\text{immediate-release stimulants}}$	-1.805		2.391	
$\underline{\nu}_{\text{drug holiday}}$	-0.160		0.025	
Std. dev., symptom match values				
$\underline{\sigma}_{\text{vyvanse}}$	1.266			
$\underline{\sigma}_{\text{extended-release stimulants \& nonstimulants}}$	1.364			
$\underline{\sigma}_{\text{immediate-release stimulants}}$	1.192			
$\underline{\sigma}_{\text{drug holiday}}$	1.392			
Std. dev., symptom signals				
σ_{vyvanse}	1.435			
$\sigma_{\text{extended-release stimulants \& nonstimulants}}$	1.894			
$\sigma_{\text{immediate-release stimulants}}$	1.602			
$\sigma_{\text{drug holiday}}$	1.030			
Std. dev., curative match values				
$\underline{\tau}$	1.097			
Std. dev., curative signals				
τ	1.034			
Price coefficient, α	1.178			
Risk-aversion parameter, r	1.047			
Discount rate, β	0.950	(fixed)		
N patients	500			
N time periods	24			
N observations				
N draws	10			

Notes: Drug prices are averaged across time and patients. Vyvanse is a new drug and its mean match values vary in time. Time periods are months.

In terms of time-varying beliefs about match values, the model estimates show that physicians felt overly optimistic about Vyvanse side effects profile and its relief effects for sick patients and slightly less optimistic for healthy patients. However, they revised their beliefs up for its curative properties for sick type and did not change them for health type.

It is possible to cross-check some of the results with the actual market parameters. Extended-release stimulants fare very well in terms of symptomatic effects for both types and also in terms of curative effects for the larger class of patients - sick type. This combined class has a dominating share of 52.6% in my sample. Non-stimulants that are rarely recommended as a first-line therapy due to the lack of evidence on their efficacy, rank low on curative properties for both patient types, but rank second for the sick type in terms of symptomatic effects. This reconciles with the fact that nonstimulant medicines do not cause agitation or sleeplessness - typical stimulant side effects, and also they are not a controlled substance and do not cause addiction.

Next, I estimate a model that accounts for interruptions in treatment, drug holidays. The parameter estimates are presented in table 3.8. To keep the choice set small, I combine extended-release stimulants with nonstimulants as they are relatively recent market entrants and are also taken about once a day.

This model estimates even higher polarization between sick and healthy patient types: most patients have severe ADHD and the chance that they will be able to function successfully without treatment after the first prescription is extremely small. Although the models are not directly comparable because of a different choice set, it is clear that when the extended-release stimulants and non-stimulants ranked next to each other in the baseline case, the relative drug ranking was preserved in the model with drug holidays.

As in previous model, patient types are heterogeneous in both their symptomatic and curative responses to treatments. Symptomatically, drug holidays feel best for healthy types. Their ability to function without drugs is highly probable, so the relief they feel from taking any drug from the choice set is not enough to compensate for the disutility from drugs side effects.

The distributions of match values are very similar to estimated distributions in the baseline model (see Figure 3.3). However, in the model with drug holidays, extended-release stimulants and nonstimulants stand out by their symptomatic properties for the sick type and drug holidays - for the healthy type. The rest of the drugs' distributions overlap significantly, pointing at their horizontal

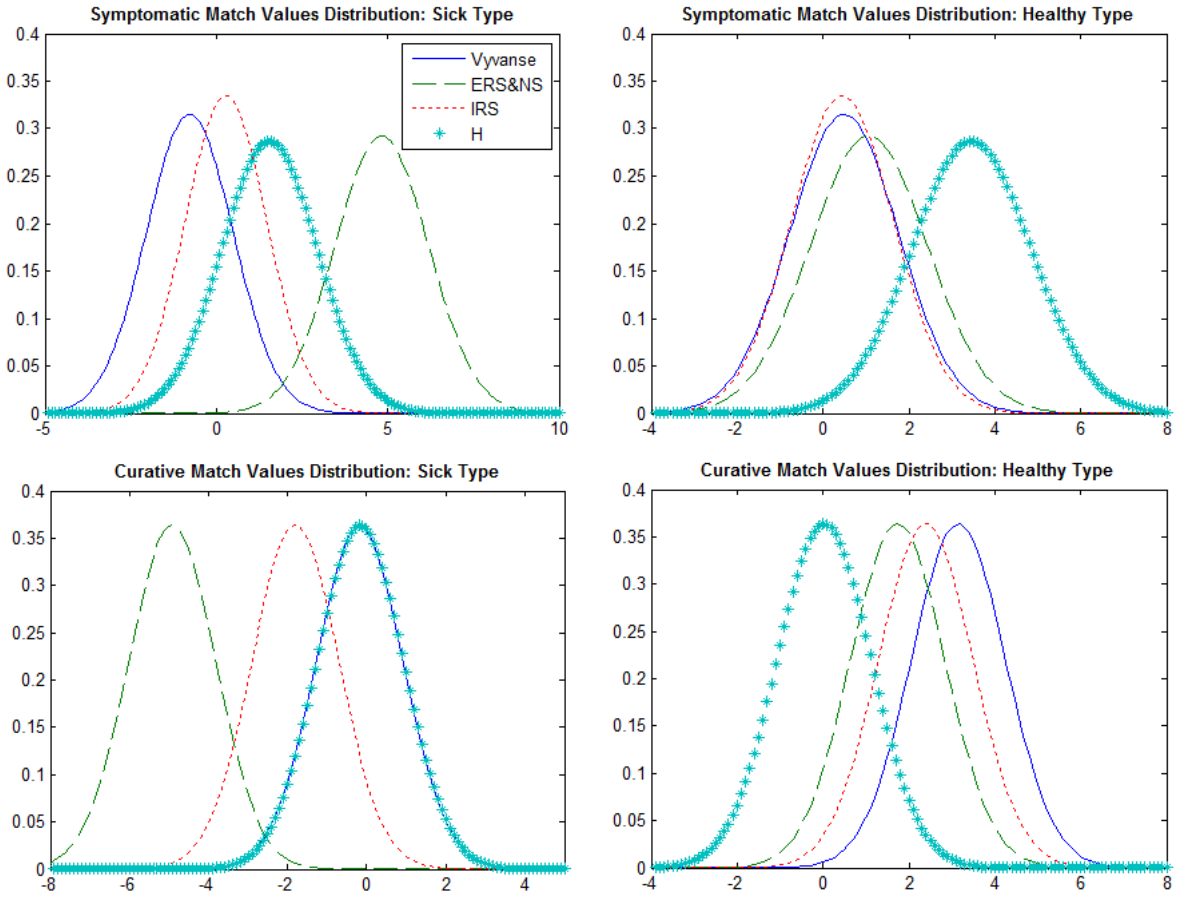

Figure 3.3: Estimated Symptomatic and Curative Match Values Distributions. Model with Drug Holidays.

differentiation. There is also relatively little uncertainty regarding drugs' curative properties for the sick type, but for healthy types, the drug distributions overlap a lot.

Finally, I estimate a variation of the baseline model that includes behavioral therapy sessions and a combination of behavioral therapy and (any) ADHD drug (Table 3.9). Consistent with the medical literature, behavioral therapy alone ranks very low for both patient types. However, behavioral therapy combined with an ADHD medicine yields better results: it ranks second by both symptomatic and curative match value for sick type, following immediate-release stimulants. Note that introduction of behavioral therapy changed relative ranking of treatments. Immediate-release stimulants, for example, ranked only third for sick types in the baseline model and in the model with drug holidays. A more precise analysis and comparison will be possible after I estimate the model on a larger sample that includes both behavioral therapy and drug holidays.

The estimates of the model with drug holidays will be used to construct a set of counterfactuals to evaluate the effect of eliminating drug holidays. It will estimate the effect of "forcing" patients to take a drug until they are cured or alternatively, adopting regular drug holidays as a treatment strategy, on the length of treatment and its cost to Medicaid.

3.8 Conclusion

The incidence of diagnosis of ADHD in the U.S. among children increased significantly over the past decade. While many believe that ADHD drugs are overprescribed, very little is known about the existing prescribing practices, physician learning processes, and relative efficacies of various ADHD treatment strategies. This paper sheds some light on it. I estimated two variants of a dynamic model of demand for ADHD drugs under uncertainty. First, I estimate a baseline model that ignores drug holidays and behavioral treatment. The choice set consists of four composite drugs for a random sub-sample of 1,000 patients. Second, I rearrange the drug choice set to add on drug holidays as a fourth option.

At this stage, the following conclusions can be made. First, patients are extremely heterogeneous in the underlying illness severity. Both models suggest that the probability of a child functioning successfully in their everyday life without ADHD treatment differs from very low to very high. Although merely suggestive, it might point at the presence of overdiagnosis and overprescription practices. Second, there is considerable uncertainty regarding patient-drug match by

Table 3.9: Dynamic Model with Behavioral Therapy: Parameter Estimates.

Parameter	**Est.**	**Std. err.**	**Est.**	**Std. err.**
Illness heterogeneity distribution	**Type 1 ("Sick")**		**Type 2 ("Healthy")**	
Recovery probability	0.025		0.808	
Type probability	0.702		0.298	
Means, symptom match values				
$\underline{\mu}_{\text{vyvanse, Period 1}}$	1.483		1.333	
$\underline{\mu}_{\text{vyvanse, Period 2}}$	-0.103		-0.911	
$\underline{\mu}_{\text{extended-release stimulants}}$	0.855		2.723	
$\underline{\mu}_{\text{immediate-release stimulants}}$	2.294		2.864	
$\underline{\mu}_{\text{non-stimulants}}$	0.207		-0.016	
$\underline{\mu}_{\text{B-therapy}}$	0.298		1.988	
$\underline{\mu}_{\text{B-therapy+drug}}$	0.920		1.823	
Means, curative match values				
$\underline{\nu}_{\text{vyvanse, Period 1}}$	1.864		0.276	
$\underline{\nu}_{\text{vyvanse, Period 2}}$	1.390		1.745	
$\underline{\nu}_{\text{extended-release stimulants}}$	2.197		1.874	
$\underline{\nu}_{\text{immediate-release stimulants}}$	4.616		-1.729	
$\underline{\nu}_{\text{non-stimulants}}$	1.574		0.237	
$\underline{\nu}_{\text{B-therapy}}$	0.903		-2.962	
$\underline{\nu}_{\text{B-therapy+drug}}$	3.061		0.358	
Std. dev., symptom match values				
$\underline{\sigma}_{\text{vyvanse}}$	1.086			
$\underline{\sigma}_{\text{extended-release stimulants}}$	1.223			
$\underline{\sigma}_{\text{immediate-release stimulants}}$	1.876			
$\underline{\sigma}_{\text{non-stimulants}}$	0.827			
$\underline{\sigma}_{\text{B-therapy}}$	0.626			
$\underline{\sigma}_{\text{B-therapy+drug}}$	1.267			
Std. dev., symptom signals				
σ_{vyvanse}	1.173			
$\sigma_{\text{extended-release stimulants}}$	1.302			
$\sigma_{\text{immediate-release stimulants}}$	0.979			
$\sigma_{\text{non-stimulants}}$	1.114			
$\sigma_{\text{B-therapy}}$	1.141			
$\sigma_{\text{B-therapy+drug}}$	1.367			
Std. dev., curative match values				
$\underline{\tau}$	0.985			
Std. dev., curative signals				
τ	1.138			
Price coefficient, α	1.143			
Risk-aversion parameter, r	1.108			
Discount rate, β	0.950	(fixed)		
N patients	1,000			
N time periods	24			
N observations				
N draws	10			

Notes: Drug prices are averaged across time and patients. Vyvanse is a new drug and its mean match values vary in time. Time periods are months.

both symptomatic and curative properties. Although some drugs are better than others for each of the patient types, their match value distributions overlap significantly. In other words, knowing a patient's type, does not resolve patient-drug match uncertainty.

Although the model with drug holidays yields overall similar results to the baseline model, in their current formulation they cannot be directly compared because of the differences in the choice set. Notably, drug holidays rank first for the healthy type by symptomatic relief properties.

Additionally, I will estimate a model that includes behavioral treatments that might influence drug choice as well as the length of therapy and the probability of a child being able to function without ADHD treatment. The model with drug holidays and behavioral therapy will be used to construct a set of counterfactuals to evaluate the effect of eliminating or alternatively, "forcing" drug holidays and/or behavioral treatment on the overall length of treatment and its cost to Medicaid.

Appendices

Appendix A Methodological Appendix for Chapter 1

A growing string of health economics literature estimates prescription drug demand using the learning model framework. Newly diagnosed patients face uncertainty regarding average drugs' efficacy (Ching, 2010), regarding best-match for each particular individual (Crawford&Shum, 2005; Dickstein, 2011; Saxell, 2013), or both (Chintagunta et al., 2009). These studies (Crawford&Shum, 2005; Dickstein, 2011; Saxell, 2013) focus on modeling individual choices every time period, using Bayesian updating in formulating the process of information acquisition. The resulting estimates of these models are moments of drug-specific distribution of match values that can be used to determine optimal treatment.

Each of these models uses individual-level claims data. One of the limitations of claims data is the scarcity of outcome variables. In contrast to survey data (Chintagunta et al., 2009), there is no information on patient satisfaction with treatment, nor self-reported health measures. Common approach in this case is to define successful treatment based on patient exit from treatment ("cure") or adherence to treatment (a number of months treated without interruption). Because of data limitations, these studies rely on a rather strong assumption of patient being cured after he or she had a medicine-free period in treatment for 90 days (Dickstein, 2011) or 180 days (Crawford&Shum, 2005; Saxell, 2013). However, the conditions they look at are chronic: gastrointestinal conditions (Crawford&Shum, 2005), cardiovascular diseases (Saxell, 2013), and depression (Dickstein, 2011) that technically cannot be cured and are likely to progress in cycles with disease manifestation episodes. Dickstein's argument is more elaborate than Crawford&Shum and Saxell: successful treatment should last for at least 6 months (adherence) and if the patient restarts her treatment after not taking prescriptions for over 90 days, it is considered a new treatment episode, separate observation.

My sample spans 10 years and allows for a much more accurate analysis than it has been done previously in the literature. In order to make an assumption about how to determine the end of treatment episode, I look at patients who have been observed for at least 1, 2, or 3 years after their last ADHD claim.[10] These time periods are listed in panels A, B, and C of Table 10 respectively. The last ADHD treatment is defined flexibly because of the lack of medical evidence. I use a range of criteria, from 90 days to 1.5 years, 3-month spaced to help guide my assumption on the duration

[10]Note that I distinguish last treatment from last claim. The former includes prescription or behavioral therapy visit, and the latter also includes any doctor visit with ADHD diagnosis.

of treatment.

As Table 10 shows, for ADHD patients it is fairly common to have a medication-free time period of at least 3 months. Over 40% of patients restart treatment within 90 days from their last ADHD treatment. Indeed, my data suggests that many children and their parents tend to suspend ADHD treatment during summer holidays.

The data suggest that a more reasonable assumption would be to take a year of medicine- and diagnosis-free period as a criterion for defining the end of episode, or "cure". If a child after her treatment went without it for a year, it means that this treatment was successful in managing her condition and she didn't need the treatment any longer. Note that this is still not perfect, because in 2 years that this sub-sample of patients is observed in the data, still 11% of children would restart their treatment after a year and 29% will do so by the end of the second year since last treatment. However, in one year, the student and her parents have time to fully evaluate short- and long-term evidence of their child being able to function successfully in life. Even if school performance is not being monitored regularly by parents, they would certainly notice poor outcome of no-treatment if a child has to repeat a grade.

This assumption is also supported by medical literature. For example, for ADHD Chen et al. (2009) use 12-month continuous Medicaid eligibility with no ADHD claims to identify first treatment episode. In terms of treatment duration, medical profession has not come up with conclusive evidence yet. Van de Loo-Neus et al.(2011) review 53 clinical studies with duration of treatment longer than 12 weeks that were published between 1990 and 2010 for no fewer than 20 subjects, who are between 6-18 years old. Out of 15 most relevant studies, only 10 span a year and just 5 of them followed patients for 36+ months. They conclude that "clinical decisions about starting, continuing, and stopping ADHD medication have to be made on an individual basis". Decision to stop taking medication is typically not a result of intolerable side effects that can be dealt with by an adjustment of dose. Based on all reviewed work, they can only say that a substantial subsample of children with ADHD continue benefitting from long-term medical treatment in terms of ADHD symptom control, while other children with ADHD fail to show beneficial effects of medication after 1 or 2 years. They suggest to implement yearly medication-free periods for "several days to one week or longer" to check the need for medication.

Table 10: Testing assumption of cure

		PANEL B										PANEL C											
		Time after last ADHD: 2 years (720 days)										Time after last ADHD: 3 years (1080 days)											
Days since last ADHD claim		No-ADHD spell, days										No-ADHD spell, days											
		90		180		270		360		450		540		90		180		270		360		450		540	
		6,490		5,210		4,651		4,248		3,943		3,716		4,595		3,652		3,230		2,909		2,655		2,476	
	0-90	H		H		H		H		H		H		H		H		H		H		H		H	
	90-180	2,751	42%	H		H		H		H		H		1,958	43%	H		H		H		H		H	
	180-270	709	11%	1,118	21%	H		H		H		H		533	12%	825	23%	H		H		H		H	
Restarted	270-360	424	7%	608	12%	711	15%	H		H		H		312	7%	447	12%	525	16%	H		H		H	
within xx	360-450	261	4%	347	7%	423	9%	477	11%	H		H		187	4%	252	7%	315	10%	357	12%	H		H	
days after	450-540	190	3%	249	5%	288	6%	315	7%	339	9%	H		142	3%	189	5%	219	7%	235	8%	251	9%	H	
the last	540-630	148	2%	199	4%	233	5%	247	6%	252	6%	260	7%	103	2%	144	4%	166	5%	177	6%	181	7%	187	8%
ADHD	630-720	122	2%	163	3%	177	4%	189	4%	198	5%	200	5%	91	2%	120	3%	129	4%	139	5%	144	5%	146	6%
claim	720-810													87	2%	103	3%	110	3%	114	4%	118	4%	122	5%
	810-900													53	1%	70	2%	78	2%	88	3%	91	3%	96	4%
	900-990													51	1%	70	2%	80	2%	84	3%	88	3%	88	4%
	990-1080													49	1%	59	2%	72	2%	77	3%	78	3%	80	3%
N restarted, in sample		**4,605**	**71%**	**2,684**	**52%**	**1,832**	**39%**	**1,228**	**29%**	**789**	**20%**	**460**	**12%**	**3,566**	**78%**	**2,279**	**62%**	**1,694**	**52%**	**1,271**	**44%**	**951**	**36%**	**719**	**29%**

Notes: "H" stands for drug holiday.

Appendix B Additional Tables for Chapter 3

B.1 Part D Plan-level Summary Statistics, 2006-2012.

	2006	2007	2008	2009	2010	2011	2012
Outcome variables							
Premium	37.36	36.69	40.31	45.81	46.17	53.62	53.41
	(12.82)	(15.08)	(20.02)	(20.70)	(19.13)	(25.27)	(26.72)
Out-of-pocket drug cost	45.12	46.99	53.36	58.02	71.25	77.05	87.48
	(12.38)	(12.91)	(12.31)	(10.77)	(9.84)	(9.42)	(13.06)
N of NDCs covered*	14,688	4,791	4,117	4,014	3,401	3,359	3,441
	(13,682)	(1,484)	(1,064)	(965)	(675)	(613)	(585)
N of top100 drugs covered	91.58	93.07	90.10	87.84	82.63	78.04	74.52
	(5.92)	(5.96)	(7.67)	(9.14)	(7.43)	(6.96)	(7.38)
Controls							
Deductible	92.51	93.57	103.73	110.02	144.18	153.50	153.40
	(115.84)	(121.81)	(128.40)	(136.56)	(135.57)	(141.97)	(152.51)
Mean tier, all drugs	0.22	0.27	0.30	0.32	0.30	0.33	0.37
	(0.13)	(0.08)	(0.08)	(0.10)	(0.08)	(0.10)	(0.10)
Mean tier, top100 drugs	0.20	0.26	0.26	0.27	0.22	0.23	0.25
	(0.14)	(0.09)	(0.09)	(0.12)	(0.08)	(0.09)	(0.10)
Mean restriction (0-3), all	0.16	0.20	0.28	0.32	0.33	0.33	0.36
	(0.19)	(0.10)	(0.13)	(0.12)	(0.12)	(0.11)	(0.13)
Mean restriction (0-3), top	0.17	0.23	0.28	0.30	0.29	0.26	0.28
	(0.19)	(0.16)	(0.17)	(0.15)	(0.12)	(0.11)	(0.12)
Mean restriction (1-3), all	1.07	1.11	1.15	1.15	1.15	1.14	1.14
	(0.05)	(0.07)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)
Mean restriction (1-3), top	1.07	1.13	1.15	1.12	1.11	1.10	1.08
	(0.10)	(0.13)	(0.14)	(0.10)	(0.12)	(0.12)	(0.08)
% of plans w/gap coverage	0.31	0.29	0.30	0.25	0.19	0.35	0.24
% of basic plans	0.58	0.52	0.48	0.45	0.51	0.57	0.52
% of benchmark plans	0.28	0.60	0.24	0.17	0.19	0.30	0.30
% of renewal plans		0.58	0.72	0.82	0.74	0.46	0.69
% of consolidated plans		0.12	0.14	0.10	0.19	0.46	0.19
% of new plans		0.30	0.10	0.03	0.06	0.01	0.09
N of observations	1,446	1,908	1,778	1,626	1,493	1,034	995

Notes: The unit of observation is a plan. All stand-alone Part D plans are included. Out-of-pocket cost of top 100 drugs assigns a 1/100 weight to each drug. In 2006, requirements on formulary listing of NDCs differ from the requirements in 2007-2012. Gap coverage and deductible standards for Part D plans were altered through 2006-2012 as described in detail in the paper. All prices are in nominal terms. Standard deviations are in parentheses.

B.2 Control and Comparison Groups, 2006-2012.

	Plans affected by M&A		Plans unaffected by M&A	
	Before	**After**	**Before**	**After**
Premium	40.27	44.81	42.54	45.16
	(16.83)	(19.51)	(19.94)	(22.03)
Out-of-pocket cost of top100 drugs	57.90	63.47	56.71	63.45
	(16.65)	(18.03)	(15.78)	(17.59)
N of NDCs covered	3,983	3,847	4,036	3,712
	(1,143)	(960)	(1,190)	(900)
N of top 100 drugs covered	88.22	86.31	88.06	85.32
	(10.54)	(11.15)	(8.35)	(9.49)
Deductible	121.46	117.78	112.52	118.14
	(139.07)	(145.21)	(130.37)	(135.06)
Mean tier, all drugs	0.30	0.32	0.29	0.31
	(0.09)	(0.08)	(0.10)	(0.09)
Mean tier, top100 drugs	0.24	0.25	0.24	0.25
	(0.13)	(0.13)	(0.10)	(0.09)
Mean restriction (0-3), all	0.23	0.29	0.27	0.30
	(0.15)	(0.14)	(0.14)	(0.13)
Mean restriction (0-3), top100	0.19	0.23	0.27	0.28
	(0.14)	(0.13)	(0.16)	(0.14)
Mean restriction (1-3), all	1.11	1.14	1.13	1.15
	(0.07)	(0.06)	(0.08)	(0.09)
Mean restriction (1-3), top100	1.13	1.10	1.12	1.13
	(0.14)	(0.06)	(0.12)	(0.13)
Plan market share	0.014	0.019	0.008	0.009
	(0.024)	(0.033)	(0.016)	(0.017)
Enrollment	15,825	22,940	9,583	11,562
	(33,560)	(47,206)	(23,577)	(25,447)
LIS enrollment	8,681	12,167	4,436	5,276
	(18,393)	(25,319)	(13,244)	(14,171)
% of plans with gap coverage	0.26	0.24	0.27	0.26
% of basic plans	0.59	0.53	0.50	0.51
% of benchmark plans	0.33	0.41	0.31	0.29
% of renewal plans		0.76		0.67
% of consolidated plans		0.21		0.18
% of new plans		0.02		0.12
% of terminated plans		0.02		0.03
N of observations		1,379		7,598

Notes: The unit of observation is a plan. Only renewal and consolidated renewal stand-alone Part D plans are included. Out-of-pocket cost of top 100 drugs assigns a 1/100 weight to each drug. Since the requirements on formulary listing of NDCs differ from the requirements in 2007-2012, the data on NDC coverage in 2006-2007 are excluded. Standard deviations are in parentheses.

B.3 Comparative Summary Statistics for Non-renewed Plans, 2006-2012.

	2006-2007		2007-2008		2008-2009		2009-2010		2010-2011		2011-2012	
	ALL	**T**	**ALL**	**T**	**ALL**	**T**	**ALL**	**T**	**ALL**	**T**	**ALL**	**T**
Monthly premium	37.36	66.44	36.69	39.81	40.31	55.38	45.81	65.73	46.17	59.21	53.62	49.92
	(12.82)	(33.32)	(15.08)	(8.22)	(20.02)	(19.32)	(20.70)	(36.78)	(19.13)	(20.30)	(25.27)	(11.25)
Deductible	92.51	83.33	93.57	113.15	103.54	73.85	110.02	110.63	144.18	49.78	153.50	129.10
	(115.84)	(144.34)	(121.81)	(131.82)	(128.35)	(90.05)	(136.56)	(147.50)	(135.57)	(109.69)	(141.97)	(91.29)
Plan enrollment	10,730	267	8,473	122	8,573	310	9,415	1,514	10,594	3,263	16,201	568
	(25,159)	(443)	(23,066)	(487)	(21,155)	(750)	(21,912)	(3,058)	(24,187)	(14,307)	(37,194)	(1,123)
LIS enrollment	5,588	58	4,196	28	4,051	143	4,377	849	5,042	2,941	7,699	355
	(13,368)	(92)	(13,820)	(119)	(11,104)	(636)	(12,387)	(2,632)	(14,401)	(13,432)	(20,340)	(1,123)
Plan market share	0.009	0.00009	0.007	0.0001	0.007	0.0003	0.008	0.002	0.008	0.002	0.012	0.0006
	(0.018)	(0.0001)	(0.016)	(0.0004)	(0.015)	(0.0009)	(0.015)	(0.003)	(0.016)	(0.004)	(0.024)	(0.001)
% basic plans	0.58	0.34	0.52	0.83	0.48	0.10	0.45	0.44	0.51	0.49	0.57	0.94
% benchmark plans	0.28	0.00	0.60	0.25	0.24	0.05	0.17	0.06	0.19	0.12	0.30	0.06
% plans w/gap cover	0.31	0.67	0.29	0.15	0.30	0.90	0.25	0.38	0.19	0.51	0.35	0.06
N plans	**1,446**	**3/0**	**1,908**	**89/2**	**1,776**	**87/0**	**1,627**	**16/0**	**1,493**	**104/2**	**1,034**	**33/27**

Notes: The table compares plan characteristics of terminated plans to the all-plan average. For example, for 2006-2007 all plans offered in 2006 are compared to the plans terminated in the end of 2006. "T" stands for terminated plans. Number of plans in "T" panels reports the total number of terminated plans/number plans terminated by merging parties. Standard errors are in parentheses.

B.4 Difference-in-Difference Estimates: Premiums.

	A		B		C	
	(1)	(2)	(1)	(2)	(1)	(2)
Merger-affected plan	1.703	3.607			2.241	3.840
	(0.363)	(2.219)			(0.400)	(2.494)
Consolidated plan			-4.221	-3.861	-3.911	-3.422
			(0.320)	(1.339)	(0.343)	(1.547)
Consolidated x Merger plan					-2.199	-2.105
					(0.827)	(2.127)
Covariates in 1st differences						
Price index	-0.189	-0.196	-0.186	-0.188	-0.177	-0.190
	(0.019)	(0.079)	(0.018)	(0.085)	(0.019)	(0.083)
Deductible	-0.023	-0.021	-0.026	-0.024	-0.026	-0.024
	(0.002)	(0.008)	(0.002)	(0.008)	(0.002)	(0.008)
Gap coverage	8.879	8.819	8.660	8.774	8.773	8.780
	(0.363)	(1.906)	(0.360)	(1.821)	(0.360)	(1.869)
LIS eligibility	-6.666	-6.557	-6.220	-6.085	-6.280	-6.224
	(0.290)	(0.852)	(0.288)	(0.823)	(0.289)	(0.846)
Benefit type	-2.645	-3.089	-1.330	-1.778	-1.235	-1.834
	(0.388)	(1.391)	(0.398)	(1.399)	(0.398)	(1.404)
Top100 drugs covariates						
N of covered drugs	0.025	-0.025	0.057	0.011	0.061	0.004
	(0.034)	(0.183)	(0.034)	(0.192)	(0.034)	(0.186)
Mean tier	0.236	-0.333	-0.069	-1.149	0.183	-0.435
	(2.254)	(14.373)	(2.235)	(14.575)	(2.233)	(14.420)
Mean number of restrictions	1.571	1.656	-2.003	0.593	-1.050	-0.388
	(2.675)	(13.715)	(2.657)	(14.532)	(2.658)	(14.028)
All drugs covariates						
N of covered drugs, per 100	-0.019	-0.019	-0.024	-0.024	-0.021	-0.021
	(0.003)	(0.008)	(0.002)	(0.007)	(0.003)	(0.008)
Mean tier	-5.334	-3.539	-7.181	-5.563	-6.938	-5.557
	(2.721)	(14.751)	(2.698)	(15.310)	(2.698)	(14.935)
Mean number of restrictions	0.349	-4.015	5.196	-1.406	2.770	-2.163
	(3.042)	(14.052)	(2.999)	(14.690)	(3.025)	(14.288)
Covariates in levels						
Lagged enrollment, in ('000)	-0.048	-0.023	-0.043	-0.022	-0.044	-0.021
	(0.006)	(0.014)	(0.006)	(0.013)	(0.006)	(0.013)
Lagged log mkt share, in fractions	0.834	0.291	0.758	0.245	0.759	0.236
	(0.077)	(0.284)	(0.076)	(0.313)	(0.077)	(0.306)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	8,839		F-test		29.7	0.6
N of M&A affected plans	1,375					
N of consolidated plans	1,994					
N of M&A consolidated plans	296					

Notes: Panel A shows estimates for the plans involved in a merger; this specification does not distinguish between mergers that consolidated plans and mergers that didn't. Panel B shows estimates for the plan consolidation effect on premiums. Panel C includes the merger-consolidated plan interaction term. The F-test null hypothesis is that the sum of the coefficients on merger dummy, consolidation dummy and their interaction term is zero. Standard errors are in parentheses, clustered by pre-merger insurer for specification with pre-merger insurer fixed effects.

B.5 Difference-in-Difference Estimates: Formulary, Top 100 Drugs.

	A		B		C	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Merger-affected plan	0.391	-0.146			-0.492	-1.081
	(0.172)	(1.872)			(0.189)	(2.025)
Consolidated plan			-0.196	-0.176	-0.866	-0.880
			(0.155)	(0.922)	(0.165)	(0.940)
Consolidated x Merger plan					4.357	4.459
					(0.396)	(2.244)
Covariates in 1st differences						
Deductible	-0.005	-0.005	-0.005	-0.005	-0.004	-0.004
	(0.001)	(0.005)	(0.001)	(0.005)	(0.001)	(0.005)
Gap coverage	0.966	1.270	0.958	1.258	0.921	1.208
	(0.170)	(1.191)	(0.171)	(1.195)	(0.169)	(1.187)
LIS eligibility	0.450	0.647	0.483	0.656	0.393	0.561
	(0.138)	(0.346)	(0.139)	(0.323)	(0.138)	(0.326)
Benefit type	1.439	1.609	1.489	1.662	1.459	1.619
	(0.186)	(1.109)	(0.192)	(1.078)	(0.191)	(1.009)
Covariates in levels						
Lagged enrollment, in ('000)	-0.013	-0.007	-0.012	-0.007	-0.013	-0.006
	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)
Lagged log mkt share, in fractions	0.084	0.088	0.086	0.082	0.035	0.017
	(0.037)	(0.129)	(0.037)	(0.132)	(0.037)	(0.130)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	8,839		F-test		77.4	1.48
N of M&A affected plans	1,375					
N of consolidated plans	1,994					
N of M&A consolidated plans	296					

Notes: Dependent variable is the change in the number of drugs ranked in top100 by prescriptions filled, in the formulary. Standard errors are in parentheses, clustered by insurer for specification with insurer fixed effects.

B.6 Difference-in-Difference Estimates: Formulary, All Drugs.

	A		B		C	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
Merger-affected plan	43.555	-182.801			-47.084	-320.229
	(25.834)	(338.649)			(29.148)	(354.326)
Consolidated plan			16.570	30.604	-45.124	-62.340
			(22.582)	(109.959)	(24.292)	(123.18)
Consolidated x Merger plan					373.068	552.925
					(56.411)	(221.745)
Covariates in 1st differences						
Deductible	0.014	-0.143	0.009	-0.084	0.081	-0.017
	(0.118)	(0.991)	(0.118)	(0.954)	(0.119)	(0.953)
Gap coverage	558.694	628.355	556.604	644.435	542.380	607.703
	(30.792)	(158.959)	(30.795)	(181.823)	(30.887)	(144.682)
LIS eligibility	123.063	158.140	124.321	149.531	108.372	136.986
	(21.228)	(82.448)	(21.222)	(74.715)	(21.295)	(75.545)
Benefit type	340.647	428.907	333.272	417.181	333.451	416.994
	(29.097)	(188.929)	(30.509)	(189.715)	(30.429)	(185.607)
Covariates in levels						
Lagged enrollment, in ('000)	-1.101	0.674	-1.085	0.733	-1.150	0.872
	(0.418)	(0.930)	(0.418)	(0.970)	(0.417)	(1.062)
Lagged log mkt share, in fractions	-17.222	4.909	-16.649	2.633	-20.676	-2.332
	(5.378)	(16.24)	(5.371)	(18.206)	(5.390)	(16.356)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	7,396		F-test		34.9	0.2
N of M&A affected plans	1,082					
N of consolidated plans	1,746					
N of M&A consolidated plans	276					

Notes: Dependent variable is the change in the number of drugs included into the formulary. 2006-2007 year-plan pairs are excluded. Standard errors are in parentheses, clustered by pre-merger insurer for specification with insurer fixed effects.

B.7 Difference-in-Difference Estimates: Price Index.

	A		B		C	
	(1)	(2)	(1)	(2)	(1)	(2)
Merger-affected plan	-0.424	1.755			0.076	2.441
	(0.311)	(2.240)			(0.344)	(2.033)
Consolidated plan			1.706	0.908	2.132	1.440
			(0.280)	(1.152)	(0.300)	(1.299)
Consolidated x Merger plan					-2.723	-3.070
					(0.722)	(3.311)
Covariates in 1st differences						
Benefit type	-2.456	-3.212	-2.978	-3.468	-2.967	-3.482
	(0.337)	(2.072)	(0.348)	(1.939)	(0.348)	(1.889)
LIS eligibility	0.609	0.087	0.440	0.078	0.504	0.079
	(0.251)	(1.237)	(0.252)	(1.213)	(0.252)	(1.237)
Deductible	0.0002	0.003	0.001	0.003	0.001	0.003
	(0.001)	(0.009)	(0.001)	(0.009)	(0.001)	(0.009)
Gap coverage	-0.176	-1.385	-0.114	-1.282	-0.092	-1.317
	(0.309)	(2.188)	(0.309)	(2.167)	(0.309)	(2.183)
Covariates in levels						
Lagged enrollment, in ('000)	0.015	-0.001	0.014	-0.003	0.014	-0.002
	(0.005)	(0.010)	(0.005)	(0.009)	(0.005)	(0.009)
Lagged log mkt share, in fractions	-0.121	-0.103	-0.095	-0.063	-0.059	-0.037
	(0.067)	(0.263)	(0.067)	(0.260)	(0.067)	(0.236)
Year & Region F.E.	Y	Y	Y	Y	Y	Y
Insurer F.E.		Y		Y		Y
N of year-pairs	8,839		F-test		0.7	0.98
N of M&A affected plans	1,375					
N of consolidated plans	1,994					
N of M&A consolidated plans	296					

Notes: Dependent variable is the change in the weighted price of the basket of top100 drugs under each plan. Standard errors are in parentheses, clustered by pre-merger insurer for specification with insurer fixed effects.

Bibliography

J. Abaluck and J. Gruber. Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *American Economic Review*, 101(4):1180–1210, 2011.

Daniel A Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination*. *International Economic Review*, 44(3):1007–1040, 2003.

J. Angrist and J. Pischke. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *The Journal of Economic Perspectives*, 24(2):3–30, 2010.

L Eugene Arnold. Sex differences in adhd: conference summary. *Journal of abnormal child psychology*, 24(5):555–569, 1996.

Russell A Barkley. *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. Guilford Press, 2006.

L. Bates, J. Hilliard, and R. Santerre. Do Health Insurers Possess Market Power? *Southern Economic Journal*, 78(4):1289–1304, 2012.

Ernst R Berndt, Linda T Bui, David H Lucking-Reiley, and Glen L Urban. The roles of marketing, product quality, and price competition in the growth and composition of the us antiulcer drug industry. In *The economics of new goods*, pages 277–328. University of Chicago Press, 1996.

M. Bertrand, E. Duflo, and S. Mullainathan. How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1):249–275, 2004.

Alok Bhargava. Identification and panel data models with endogenous regressors. *The Review of Economic Studies*, 58(1):129–140, 1991.

Ellen Bouchery, Rick Harwood, Rosalie Malsberger, Emily Caffery, Jessica Nysenbaum, and Kerianne Hourihan. Developing medicare and medicaid substance abuse treatment spending estimates. Technical report, Mathematica Policy Research, 2012.

Pedro Carneiro, Claire Crawford, and Alissa Goodman. The impact of early cognitive and non-cognitive skills on later outcomes. 2007.

A. Chorniy and L. Kitashima. The effect of adhd medications on health and behavioral outcomes. *Working Paper*, 2014a.

A. Chorniy and L. Kitashima. Effect of adhd treatment on educational outcomes. *Working Paper*, 2014b.

Gabriella Conti and James J Heckman. Understanding the early origins of the education–health gradient a framework that can also be applied to analyze gene–environment interactions. *Perspectives on Psychological Science*, 5(5):585–605, 2010.

Gabriella Conti and James J Heckman. *Economics of Child Well-Being*. Springer, 2014.

Gregory S Crawford and Matthew Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, 2005.

Flavio Cunha and James Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, 2007.

Flavio Cunha, James J Heckman, and Susanne M Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Janet Currie, Mark Stabile, and Lauren Jones. Do stimulant medications improve educational and behavioral outcomes for children with adhd? *Journal of Health Economics*, 37:58 – 69, 2014.

L. Dafny. Are Health Insurance Markets Competitive? *American Economic Review*, 100(4):1399–1431, 2010.

L. Dafny, M. Duggan, and S. Ramanarayanan. Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry. *American Economic Review*, 102(2):1161–85, 2012.

Soren Dalsgaard, Helena Skyt Nielsen, and Marianne Simonsen. Consequences of {ADHD} medication use for children's outcomes. *Journal of Health Economics*, 37:137 – 151, 2014.

F. Decarolis. Pricing and Incentives in Publicly Subsidized Health Care Markets: the Case of Medicare Part D. *working paper, Penn Institute for Economic Research*, 2012.

Michael Dickstein. Efficient provision of experience goods: Evidence from antidepressant choice. *Manuscript*, 2014a.

Michael Dickstein. Physician vs. patient incentives in prescription drug choice. *Manuscript*, 2014b.

M. Duggan and F. Scott-Morton. The Effect of Medicare Part D on Pharmaceutical Prices and Utilization. *The American Economic Review*, 100(1):590–607, 2010.

R. Ellis. Five Questions for Health Economists. *International Journal of Health Care Finance and Economics*, pages 1–17, 2012.

Sara Fisher Ellison, Iain Cockburn, Zvi Griliches, and Jerry Hausman. Characteristics of demand for pharmaceutical products: an examination of four cephalosporins. *The Rand journal of economics*, pages 426–446, 1997.

Tülin Erdem and Michael P Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, 15(1):1–20, 1996.

K. Ericson. Consumer Inertia and Firm Pricing in the Medicare Part D Prescription Drug Insurance Exchange. *American Economic Journal: Economic Policy*, 6(1):38–64, 2014.

C. Fee and S. Thomas. Sources of gains in horizontal mergers: evidence from customer, supplier, and rival firms. *Journal of Financial Economics*, 74(3):423–460, 2004.

Lawrence B Finer and Mia R Zolna. Unintended pregnancy in the united states: incidence and disparities, 2006. *Contraception*, 84(5):478–485, 2011.

Betty Tao Fout and Donna B Gilleskie. Does health insurance encourage or crowd out beneficial non-medical care? a dynamic analysis of insurance, health inputs, and health production. *Manuscript*, 2014.

R. Frank and J. Newhouse. Should drug prices be negotiated under Part D of Medicare? And if so, how? *Health Affairs*, 27(1):33–43, 2008.

Larry S Goldman, Myron Genel, Rebecca J Bezman, Priscilla J Slanetz, et al. Diagnosis and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Jama*, 279(14): 1100–1107, 1998.

Alissa Goodman, Robert Joyce, and James P Smith. The long shadow cast by childhood physical and mental problems on adult life. *Proceedings of the National Academy of Sciences*, 108(15): 6032–6037, 2011.

G. Gowrisankaran, A. Nevo, and R. Town. Mergers When Prices are Negotiated: Evidence from the Hospital Industry. *National Bureau of Economic Research*, 2013.

Brady E. Hamilton, Joyce A. Martin, Michelle J.K. Osterman, and Sally C. Curtin. Births: Preliminary data for 2013. *Centers for Disease Control and Prevention National Vital Statistics Reports*, 63(2), 2014.

J. Hastings. Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California. *American Economic Review*, 94(1):317–328, 2004.

James J. Heckman, Jora Stixrud, and Sergio Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482, 2006.

F. Heiss, A. Leive, D. McFadden, and J. Winter. Plan selection in Medicare Part D: Evidence from administrative data. *Journal of Health Economics*, 32(6):1325–1344, 2013.

Judith K Hellerstein. The importance of the physician in the generic versus trade-name prescription decision. *The Rand journal of economics*, pages 108–136, 1998.

K. Ho. Insurer-provider Networks in the Medical Care Market. *The American Economic Review*, 99(1):393–430, 2009.

K. Ho and R. Lee. Insurer competition and negotiated hospital prices. *National Bureau of Economic Research*, 2013.

P. Karaca-Mandic, J. Abraham, and C. Phelps. How Do Health Insurance Loading Fees Vary by Group Size?: Implications for Healthcare Reform. *International Journal of Health Care Finance and Economics*, 11(3):181–207, 2011.

Michael P Keane and Kenneth I Wolpin. The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics*, pages 648–672, 1994.

Ronald C Kessler, Patricia A Berglund, Cindy L Foster, William B Saunders, Paul E Stang, and Ellen E Walters. Social consequences of psychiatric disorders, ii: Teenage parenthood. *American Journal of Psychiatry*, 154(10):1405–1411, 1997.

J. Ketcham, C. Lucarelli, E. Miravete, and M. Roebuck. Sinking, Swimming, or Learning to Swim in Medicare Part D. *Working Paper*, 2011.

J. Kling, S. Mullainathan, E. Shafir, L. Vermeulen, and M. Wrobel. Comparison Friction: Experimental Evidence From Medicare Drug Plans. *Quarterly Journal of Economics*, 127(1):199–235, 2012.

Pat Levitt. Structural and functional maturation of the developing primate brain. *The Journal of pediatrics*, 143(4):35–45, 2003.

M. Lewis and K. Pflum. Diagnosing Hospital System Bargaining Power in Managed Care Networks. *unpublished manuscript, Ohio State University*, 2011.

Alison Looby. Childhood attention deficit hyperactivity disorder and the development of substance use disorders: valid concern or exaggeration? *Addictive behaviors*, 33(3):451–463, 2008.

C. Lucarelli, J. Prince, and K. Simon. The Welfare Impact of Reducing Choice in Medicare Part D: A Comparison of Two Regulation Strategies. *International Economic Review*, 53(4):1155–1177, 2012.

V. Maksimovic, G. Phillips, and N. Prabhala. Post-merger Restructuring and the Boundaries of the Firm. *Journal of Financial Economics*, 102(2):317–343, 2011.

Steven C Marcus, George J Wan, Huabin F Zhang, and Mark Olfson. Injury among stimulant-treated youth with adhd. *Journal of attention disorders*, 12(1):64–69, 2008.

Kathleen Ries Merikangas, Jian-ping He, Marcy Burstein, Sonja A Swanson, Shelli Avenevoli, Lihong Cui, Corina Benjet, Katholiki Georgiades, and Joel Swendsen. Lifetime prevalence of mental disorders in us adolescents: results from the national comorbidity survey replication–adolescent supplement (ncs-a). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(10): 980–989, 2010.

D. Miller and J. Yeo. Estimating Dynamic Discrete Models of Product Differentiation: An Application to Medicare Part D with Switching Costs. *Working Paper Clemson University, Singapore Management University*, 2012.

D. Miller and J. Yeo. The Consequences of a Public Health Insurance Option: Evidence from Medicare Part D Prescription Drug Markets. *Working Paper Clemson University, Singapore Management University*, 2013.

Ted R Miller and Delia Hendrie. *Substance abuse prevention dollars and cents: A cost-benefit analysis*. US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Prevention, 2009.

W. D. Mosher, J. Jones, and J.C. Abma. Intended and unintended births in the united states: 1982–2010. *National Health Statistics Reports*, 55, 2012.

Thomas A Mroz and Timothy H Savage. The long-term effects of youth unemployment. *Journal of Human Resources*, 41(2):259–293, 2006.

David L Olds. Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science*, 3(3):153–172, 2002.

Monica Payne. High school girls with adhd. 2014. Accessed on May 16^{th}, 2014.

Robert J Resnick. Attention deficit hyperactivity disorder in teens and adults: They don't all outgrow it. *Journal of clinical psychology*, 61(5):529–533, 2005.

Tanja Saxell. Private experience and observational learning in pharmaceutical demand. *Manuscript*, 2013.

A. Starc. Insurer Pricing and Consumer Welfare: Evidence from Medigap. *Working Paper University of Pennsylvania*, 2012.

STD Fact Sheet. *STD trends in the United States*. 2013.

Andrine Swensen, Howard G Birnbaum, Rym Ben Hamadi, Paul Greenberg, Pierre-Yves Cremieux, and Kristina Secnik. Incidence and costs of accidents among attention-deficit/hyperactivity disorder patients. *Journal of Adolescent Health*, 35(4):346–e1, 2004.

R. Town and M. Park. Market structure beliefs and hospital merger waves. *Working Paper, University of California, Berkeley*, 2011.

M. Weinberg and D. Hosken. Evidence on the accuracy of merger simulations. *Review of Economics and Statistics*, 95(5):1584–1600, 2013.

Gabrielle Weiss and Lily Trokenberg Hechtman. *Hyperactive children grown up: ADHD in children, adolescents, and adults* . Guilford Press, 1993.

Timothy E. Wilens, Stephen V. Faraone, Joseph Biederman, and Samantha Gunawardene. Does stimulant therapy of attention-deficit/hyperactivity disorder beget later substance abuse? a meta-analytic review of the literature. *Pediatrics*, 111(1):179, 2003.

Zhou Yang, Donna B Gilleskie, and Edward C Norton. Health insurance, medical care, and health outcomes a model of elderly health dynamics. *Journal of Human Resources*, 44(1):47–114, 2009.