ECONOMETRIC ANALYSES OF HEALTH CARE MARKETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Joseph Robert LeCates III

May 2012

ECONOMETRIC ANALYSES OF HEALTH CARE MARKETS

Joseph Robert LeCates III, Ph.D.

Cornell University 2012

This dissertation uses novel econometric methods to correct for biases stemming from imperfect information in health care markets. Because of the nature of health and the endogeneity of treatment choices, modeling the demand for health and outcomes in the health care market can be difficult, and special econometric approaches are needed to identify causal effects.

The first essay extends empirical models of consumer behavior to allow individuals to learn across bundled goods, and applies the framework to physician treatment choice. Such perception spillovers can affect the adoption rate and steady-state market shares of goods, particularly for market entrants which are strong complements to available products. I develop and estimate a Bayesian model of learning which allows the quality signal arising from one particular bundle to be correlated with physician beliefs about the quality of all other potential bundles in the context of oncology. Using cancer surveillance data with patient and physician characteristics, I find that physicians do learn across bundles, though the magnitudes of between-regimen correlations in quality vary substantially.

The second essay investigates how the atypically heaped smoking cessation rate computed from retrospective survey questions biases coefficient estimates in discrete time hazard models. If this characteristic of the data is ignored, estimates of the effect of price on smoking cessation decisions can be attenuated up to sixty-five percent. This paper uses Monte Carlo simulation to compare an adjusted likelihood approach and several ad hoc approaches to correct the bias and finds that the adjusted likelihood approach and one of the ad hoc approaches perform well. The methods are then applied to the Tobacco Use Supplement to

the Current Population Survey and the results are consistent with the simulation analysis.

Returning to the treatment of cancer, the final essay explores how average costs and outcomes change in the United States as physicians alter chemotherapy treatment patterns over time. Considering chemotherapy for metastatic colorectal cancer and metastatic non-small cell lung cancer, and adjuvant therapy for breast cancer, I find that between 2003 and 2006 quality-adjusted life years of survival increased by 11.2 percent on average, but costs increased by 143.8 percent.

BIOGRAPHICAL SKETCH

Joseph Robert LeCates III was born in Jacksonville, Florida, the son of Joseph Robert LeCates Jr. of Philadelphia, Pennsylvania, and Pamela McCarthy LeCates of Savannah, Georgia. With younger brother Daniel Patrick LeCates, the family settled in McDonough, Georgia, in 1996. Joseph graduated valedictorian from Henry County High School in 2002 and then moved to Athens, Georgia, to attend the University of Georgia. Selected as a Foundation Fellow, the University's support afforded him educational opportunities across the United States, in Taiwan, Cuba, Bolivia, and the Galápagos Islands.

In his first economics course, Joseph developed a fascination for the blend of social science and mathematics. He declared an economics major, with minors in statistics and Spanish, to prepare for graduate school. In coordination with the Honors Program, he further developed a program in the Terry College of Business to complete the requirements for a master's degree in economics concurrently. Professors Angela Fertig, Christopher Cornwell, and David Mustard advised his thesis on the effect of local traffic congestion on the inducement of normal-term births. In May 2006, Joseph graduated the top of his class, First Honor Graduate, *summa cum laude* with Bachelor of Arts and Master of Arts degrees in economics.

Joseph entered doctoral studies in August 2006 at Cornell University in Ithaca, New York. His collaboration with Professors Sean Nicholson and Claudio Lucarelli started the following year and developed his interests in physician decision-making, structural econometrics, and the economics of cancer. He was a teaching assistant to Professor Donald Kenkel in the fall of 2007 and their mutual interest in applied econometrics developed into joint research modeling smoking cessation. Together, Professors Nicholson, Kenkel, and Lucarelli comprised his graduate committee and guided the development of his dissertation research. Joseph graduated with his Doctor of Philosophy degree in economics in May 2012 before moving to Boston, Massachusetts, to begin his career with Analysis Group.

To My Family

ACKNOWLEDGMENTS

This dissertation, and I personally, have benefited incalculably from the excellent aid of my committee, Sean Nicholson, Don Kenkel, and Claudio Lucarelli. They have my ardent thanks for their expert help, unwavering patience, and personal support throughout. Together we have benefited from the gracious financial support of Cornell University's Department of Policy Analysis and Management, the College of Human Ecology, the Graduate School, Mr. Larry Stern, the National Institutes of Health, and Pfizer.

My experience with the faculty, staff, and students of the field of economics and department of Policy Analysis and Management have made graduate school rewarding and memorable. John Cawley and Jenny Gerner have my sincere appreciation, as do Angelica Hammer, Eric Maroney, and Geysa Smiljanic. My fellow PAMmies, past and present, have provided invaluable feedback on my research and the best camaraderie. For their constant encouragement, friendship, and kindness I thank Ed Cobb, Sam Crowell, Andrew Hunter, Sharon Kim, Jeff Larrimore, Laura Larrimore, Karl Niklas, Tom Payne, Emily Rosenzweig, Kevin Roth, and John Yen.

Finally, I give my greatest thanks to my mom, dad, and brother for propelling me to this point in life. Between cheering on my successes and reassuring me through failures, they have provided more love and support than I could ever adequately thank them for.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PERCEPTION SPILLOVERS BETWEEN BUNDLED GOODS: A MODEL OF PHYSICIAN LEARNING FROM CHEMOTHERAPY CHOICE

1.1 Introduction

When new products are launched, or when consumption does not fully reveal the quality of an experience good, consumers face uncertainty about the expected utility value of available products. The natural formation of quality beliefs aids their decision making and provides a framework for individuals to revise their choices in response to new information. Empirical models of learning have shown that information gleaned from experience with a particular good strongly impacts subsequent consumption decisions. A growing literature demonstrates that physicians in particular, acting as agents for their patients, prescribe a specific treatment using previous patients' feedback as a source of information about the value of that therapy.[1] When treatments are combinations of two or more drugs, as is increasingly common in pharmaceutical markets, a physician's experience with any component drugs can also influence her beliefs about the quality of a combination regimen, so that information spills over across treatment options. Such perception spillovers have implications for new treatment adoption, optimal patient care, and product market share whenever physicians evaluate products through joint consumption with other goods and use the quality signals to update their beliefs across all potential bundles.

This research extends empirical models of consumer behavior to allow individuals to learn

[1]The influence of physician experience has been demonstrated by Coscelli and Shum (2004) and Crawford and Shum (2005) with anti-ulcer drugs, Narayanan et al. (2005) for prescription antihistamines, Chintagunta et al. (2009) with Cox-2 inhibitors, Janakiraman et al. (2009) and Dickstein (2011) for antidepressants, Narayanan and Manchanda (2009) in the treatment of erectile dysfunction, Ching and Ishihara (2010) with ACE-inhibitors, and Camacho et al. (2011) for asthma and COPD medications.

about quality across all consumption experiences. Specifically, I develop a dynamic model of physician beliefs wherein patient feedback from a particular treatment regimen is used by the physician to update her beliefs about the quality of all potential treatment regimens, not just the prescribed regimen. Starting with Erdem and Keane (1996) and continuing with recent research, learning models have examined consumers' changing beliefs as they incorporate information from noisy quality signals arising from consuming a particular good.[2] That is, consumers update their beliefs about the true quality of a good from repeated consumption of the good. Existing models, however, allow beliefs to change only for the particular product being consumed, regardless of whether the revealed information might alter beliefs about other products as well. I generalize a model originating with Coscelli and Shum (2004) to allow feedback from one regimen to inform the perceived value of all regimens, and thereby capture learning across choice options.

Numerous studies in the empirical literature of economics and marketing have demonstrated that consumer perceptions respond to market conditions and experiential information across a variety of settings. Erdem and Keane (1996) originally estimated structural models of brand choice to explain endogenous purchase behavior by allowing both consumption experience and advertising to act as sources of information for uncertain brand attributes. Their model of changing consumer beliefs as Bayesian learning has since been used in diffusion models of new product adoption (Ackerberg, 2003; Coscelli and Shum, 2004; Chintagunta et al., 2009), to study the influence of the extent and timing of advertising (Narayanan et al., 2005; Ching and Ishihara, 2010; Narayanan and Manchanda, 2009), empirical evaluations of branding (Erdem, 1998; Erdem and Sun, 2002; Janakiraman et al., 2008), and models of consumers' behavioral inconsistencies (Lovett, 2008; Camacho et al., 2011).

Absent from this literature is an empirical model allowing for broad perception spillovers across horizontally differentiated products. Information transfers across goods has been in-

[2]For an example of recent research, see Osborne (2011).

vestigated as it relates to branding. Erdem and Sun (2002) expanded Erdem (1998) to show that both experience and advertising for a particular brand influences consumer beliefs across products (toothbrushes and toothpaste) of the same brand. The common component, the brand name, is the channel through which information flows. Janakiraman et al. (2009) look for evidence of perception spillovers across competing brands by comparing models of learning in which the quality beliefs of similar and dissimilar products (SSRI and TCA antidepressants, respectively) are used separately as initial beliefs for new entrants (novel SSRI drugs).[3] They conclude that perception spillovers occur only when products are sufficiently similar without including a channel for spillovers between similar and dissimilar products simultaneously. In the context of patient-drug matches, patient responsiveness to one treatment is informative about the effectiveness of other treatments via correlated quality beliefs, as shown specifically within the learning literature by Crawford and Shum (2005) and Dickstein (2011). A correctly specified choice model must therefore allow for informational spillovers between all products concurrently if consumers indeed update beliefs across multiple alternatives from experience with one.

When goods are consumed jointly, the potential to learn about the value of alternative choices is significant because of the shared components. A variety of consumer experiences are characterized by bundled consumption, such as computer hardware and software. This type of learning is increasingly important to physicians because many new treatments, especially for cancer, HIV/AIDS, and cholesterol, are combinations of drugs. Since 1991, several new molecules have been approved by the Food and Drug Administration (FDA) for the chemotherapeutic treatment of colorectal cancer, most as part of combination therapy. Today nearly all new cases are treated with a multi-drug regimen. The intriguing adoption patterns of early entrants to the colorectal chemotherapy market provide evidence that

[3]Although Janakiraman et al. (2009) focus on the branding implications of perception spillovers, the market for antidepressants may instead be suited to an analysis of horizontally differentiated goods.

physicians have particular beliefs about how information spills over to alternatives. The first new drug was adopted slowly despite being a strong complement in combination with the incumbent drug. In contrast, the second entrant, although approved as a stand alone treatment, was adopted primarily as part of a combination regimen with existing chemotherapy drugs. Broadening the scope of information spillovers can explain some of this behavior.

A physician's first choice of chemotherapy regimen for the treatment of colorectal cancer is an ideal setting for a test of perception spillovers. An oncologist's experience is the primary source of information about regimen quality because only a fraction of chemotherapy patients meet the specific conditions for which each of the available drugs was approved, and because pharmaceutical companies can only market a drug for approved uses. When the FDA approved new colorectal chemotherapy drugs in 1990 and 1991, the first time since 1962, physicians faced uncertainty regarding which regimen was best suited for a new patient. As they learned about the quality distributions of the regimens through use, the market leading regimen was supplanted by the entrant, a regimen that would eventually hold more than 80 percent of the market share. Another chemotherapy drug introduction in 1996 repeated the need for learning regarding regimen qualities, particularly since it was primarily adopted in combination with the other available drugs. The later entrant also faced a greater tradeoff between efficacy and severity of side effects, which additionally required physicians to learn about balancing these drug characteristics.

Finding empirical evidence of perception spillovers requires observing repeated choices by consumers as they encounter new products and learn about their qualities. I hypothesize that consumers have a joint distribution of quality beliefs for the available products and update those beliefs in a Bayesian fashion after experience with a particular bundle. If consumers hold nonzero beliefs about the covariances of quality, feedback from one product experience will influence the expected quality of every alternative choice option and thereby exhibit perception spillovers. By allowing those initial covariances to be determined within an

econometric model, the existence of perception spillovers, even across differentiated products, can be tested empirically.

I estimate the model of physician learning using linked cancer surveillance and Medicare claims data on new colorectal cancer diagnoses. Covering an area representing 26 percent of the U.S. population, the data include detailed medical information, patient demographics, and a consistent physician identifier. I construct the complete Medicare patient load of physicians using that identifier and estimate a learning model with perception spillovers for their chemotherapy choices between 1991 and 2001. Identification of the model comes from the changing regimen shares within physician as they treat a sequence of patients over the decade.

The results confirm that physicians hold joint distributional beliefs about regimen qualities. Physicians update their beliefs for each regimen from all chemotherapy experience, a conclusion eluding the previous literature. In accordance with intuition from the adoption rates, regimens entering the market in 1991 were not initially considered by physicians to be significantly correlated ($\rho = 0.011$). Although the regimens shared their major component drug, the new molecule entered after 28 years without innovation and was not proven superior in a large study until years after introduction. In this case, perception spillovers were not a significant source of information for physicians. In contrast, the initial beliefs for those two regimens with irinotecan, introduced in 1996, exhibited considerable correlation ($\rho = 0.702$); physicians updated their beliefs for both regimens from experiences with either. However, neither of the new regimen qualities was strongly correlated with the incumbent regimens' qualities. These results together provide evidence that physicians believed that any new drug was sufficiently different from the existing drugs during the period that information spilled over only between regimens sharing the new drug. This type of learning behavior slows the adoption rate of complementary drugs such as those entering in the years after the sample.

To explore the magnitude of these results, I simulate three counterfactual scenarios. First, to evaluate the importance of spillovers to the dynamics of learning, I simulate regimen shares when the covariances between the beliefs is set to zero. The counterfactual regimen shares indicate that learning from experience is primarily informative about the prescribed drug, but that spillovers to other alternatives do accelerate adoption mildly. The second and third counterfactuals demonstrate the influence of learning on regimen use. In the second simulation I force physicians to maintain their initial quality beliefs regarding the regimens and I find that the adoption of the dominant regimen is delayed by roughly one and a half years. In contrast, when physician beliefs are set at their ex post levels, estimated as the final posterior in the model, the third counterfactual shows that learning only partially mitigates the delayed adoption resulting from quality uncertainty. Together, these results provide a richer perspective for consumer behavior; namely, physicians form beliefs regarding how information spills over between bundled goods and use patient feedback to update their beliefs across treatment options.

This chapter proceeds by first presenting an overview of the chemotherapy market for colorectal cancer. I then introduce the model in two parts: prescription choice and physician learning. A description of the data then allows for a discussion of the identification and estimation of the model. Finally I present the results and discuss simulated counterfactuals before concluding.

1.2 Chemotherapy for Colorectal Cancer

In 2001, colorectal cancer was third most diagnosed and fatal cancer with an estimated 135,400 new cases and 56,700 deaths in the United States (Greenlee et al., 2001). It has since become the second most diagnosed and fatal cancer among cancers affecting both men and women, with diagnosis expected in one of twenty people born today (Centers for

Disease Control and Prevention, 2011).[4] The extent of disease is determined by the tumors' increasing invasion of local tissue, involvement of lymph nodes, and metastasis, the spread of cancer cells to other parts of the body. Between 1995 and 2000, the median age at diagnosis with colorectal cancer was 70 years old and patients had a 63.4 percent chance of five year survival overall (Ries et al., 2004). The probability a patient survived for five years ranged from 89.9 percent among those with localized cancers, to 9.6 percent for patients with metastatic cancer.

The three general methods of cancer treatment are surgical removal, radiation therapy, and chemotherapy, the total cost for which has more than doubled in the U.S. since 1987, and exceeded $20 billion in 2009 (Caplan, 2011). The primary treatment for any colorectal cancer is resection, removing the tumor or sections of the colon with cancer cells, if possible. Chemotherapy is generally considered adjuvant, or supplementary, to resection before metastasis, but is the standard of care after spread has occurred. The National Comprehensive Cancer Network (NCCN) maintains treatment guidelines which update the standard of care with results from new medical research. Most oncology drugs are infused intravenously into patients within their doctors' offices. Hospitals and physicians directly purchase the individual drugs, storing and administering the drugs to their patients as chosen, creating any cocktail treatment regimens themselves. For patients 65 and older, Medicare reimburses the physician for the purchase and administration of intravenous drugs and any oral equivalents if they exist. Particularly as new drugs later entered the market, physician profits varied by regimen, altering treatment incentives and leading to Medicare reimbursement reform in 2003.

Chemotherapy for colorectal cancer is primarily dichotomized into the treatment of metastatic disease and adjuvant therapy for localized cancer. FDA approvals for new

[4]The change in rank is due to increased screening for colorectal cancer and improved screening and therapies for prostate and breast cancers.

chemotherapy drugs, whether the drug is used alone or in combination, indicate for which type of therapy the approval is issued: metastatic or adjuvant. To the extent a drug's use represents a legitimate medical practice, physicians legally may use any FDA-approved (ethical) drug for therapy; however, pharmaceutical firms are only allowed legally to market the drug for its indicated, "on-label" use. Those firms may later apply for other indications to be added, such as the other of metastatic or adjuvant treatment, or in a different combination regimen, by providing supporting clinical evidence of effectiveness for the new indication. All currently available colorectal chemotherapy regimens were initially approved for metastatic therapy, despite only 19 percent of new diagnoses being for metastatic cancer and only 25 percent of chemotherapy patients having such extensive disease. Not surprisingly, the National Comprehensive Cancer Network (NCCN) estimates that 50 to 75 percent of drugs and biologics used to treat all cancers are used "off-label," for conditions not approved by the Food and Drug Administration (Soares, 2005). These circumstances leave physician experience as the primary source of information for 75 percent of all colorectal chemotherapy patients.

The FDA's first approved molecule for the treatment of metastatic colorectal cancer was fluorouracil (5FU) in 1962, predating the current era requiring controlled clinical trials in oncology (Ibrahim, 2003). Before 1962 and well into the 1980s, the nonspecific anticancer drug methotrexate, approved in 1953, was prescribed in some circumstances, but chemotherapy remained an adolescent practice.[5] The use of 5FU was still sporadic through the 1980s, when oncologists tried using 5FU for adjuvant therapy. Since the drug had passed patent protection, no clinical trials were performed so that no pharmaceutical firm applied to the FDA to include adjuvant use as an approved indication.

Levamisole, a veterinary deworming drug, was approved in 1990 for the adjuvant treat-

[5]As one medical oncologist described early chemotherapy, "When I graduated from residency in 1972 there were five chemotherapy drugs, and you wouldn't wish any of them on your worst enemy."

Table 1.1: FDA Drug Approvals for Colorectal Chemotherapy

Approval Year	Drug, In Combination	Indication	Type of Approval	Infusions
1962	Fluorouracil, 5FU	Metastatic, First Line	Regular	30
1990	Levamisole	Adjuvant	Regular	30
1991	Leucovorin, 5FU/LV	Metastatic, First Line	Regular	30
1996	Irinotecan, IRI	Metastatic, Recurrent	Accelerated	8
2000	IRI+5FU/LV	Metastatic, First Line	Regular	24

'Infusions' are the expected number of scheduled visits at the outset of treatment to receive therapy. Sources: Ibrahim (2003); National Comprehensive Cancer Network (2004).

ment of colorectal cancer after two successful studies. On the theory it produced an immune response to the tumor, the National Institutes of Health (NIH) published a consensus statement recommending it in 1990 as the first new therapy in 28 years (NIH Consensus Conference, 1990). Scientific evidence for levamisole was weak, however, and continued to be debated in the medical literature. Levamisole never became an orthodox treatment option and by 1998 it had definitively been shown to produce no benefit while causing unwanted side effects, prompting its withdrawal from the market.

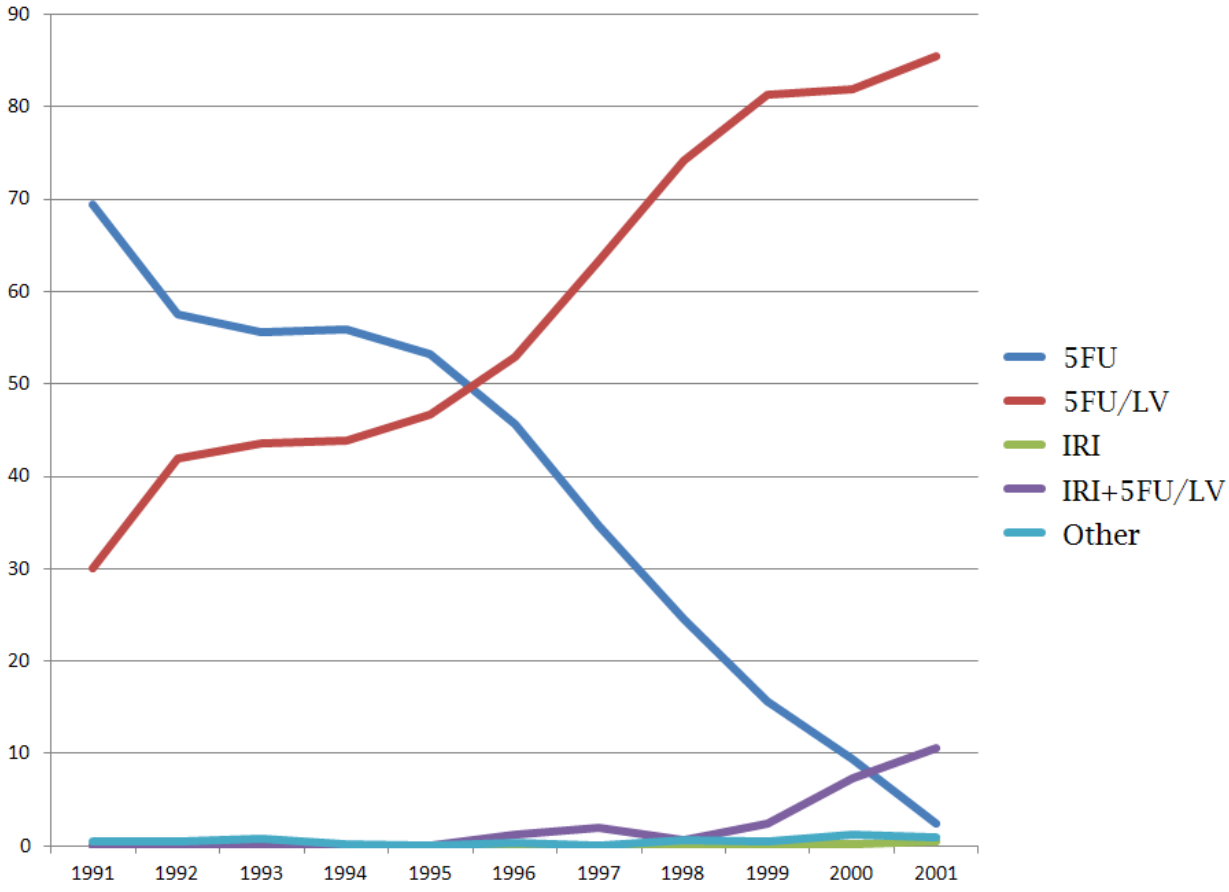
Concurrent with research on levamisole, leucovorin (LV) was tested with 5FU and shown to be effective against metastatic cancer, receiving FDA approval in 1991. LV alone is not chemotherapeutic, but like levamisole, 5FU combined with leucovorin (5FU/LV) failed to consistently produce benefits in small samples in the early 1990s. However, as larger clinical trials developed, 5FU/LV demonstrated superior benefits and would comprise nearly 90 percent of colorectal chemotherapy use over the middle of the decade. As an adjuvant treatment, the regimen quickly became the standard of care and continued to be the NCCN's preferred adjuvant treatment into the next decade (National Comprehensive Cancer Network, 2004).

In 1996 Pfizer won accelerated approval for its molecule irinotecan (IRI) under the brand name Camptosar, with full approval coming in 1998. The FDA indicated approval for the drug by itself as second line therapy for metastatic cancer; physicians, however, adopted the drug more quickly in combination with 5FU/LV for patients with advanced disease. The first NCCN practice guidelines for colorectal cancer in 1996 did not endorse IRI (National Comprehensive Cancer Network, 1996), although the combination IRI+5FU/LV was recommended in 2000. The bundled therapy was added as an approved indication in 2000 because it was demonstrated to extend patients' lives by an average 3 months despite having more significant side effects. In addition to effectiveness, the introduction of IRI required physicians learn about the trade-offs between greater toxicity, which promotes efficacy in cancer treatment, and more severe side effects.

By the end of the decade, there were four main regimens for use in the treatment of colorectal cancer: 5FU, 5FU/LV, IRI, and IRI+5FU/LV. A very small fraction of patients received other regimens, typically chemotherapies approved for other types of cancer. In the analysis, this non-standard treatment regimen forms the outside option available to physicians. Table 1.1 provides an approval time line, approved indications, and some facts for each regimen, and Figure 1.1 illustrates the share of prescriptions for each regimen within the data.

As noted earlier, a significant portion of learning research has focused on physician choice and pharmaceutical markets because of their impact on individuals and the broader national accounts. Considering generic prescription drugs cost an average 40 percent less than brand name products, Ching (2010) simultaneously models aggregate consumer learning and firm pricing policies to show that slow diffusion of generics can at least be partially explained by learning. Coscelli and Shum (2004) and Crawford and Shum (2005) use panel data for the Italian anti-ulcer market to study within-patient and across-diagnosis learning, respectively, showing that each can substantially reduce consumer uncertainty and influence choice

Figure 1.1: Regimen Shares for Colorectal Chemotherapy Prescriptions



probabilities. Other, ongoing research investigates the impact of supply-side behavior stemming from product bundling in the colorectal chemotherapy market with results showing a less competitive equilibrium resulting from bundling (Lucarelli et al., 2010). Although firm profits are subject to competitors' behaviors, their effect on information production and dissemination to physicians is uncertain.

1.3 Model

My model of physician treatment behavior consists of two parts: chemotherapy choice, and physician learning. The choice problem follows the standard conditional logit framework with assumptions reflecting the context of chemotherapy. Physician learning occurs between choices, so that treatment decisions are made using feedback from all previous patients.

In the model, physicians act as consumer agents for their patients and maximize their own utilities by optimally choosing among the available regimens to treat their patients. I assume physicians are myopic, risk-neutral decision-makers, so that in any period, a doctor chooses the regimen with the highest expected utility given her current information and without regard to the dispersion of her beliefs. In contrast to a myopic physician, a forward-looking physician would consider the potential gain to future patients by experimentation with regimens on the current period's patients. The risk of malpractice litigation and high cost of liability insurance are significant deterrents to such behavior. A form of single-patient experimentation, "N-of-1 trials," has been established since 1981 but remains extremely rare (Vohra et al., 2011). Interviews with practicing oncologists confirm that physicians prefer to prescribe regimens well known through practice, research, or communication with pharmaceutical companies, and to practice "defensive medicine" to decrease the probability of a lawsuit (Janakiraman et al., 2008; Kessler and McClellan, 1996). Oncologists, in particular, face greater than average risk of malpractice suits, though payments are sufficiently rare to

obscure statistics about the amount (Jena et al., 2011).

Risk-neutrality is an assumption I must make to allow initial quality perceptions to differ from their final values. In a learning model similar to that described below, Coscelli and Shum (2004) demonstrate the inability to separately identify a risk aversion parameter from the initial quality perceptions without additional assumptions. Namely, physicians' initial prior means of the quality perceptions must coincide with the "true" qualities across all patients. In the context of an expanding choice set, however, quality perceptions should change to reflect the changing relative values of regimens. As a result, no constant "true" quality ranking exists across patients of different times. Instead I assume risk neutral physicians, allowing initial and final quality perceptions to differ and quality rankings to change after new regimen introductions.

Finally, as described above, the set of possible regimens to treat colon cancer has experienced significant growth. Because the set of available regimens grows over time, define an "era" as the time span over which the regimen choice set is constant. The introduction of a new drug will increase the choice set of regimens and thus start a new era. Notationally, time and eras will pass separately.

1.3.1 Chemotherapy Choice

Let the utility of doctor i treating patient j with drug regimen $g \in [1, ..., G_p]$ at time t (within era p) be given by the following:

$$\begin{aligned} U^*_{ijg} &= \beta_1 \text{Profit}_{gt} + \beta_2 \text{OnLabel}_{jgt} + \beta_3 \text{Recommended}_{jgt} + \beta_4 \text{NumVisits}_g \\ &\quad + \gamma_{1,g} \text{Metastatic}_j + \gamma_{2,g} \text{Comorbidity}_j + \gamma_{3,g} \text{Age}_j + \delta^*_g + \xi^*_g(t) + \varepsilon^*_{ijg} \qquad (1.1) \\ &\equiv V^*_{ijg} + \varepsilon^*_{ijg} \end{aligned}$$

where

- $Profit_{gt}$ is the average reimbursement for regimen g at time t less the physician's acquisition costs of its components,[6] $OnLabel_{jgt}$ indicates if the use of regimen g on patient j falls within the official FDA indication for the regimen at the time of treatment, $Recommended_{jgt}$ indicates whether the use is recommended by the NIH or NCCN, and $NumVisits_g$ is the minimum number of office visits a patient would need to make to receive the regimen.
- $Metastatic_j$ indicates if the patient has metastatic cancer, $Comorbidity_j$ gives the Charleson comorbidity weight of patient j derived from his Medicare claims as developed by Klabunde et al. (2000), and Age_j is the patient's age at treatment.
- δ_g^* parameterizes the core quality perception of regimen g. Although completely unobserved by the econometrician, physicians learn imperfect information about the values both within and across prescriptions as described below.[7]
- $\xi_g^*(t)$ is a flexible, drug-specific time trend summarizing those aspects of period t which affect the perception of regimen g but are outside the available data. Namely, $\xi_g^*(t)$ captures learning from medical journals, participation in clinical trials, medical congresses, and detailing by pharmaceutical companies. The drug-specific trend is constant across patients and doctors.
- ε_{ijg}^* are independent and identically distributed (across patients, doctors, regimens, eras, and time) shocks to patient-drug match observed (perfectly) by the doctor but not the analyst.

If the outside option, the "other" regimen is chosen, let the physician's utility be

$$U_{ij0}^* = \delta_0^* + \xi_0^*(t) + \varepsilon_{ij0}^* \tag{1.2}$$

[6]Chemotherapy is traditionally dosed according to the patient's body surface area since it is believed to reduce variability in both drug exposure and side effects between patients (Gurney, 1996; Baker et al., 2002). Since patient height and weight is unknown, this variable has been constructed using a representative patient based on the mean body surface area and kilograms from patient-level IntrinsiQ data.

[7]Within the utility model, the core quality perception is constant across patients. While patient response to particular regimens is indeed variable, the included covariates, taken together, form a particular patient-drug match value.

where δ_0^* is known to the physician. $\xi_0^*(t)$ and ε_{ij0}^* are consistent with the above.

Because the utility in Equation 1.1 has a stochastic component, a physician chooses that regimen g which has the greatest expected utility: $\mathrm{E}_t(U_{ijg}^*) > \mathrm{E}_t(U_{ijh}^*)$ for all $h \in [0, ..., G_p]$. Since only differences in utility matter, I normalize the value of the "other" regimen to be zero without loss of generality. The expected utility for any regimen $g \in [1, ..., G_p]$ may therefore be given as

$$\begin{aligned}\mathrm{E}_t(U_{ijg}) =& \beta_1 \mathrm{Profit}_{gt} + \beta_2 \mathrm{OnLabel}_{jgt} + \beta_3 \mathrm{NCCN}_{jgt} + \beta_4 \mathrm{NumVisits}_g \\ &+ \gamma_{1,g} \mathrm{Metastatic}_j + \gamma_{2,g} \mathrm{Comorbidity}_j + \gamma_{3,g} \mathrm{Age}_j + \mathrm{E}_t(\delta_g) + \xi_g(t) + \varepsilon_{ijg} \qquad (1.3) \\ \equiv& V_{jgt} + \varepsilon_{ijg}\end{aligned}$$

where $\delta_g = \delta_g^* - \delta_0^*$, $\xi_g(t) = \xi_g^*(t) - \xi_0^*(t)$, and $\varepsilon_{ijg} = \varepsilon_{ijg}^* - \varepsilon_{ij0}^*$. Letting g_{ij} denote the chosen regimen and assuming ε_{ijg} is distributed i.i.d. with the type 1 extreme value distribution, the probability of choosing a specific regimen g is

$$\begin{aligned}Prob(g_{ij} = g|X, \Theta) =& \int \frac{\exp(V_{ijg})}{1 + \sum_{h=1}^{G_p} \exp(V_{ijh})} \mathrm{d}F(\Theta) \\ \equiv& \int P_{ijg}(X, \Theta) \mathrm{d}F(\Theta) \qquad (1.4)\end{aligned}$$

for any $g \in [1, ..., G_p]$. In the standard conditional logit decision model, the choice probability is simply the integrand P_{ijg} above. Since physician perceptions δ_g vary with experience, however, integration recovers the unconditional probability being modeled.

1.3.2 Physician Learning

The limited information sources available to physicians make experience and learning the primary driver of the treatment and health of future patients. As discussed earlier, no approved FDA therapy existed for nearly 75 percent of colorectal chemotherapy patients over the period. Although the FDA had approved three regimens before 2000 for metastatic

cancer, clinical trials provide a limited amount of information to physicians for at least the reason that an individual's patient load will differ from the trial group. Nevertheless, a physician begins the era with initial perceptions about each regimen's quality. In the second and all subsequent eras, these initial beliefs will be the posterior beliefs from the previous era. For clarity, let t denote the beginning of the time period.

I assume that at the beginning of time $t = 1$, doctor i has the following initial beliefs about $\vec{\delta}$, the vector of regimen quality perceptions:[8]

$$\vec{\delta}^i \sim N\left(\vec{\delta}_1^i \equiv \begin{bmatrix} \mathrm{E}_1\delta_1^i \\ \vdots \\ \mathrm{E}_1\delta_{G_p}^i \end{bmatrix},\ \Sigma_{\delta,1}^i \equiv \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots \\ \sigma_{1,2} & \sigma_{2,2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}\right). \tag{1.5}$$

Her beliefs consist of average regimen qualities and their variances, as well as covariances between the regimens. Those covariances capture the extent to which regimens may be related because of shared component drugs and, like the mean quality beliefs, will be updated based on the physician's experiences. The Bayesian learning framework formalizes this.

Each time a physician prescribes a regimen, she receives feedback from the patient's experience. That feedback comes in the form of a quality signal. That is, for each patient receiving regimen g, she observes a signal which she characterizes by

$$\mu_{gt} = \mathrm{E}_t\delta_g^i + \nu_{gt}, \tag{1.6}$$

with $\mathrm{E}_t\delta_g^i$ coming from her prior and $\nu_{gt} \sim$ i.i.d. $N(0, \omega^2)$ over j, g and t, resulting from patient idiosyncrasies.

Supposing doctor i begins period t with beliefs on $\vec{\delta}^i$ as her priors and receives signal

[8] I present the general case for era $p = 1$; the same structure will hold true for subsequent eras by taking the posteriors of the previous era, $\vec{\delta}_{t-1}^i$ and $\Sigma_{\delta,t-1}^i$, and extending them by the appropriate number of rows (and columns) for the new regimens available.

vector $\vec{\mu}_t$, the joint distribution of beliefs and signals in t is given by

$$\begin{pmatrix} \vec{\delta}^i \\ \vec{\mu}_t \end{pmatrix} \sim N \left(\begin{bmatrix} \vec{\delta}^i_t \\ \vec{\delta}^i_{\mu t} \end{bmatrix}, \begin{bmatrix} \Sigma_{\delta t} & \Sigma_{\delta \mu t} \\ \Sigma'_{\delta \mu t} & \Sigma_{\mu t} \end{bmatrix} \right), \tag{1.7}$$

where (i) $\vec{\delta}^i_t$ and $\Sigma_{\delta t}$ are the expected mean and variance-covariance matrix of $\vec{\delta}^i$, conditional on the previous signals received before t; (ii) $\vec{\delta}^i_{\mu t}$ and $\Sigma_{\mu t}$ are the mean and variance-covariance matrix of the physician's signals in period t; and (iii) $\Sigma_{\delta \mu t}$ is the covariance matrix between $\vec{\delta}^i$ and $\vec{\mu}_t$ which allows for across-regimen learning. Given the form of a signal μ_{gt} and the beliefs $\vec{\delta}^i_t$, the elements of the arrays above derive from the following facts:

- $\mathrm{E}(\mu_{gt}) = \mathrm{E}_t \delta^i_g$,
- $\mathrm{Var}(\mu_{gt}) = \Sigma_{\delta t(g,g)} + \omega^2$,
- $\mathrm{Cov}(\mu_{gt}, \mu'_{gt}) = \Sigma_{\delta t(g,g)}$, that is, for two signals on the same regimen,
- $\mathrm{Cov}(\mu_{gt}, \mu_{g't}) = \Sigma_{\delta t(g,g')}$, for two signals of different regimens, and
- $\mathrm{Cov}(\delta^i_g, \mu_{g't}) = \mathrm{Cov}(\mathrm{E}_t \delta^i_g, \mathrm{E}_t \delta^i_{g'} + \nu_{g't}) = \Sigma_{\delta t(g,g')}$

with $\Sigma_{\delta t(x,y)}$ giving the (x, y) coordinate of $\Sigma_{\delta t}$. Thus $\Sigma_{\delta \mu t}$ consists of the elements of $\Sigma_{\delta t}$ ordered according to the number of received signals for each regimen in each period.

I model physician learning from experience as Bayesian belief updating by use of the best linear predictor of $\vec{\delta}^i_t$ given $\vec{\mu}_t$. The distribution of quality beliefs conditioned on the signals received in period t is the posterior distribution for period t. This posterior then becomes the prior belief distribution for the next period. Following Amemiya (1985), the conditional distributions are calculated as

$$\begin{aligned} \vec{\delta}_{t+1} \equiv \mathrm{E}(\vec{\delta}|\vec{\mu}_t) &= \mathrm{E}_t \vec{\delta} - \Sigma_{\delta \mu t} \Sigma^{-1}_{\mu t} \vec{\delta}_{\mu t} + \Sigma_{\delta \mu t} \Sigma^{-1}_{\mu t} \vec{\mu}_t \\ &= \mathrm{E}_t \vec{\delta} + \Sigma_{\delta \mu t} \Sigma^{-1}_{\mu t} (\vec{\mu}_t - \vec{\delta}_{\mu t}) \\ \Sigma_{\delta, t+1} \equiv \mathrm{Var}(\vec{\delta}|\vec{\mu}_t) &= \Sigma_{\delta t} - \Sigma_{\delta \mu t} \Sigma^{-1}_{\mu t} \Sigma'_{\delta \mu t}. \end{aligned} \tag{1.8}$$

so that the updating behavior weights the new information by the inverse of the signal's variance, just as in ordinary least squares. Physicians use these updated beliefs in their therapy choice decisions in the following period.

A standard concern with logit choice models is the restrictive substitution patterns resulting from functional form. The independence from irrelevant alternatives (IIA) exhibited by the logit model implies proportional substitution across alternatives; the model is misspecified if the choices exhibit varying degrees of substitutability because of unobserved correlation between them. Because the learning model here explicitly accounts for correlations between the choices, proportional substitution must only hold within physician at any time's information set. When the physician updates her beliefs and their covariances from a new feedback signal, the same proportionality need not hold, reducing the restrictiveness of IIA. The incorporation of learning into the choice model thus additionally provides for more flexible substitution patterns in applied choice analysis.

In summary, the sequence of events in the models is as follows. The physician begins by holding initial beliefs about the value of all regimens available. When the first patient needs treatment, she chooses the regimen with the greatest expected utility, which is a function of her beliefs, the patient's characteristics, and national recommendations. After treatment, she receives a signal of the value of the chosen regimen. Because the quality values across regimens is correlated, she uses the information from the signal to update her beliefs for all regimens. When the next patient arrives, she uses these updated beliefs in her decision. Therefore, from the model, the parameter vector Θ consists of β, γ, and $\xi(t)$ from the choice problem, the initial mean and variance-covariance matrix of the regimens' quality perceptions $(\mathrm{E}_1\delta_1, ..., \mathrm{E}_1\delta_G; \sigma_{1,1}, ..., \sigma_{G,G})$, and the signal variance ω^2.

1.4 Data

The Surveillance Epidemiology and End Results (SEER) Program of the National Cancer Institute (NCI) coordinates the collection of the universe of cancers within several cancer registries covering 26 percent of the US population. This analysis uses the cases reported

Table 1.2: Sample Statistics

		Mean	Std. Dev.	Median
Physicians				
	Patient Load	25.017	17.453	19
	Number of Regimens Used	2.380	0.693	2
Patients				
	Metastatic Cancer	0.251		
	Comorbidity Weight	0.422	0.791	0
	Age at Therapy	73.404	5.419	73
Regimens				
	Prescribed Regimen			
	FDAA Approved	0.633		
	Recommended	0.276		
	5FU	0.390		
	5FU/LV	0.584		
	IRI	0.001		
	IRI+5FU/LV	0.020		
	Other	0.005		

N = 411 Physicians, 10,283 Patients, 11 Years

in the SEER-13 registries from 1991 to 2001: San Francisco, CA; Connecticut; Detroit, MI; Hawaii; Iowa; New Mexico; Utah; Atlanta, GA; San Jose, CA; the Arizona Indian System; Los Angeles, CA; and Rural Georgia.[9] In addition to information about the tumor, SEER collects demographic and socioeconomic information about the patient and, through a unique patient identifier, follows patients after their diagnosis and surgery. For those in the Medicare population, the SEER-Medicare linked database provides all Medicare claims linked with the patient before and after diagnosis. These claims also contain a consistent physician identifier. Because SEER collects the universe of cancer patients, the physician identifier enables me to construct each physician's complete Medicare patient load.

The analysis sample consists of 411 doctors treating 10,283 cases of colorectal cancer between 1991 and 2001. The sample of physicians was limited to those who treat at least

[9]Data requests including ZIP code information must be approved by both the NCI and each individual registry. All registries except Seattle, WA, approved this request.

10 Medicare patients, aged 65 or older, over the period to select a sample for which learning could significantly enhance a physician's treatment patterns. Using drug-specific codes on physician claims, I deduce the regimen first chosen after diagnosis following Warren et al. (2002). The regimen specific characteristics come from Thomson Reuters/Medical Economics Red Book, the FDA's *Drugs@FDA* web site,[10] the NIH and NCCN practice guidelines, and individual drug package inserts.

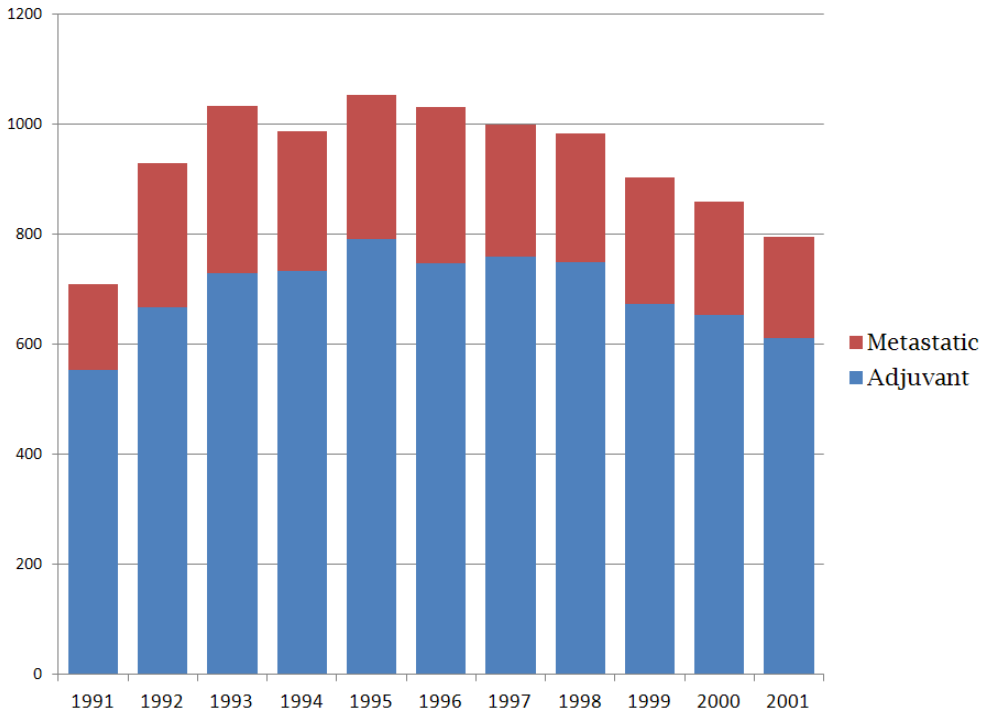
Physicians in the sample had a median Medicare patient load of 19 colorectal cancer cases, with a mean of 25 cases.[11] They used an average of 2.38 regimens in chemotherapy, with the number of those choosing the "other" regimen roughly equal between physicians who used 2 and those who used 3 regimens. Table 1.3 shows how quickly each regimen was adopted by physicians. Without controlling for patient arrival, 75 percent of physicians had prescribed 5FU within the first three years, as compared to the nearly 5 years until the same percentage had used 5FU/LV. Even after 6 years, however, only 35 percent of physicians treated with IRI+5FU/LV, and less than two percent ever prescribed IRI alone. Given the FDA approval patterns, such adoption rates imply calculated behavior by individual physicians.

The median age of the patients in the sample is 73 years old, slightly older than the median of the entire population, and 25 percent have metastatic cancer. Figure 1.2 shows how patients are distributed over time and by metastatic versus adjuvant therapy. Since the number of patients is evenly distributed, and the proportion with metastatic disease is constant, patient composition alone cannot explain the changes in regimen shares over time. In estimation, age is scaled down so that the Medicare entry age of 65 corresponds to 0, easing later calculations. Using each patient's claims for the year preceding the month of diagnosis, the Charleson comorbidity weight, a measure of potentially confounding factors to

[10]Available: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm. Approval information for levamisole and leucovorin were not available on the web site but were obtained through a Freedom of Information Act request to the FDA.

[11]Industry publications from the American Society of Clinical Oncology suggest these numbers are representative of a typical oncologist's Medicare patient load for colorectal cancer.

Figure 1.2: Patients by Year and Extent of Disease



the treatment of cancer, shows no other major comorbidities for 71 percent of the patients in the sample. Controlling separately for age and comorbidity is necessary since the elderly are absolutely less likely to receive chemotherapy than younger patients (Schrag et al., 2001).

Despite the broad coverage of the SEER-Medicare data, the specifics of each program prevents some information from being available. The SEER data identify cases of colorectal cancer based on a physician's location. If a patient is diagnosed or receives therapy outside of the registry, his data may be incomplete. Additionally, because levamisole was an oral medication, physicians could not submit reimbursement claims for its use in chemotherapy.[12] Thus, the patients receiving the regimen labeled "5FU" may be receiving 5FU alone or in combination with levamisole. The bias from this measurement error should be small: leucovorin was an unproven therapy and used infrequently over the period, declining even more rapidly than 5FU.

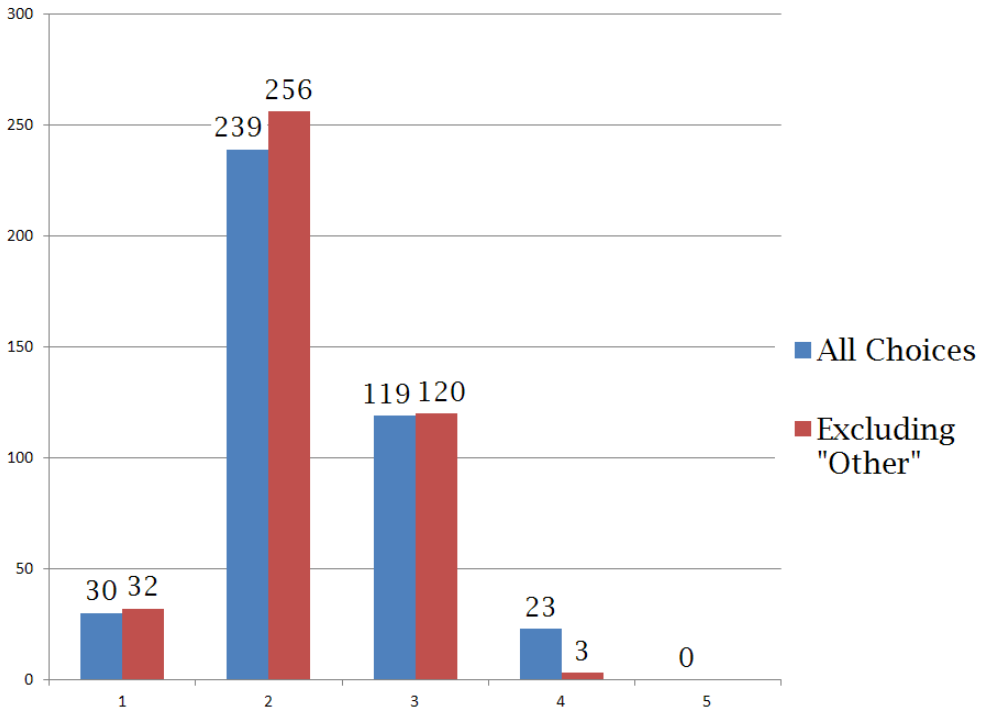
1.5 Identification

The estimated parameters from the model fall into two groups: the learning parameters, consisting of the initial quality perceptions, their variance-covariance matrix, and the signal variance; and the choice parameters, which apart from the doctor's expectation of the regimen quality at time t, affects the choice probabilities. The choice parameters are identified from variation in the choice characteristics across patients, regimens, and time as in a standard conditional logit model. The identification of the learning parameters comes from the physicians' choice patterns and signaling mechanism.

Over the sample period, each of the regimens experiences variation across patients and time. The FDA approvals and NIH and NCCN recommendations are matched by a patient's stage and quarter of treatment, so that each doctor faces variation across her patient load as

[12]The newest data available from the SEER-Medicare linkage does include Medicare Part D claims, made available after the program's 2006 expansion to prescription drug coverage.

Figure 1.3: Number of Regimens Used by Sample Physicians



to the approval and recommendation status of the regimens.[13] Further, the NIH recommendations are updated by the NCCN in 1996 and 2000. The number of visits for a regimen, the metastatic indicator, comorbidity weight, and age variables are all regimen specific to satisfy the conditional logit requirements of variation across alternatives.

The initial quality values are identified by both the choice model for the initial period of prescription, as well as by the subsequent pattern of a physician's prescriptions. In the first instance of treatment, all physicians share the same initial beliefs on 5FU and 5FU/LV, so that a conditional logit estimated on that subset of the data would identify those parameters as regimen-specific constants. In 1996, at the introduction of IRI, the beliefs on 5FU and 5FU/LV would have updated, but again all physicians then share the same beliefs for IRI and IRI+5FU/LV so that their values act as regimen-specific constants. Beyond these initial

[13]Of the 411 physicians, 5 do not treat any metastatic cancer patients.

conditions, however, the learning process itself provides identification. Given a physician's prescription choices and any covariance matrix for the beliefs, every value of $E_t\vec{\delta}_g$ within the choice model is a function of the initial conditions because of the Bayesian updating process. Every choice within the data, therefore, provides information on the initial beliefs, albeit with different, decreasing signal to noise ratios as time progresses.

The variance-covariance matrix of the initial beliefs and the variance of the feedback signals are determined across physician prescription patterns. Since only the expected value of the regimen quality enters the choice problem, the variances affect only the degree to which those means change given the quality signals from a prescription. It is thus necessary to observe a sufficient number of prescriptions both within and across regimens by the same doctor to learn about how the updating process occurs. Limiting the sample to only those physicians who treat at least ten people ensures sufficient information to capture that process. Figure 1.3 shows the distribution of the number of regimens used by physicians in the sample. The signal variances are a measure of how responsive physicians are to patient feedback. A small signal variance will discourage physicians from updating their beliefs since signals are centered at their perceived mean. As that variance grows larger, physicians will update their beliefs toward the value of the signal as well as update the regimen's perceived variance. Repeatedly noisy signals result in persistently high belief variances over the period.

1.6 Estimation

The data set contains observations on patients, their physicians, and the regimen prescription for treating colorectal cancer. If I observed the feedback signal from each treatment episode I could calculate the revised belief distribution for each physician and substitute their expectations into each of her treatment choice problems. That signal, however, is

Table 1.3: Percentage of Physicians Who Prescribe A Regimen for the First Time in a Particular Year

All Patients	5FU	5FU/LV	IRI	IRI+5FU/LV	Other
1991	44.28	26.03			0.97
1992	21.90	27.01			0.73
1993	9.00	12.17			0.97
1994	5.35	6.08			0.49
1995	4.14	6.33			0.24
1996	4.14	3.89	0.00	2.19	0.49
1997	0.97	4.38	0.00	4.14	0.00
1998	0.73	4.62	0.24	1.22	1.22
1999	0.49	3.41	0.49	4.62	0.97
2000	0.00	1.22	0.24	10.95	2.43
2001	0.24	2.68	0.97	11.44	1.70
Never Rx	8.76	2.19	98.05	65.45	89.78
Metastatic	5FU	5FU/LV	IRI	IRI+5FU/LV	Other
1991	14.29	14.29			0.00
1992	11.58	23.40			0.00
1993	7.14	14.53			0.00
1994	4.68	8.87			0.00
1995	3.45	8.37			0.00
1996	4.19	4.68	0.00	0.74	0.00
1997	4.19	3.94	0.00	2.22	0.00
1998	1.23	5.42	0.25	1.23	0.00
1999	0.74	3.69	0.49	3.94	0.00
2000	0.74	2.71	0.25	7.88	0.00
2001	0.00	1.23	0.99	10.34	0.00
Never Rx	47.78	8.87	98.03	73.65	100.00
Adjuvant	5FU	5FU/LV	IRI	IRI+5FU/LV	Other
1991	43.07	17.76			0.97
1992	21.65	17.03			0.73
1993	9.25	12.41			0.97
1994	5.84	9.00			0.49
1995	3.89	6.08			0.24
1996	3.89	8.52	0.00	2.19	0.49
1997	1.22	9.00	0.00	2.19	0.00
1998	1.22	5.84	0.00	0.00	1.22
1999	0.24	4.14	0.00	0.97	0.97
2000	0.00	3.16	0.00	4.62	2.43
2001	0.24	3.65	0.00	3.65	1.70
Never Rx	9.49	3.41	100.00	86.37	89.78

'Never Rx' indicates the group of physicians within the sample who never prescribe the regimen over the sample period.

unobserved.[14] Nevertheless, since the quality perception of a regimen, conditional on the covariates, is homogeneous across patients and normally distributed for a given physician, I can simulate a sequence of physician signals before substituting the quality perception values into the treatment choice model. Utilizing the Bayesian assumption on updating along with the physician's prescription pattern, I use simulated maximum likelihood estimation to estimate the model parameters as suggested by Train (2009).

Given some initial beliefs on $\vec{\delta} \sim N(\vec{\delta}_1, \Sigma_{\delta 1})$, I can forward construct the signals and updated beliefs as follows. For a given period's treatment choices, signal vector $\vec{m}_t$ has variance-covariance matrix M made up of parameters from $\Sigma_{\delta t}$ and ω^2 according to the equations on page 17. Just as random variables are transformed to match known distributions, I use $\vec{\delta}_t$, M, and a vector $\vec{z}_m$, with the appropriate dimension and randomly drawn from $N(0, \mathrm{I})$, according to

$$\vec{m}_t = \vec{\delta}_t + \check{M}\vec{z}_m \tag{1.9}$$

where $\check{M}$ is the lower triangular matrix of strictly positive values which results from Cholesky decomposition of M. With $\vec{\delta}_t$ and $\vec{m}_t$ for a period, I use the conditional distribution of the learning model in Equation 1.8 to create $\vec{\delta}_{t+1}$, the period's posterior and following period's prior, and its variance-covariance matrix $\Sigma_{\delta,t+1}$. Starting with the first period and running through each of the eras, I simulate the beliefs by doctor for all periods. The beliefs are then matched by period to the treatment choice decision for each patient and a conditional logit analysis is performed, resulting in an estimated $\hat{\Theta}$ conditional on the randomly drawn $\vec{Z}$.

Any one draw of $\vec{Z}$ and simulation of signals is unlikely to approximate the actual signal well. To prevent bias from entering through the signal construction, simulating multiple signal sequences and averaging their outcomes is necessary. Simulating signals R times

[14]The SEER-Medicare data are rich enough to measure overall survival, progression-free survival, and potentially even severe adverse events. Incorporating these measures into the value of the signal stands at the front of future work.

results in the overall simulated log-likelihood function to be estimated:

$$\mathcal{SL}(\Theta|X) = \sum_{i=1}^{N}\sum_{j=1}^{J_i} \log\left(\frac{1}{R}\sum_{r=1}^{R}\left[1[g_{ij}=g]P_{ijg}(X,\Theta^r)\middle|\vec{m}^r\right]\right). \tag{1.10}$$

That is, for any randomly drawn $\vec{Z}^r$, I calculate the sequence of signals and update beliefs for all physicians before substituting into the logit choice probabilities for each choice. Finally I average the probabilities across the R draws and then calculate the likelihood as in a standard conditional logit using the averaged probabilities; they are consistent estimates of the true probabilities (Train, 2009). For the results in this paper, $R = 10$ simulations, as in Coscelli and Shum (2004).

1.7 Results

Tables 1.4 and 1.5 contain the learning and choice parameter estimates, respectively, from maximum simulated likelihood estimation. The standard errors are derived from finite differences estimates of the Hessian at the convergent value of the parameter vector. As a measure of model fit, Figure 1.4 plots the actual and predicted market shares from the data and model.

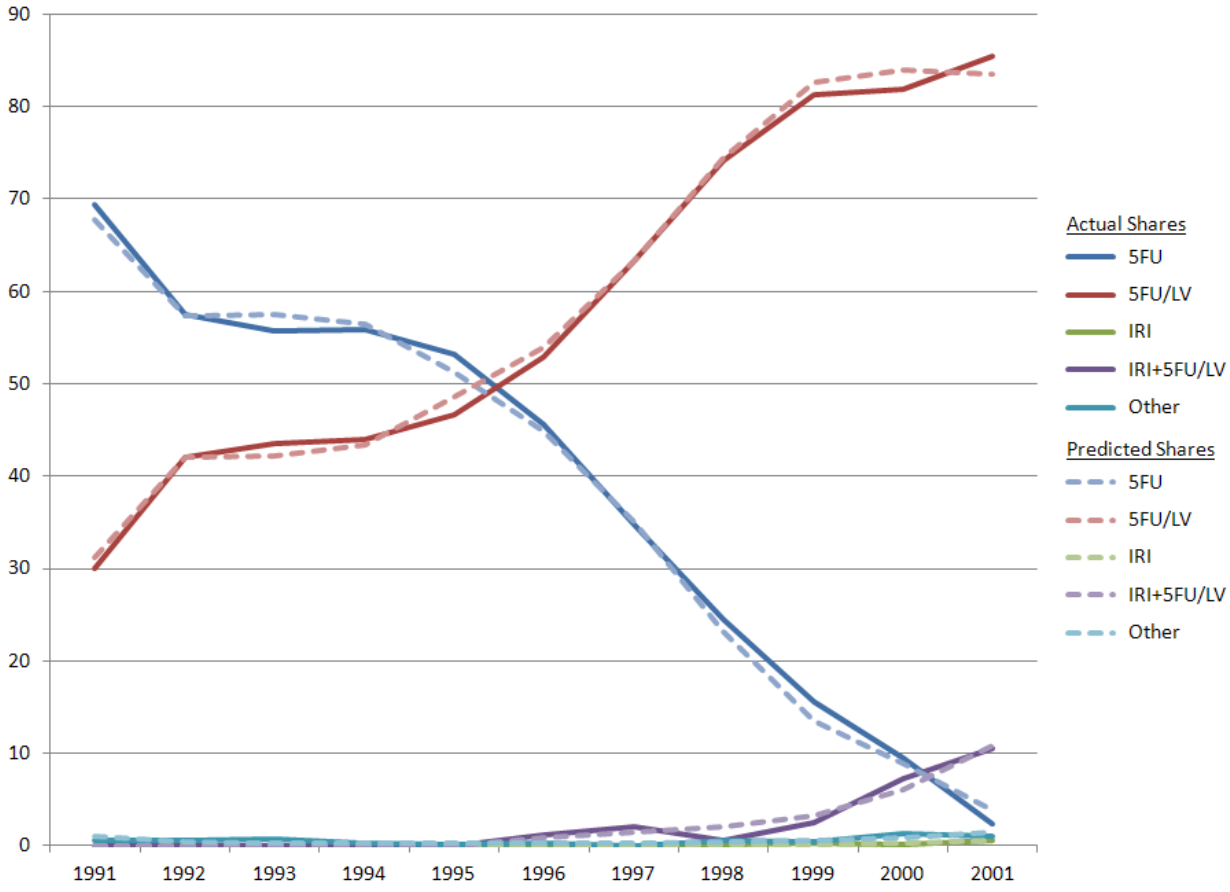
Since the initial perceptions estimates are relative to the outside "other" regimen's value of zero, the initial beliefs are estimated on the order we would expect: 5FU was initially the most popular regimen used and hence has the highest estimate, followed by 5FU/LV. When IRI is introduced, physicians hesitate to adopt it so that its initial quality value is large and negative. IRI+5FU/LV offers more promise, however, and thus has a small but positive value. Each of these parameters is statistically significantly different from zero so that physicians were clearly differentiating these regimens from the "other" option. Although the large negative value on IRI holds that the outside option is conditionally preferred to an FDA approved therapy, this result is only indicative that IRI was not considered a legitimate

Table 1.4: Learning Parameter Estimates from SMLE

	Estimate	Std. Errors
Initial Perceptions		
5FU	6.8894 *	0.2122
5FU/LV	5.7422 *	0.8929
IRI	-56.8409 *	1.0703
IRI+5FU/LV	3.0034 *	0.0400
Initial Covariance Matrix		
Var(5FU)	1.7802 *	0.0241
Cov(—, 5FU/LV)	0.1776 *	0.0226
Cov(—, IRI)	0.7773 *	0.0356
Cov(—, IRI+5FU/LV)	0.3197 *	0.0631
Var(5FU/LV)	67.8321	85.1048
Cov(—, IRI)	-0.1811 *	0.0161
Cov(—, IRI+5FU/LV)	0.9446 *	0.0231
Var(IRI)	76.0225 *	5.3168
Cov(—, IRI+5FU/LV)	9.1148 *	0.0416
Var(IRI+5FU/LV)	2.2152 *	0.0253
Signal Variance	44.3427 *	0.6587
Simulated Log Likelihood	-6488.478	

* $t > 1.96$. $R = 10$ Simulations. $N = 411$ Physicians, 10,283 Patients

Figure 1.4: Actual and Predicted Regimen Shares



treatment option for most patients despite its approval. In particular, estimates from the choice model show that IRI was considered only a treatment for metastatic cancer on patients with no comorbidity. Because these characteristics match patients likely to enter a clinical trial, physicians may be attempting to match clinical trial populations for this regimen whereas the absence of clinical trial data for IRI+5FU/LV made it more appealing for more general use.

Turning to the estimates of the initial variance-covariance matrix of the physicians' beliefs, the variance for the 5FU regimen is the smallest; that physicians were relatively certain of the quality value of the oldest drug on the market is reassuring. As entrants approved

Table 1.5: Choice Parameter Estimates from SMLE

	Estimate	Std. Errors
FDA Approved	0.8016 *	0.0062
Recommended	-0.4588 *	0.0008
Number of Visits	-0.1305 *	0.0187
Metastatic		
5FU	-3.1366 *	0.1651
5FU/LV	-1.5172 *	0.0245
IRI	60.4247 *	1.5394
IRI+5FU/LV	0.4744 *	0.0066
Comorbidity Weight		
5FU	0.1366 *	0.0010
5FU/LV	0.4079 *	0.0327
IRI	-25.2968	113.0704
IRI+5FU/LV	0.1990 *	0.0222
Age		
5FU	0.0673 *	0.0033
5FU/LV	0.0539 *	0.0054
IRI	0.0064 *	0.0083
IRI+5FU/LV	-0.0204 *	0.0009
Time		
5FU	0.7521 *	0.0478
5FU/LV	0.4570 *	0.0282
IRI	0.2961 *	0.0690
IRI+5FUL/LV	0.1288 *	0.0525
Time^2		
5FU	-1.2615 *	0.0035
5FUL/LV	-0.5488 *	0.0034
Simulated Log Likelihood	-6488.478	

* $t > 1.96$. $R = 10$ Simulations. $N = 411$ Physicians, 10,283 Patients

on small samples and short-run endpoints, the beliefs around 5FU/LV and IRI are quite uncertain, and only the latter are precisely estimated. Leucovorin's (LV) initial approval was based on its marginal superiority in a number of small studies when researchers were desperate to find more effective therapies. The benefits of LV continued to be debated in the literature until large, widespread clinical trials finished years after its introduction. Similarly, the large clinical trials for IRI alone concluded after its accelerated approval. In comparison to the two other new regimens, the initial variance on IRI+5FU/LV is much smaller but also statistically significant. Why physicians appear to have been so much more confident about the quality of IRI+5FU/LV at its launch is unclear. Finally, the variance of the signal is less than that of 5FU and IRI, and its standard deviation in comparison to the size of the initial quality perceptions is not so large as to prevent inference by physicians on the value of regimens.

The estimates of the covariances between the initial beliefs of the regimens show to what extent physicians initially believed the regimens to be correlated, and thus to what extent information spilled over across alternative regimens. Results indicate that physicians do hold joint distributional beliefs across bundles in the form of statistically significant covariances. Although most covariances are small relative the large variances of 5FU/LV and IRI, they are all statistically significantly different from zero. Between 5FU and 5FU/LV, the correlation is 0.011, indication an initial perception of only weak similarity. If 5FU was the standard regimen in 1991 when LV was added under weak medical evidence, then physicians could reasonably have considered their effectiveness to be related only mildly, with the new drug providing an uncertain degree of complementarity. 5FU/LV remains only weakly correlated to IRI and IRI+5FU/LV. The exception to the small magnitudes of the correlations, however, is that between the regimens containing IRI. The initial covariance estimates from the model imply an initial belief correlation between IRI and IRI+5FU/LV of 0.702. The strong correlation between IRI and IRI+5FU/LV implies that information from

the use of either regimen would considerably inform the beliefs of the other simultaneously. Given the pattern of sizable correlations, physicians seem to believe that any new drug's addition to a regimen considerably alters the quality of the regimen, but regimens that share a new drug are initially perceived to be very similar.

Like the learning parameters, the choice parameter estimates in Table 1.5 reinforce the anecdotal evidence. Being FDA approved is positively related to the probability a regimen is chosen, although the model indicates that being recommended by the NIH or NCCN is negatively related to a regimen being chosen. This latter result is not surprising given the current knowledge of efficacy for these drugs; that is, the early recommendations were incorrect. The NIH recommended 5FU as adjuvant therapy in its 1990 publication, only mentioning ongoing research into 5FU/LV. When the NCCN updated standard of care recommendations in 1996, it gave the two regimens equal recommendation, not settling on 5FU/LV until 2000. In several large clinical trials publishing later in the decade, 5FU/LV was been demonstrated to be superior in every dimension, so that the recommendations stood wrong for over most of the sample period. IRI+5FU/LV, although later shown to be superior to IRI alone, was not recommended either. In sum, the "recommendation" variable is signed appropriately. The remaining covariates suitably show that IRI regimens were favored for metastatic patients over those with localized diseases, but that IRI alone was intended for otherwise healthy individuals. Finally, although age has been shown to affect whether or not chemotherapy is prescribed for any colorectal cancer patient, these results show that age has a small impact on which regimen is ultimately chosen.

To assess the importance of learning and perception spillovers to physician treatment practice, I simulate regimen shares over the sample period under three counterfactual scenarios. First, to test the empirical importance of perception spillovers, the initial covariances between regimen beliefs are set equal to zero to prevent information from experience to spill over between regimens in the updating process. Given the small correlations from

Figure 1.5: Counterfactual Regimen Shares Without Spillovers

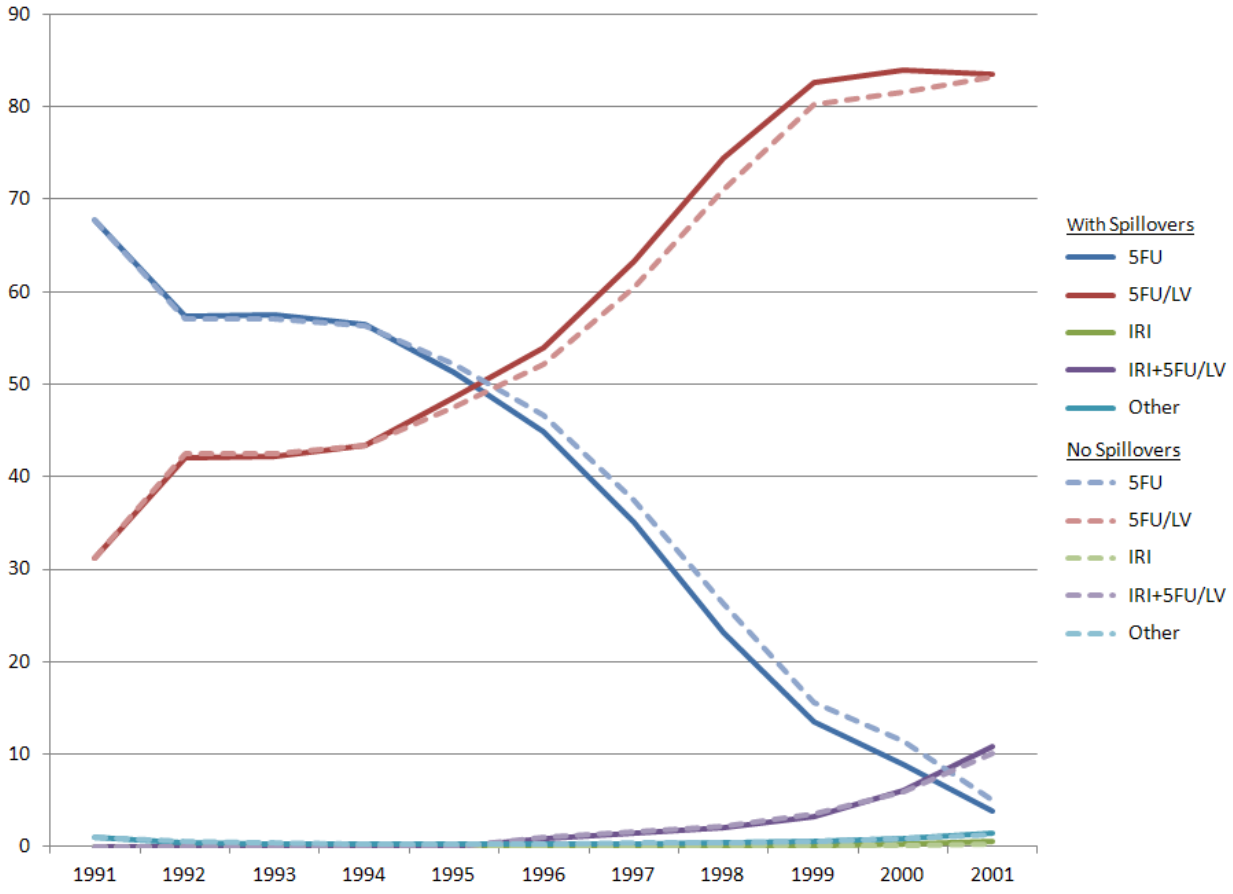
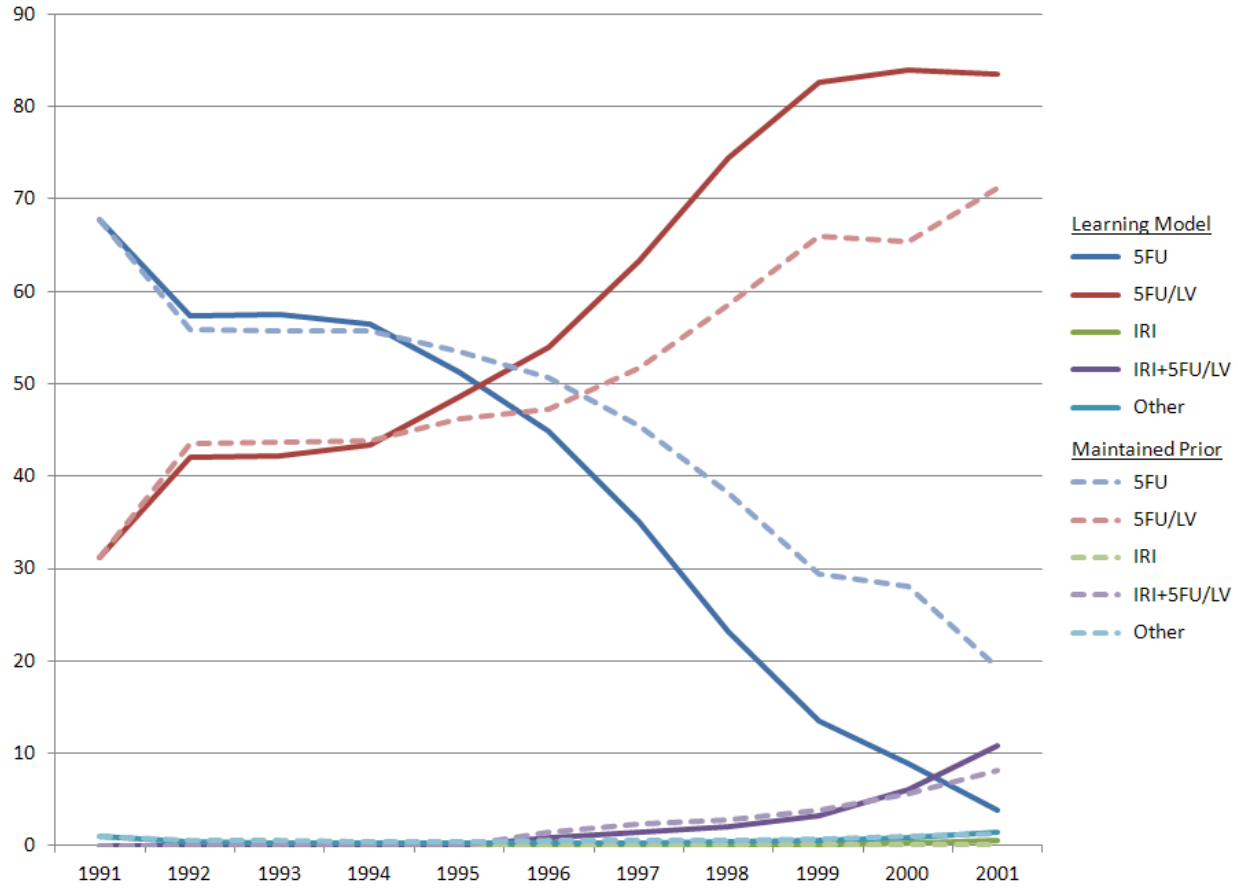


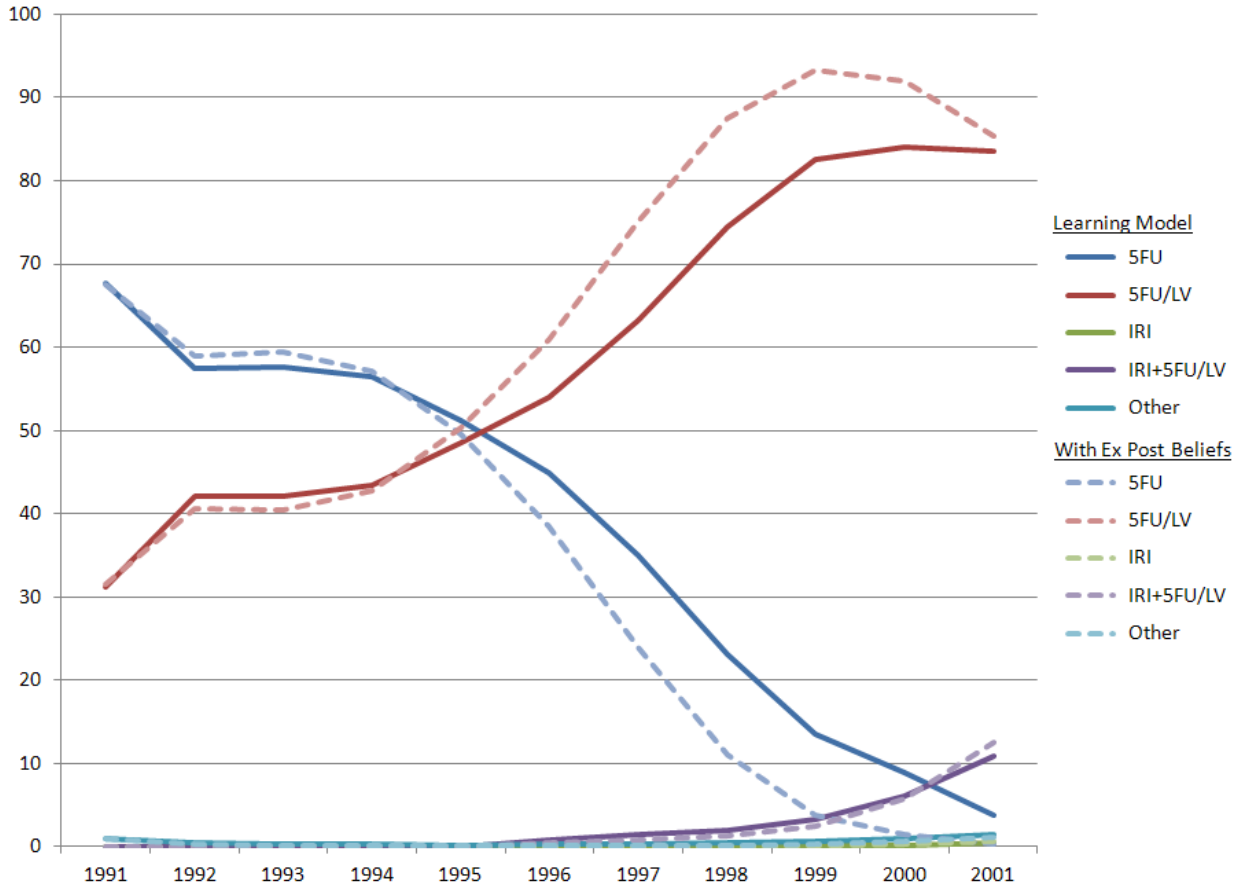
Figure 1.6: Counterfactual Regimen Shares With Maintained Prior Beliefs



the estimates, and as Figure 1.5 shows, the difference between strict product learning and across-alternative learning is small with respect to how regimen shares shift. Although the market for colorectal chemotherapy used a variety of single- and multiple-drug regimens for therapy between 1991 and 2001, physicians' beliefs about how those drugs were related reduces the magnitude of the effect of spillovers. Nevertheless, since physicians do hold statistically significant beliefs about the initial covariances of the quality of regimens, models of treatment choice, and consumer choice in general, should capture these covariances to preclude estimation bias from misspecification.

Although the market for colorectal chemotherapy displayed limited use of perception

Figure 1.7: Counterfactual Regimen Shares With Ex Post Knowledge



spillovers between 1991 and 2001, the second and third counterfactual simulations confirm that experience was a significant source of information over the period. As presented in Figure 1.6, if physicians do not learn from experience, but maintain their initial beliefs over the entire period, a substantial fraction of patients continue to be prescribed 5FU despite its inferiority. Because physicians are not learning about 5FU/LV at the same time, IRI is briefly adopted more quickly than if learning occurs, but later falls below the actual rate because physicians are so hesitant to move away from 5FU. Although learning from experience greatly speeds up the adoption rate, it is not a substitute for greater knowledge. In Figure 1.7, regimen shares are simulated under the assumption that physicians start the treatment period with the ex post beliefs estimated in the learning model. 5FU and 5FU/LV reach their actual market shares roughly a year before they are predicted in the learning model. By the end of the data, IRI is also predicted to have a 16 percent greater market share. These counterfactuals require the assumption that the extensive margin of chemotherapy would have remained the same under the alternative belief regimes, though the data suggest the extensive trend does not change over the period.

Using overall survival data from clinical trials and the FDA's original drug approvals, I can estimate the survival value of learning within each of these counterfactuals based on market share. Given the prescribed regimens and their average survival duration in clinical trials, the sample patients survived an estimated 6913.577 years, or 8.068 months per patient. Disallowing perception spillovers as in the first counterfactual shows that patients would live 0.045 months less if spillovers were not a source of information to physicians. The bigger survival difference is a 236.979 year increase learning gives over the second counterfactual of maintained priors, amounting to a 0.277 month, or 3.43 percent, increase in survival per patient from learning. If instead, physicians had their ex post beliefs at the beginning, the sample patients would have survived an additional 145.360 years, or 0.170 months, under the third counterfactual. In sum, changes to a physician's practice pattern

from learning reduces the cost of uncertainty by 62.0 percent. Learning from experience therefore enables physicians to overcome their initial uncertainty about a regimen due to imperfect information, a relatively small portion of which comes from perception spillovers. Nevertheless, shares would still be greater for the superior regimens if physicians had even more information available.

1.8 Conclusions

In this paper, I identify a market in which knowledge gleaned from experience with a particular bundle of goods is informative about the value of other alternative bundles. I estimates a model of learning allowing for these perception spillovers using physicians and their chemotherapy choices during a period of expanding treatment alternatives. As a particular example of consumer learning, this work contributes to the understanding of physician decision-making and consumer information processing. If information flow across alternatives indeed influences subsequent choices, empirical models which ignore this behavior will be misspecified and suffer from unobserved correlation between choices. Since a restricted spillover model can be nested within my proposed framework, however, this model may serve as a starting point for future research.

Employing a data set of the complete Medicare patient load for a set of physicians treating colorectal cancer between 1991 and 2001, I estimate the model of Bayesian learning and find physicians do learn across consumption alternatives, evidenced by statistically significant non-zero covariances within their belief distributions. The results show that physicians are initially skeptical about the spillovers between incumbent goods and market entrants even when those goods are used in combination, as seen with the low correlation between 5FU and 5FU/LV, and 5FU/LV and IRI+5FU/LV. Regimens which shared the new drug IRI were nevertheless considered highly correlated by physicians, indicating that quality beliefs

for both regimens were updated when new information for either regimen was received.

Together these results present a picture of how physicians reacted to pharmaceutical innovation: they considered the addition of a new drug to an established regimen as a completely distinct therapy, although regimens sharing the a new drug were informative for each other. After the sample period, the market for colorectal chemotherapy continued to grow with the approval of two additional drugs over the next three years and two more two years later, bringing the total number of treatment regimens to 14 by the end of 2006. In contrast to the regimens introduced in the current sample, one of subsequent entrants, oxaliplatin, is used exclusively in combination with previously available regimens. Estimation is currently in progress to include this other type of entrant in the model to further understand the adoption patterns of physicians.

The importance of learning and perception spillovers are explored in several counterfactual simulations after estimation of the model. From these, results indicate that learning about the individual product prescribed accounts for most of the changes to regimen shares over time. Perception spillovers do accelerate adoption slightly in this market, but the difference between a model which account for learning and one holding prior beliefs constant is far greater. In fact, comparing this difference to the difference between the learning model I employ and simulated market shares using physician's ex post beliefs over the entirety of the treatment period reveals that learning accelerates regimen share use by more than 60 percent.

Despite the rich dataset used in this analysis, however, not all potential sources of variation are captured within the model. The sample is limited to chemotherapy after a patient's first cancer diagnosis to reduce informational heterogeneity between cases. Physicians may also learn from recurrent cancer in patients, but how the information available to physicians in such cases would influence beliefs for initial patient treatment is unclear. An additional limitation is the absence of information on younger patients with private insurance. Previous

research has shown that older patients, that is, those on Medicare, are treated differently on the extensive margin of chemotherapy. If they are treated differently on the intensive margin as well, the estimates presented here for physician learning on older Americans are not biased by the exclusion of younger patients as sources of information for physicians. Finally, a potential source of unobserved information is likely to come from detailing. If physicians experience differential detailing so that heterogeneity is missed in the time trend then pharmaceutical advertising remains a concern. Future research could address these issues and further explain how consumers learn and process information.

Criticisms of the model itself include the lack of complete profit data and reliance on a time trend to capture relevant changes to universal medical knowledge. Ongoing work will address these issues by incorporating complete profit data as well as journal article counts for the Journal of Clinical Oncology, Journal of the American Medical Association, and the New England Journal of Medicine. As the number of articles using regimens as either the experimental or control arm of a clinical trial increases, I will assume the increased knowledge of outcomes will translate into changed probabilities of receiving each regimen.

Two final concerns regard the estimation of the model. First, the estimated standard errors here are almost surely attenuated, leading to false results of statistical significance. This attenuation stems from need to simulate the likelihood. A working paper by Lee and Song presents a modified simulated likelihood which improves inference and future work will incorporate this method. Second, an implied assumption of the model is that all physicians begin the treatment of their patients with the same set of beliefs regardless of when they actually begin practicing. That is, a physician who starts practicing in 1998 is assumed to have the same initial beliefs as a physician who begins practicing in 1992, despite significant changes in to the market structure and knowledge base. This concern will be partially mitigated by the inclusion of medical article counts, but also by the creation of an artificial physician that sees an average number of randomly selected patients and "learns" over

time as any actual physician would. The artificial physician's beliefs are then assigned to the new physician at the time of her first practice. This procedure will allow for the differentiation of initial beliefs by time but preserve the inherent uncertainty of beliefs. With these concerns addressed, this research will contribute an improved empirical method for estimating consumer behavior and suggest policies that might improve patient welfare through changes to physician decision-making.

CHAPTER 2

ERRORS IN RETROSPECTIVE DATA IN SMOKING: COMPARING MAXIMUM LIKELIHOOD AND AD HOC APPROACHES

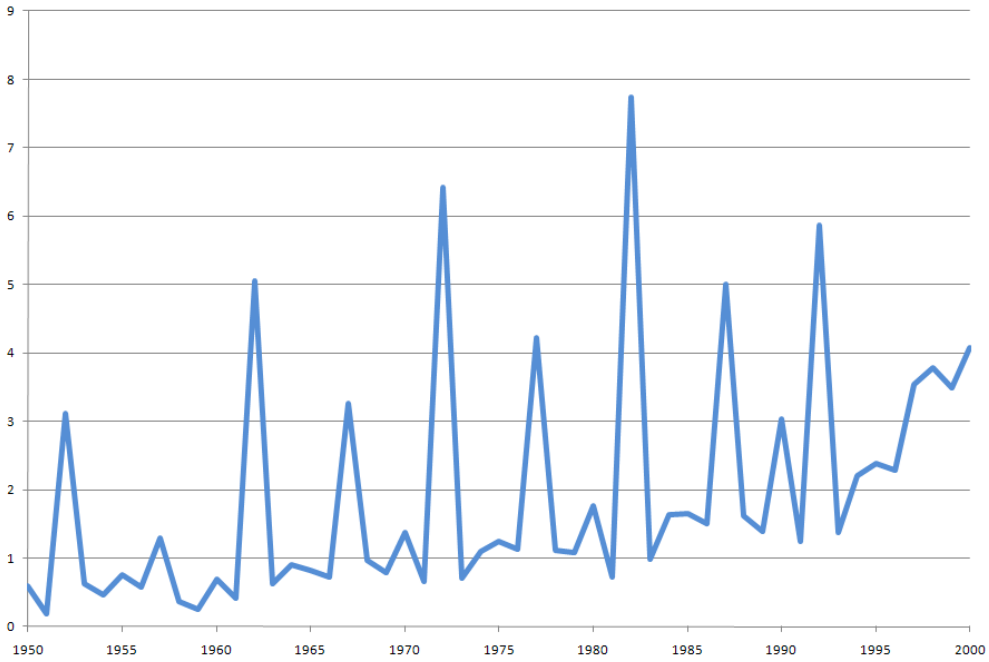
This research is joint with Donald Kenkel (Cornell University and NBER) and Feng Liu (Canadian Institute for Health Information), and supported by Award # R01 HD048828 from the National Institutes of Health.

2.1 Introduction

A number of cross-sectional and longitudinal surveys include retrospective questions about the timing of smoking initiation and cessation.[1] These data appear to offer the opportunity to explore the dynamics of smoking over relatively long time periods. For example, the current prevalence of U.S. adult smoking reflects smoking initiation and cessation decisions made over multiple decades and over very wide ranges of cigarette taxes. However, this research opportunity might be partly illusory because of errors in retrospectively reported data on smoking. On the one hand, contemporaneous self-reports of smoking status are generally reliable (Patrick et al., 1994), and other studies find substantial agreement between contemporaneous and retrospective reports of smoking status (Machlin et al., 1989; Kenkel et al., 2003). On the other hand, as in other types of retrospectively reported data, retrospective data on smoking cessation show heaping on round numbers. A smoker is much more likely to report having quit a round number of years ago (such as 10) than to report an odd number

[1]Retrospective questions about starting and quitting smoking are included in the cross-sectional U.S. National Health Interview Survey and the Tobacco Use Supplements to the Current Population Survey. Retrospective questions about smoking are also included in recent waves of: three of the samples of the National Longitudinal Surveys Original Cohorts; the National Longitudinal Survey of Youth 1979; the Panel Study of Income Dynamics; and the Health and Retirement Study. Similar questions are included in the British Household Panel Survey, the German Socio-economic Panel Survey, the Russian Longitudinal Monitoring Survey, and the China Health and Nutrition Survey.

Figure 2.1: Cessation Rate Over Time, TUS Survey Year 2002



of years ago (such as 9 or 11).[2]

Data from the 2002 Tobacco Use Supplement to the Current Population Survey (TUS-CPS) provide a typical example of heaping in retrospective reports on smoking cessation. The TUS-CPS includes questions that identify ever smokers, current smokers, and former smokers. Former smokers are asked: "About how long has it been since you last smoked cigarettes every day?" We convert these responses to measure the calendar year of reported cessation. Figure 2.1 shows the implied cessation rate over time. The data show pronounced

[2]The evidence on the reliability of contemporaneous reports of smoking status comes from studies that compare self-reports to biochemical markers of smoking (Patrick et al., 1994). Note that it is possible for retrospective reports of smoking status to be fairly accurate, as Machlin et al. (1989) and Kenkel et al. (2003) find, even when heaping in the reported date of cessation is common. Consider for example someone who: smoked for 20 years; is surveyed 11 years after he quit; and in the survey retrospectively heaps his date of cessation as being 10 years ago. Out of the 30 year period after he started smoking, his smoking status is correctly classified 97% of the time – his smoking status is only misclassified in the one year he heaped away from. Whether retrospective data on smoking is accurate enough depends on the research question. Below we show that heaping at rates comparable to observed data results in substantial bias in the estimated parameters of a discrete-time hazard model of smoking cessation.

heaping at 5 and 10 years intervals before the survey year of 2002. For example, the cessation rate in 1992 — exactly 10 years before the survey — is almost 6%, compared to less than 1.5% in 1991 and 1993. The cessation rate suggests similar heaping in 1987, 1982, 1977, and so on. In data from the 2003 TUS-CPS (not shown but available upon request), the heaping is shifted forward one year, with much higher cessation rates in 1993, 1988, 1983, 1978, and so on.

Because we are interested in modeling the determinants of smoking cessation, the heaping problem is an example of a mismeasured dependent variable. Our study contributes to a line of applied econometrics research on similar measurement problems in labor and health economics. In the ordinary least squares context, classical measurement error in the dependent variable is fairly innocuous; it simply leads to less statistical precision in estimation (Hausman, 2001). In non-linear models, however, mismeasurement of the dependent variable leads to estimators that are biased and inconsistent. Poterba and Summers (1995) study misclassification errors in a multinomial logit model of employment status. Hausman, Abrevaya, and Scott-Morton (1998) consider misclassification error in a probit model of job change. Kenkel, Lillard, and Mathios (2004) apply the Hausman et al. (1998) approach to a probit model of smoking participation. Keane and Sauer (2009) incorporate misclassification error into a dynamic discrete choice model of female labor supply. Turning to non-linear models with continuous dependent variables, Torelli and Trivellato (1993) and Abrevaya and Hausman (1999) consider measurement error in data on the duration of unemployment. Forster and Jones (2001) apply the Torelli and Trivellato (1993) approach to data on the duration of smoking.

Our key insight is that in a discrete-time hazard model, heaping can be viewed as a special case of misclassification error. The probabilities of a true 0 being misclassified as a 1, and vice versa, depend upon whether or not the discrete-time period is a round number of years before the survey. Hausman et al. (1998) propose an adjusted maximum likelihood estimator

(MLE) that corrects for misclassification error in a probit model. We tailor their approach to fit the problem of estimating a discrete-time hazard model with misclassification error due to heaping. Torelli and Trivellato (1993) and Forster and Jones (2001) consider heaping in continuous-time duration models. We connect this line of research to the Hausman et al. (1998) model of misclassification error, moving from a continuous-time duration model to a discrete-time hazard model (Allison, 1982; Jenkins, 1995). In our application to retrospective data on smoking cessation, the event of interest seems more naturally modeled in discrete time.[3] Moreover, after the at-risk sample is formed, the discrete-time hazard model only requires estimating a probit model, which is familiar ground for empirical microeconomists. We contribute a straight-forward extension of the familiar probit model that allows for misclassification error due to heaping.

We compare the maximum likelihood approach to several ad hoc approaches with intuitive appeal. The first ad hoc approach includes an indicator for years in which heaping is likely as an additional control variable in the discrete-time hazard model. The second ad hoc approach "coarsens" the data by changing the unit of analysis in the discrete-time model from a period of one year to five years. The third ad hoc approach "decimates" the data by eliminating all observations from respondents who report cessation in a heaping year.[4]

We conduct Monte Carlo simulations to compare the relative performance of the adjusted MLE to the ad hoc approaches. We simulate data with reported cessation heaped on years that are divisible by five. We assume that 20% of people who quit smoking in any non-round year heap their reported quit to the nearest 5-year period. With this degree of heaping, a naïve probit model that fails to account for heaping yields coefficients that are on average

[3] As Allison (1984) notes, "when the time units are large - months, years, or decades - it is more appropriate to use discrete-time methods."

[4] The word "decimate" originates from a form of military discipline used in the Roman army, where one in every ten soldiers in an offending unit would be chosen by lot to be executed, often by bludgeoning. In our application reported below, using the 2002 TUS-CPS data our version of decimation eliminates about four of every ten person-year observations.

biased downwards (towards zero) by 20–65% of the true values used to generate the data. The adjusted MLE yields coefficients that are on average very close to the true values. Of the ad hoc methods, the decimation approach yields coefficients that are on average close to their true values. The decimation approach does less well in recovering the time trend. The other ad hoc approaches perform poorly across the board.

We also conduct two additional exercises. First, we apply the alternative approaches to real data on smoking cessation from the 2002 TUS-CPS. Although we no longer have benchmark true values for the parameters, we note that a naïve probit that fails to account for heaping yields coefficients on price, schooling, and income that are 20% smaller than the coefficients from the adjusted MLE. However, other results suggest the actual heaping in the real data might be more complicated than allowed for in the assumptions of the adjusted MLE. In our last exercise, we compare the alternative approaches in simulated data with more complicated heaping patterns than allowed for in the adjusted MLE. The MLE and the decimation corrections perform well in the presence of uneven heaping. If there is additional heaping on calendar-decades, these approaches only improve on the bias by roughly half.

2.2 A Maximum Likelihood Approach to Misclassification Error Due to Heaping

Misclassification Error in a Cross-Sectional Probit Model

We begin by reviewing Hausman, Abrevaya, and Scott-Morton's (1998) adjusted maximum likelihood estimator that corrects for misclassification error. Equation (2.1) describes an empirical demand function for smoking cessation. Let y_i^* be the latent variable showing the net benefits of smoking cessation as a function of observable determinants $\mathbf{x}_i$ and a random

disturbance term ε_i:

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \tag{2.1}$$

The net benefits of smoking cessation y_i^* could be formally defined as the difference in lifetime utility from quitting and the lifetime utility from continuing to smoke (Becker and Murphy, 1988; Suranovic et al., 1999; Jones, 1999). In contrast to equation (2.1), a standard approach in health economics is to estimate a model of smoking participation, and sometimes a second-part model of the quantity of cigarettes smoked conditional upon smoking participation (see Chaloupka and Warner (2000) for a review). As DeCicca, Kenkel, and Mathios (2008) discuss in more detail, the addictive nature of smoking makes it important to model smoking initiation and cessation as distinct behaviors.[5]

The individual quits smoking when the net benefits of quitting are positive, so letting $\tilde{y}_i$ be the true response (true smoking cessation)

$$\tilde{y}_i = \begin{cases} 1, & \text{if } y_i^* > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

Let y_i be the observed dependent variable which takes the value of 1 if the individual reports smoking cessation and 0 otherwise. Let α_0 be the probability that a true 0 is misclassified as a 1,

$$\alpha_0 = \Pr(y_i = 1 | \tilde{y}_i = 0) \tag{2.3}$$

and let α_1 be the probability that a true 1 is misclassified as a 0,

$$\alpha_1 = \Pr(y_i = 0 | \tilde{y}_i = 1). \tag{2.4}$$

[5]DeCicca et al. (2008) show that a myopic addiction model leads to distinct models of smoking initiation and cessation. The specification of the standard model of smoking cessation is correct only if smoking is not addictive: it implicitly imposes the testable restriction that the explanatory variables should have symmetric effects on initiation and cessation. Not surprisingly, in their empirical analysis DeCicca et al. reject the hypothesis that smoking is not addictive. DeCicca et al. provide some discussion of how the approach could be extended to allow for rational addiction or different degrees of addiction. These extensions are beyond the scope of this paper, but in the conclusion we discuss how future work might usefully explore duration dependence, which is related to the degree of addiction.

The expected values of the true smoking cessation, $\tilde{y}_i$, and observed smoking cessation, y_i, variables are

$$E(\tilde{y}_i|\mathbf{x}_i) = \Pr(\tilde{y}_i = 1|\mathbf{x}_i) = F(\mathbf{x}_i'\boldsymbol{\beta}) \tag{2.5}$$

$$E(y_i|\mathbf{x}_i) = \Pr(y_i = 1|\mathbf{x}_i) = \alpha_0 + (1 - \alpha_0 - \alpha_1)F(\mathbf{x}_i'\boldsymbol{\beta}). \tag{2.6}$$

Typically, we are interested in the partial derivative of equation (2.5) with respect to a particular independent variable x_{ij}; for example, we might be interested in the effect of a marginal change in the price of cigarettes on the probability of true smoking cessation. However, comparing the partial derivative of equation (2.5) with the partial derivative of equation (2.6) shows that with misclassification error, the estimated coefficients of interest will be biased towards zero:

$$\frac{\partial \Pr(\tilde{y}_i = 1|\mathbf{x}_i)}{\partial x_{ij}} = f(\mathbf{x}_i'\boldsymbol{\beta})\beta_j \tag{2.7}$$

$$\frac{\partial \Pr(y_i = 1|\mathbf{x}_i)}{\partial x_{ij}} = (1 - \alpha_0 - \alpha_1)f(\mathbf{x}_i'\boldsymbol{\beta})\beta_j. \tag{2.8}$$

Following Hausman et al. (1998), assuming that the distribution function $F()$ is known (e.g. normal or logistic) we can estimate $(\alpha_0, \alpha_1, \boldsymbol{\beta})$ by maximum likelihood estimation (MLE), based on the log-likelihood function

$$\begin{aligned}\mathcal{L}(a_0, a_1, \mathbf{b}) =& \frac{1}{n}\sum_{i=1}^{n}\big\{y_i \ln(a_0 + (1 - a_0 - a_1)F(\mathbf{x}_i'\mathbf{b})) \\ &+ (1 - y_i)\ln(1 - a_0 - (1 - a_0 - a_1)F(\mathbf{x}_i'\mathbf{b}))\big\}.\end{aligned} \tag{2.9}$$

In this approach, the misclassification probabilities α_0 and α_1 are estimable parameters: significance tests on a_0 and a_1 provide tests for misclassification error.

Hausman, Abrevaya, and Scott-Morton (1998) point out that identification of the model parameters stems from the nonlinearity of the $F()$, so that in the linear probability model the parameters are not separately identified. They discuss an additional monotonicity condition

required for identification because for a symmetric $F()$ like the normal or logistic, the MLE estimator cannot distinguish between the parameter values $(\alpha_0, \alpha_1, \boldsymbol{\beta})$ and $(1-\alpha_0, 1-\alpha_1, -\boldsymbol{\beta})$. Imposing the monotonicity assumption that the sum of the misclassification probabilities must be less than one $(\alpha_0 + \alpha_1 < 1)$ rules out the second possible set of parameter values.

To shed light on the empirical importance of even modest misclassification, Hausman et al. (1998) report Monte Carlo simulation results. They consider symmetric misclassification probabilities $(\alpha_0 = \alpha_1)$ of 2%, 5%, and 20%. Even with $\alpha_0 = \alpha_1 = 0.02$, ordinary probit estimates are under-estimated by 15–25%. When misclassification error is more common, the ordinary probit estimates are underestimated by 62–81%. In contrast, their modified MLE results based on equation (2.9) yield estimates of α_0 and α_1 (restricted to be equal) and $\boldsymbol{\beta}$ that are very close to the parameters used to generate the simulated data.

Misclassification Error Due to Heaping

We focus on estimating a discrete-time hazard model of smoking cessation. The basic model of the demand for smoking cessation is of the same form given by equation (2.1). The dependent variable equals 0 in every year a person smokes and equals 1 in the year a smoker quits. A person is in the sample every year she is at risk of quitting.

Heaping leads to a specific form of misclassification error in a discrete time hazard model. We assume that there are two types of years: heaped years (such as 10) where reported cessation over-states true cessation; and nonheaped years (such as 9 and 11) where reported cessation under-states true cessation. Using the notation introduced above and subscripting with "H" and "N" to denote heaped and nonheaped years, respectively, we specifically assume that in a nonheaped year there is a positive probability that a true 1 is misclassified as a 0: $\alpha_{1,N} > 0$. We further assume that there is no probability that a true 0 is misclassified as a 1: $\alpha_{0,N} = 0$. This is based on the idea that no one mistakenly reports that he or she quit

smoking an odd number of years ago, for example 9 or 11 years ago. Conversely, for a heaped year we assume that $\alpha_{1,H} = 0$ but $\alpha_{0,H} > 0$. The former assumption is based on the idea that no one mistakenly misclassifies that he or she quit smoking a heaped number of years ago, for example, exactly 10 years ago, while the latter is positive to account mathematically for the misreported quits from nonheaped years. An intuitive example is described in section 2.4.

To augment the likelihood function in equation (2.9) for this setting, let T_i be the number of periods person i is observed at risk of quitting and n be the total number of individuals. Further, define

$$a_{0,H} = a_0 \times 1\lceil t \text{ is a heaped year} \rfloor \tag{2.10}$$

$$a_{1,N} = a_1 \times 1\lceil t \text{ is a nonheaped year} \rfloor \tag{2.11}$$

where $1\lceil\rfloor$ is the indicator function equal to 1 if its argument is true and 0 otherwise. The estimated log-likelihood is thus

$$\begin{aligned}\mathcal{L}(a_0, a_1, \mathbf{b}) = \left(\frac{1}{\sum_{i=1}^{n} T_i}\right) \sum_{i=1}^{n} \sum_{t=1}^{T_i} \big\{ & y_{it} \ln(a_{0,H} + (1 - a_{0,H} - a_{1,N}) F(\mathbf{x}_{it}'\mathbf{b})) \\ & + (1 - y_{it}) \ln(1 - a_{0,H} - (1 - a_{0,H} - a_{1,N}) F(\mathbf{x}_{it}'\mathbf{b})) \big\}\end{aligned} \tag{2.12}$$

with the index function containing any time characteristics needed for modeling the hazard function.[6]

Equation (2.12) is a special case of the adjusted MLE approach of Hausman et al. (1998), tailored to fit the problem of estimating a discrete time hazard model with misclassification error due to heaping. Where the Hausman et al. model is technically identified through non-linearities, we exploit exclusion restrictions based on patterns of heaping. In our ap-

[6]To ensure the maximization routine satisfies the monotonicity assumption in practice, the misclassification probabilities are maximized within a logistic framework. That is, for both $a_{0,H}$ and $a_{1,N}$, the estimated parameter is λ_k such that $a_k = \frac{\exp(\lambda_k)}{1+\exp(\lambda_{0,H})+\exp(\lambda_{1,H})}$. λ_k may take any real value while ensuring that both misclassification probabilities are positive and sum to less than one.

plication, identification is based on the argument that heaping changes the probabilities of misclassification error in predictable ways, but should not enter the true hazard rate.

2.3 Ad Hoc Approaches to Misclassification Error Due to Heaping

In addition to the adjusted maximum likelihood approach, we evaluate two ad hoc approaches previously suggested, as well as a novel ad hoc approach to limit the potential influence of misclassification. The first previously suggested ad hoc approach creates an indicator for years in which heaping is more likely and includes the dummy variable as an additional covariate in the discrete time hazard model. For example, Torelli and Trivellato (1993) and Forster and Jones (2001) suggest this approach in the context of a continuous-time duration model. Intuitively, the hope is that by controlling for the increase in the unconditional hazard for those periods, this approach will reduce any bias resulting from heaping. The hazard equation that includes an indicator for a heaping year can be thought of as a reduced-form version of the structural model described by equations (2.1), (2.3), and (2.4). The indicator for a heaping year serves as an instrumental variable for the misclassification probabilities given by equations (2.3) and (2.4). Under this interpretation, the ad hoc approach relies on the hope that it might be possible to approximately recover the parameters of the structural equation (2.1) directly from the reduced-form. However, in addition to the non-linearity of the hazard probability, from equation (2.6) the misclassification probabilities enter non-linearly as well.

The next previously suggested approach "coarsens" the data by changing the unit of analysis from an annual basis to the period of years around each heaping point. The goal of this approach is to eliminate the effect of heaping by eliminating heaps and instead focusing on the cessations over wider intervals, made simple by the discreteness of the hazard function. A line of statistics research focuses on the data coarsening that results from rounding in

survey responses (e.g. Heitjan and Rubin, 1990; Manski and Molinari, 2010). In our context, although heaping results in coarse data on the timing of smoking cessation, we typically have finer annual data on independent variables such as cigarette prices and income. Our ad hoc approach is to coarsen the independent variables to match the apparent coarseness of the dependent variable (smoking cessation).[7] Given the data, however, this may be accomplished in two ways. If the data end at the event, then no further information can be observed and averaged over the time period around the heaped year. For example, income is unobservable after an event such as death, so only observations in the years up to the event are averaged. This procedure will tend to under- or over-state the average of trending variables such as price. However, in the case more typical of smoking cessation, variables from panel data sets, such as income, are still observable after quitting, so they may be included in the average for the period. In our analysis, we refer to the latter case as "complete data, coarsened," whereas "heaped data, coarsened" denotes the case where data are unseen after the event.

Our novel ad hoc approach "decimates" the data by eliminating all observations from respondents who report cessation in a heaped year. Again the hope is to reduce bias at the cost of information. While this approach distorts the observed rate of smoking cessation, it eliminates all misclassification error due to heaping and so might reduce bias in the estimated model coefficients.[8] Decimation can be viewed as producing an endogenously stratified sample from the population. The strata corresponding to heaped quit years are

[7]The coarseness of the typical data on smoking cessation is perhaps not surprising, because a typical form of the survey question asks "*About* how long has it been since you last smoked cigarettes every day?" (emphasis added). To further complicate matters, in typical smoking cessation data it appears that some respondents round to annual data while others round to half-decades. Manski and Molinari (2010) propose to use response patterns across different survey questions to infer respondents' rounding practice. Compared to more complicated and sophisticated approaches, the ad hoc approach of coarsening discards potentially useful information.

[8]To deal with measurement error in reported birth weights, Barreca et al. (2010) propose an approach similar to what we call decimation. They find that birth weights are disproportionately represented at multiples of round numbers, especially for children of lower socioeconomic status mothers. Their focus is on the implications for regression discontinuity (RD) designs where birth weight is the running variable. They suggest that a straightforward approach to deal with the problem is what they call a "donut RD": "dropping observations coinciding with heaps in the running variable (i.e. 100-gram and ounce multiples)."

missing from that sample though all non-heaped years are perfectly observed. Because the subset of outcomes corresponding to the heaped years are omitted by decimation, the distribution of independent variables around heaped years approximates the density of those variables across all years. This limits the bias introduced by treating the decimated sample as if it were random when we estimate the hazard model on the remaining sample.[9]

2.4 Monte Carlo Methods

We conduct a Monte Carlo simulation to assess the impact of misclassification due to heaping, and to compare the adjusted MLE approach to the ad hoc approaches. The goal of the Monte Carlo design is to mimic data on smoking cessation that might come from a cross-sectional survey that asked former smokers: "About how many years ago did you quit smoking?" We start with a simulated sample of 7,500 smokers of different ages who are in the sample at-risk of smoking cessation for up to 27 years. We assume each person started smoking, and thus entered the at-risk sample, in the same calendar year, but at different ages.[10] Because we follow the discrete-time event analysis developed by Allison (1982), a traditional latent dependent variable model as given by equation (2.1) is constructed for each of the possible 27 periods. An individual "quits" when the latent index first exceeds zero. The individual remains in the at-risk sample for every period until the quit occurs. We perform 250 Monte Carlo simulations.[11]

[9]The statistical literature has developed estimation techniques to analyze stratified data and Cosslett (1993) provides a thorough review. The general solutions are (i) supplementing the sample with exogenous information, and (ii) jointly estimating the marginal probabilities of the outcomes as part of a pseudo-likelihood. We appeal to one which introduces exogenous information on the distribution of independent variables and reweights the likelihood to obtain consistent estimates. Using the methods in Cosslett (1993), one could fully correct the likelihood by replacing the sampling density with its nonparametric maximum likelihood estimator. The current analysis seeks approaches readily available to applied economists to reduce the bias caused by heaping while acknowledging more rigorous, though potentially also more difficult, solutions exist.

[10]This is a simplifying assumption to avoid simulating the calendar year of initiation.

[11]Results will be from the 248 simulations which achieve convergence.

The two panels of Figure 2.2 show some of the main features of the simulated data on smoking cessation. Most smoking initiation occurs in the late teens and early 20s, while smoking cessation is most common in middle age. Although the simulated data are not calibrated to actual data, they roughly mimic known patterns of smoking initiation and cessation over the life cycle.

We assume that the latent index that determines smoking cessation is a function of six covariates, $x1 - x6$. We use a variety of distributions to simulate the covariates; we include time-varying and time-invariant variables, and some variables are chosen to mimic basic determinants of smoking cessation. The first two covariates are time-varying: $x1_{it}$ is normally distribution with mean 2 and variance 1; $x2_{it}$ is drawn from a uniform distribution on the interval $[0, 5]$. The third covariate mimics age: $x3_{it}$ increases by one each period, with initial distribution $25 \times \beta\,[1.25, 3.5] + 16$.[12] The next two covariates are time-invariant; $x4_i$ is drawn from a lognormal distribution with parameters $\mu = 1$ and $\sigma^2 = \frac{1}{9}$; $x5_i$ is an indicator, like gender, taking on value 1 with probability $p = 0.45$. The last covariate, $x6_{it}$, is intended to mimic the price of cigarettes and stochastically grows over time. $x6_{it}$ is generated in two steps: in the first step, period 1 prices are distributed uniformly over $[2.5, 6]$, and in the second step, the price is updated each subsequent period by adding a random amount distributed uniformly over the interval $[-0.1, 0.5]$. The error ε_{it} is drawn from a standard normal distribution.

Including a cubic time function, the latent dependent variable is

$$\begin{aligned} y^*_{it} = {} & 1.5x1_{it} - 2.0x2_{it} + 0.1x3_{it} - 2.5x4_i + 0.3x5_i + 0.7x6_{it} \\ & - 3.420 + 0.1900t - 0.0123t^2 + 0.000384t^3 + \varepsilon_{it} \end{aligned} \tag{2.13}$$

The coefficients on the time function were chosen so as to have an unconditional hazard function increasing over time, as we believe the unconditional hazard for smoking cessation

[12] A floor function is used to ensure ages are whole numbers. The resulting distribution of start ages ranges from 16 to 40.

Figure 2.2: Distribution of Simulated Covariates

does. Here, the selected distribution of unconditional event times is $28B$, where B is drawn from beta distribution $\beta(1.2, 1.8)$.

The example misclassification rule we use, labeled "random heaping," posits that individuals correctly report if their quit occurred in the past two years or any year which is a multiple of 5, but that 20% of all people who quit in years which are not a multiple of 5 randomly heap to the nearest 5-year period. That is, in the 27 year window, someone who quits in years 1, 2, 5, 10, 15, 20, or 25 correctly reports the year in which they quit, but there is a 20% chance that a person who quits, for example, in year 13 will report quitting in year 15, or that someone quitting in year 12 will report quitting in year 10. In the parlance of Hausman et al., this is akin to assuming $\alpha_0 = 0$ and $\alpha_1 = 0.20$ in nonheaped years, and $\alpha_1 = 0$ in heaped years. An example of how this misclassification scheme affects the observed data can be seen in figure 2.3.

When applying the methods to the data, the "naïve probit," "MLE correction," and "heaping indicator" models use all of the observed periods until the individual reports quitting in accordance with the misclassification rule. As described above, the "decimated sampling" model drops any person reporting quitting in a heaped year. The coarsening models both average all of an individual's covariates from the five-year period, the "complete data" model using all five years since individual observation is possible after quitting, whereas the "heaped data" model only averages up to the observed quit period in the raw data. In graphs these are "Full Coarsening" and "Coarsened Seen" respectively.

2.5 Simulation results

Tables 2.1 – 2.4 present the Monte Carlo simulation results for the different approaches to estimating the discrete-time hazard model with misclassification error due to heaping. Table 2.1 presents the results for the coefficients on the main covariates of interest $x1 - x6$,

Figure 2.3: Monte Carlo Simulated Quits Under Random Heaping

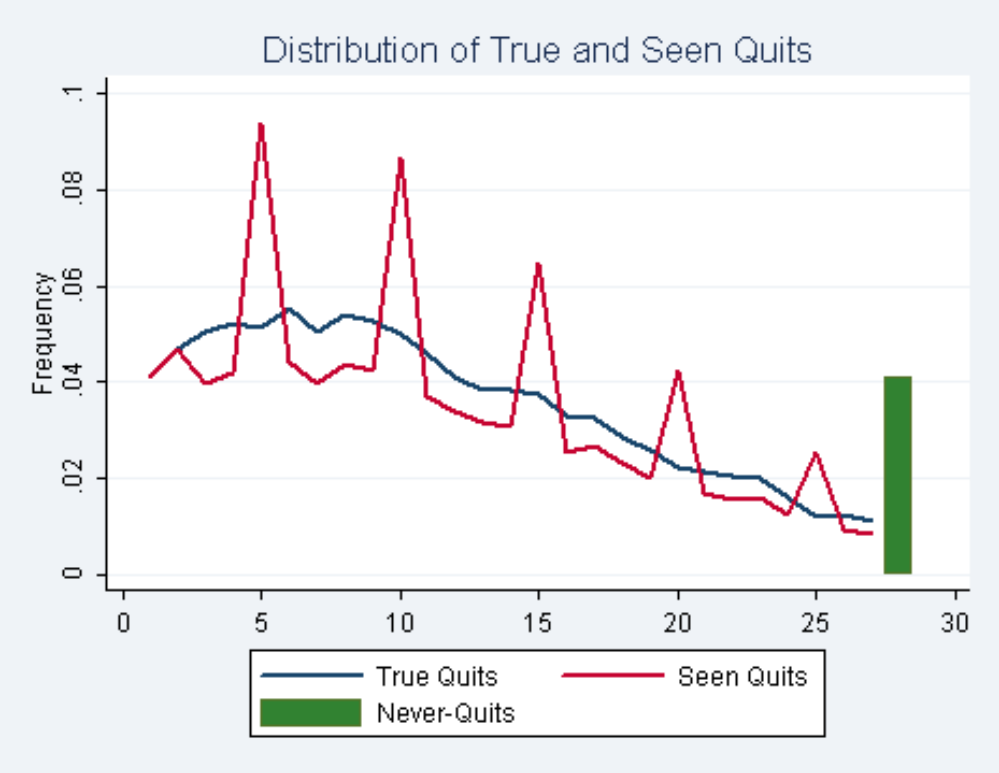


Table 2.1: Monte Carlo Simulation Results, Covariates

Model	Coefficient					
	β_1	β_2	β_3	β_4	β_5	β_6
True Value	1.5	-2.0	0.1	-2.5	0.3	0.7
True Probit	1.4994	-2.0000	0.1000	-2.5013	0.2995	0.7000
	(0.0192)	(0.0192)	(0.0030)	(0.0328)	(0.0266)	(0.0152)
	[0.0%]	[0.0%]	[0.0%]	[0.1%]	[0.2%]	[0.0%]
Naïve Probit	0.5946	-0.6928	0.0495	-1.2466	0.1485	0.3458
	(0.0133)	(0.0133)	(0.0018)	(0.0198)	(0.0182)	(0.0084)
	[60.4%]	[65.4%]	[50.5%]	[50.1%]	[50.5%]	[50.6%]
Heaping Indication	0.6618	-0.7812	0.0532	-1.3412	0.1596	0.3718
	(0.0131)	(0.0131)	(0.0020)	(0.0205)	(0.0195)	(0.0088)
	[55.9%]	[60.9%]	[46.8%]	[46.4%]	[46.8%]	[46.9%]
Decimated Sample	1.5033	-2.0042	0.1006	-2.5213	0.3016	0.7045
	(0.0249)	(0.0249)	(0.0036)	(0.0410)	(0.031)	(0.0178)
	[0.2%]	[0.2%]	[0.6%]	[0.9%]	[0.5%]	[0.6%]
MLE Correction	1.4683	-1.9469	0.1003	-2.5113	0.3005	0.7026
	(0.0244)	(0.0244)	(0.0035)	(0.0397)	(0.0293)	(0.0176)
	[2.1%]	[2.7%]	[0.3%]	[0.5%]	[0.2%]	[0.4%]
Complete Data,	0.8999	-1.0209	0.0444	-1.5323	0.1830	0.3775
Coarsened	(0.0204)	(0.0204)	(0.0026)	(0.0252)	(0.0259)	(0.0122)
	[40.0%]	[49.0%]	[55.6%]	[38.7%]	[39.0%]	[46.1%]
Heaped Data,	0.9738	-1.1076	0.0627	-1.6202	0.1927	0.4549
Coarsened	(0.0235)	(0.0235)	(0.0029)	(0.0235)	(0.0253)	(0.0123)
	[35.1%]	[44.6%]	[37.3%]	[35.2%]	[35.8%]	[35.0%]

See text for descriptions of the methods and Monte Carlo design. Results are from 248 simulations with $n = 7,500$. Standard deviations of the simulation results in parentheses; percent difference from true value in brackets.

Table 2.2: Monte Carlo Simulation Results, Time Trend

Model	Coefficient			
	β_t	β_{t^2}	β_{t^3}	β_0
Truth	0.1900	-0.0123	0.000384	-3.420
True Probit	0.1903	-0.0123	0.000384	-3.4171
	(0.0168)	(0.0015)	(0.00004)	(0.1092)
	[0.2%]	[0.1%]	[0.0%]	[0.1%]
Naïve Probit	0.1433	-0.0098	0.000272	-2.0550
	(0.0102)	(0.0010)	(0.00002)	(0.0659)
	[24.6%]	[20.6%]	[29.2%]	[39.9%]
Heaping Indication	0.0602	-0.0034	0.000135	-2.0771
	(0.0116)	(0.0011)	(0.00003)	(0.0711)
	[68.3%]	[72.3%]	[64.9%]	[39.3%]
Decimated Sample	0.1058	-0.0055	0.000220	-3.1153
	(0.0208)	(0.0019)	(0.00005)	(0.1281)
	[44.3%]	[55.1%]	[42.7%]	[8.9%]
MLE Correction	0.1717	-0.0109	0.000354	-3.3439
	(0.0181)	(0.0017)	(0.00004)	(0.1245)
	[9.6%]	[11.4%]	[7.9%]	[2.2%]

See text for descriptions of the methods and Monte Carlo design. Results are from 248 simulations with $n = 7,500$. Standard deviations of the simulation results in parentheses; percent difference from true value in brackets. The coarsened methods include period fixed effects instead of the time trend.

Table 2.3: Simulation Results, Additional Parameters

Model	Parameters	
MLE	$\alpha_{0,H}$	$\alpha_{1,N}$
Correction	0.0632	0.2157
	(0.0023)	(0.0121)
Heaping	$\beta_{HeapYear}$	
Indication	0.9441	
	(0.0191)	

Note: See text for descriptions of the methods and Monte Carlo design. Results are from 248 simulations with $n = 7,500$. The standard deviations of the simulation results are in parentheses.

Table 2.4: Monte Carlo Simulation Results, Marginal Effects

Model	Marginal Effect					
	x_1	x_2	x_3	x_4	x_5	x_6
True Probit	0.0619	-0.0826	0.0041	-0.1032	0.0123	0.0289
	(0.0007)	(0.0008)	(0.0001)	(0.0011)	(0.0010)	(0.0005)
Naïve Probit	0.0505	-0.0587	0.0041	-0.1056	0.0126	0.0293
	(0.0008)	(0.0008)	(0.0002)	(0.0012)	(0.0015)	(0.0007)
	[18.4%]	[29.0%]	[0.16%]	[2.28%]	[2.39%]	[1.19%]
Heaping Indication	0.0509	-0.0599	0.0041	-0.1028	0.0123	0.0285
	(0.0008)	(0.0008)	(0.0002)	(0.0012)	(0.0014)	(0.0007)
	[17.9%]	[27.5%]	[0.11%]	[0.46%]	[0.39%]	[1.44%]
Decimated Sample						
Estimation	0.0614	-0.0819	0.0041	-0.1027	0.0123	0.0287
Sample	(0.0009)	(0.0011)	(0.0002)	(0.0015)	(0.0013)	(0.0007)
	[0.86%]	[0.90%]	[0.76%]	[0.52%]	[0.00%]	[0.71%]
Full	0.0617	-0.0822	0.0041	-0.1032	0.0124	0.0289
Sample	(0.0008)	(0.0010)	(0.0002)	(0.0013)	(0.0013)	(0.0006)
	[0.40%]	[0.44%]	[0.03%]	[0.06%]	[0.46%]	[0.25%]
MLE Correction	0.0604	-0.0800	0.0041	-0.1031	0.0123	0.0289
	(0.0008)	(0.0009)	(0.0001)	(0.0013)	(0.0012)	(0.0176)
	[2.51%]	[3.16%]	[0.29%]	[0.16%]	[0.12%]	[0.23%]
Complete Data,	0.1202	-0.1363	0.0077	-0.1994	0.0241	0.0560
Coarsened	(0.0024)	(0.0020)	(0.0004)	(0.0021)	(0.0031)	(0.0014)
	[94.1%]	[65.0%]	[87.2%]	[93.1%]	[95.1%]	[93.5%]
Heaped Data,	0.1079	-0.1222	0.0053	-0.1832	0.0222	0.0451
Coarsened	(0.0021)	(0.0018)	(0.0003)	(0.0022)	(0.0031)	(0.0014)
	[74.3%]	[47.8%]	[28.8%]	[77.4%]	[79.9%]	[55.8%]

See text for descriptions of the methods and Monte Carlo design. Results are from 248 simulations with $n = 7,500$. Standard deviations of the simulation results in parentheses; percent non-rounded difference from true probit marginal effect in brackets.

Table 2.2 presents the coefficients on the time trend, Table 2.3 reports additional parameters estimated in two of the approaches, and Table 2.4 reports average marginal effects for the covariates $x1 - x6$. The average marginal effects are computed for each individual in the sample and the means are presented in the table.

The naïve probit that ignores heaping yields coefficient estimates which are attenuated by 20–65%, compared to the true values used to create the simulated data. The coefficients on the main covariates of interest $x1 - x6$ are attenuated by 50% or more, while the time trend coefficients are somewhat closer to the true time trend (20–30% bias). The coefficients on the time-varying variables $x1$ and $x2$ are the most attenuated, but even the coefficients on the time-invariant variables $x4$ and $x5$ are attenuated by about 50%. That is, heaping is not just a problem because it means "the timing is wrong" between the dependent variable and time-varying covariates. Instead, like misclassification error in a cross-sectional probit model, it generally leads to biased coefficients. The additional parameters in the MLE approach, $\alpha_{0,H}$ and $\alpha_{1,N}$ satisfy the monotonicity assumption and are about the magnitude we expect given the rates of heaping assumed in the simulation. Torelli and Trivellato (1993) find large biases due to heaping in their simulation results for continuous time duration models. Hausman et al. (1998) find large biases due to misclassification in their simulation results for cross-sectional probit models with misclassification. Our simulation results confirm that misclassification due to heaping leads to potentially serious bias in a discrete-time hazard model.

Of the approaches to correct for heaping, the adjusted MLE approach and the ad hoc approach of decimation both appear to work well in the simulations. As in the Hausman et al. (1998) application to a cross-sectional probit, the adjusted MLE approach to estimating the discrete-time hazard model yields coefficients that are on average very close to the true parameters on $x1 - x6$ used to generate the data (0.2–2.7% bias). The adjusted MLE approach yields somewhat more biased estimates of the time trend (7.9–11.4%). The

decimation approach also yields coefficients for the covariates $x1 - x6$ that are very close to the true values (0.2–0.9% bias). However, the decimation approach yields substantially biased estimates of the time trend (43–55% bias).

The other ad hoc approaches to deal with heaping – adding a heaping indicator or coarsening the data – yield substantially biased estimates of the coefficients on the covariates of interest and of the time trend.[13] A weak argument might be made that accounting for heaping in any way is an improvement over the naïve probit model that ignores heaping. However, our simulation results suggest that neither of these ad hoc approaches should be recommended as a solution to heaping.

A more intuitive comparison of these methods may be seen in the densities of the estimated parameters from the simulations. Figure 2.4 graphs the distributions for the estimated coefficients on the covariates $x1 - x6$. Note that "Full Coarsening" refers the the coarsening method applied to complete period data, and "Coarsened Seen" refers to that when data ends at the event. Solid vertical lines indicate the true values used in the simulation. These figures help illustrate the relative accuracy of the methods as compared to the true value, as well as the variation of the estimates across simulations from the different approaches.

Because all of the covariates' coefficients are jointly determined within maximum likelihood estimation, differences between the estimated coefficients and true parameters may be less important than the estimates of the marginal effects, which may also be of greater concern to policymakers. Table 2.4 and Figure 2.5 present the simulation estimates of the marginal effects of the primary covariates. In the case of the decimated sample, it is possible to obtain two sets of marginal effects: one for the decimated sample used in estimation, and another using the full sample including heapers. The results for both are presented here.

Considering the marginal effects, the structural MLE correction and the decimation results most closely estimate those marginal effects given from the true probit model. However,

[13]For the coarsening approach the simulation uses year dummies instead of the time trend variables.

Figure 2.4: Distributions of Estimated Coefficients for the Covariates in Simulation

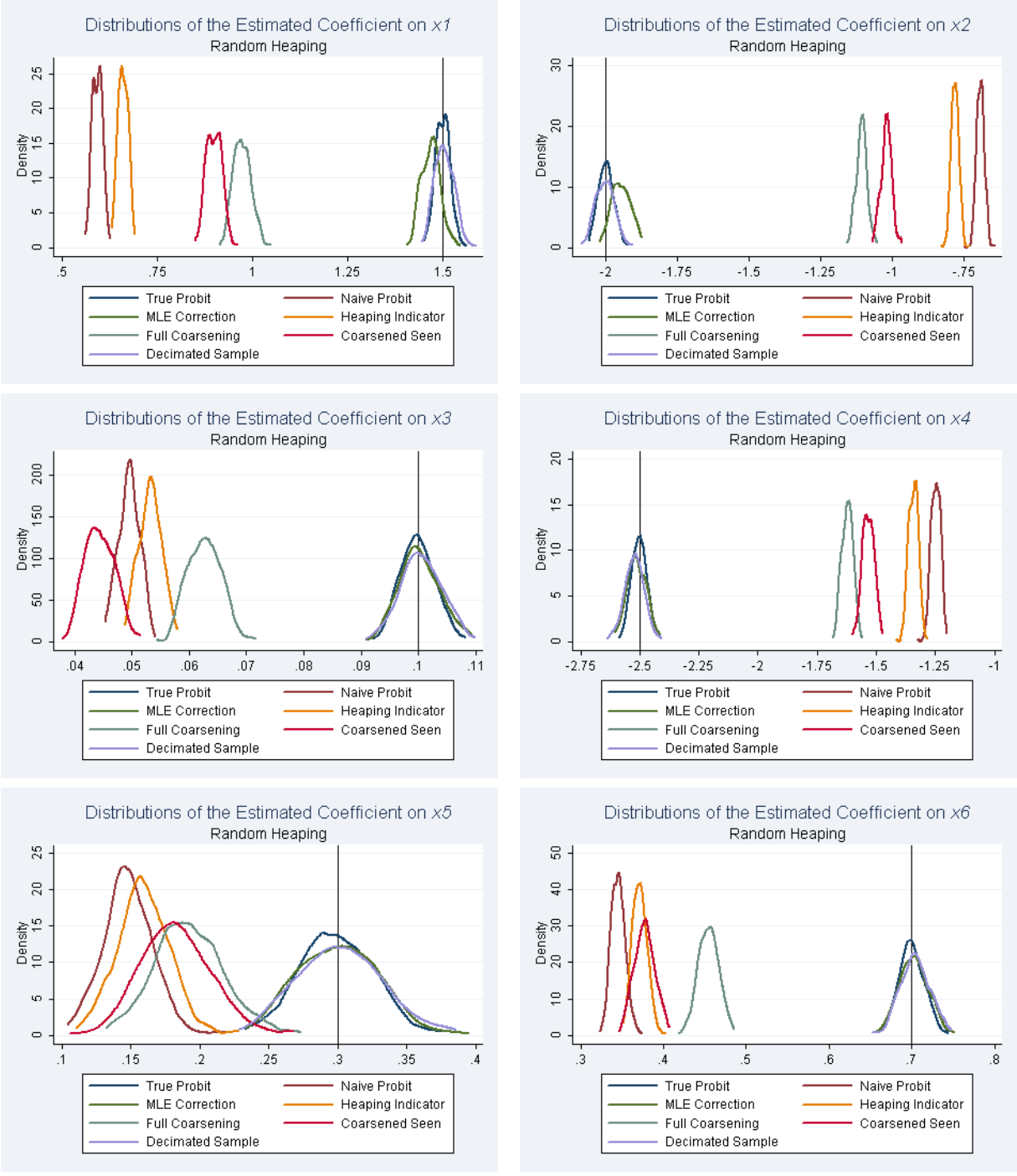
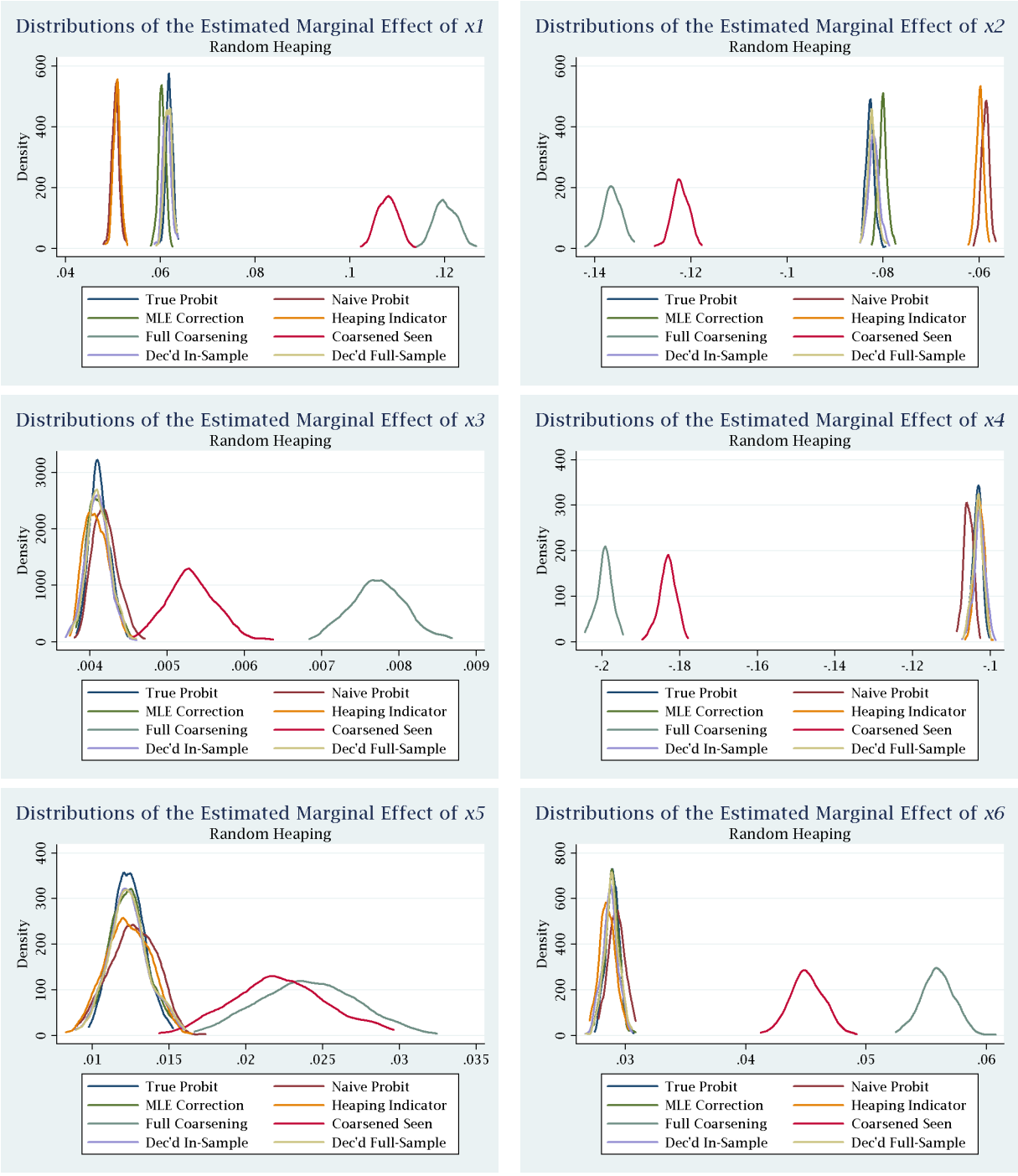


Figure 2.5: Distributions of Estimated Marginal Effects for the Covariates



the decimated results outperform the MLE correction slightly, with the full sample decimation results being the most accurate. Although those results are close to their respective true values, the coarsened sample results are strikingly different. Whereas decimation produces marginal effects all less than one percent different from the true values, coarsening produces marginal effects no better than thirty percent different. At those levels, its use should surely be discouraged.

2.6 Real Data

In this section we apply the alternative approaches to real data on smoking cessation from the February 2002 TUS-CPS. The 2002 TUS-CPS asked all current and former smokers how long it had been since they last smoked regularly. We use data from 8,055 former smokers, i.e. respondents with completed spells of smoking.[14] Because smoking cessation is uncommon among young adults, we restrict the sample to former smokers over the age of 30. Because each former smoker remains in the at-risk sample until he or she quits, the 8,055 former smokers provide an at-risk sample size of 74,153 person-years. The average smoking cessation rate is about 7%.

The explanatory variables in our discrete time hazard model of smoking cessation include the real price of cigarettes, age, sex, race, schooling, and income. Cigarette prices and age are time-varying, while the remaining variables are time-invariant. The real price of cigarettes is merged to the TUS-CPS data based on the individual's state of residence in 2002, which introduces some measurement error in assigned prices due to cross-state movers. Schooling and income are also measured as of the 2002, and so should be viewed as proxies for the individual's level of schooling and income in all previous years.

Table 2.5 presents the results for the different approaches to estimate the discrete time

[14]Focusing on the completed spells of former smokers makes the real-data sample more similar to our simulated data, where almost all spells are completed.

Table 2.5: Results from TUS-CPS Analysis

	Naïve Probit	Heaping Indicator	Multiple Indicators	Decimated Sample	MLE Correction	Heaped, Coarsened
Price	0.1494	0.1578	0.1156	0.1997	0.1908	-0.0204
	(0.0262)	(0.0267)	(0.0267)	(0.0327)	(0.0313)	(0.0229)
Age	0.0066	0.0064	0.0063	0.0058	0.0057	0.0130
	(0.0009)	(0.0009)	(0.0009)	(0.0011)	(0.0011)	(0.0011)
Female	-0.0525	-0.0477	-0.0476	-0.0283	-0.0249	-0.0873
	(0.0149)	(0.0154)	(0.0155)	(0.0200)	(0.0188)	(0.0193)
NonWhite	-0.0052	-0.0114	-0.0113	-0.0206	-0.0343	0.0187
	(0.0238)	(0.0246)	(0.0247)	(0.0323)	(0.0302)	(0.0304)
Education	0.0135	0.0139	0.0143	0.0202	0.0163	0.0174
	(0.0029)	(0.0030)	(0.0030)	(0.0040)	(0.0037)	(0.0038)
Income	0.0014	0.0015	0.0015	0.0017	0.0017	0.0023
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0004)
Year	0.1864	0.2321	0.2670	0.3350	0.2979	0.0338
	(0.0206)	(0.0213)	(0.0214)	(0.0329)	(0.0314)	(0.0014)
Year^2	-0.7014	-0.8593	-0.9919	-1.1993	-1.0510	
	(0.0682)	(0.0705)	(0.0709)	(0.1041)	(0.0997)	
Year^3	0.0900	0.1077	0.1234	0.1421	0.1274	
	(0.0072)	(0.0074)	(0.0075)	(0.0105)	(0.0101)	
Constant	-4.1689	-4.851	-5.102	-5.8480	-5.7382	-2.5850
	(0.2057)	(0.2142)	(0.2145)	(0.3388)	(0.3243)	(0.0722)
5-Yr Heap Indicator		0.713	0.566			
		(0.0161)	(0.0218)			
10-Yr Heap Indicator			0.340			
			(0.0260)			
Calendar/Decade Indicator			0.219			
			(0.0286)			
$\alpha_{0,H}$					0.1037	
					(0.0030)	
$\alpha_{1,N}$					0.0280	
					(0.0363)	

Results are from 2002 TUS-CPS. Standard deviations in parentheses. $n = 74153$ person-year observations.

hazard model of smoking cessation using the 2002 TUS-CPS data. Although we no longer have benchmark true values for the parameters, we note that a naïve probit that fails to account for heaping yields coefficients on price, schooling, and income that are 20% smaller than the coefficients from the adjusted MLE. If the adjusted MLE coefficients are closer to the true effects, this pattern is consistent with the bias we find in the simulated data.[15] Similar to our results, previous applications of the adjusted MLE approach that use real data to estimate cross-sectional probit models also find that correcting for misclassification error appears to be important (Hausman et al., 1998; Kenkel et al., 2004). However, in contrast to our results, Torelli and Trivellato (1993) and Forster and Jones (2001) find that correcting for heaping in real data does not change much the coefficient estimates in a continuous time duration model. Torelli and Trivellato caution that this might reflect specification errors of the duration model.

Unlike both the results with simulated data and the results using real data for price, schooling and income, for some of the other variables the coefficients from the naïve model are actually larger than the coefficients from the adjusted MLE model. Intuitively, this raises the possibility that the heaping in the real data might be more complicated than allowed for in the assumptions in the adjusted MLE model.

We explore one complication which is motivated by Figure 2.1 above. In Figure 2.1, the heaping in reported smoking cessation in the 2002 TUS-CPS data does not appear to be even across all 5-year increments. Instead, people appear to be more likely to heap on the 10-year increments (10, 20, 30, 40, and 50) before the survey than on the 5-year increments in-between. In addition, there appears to be slight heaping on the calendar-year decades of 1990, 1980, 1970, 1960, and 1950. To explore this, we extend the probit model with

[15] Also in accordance with what we find in the simulated data, the decimation approach yields coefficients that tend to be more similar to those from the adjusted MLE. The other ad hoc approaches yield results like those from the naïve probit. Coefficients estimated using coarsening were particularly susceptible to the form of the time trend used. Given the strong correlations between price, age, and time, we use only a linear time trend in that analysis.

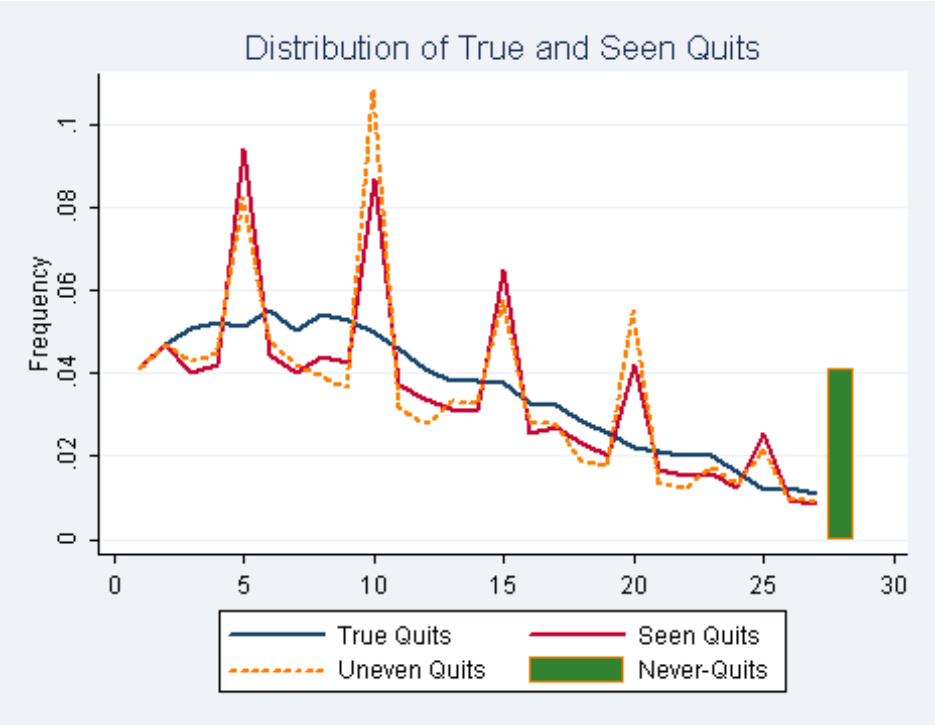
the indicator for 5-year heaping to include additional heaping indicators for 10-year heaping and calendar-year decade heaping. Although our simulation results presented above do not support this approach as a solution for heaping, it should provide a useful diagnostic for the presence of heaping. The coefficients on the additional heaping indicators are statistically significant and positive (third column of results in Table 2.5). Based on the argument that true smoking cessation should not be more likely in the years indicated by the new variables, these results suggest more complicated heaping in the real data than assumed in the adjusted MLE model.

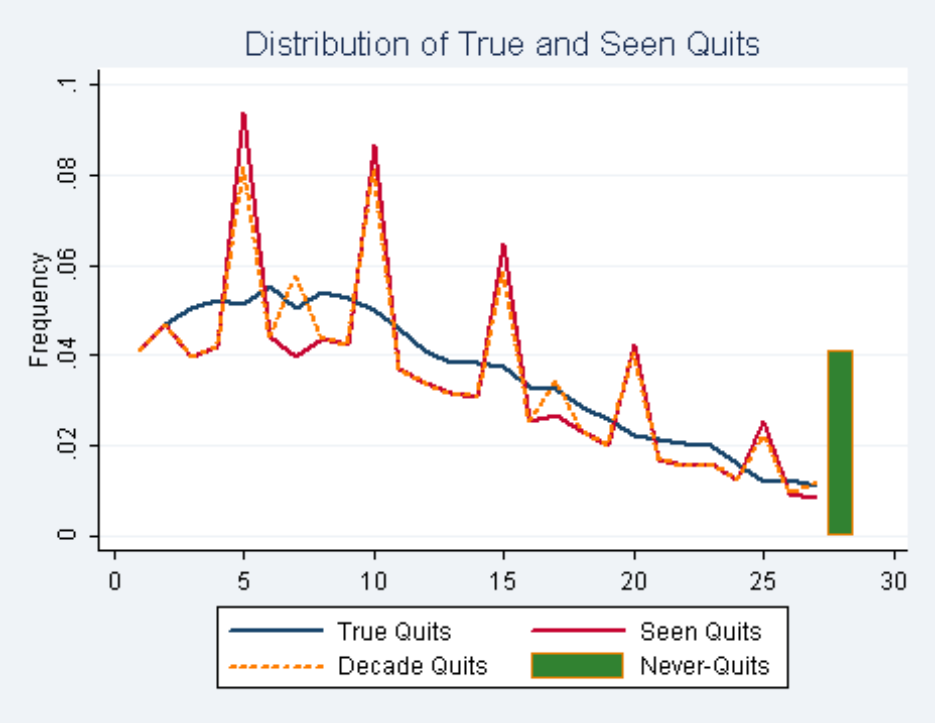
2.7 Robustness to Alternative Forms of Heaping

In this section, we return to our simulation exercise to explore how well our proposed approaches perform when the heaping in the data takes a more complicated form. We explore the two forms of heaping that appear to be present in the 2002 TUS-CPS data: uneven heaping, with different rates of heaping on 5-year and 10-year increments; and calendar-decade heaping. We use the same simulated data we used above on true smoking cessation.

We adapt our previous "random heaping" rule over the 27 period Monte Carlo simulation to incorporate these alternative forms of heaping. Under random heaping, 20% of quitters heap to the nearest 5-year increment: quitters in year 3-7 round to 5 (or 15 or 25) and quitters in years 8-12 round to 10 (or 20). Under uneven heaping, 30% of quitters in years 8-12 heap to 10 while only 15% of quitters in years 3-7 heap to 5. As such, quitters near a 10-year increment are twice as likely to heap as those near a 5-year increment. Alternatively, when calendar-decade heaps are introduced, those who heap are evenly split between rounding to a 5-year increment versus rounding to a nearby calendar-decade year. To illustrate, suppose a calendar-decade corresponds to the year 7 before the retrospective question was asked. (This would be the case, for example, for a survey conducted in 1997.) Of the quitters in

Figure 2.6: Observed Quit Rates Under Alternative Forms of Heaping





years 3 and 4, 20% round to year 5. Of the quitters in year 6, 10% round to year 5 and 10% round to the calendar-decade year 7. Where under random heaping we assumed that 20% of year 7 quitters round to year 5, because year 7 is also a calendar-decade we now assume that only 10% of quitters in year 7 round to year 5. As under random heaping, quitters in year 8 are again split between rounding to years 7 and 10. Finally, again as under random heaping, 20% of quitters in years 9, 11, and 12 round to year 10, because none of those years is a calendar-decade. Figure 2.6 shows how the observed quit rate under uneven heaping and calendar-decade heaping differs from the true quit rate and the quit rate observed with simple (5-year) random heaping, labeled "Seen Quits."

Figures 2.7 and 2.8 show the distributions of the estimated coefficients on the covariates from each of the methods when the simulated data are transformed under the uneven and calendar decade heaping rules.[16] To be clear, all of the methods use the earlier assumption of random heaping; the goal of the exercise is to see how the approaches perform when the heaping rule is mis-specified. The adjusted MLE approach uses the log-likelihood given by equation (2.12) that incorporates misclassification error only due to random 5-year heaping. The ad hoc approach of including a heaping indicator only includes an indicator for 5-year heaping. The coarsening approach continues to only coarsen around 5-year increments, not around calendar-decades. Similarly, the decimation approach only decimates observations from 5-year heapers and retains observations from calendar-decade heapers.

Despite mis-specifying the heaping rule, with uneven heaping the adjusted MLE approach and the decimation approach continue to perform quite well. Both approaches yield coefficients on the covariates $x1 - x6$ that are very close to the true values assumed to generate the data. The densities of these approaches' coefficient estimates from the 248 simulations are very similar to the densities of the coefficient estimates from the probit model estimated using the true data with no heaping. Correcting for the pattern of misclassification errors –

[16]Because of their similarity, only the coarsening of the heaped data, rather than the full panel, is graphed.

Figure 2.7: Distributions of Estimated Coefficients Under Uneven Heaping

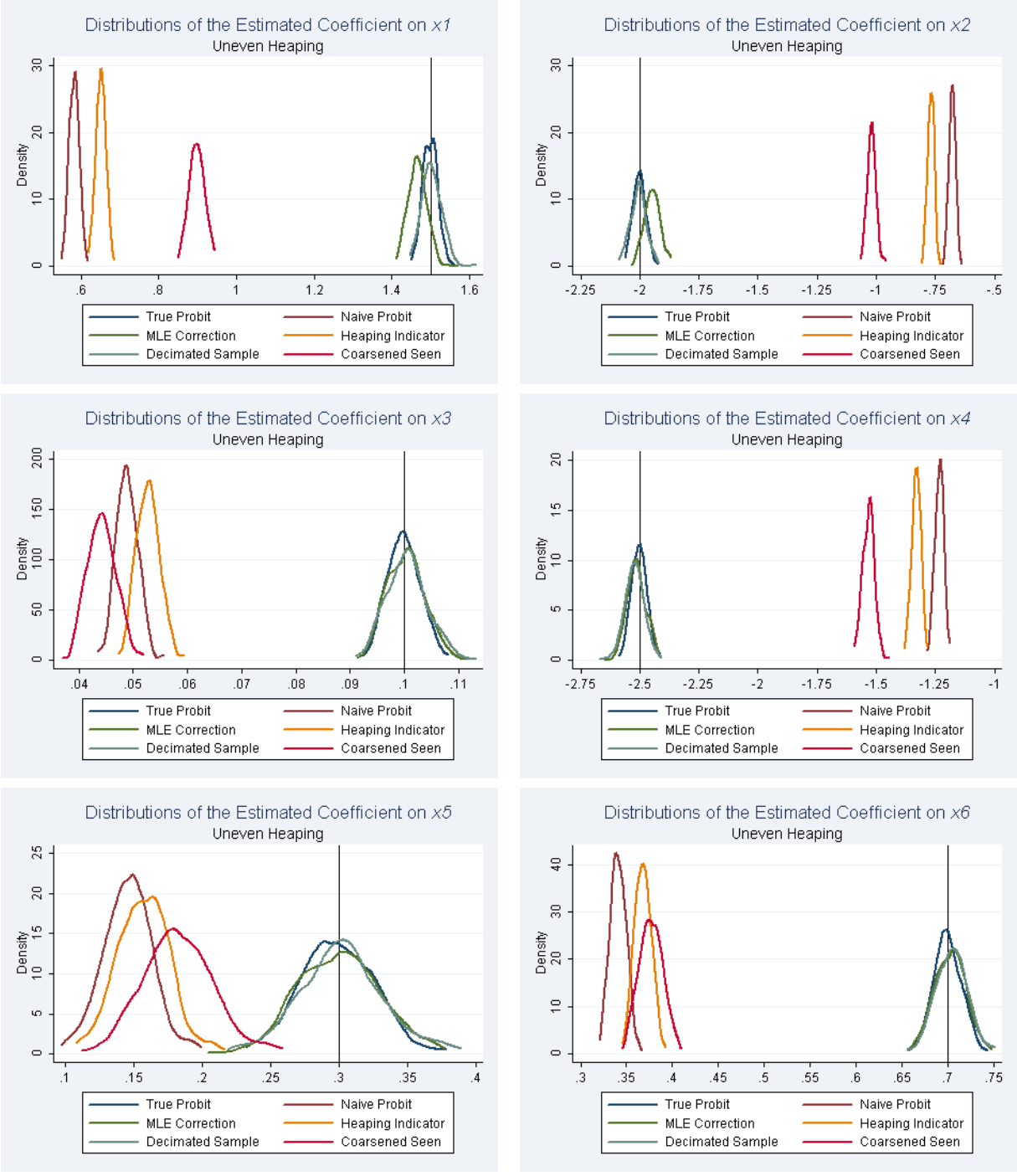
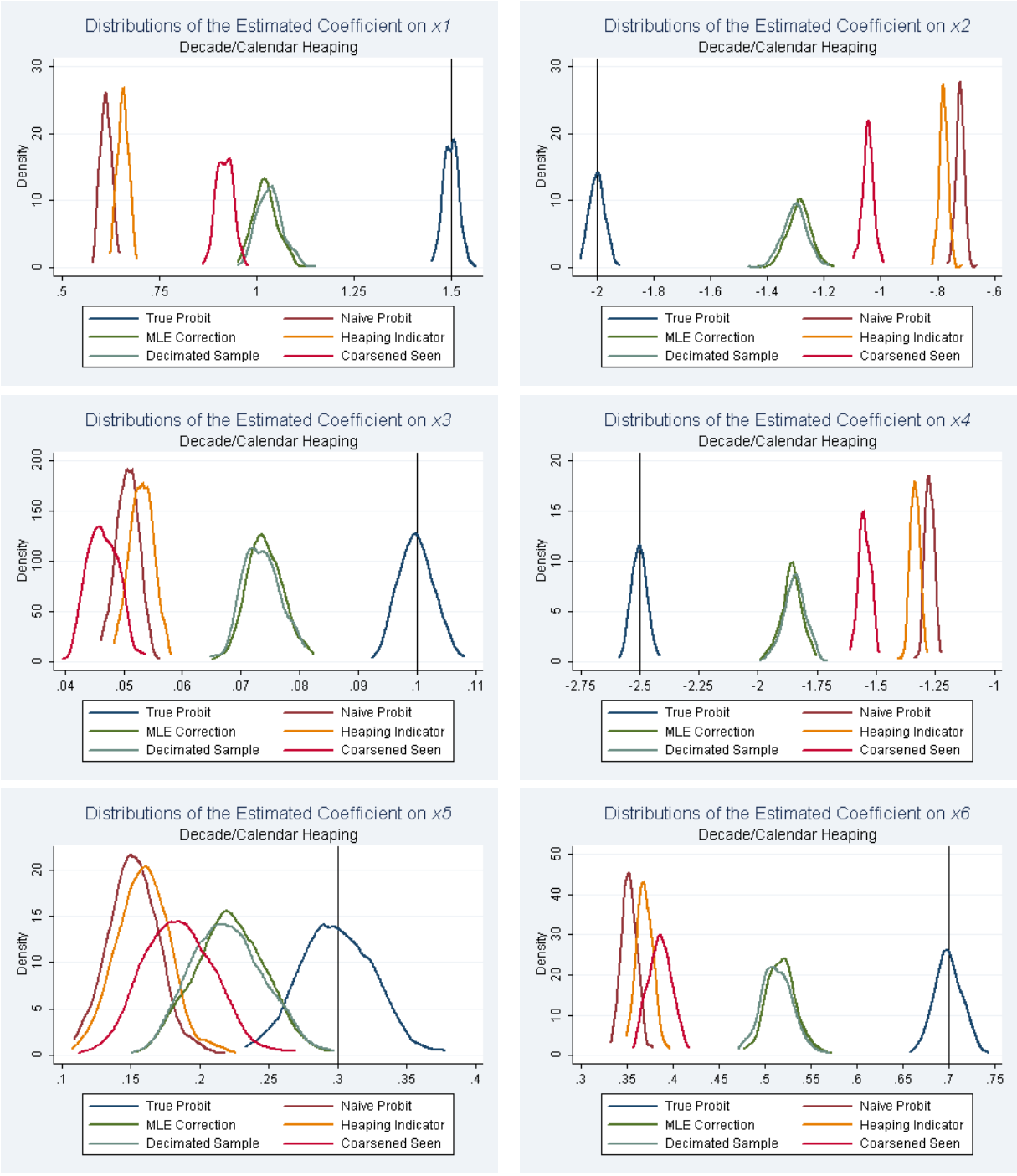


Figure 2.8: Distributions of Estimated Coefficients Under Calendar/Decade Heaping



with false 1s in years 5, 10, 15 and so on, and false 0s in odd years near the 5-year increments – appears to be more important than accounting for uneven heaping on 5-year versus 10-year increments. Above, we suggest an analogy with the method of instrumental variables, where the indicator for a 5-year increment serves as an instrumental variable for the misclassification probabilities in the structural equation describing observed quitting. Under this interpretation, the 5-year indicator would remain a valid IV even if there is additional 10-year heaping, so it is perhaps not surprising that the adjusted MLE approach continues to yield unbiased coefficient estimates. Further adjusting the MLE approach to incorporate uneven heaping is straight-forward, but might not be that important in practice. The robustness of the decimation approach to uneven heaping is easy to understand: whether the heaping is even or uneven, observations from reported quitters on 5-year and 10-year increments have been discarded.

With calendar-decade heaping, the adjusted MLE approach and the decimation approach yield less accurate estimates of the coefficients on the covariates $x1 - x6$. The estimates are attenuated towards zero, although to a lesser extent than the estimates from the naïve probit that ignores heaping or the other ad hoc approaches. Essentially, the adjusted MLE and decimation approaches address bias from the form of heaping they incorporate, but retain some bias from the calendar-decade heaping they ignore.

2.8 Conclusions

It is well-documented that in retrospective reports of past events, survey responses tend to be heaped on round numbers of years, such as 5, 10 and so on. For important applications in health and labor economics, such heaping leads to a specific form of measurement error in the dependent variable of interest, e.g. smoking cessation or a job transition. This paper is an applied econometrics exercise to compare alternative approaches to deal with

heaping. In the context of a discrete-time hazard model of the event in question, we show that heaping leads to specific patterns of misclassification error in the dependent variable. Our first approach to deal with heaping uses an adjusted MLE approach based on Hausman et al. (1998). We conduct Monte Carlo simulations to compare this approach to several ad hoc approaches to deal with heaping. Our results confirm that ignoring heaping can lead to substantial bias, and that the adjusted MLE approach recovers the true coefficients used to generate the simulated data. We also find that one of the ad hoc approaches – eliminating all observations from respondents who report the event in a heaped year, which we term decimation – does about as well as the adjusted MLE approach. Both of these approaches are robust to uneven heaping across 5- and 10-year increments, but perform less well if there is additional heaping on calendar-decade years. Moving from estimated coefficients to marginal effects, we see the models exhibit a wide range of differences from the true value. Overall, the differences are not as striking as those from the estimated coefficients. However, those variables which are "more random," as a de-meaned, de-trended price may appear are those with the greatest deviation. Based on our results, the indicator and coarsening approaches cannot be recommended as solutions to heaping. However, we note that the simplest ad hoc approach of adding a dummy indicator variable for heaped years is a useful diagnostic to detect heaping.

In our simulated data and in real data from the 2002 TUS-CPS, we find that modeling a time trend while correcting for heaping can be problematic. Because the simulated data mimic data collected retrospectively from a single cross-section like the 2002 TUS-CPS, there is an inherent identification problem in both the simulated and real data we use. In responses from a single year, there is a necessary connection between calendar time - which drives the trends in true outcomes - and the number of years before the survey - which due to heaping drives the observed outcomes. In future work we intend to explore the use of data from repeated cross-sections to model time trends while correcting for heaping. The

repeated cycles of the TUS-CPS included retrospective questions on smoking cessation asked in 1992, 1993, 1995, 1996, 1998, 1999, 2001, 2002, 2003, 2006, and 2007. The ability to model time trends is important for studying the determinants of smoking cessation, because there have been strong time trends in smoking rates and in many policy variables such as cigarette taxes. In addition, it might even be possible to explore the impact of more discrete "shocks," such as the 1998 Master Settlement Agreement (MSA) – the major legal settlement reached with the tobacco industry. The public debate and publicity surrounding the MSA might be expected to prompt higher smoking cessation. The TUS-CPS provide data before, during, and after the MSA. Important for correcting for heaping, the repeated waves of the TUS-CPS provide retrospective data collected 1, 2, 4, 5, 8, and 9 years after the MSA. Although in any one wave the retrospective data on smoking cessation in 1998 will be distorted due to heaping, careful analysis of the pooled data might uncover the impact of the MSA on smoking cessation.

Another direction for future work is to extend the discrete time hazard model to allow for duration dependence. The addictive nature of smoking suggests that the probability of quitting might fall with the duration of the habit, if longer durations are associated with a larger stock of addictive capital (Becker and Murphy, 1988). The discrete-time hazard model can flexibly allow for the probability of quitting to vary with duration, and does not impose positive or negative duration dependence. While we think that this makes the discrete-time hazard model an attractive approach to study smoking cessation, it will again be important to use a method like the adjusted MLE approach or decimation to handle heaping.

Future plans for this line of research include these extensions, as well as other corrections to the analysis. First, although most simulated covariates were chosen to mimic those in standard cessation research, with the remaining chosen to capture various types of regressor distributions, the simulation does not reflect actual data availability. In addition to simulating these with a joint distribution, the variables will be simulated to ease comparison

with CPS-TUS results. Second, a greater analysis of marginal effects is necessary to ensure recommendations for applied economists remain consistent with results from the analysis on estimated coefficients. The results presented here demonstrate a need to better understand why some marginal effects differ more than others. With these improvements made, we hope to provide a full analysis of how heaping in retrospective data may affect estimated results as well as suggest particular methods to minimize bias in those estimates.

CHAPTER 3

U.S. CHEMOTHERAPY USE, COST, AND VALUE

3.1 Introduction

The provision and cost of health care have become a central policy issue in the United States, which spends more than one-sixth of gross domestic product on detecting, diagnosing, treating, and preventing illness. Beyond spending more per person than any other industrialized nation on health care, the U.S. also spends more per person on cancer and cancer drugs than those nations (Jönsson and Wilking, 2007). Nevertheless, in 2011, cancer was the second leading cause of death, with the American Cancer Society predicting 1,596,670 new cases of cancer and 571,950 cancer-related deaths that year alone (Murphy et al., 2012; Centers for Disease Control and Prevention, 2011). Burstein (2012) notes that while cancer treatment guidelines contain "exhaustive references, highly detailed treatment algorithms, [and] careful delineation of treatment doses," they do not contain any measure of price. Thus the question of what the U.S. is getting for its health care spending, particularly surrounding cancer treatment, remains unanswered.

The costs of health care in the United States proved to be a significant part of recent health care reform. When the U.S. Congress passed the Patient Protection and Affordable Care Act in March 2010, it created the Patient-Centered Outcomes Research Institute and the Independent Payment Advisory Board. The federally funded yet independent institute, PCORI, began development immediately, charged with examining "the relative health outcomes, clinical effectiveness, and appropriateness of...medical treatments" while the government agency tasked with implementing reforms that "result in a net reduction in total Medicare program spending," IPAB, does not begin work until 2014 (PPACA, 2010). The joint goal of these entities is to reduce overall health care spending while maintaining current

quality levels.

The primary endpoint and central marker of quality in cancer therapy is overall survival. Since remission after therapy is not universal, a measure of quality-adjusted survival, which takes into account tumor shrinkage, the negative side effects of chemotherapy, and likelihood of eventual recurrence, more accurately reflects the benefits from a particular chemotherapy regimen. This paper investigates which regimens are used in practice, the average cost of a patient's therapy, survival in terms of quality-adjusted life years (QALYs), and the cost per QALY as they relate to three types of cancer chemotherapy: treatment for metastatic colorectal cancer, treatment for metastatic non-small cell lung cancer, and adjuvant treatment for localized breast cancer.

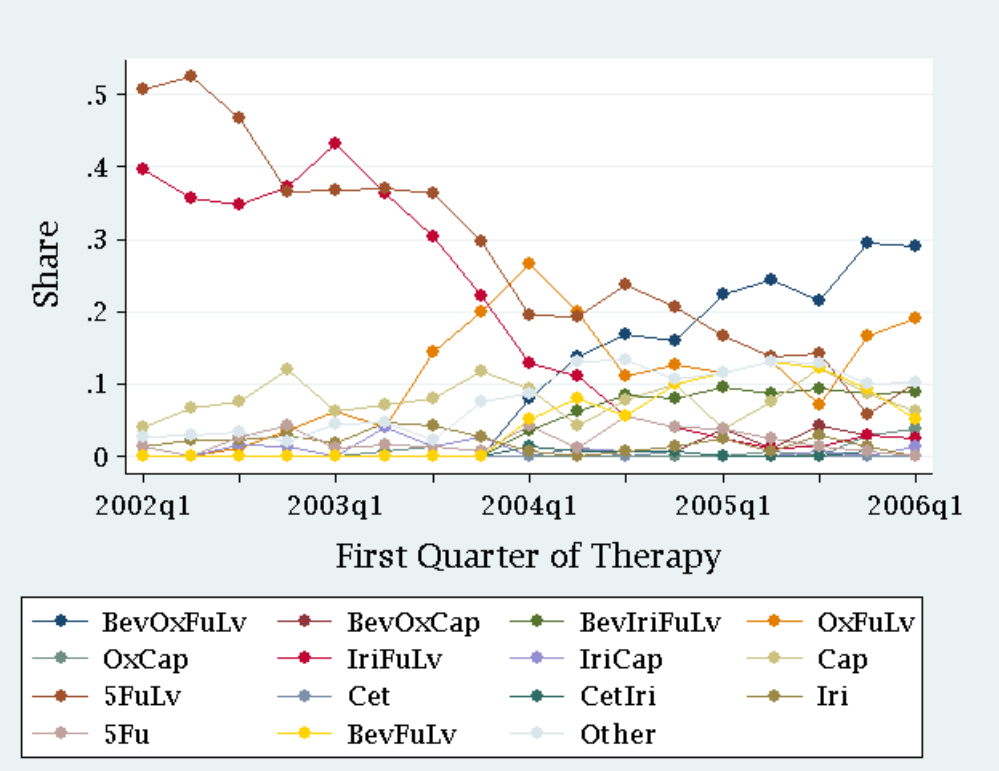
3.2 Data and Methods

Starting in the 1990s and gaining momentum in the early 2000s, numerous cancer drugs were released to treat each of the chosen therapy types. Most commonly, these new drugs are used in predefined combination therapy regimens consisting of multiple component drugs. Among cancer therapy types, treatment for metastatic colorectal cancer, treatment for metastatic non-small cell lung cancer, and adjuvant treatment for localized breast cancer were chosen because of their high degree of consistency to nationally recommended regimens.[1] Because patients may be treated with a number of distinct regimens, the comparison of use, average cost, expected QALYs of survival, and cost per QALY over time require measures of each regimen's market share, average cost per patient, and expected QALYs of survival.

Market share data were constructed using the Survey of Epidemiology and End Results, an agglomeration of several cancer registries nationwide produced by the National Cancer Institute, linked to Medicare claims for chemotherapy drug services. This limits the relevant

[1]The National Comprehensive Cancer Network (NCCN) provides clinical treatment guidelines for these and a number of other cancers and were the source of the regimen combinations in this study.

Figure 3.1: Metastatic Colon Cancer Chemotherapy Market Shares, 2002-2006

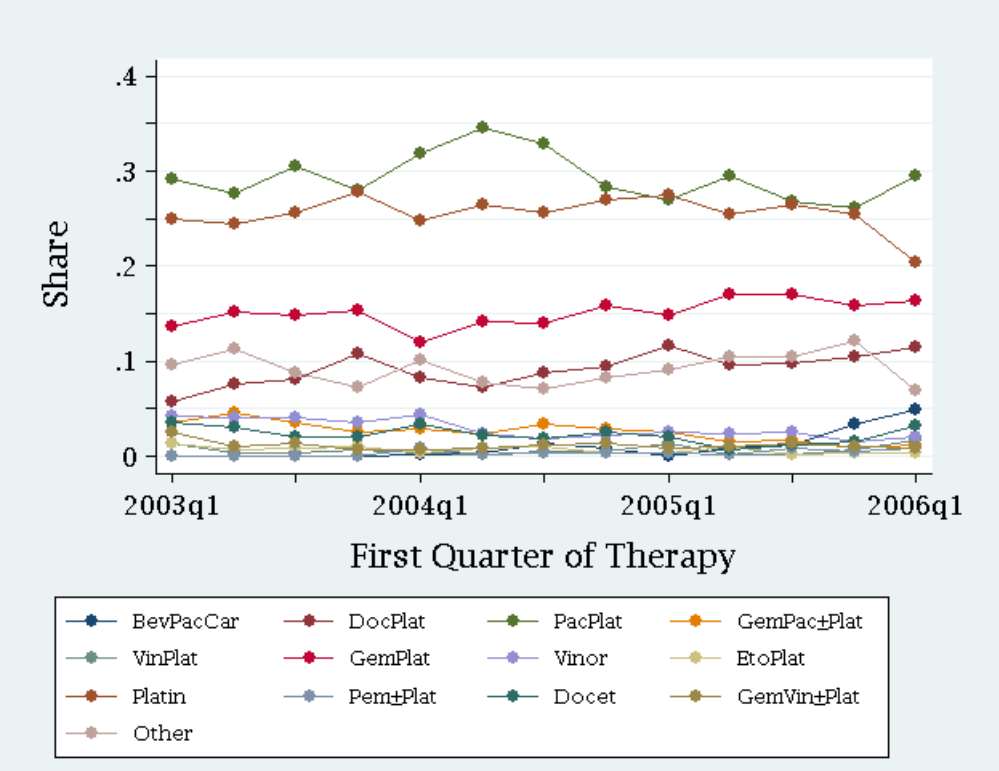


population to those aged 65 and older in the United States, but the SEER-Medicare data remains the gold standard of national chemotherapy use. I obtained physician and outpatient hospital claims for first-time chemotherapy between 2002 and 2006 for colorectal cancer, and between 2003 and 2006 for breast and non-small cell lung cancer, and identified regimens based on those claims.

For metastatic colorectal cancer, the relevant drugs are fluorouracil (5Fu), leucovorin (Lv), irinotecan (Iri), capecitabine (Cap), oxaliplatin (Ox), bevacizumab (Bev), and cetuximab (Cet). For metastatic non-small cell lung cancer, the relevant drugs are cisplatin and carboplatin (Plat), paclitaxel (Pac), vinorelbine (Vin), gemcitabine (Gem), etoposide (Eto), docetaxel (Doc), and bevacizumab.[2] The relevant drugs for the adjuvant treatment of breast

[2]Over the analysis period, two additional drugs were used to treat NSC lung cancer: gefitinib and erlotinib.

Figure 3.2: Metastatic NSC Lung Cancer Chemotherapy Market Shares, 2003-2006

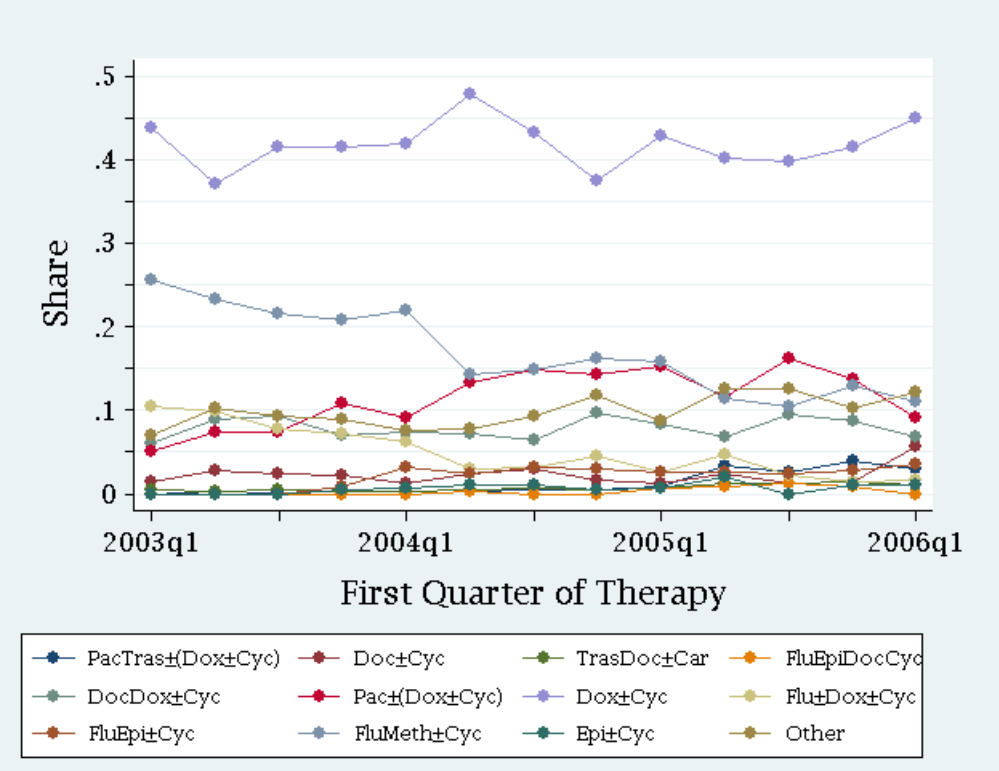


cancer are doxorubicin (Dox), cyclophosphamide (Cyc), paclitaxel, trastuzumab (Tras), docetaxel, carboplatin (Car), epirubicin (Epi), fluorouracil, and methotrexate (Meth). In some combinations, one or more drugs may be optional and these are denoted with a ± sign in Figures 3.1, 3.2, and 3.3 which give the market share of each regimen as seen in the SEER-Medicare data.

Average cost of therapy per regimen is calculated using data from IMS Health and NCCN regimen dosing guidelines. The NCCN specifies treatment length and quantity for typical patient size, which is then converted to cost using average price per milligram of active ingredient as reported by IMS Health. These prices should be close to those actually paid

Because these were oral medications without infusion equivalents, they do not appear in Medicare Part A nor B claims. Nevertheless, these second-line regimens were unlikely to be used on the chosen sample and thus affect subsequent calculations.

Figure 3.3: Adjuvant Breast Cancer Chemotherapy Market Shares, 2003-2006



by physicians since IMS Health reports the invoice amount rather than average wholesale price.

Measures of QALYs of survival are taken from Nicholson et al. (2011), which calculates the amount of survival time in each of five health states and weights them according to published QALY weights. Those health states are (1) stable disease without experiencing a side effect from chemotherapy, (2) tumor is responding to treatment with no side effects, (3) responding to treatment with side effects, (4) not responding to treatment with a side effect, and (5) disease progression to a more advanced state. Because survivors of localized breast cancer live for multiple years, the expected QALYs of survival are discounted at three percent per year. The amount of time spent in each of those states is estimated from the FDA-approved package inserts that accompany each drug, journal publications of clinical

trials referenced by the 2008 NCN Clinical Practice Guidelines in Oncology publications, and journal articles and conference abstracts identified by searching in PubMed and Google Scholar for the name of each drug in a regimen with at least one percent market share.

3.3 Results

Although physicians consistently prescribe according to established regimens, some patients do receive non-established, "other" regimens. Further, some regimens are specifically approved by the FDA as second-line treatments, to be administered after a patient fails to adequately improve from first therapy. These therapies tend to be more expensive than first-line therapies as well as have worse overall outcomes due to the test population. To accurately describe the changes in cost and outcomes over time, the market shares of regimens are normalized in two ways. Because cost and outcome measures are unavailable for the "other" regimen, the first normalization drops the portion of patients receiving it and rescales those other regimens appropriately. In the second normalization, I reapportion market shares excluding these second-line therapies in the sample from analysis because my sample consists of first time cancer patients. These measures contrast to those obtained directly from the data wherein the cost and benefits, in terms of QALYs of survival, of the "other" regimen are both assumed to be zero. It is included in the following figures to demonstrate how inclusion of the "other" regimen affects analysis.

3.3.1 Colorectal Cancer

Of the three types of therapy considered here, spending is by far the greatest on metastatic colorectal cancer. Figure 3.4 details the high levels of spending per patient as well as the explosive increase. Although there was an upward trend between 2002 and 2004 as physicians

Figure 3.4: Average Patient CRC Chemotherapy Costs, 2003-2006

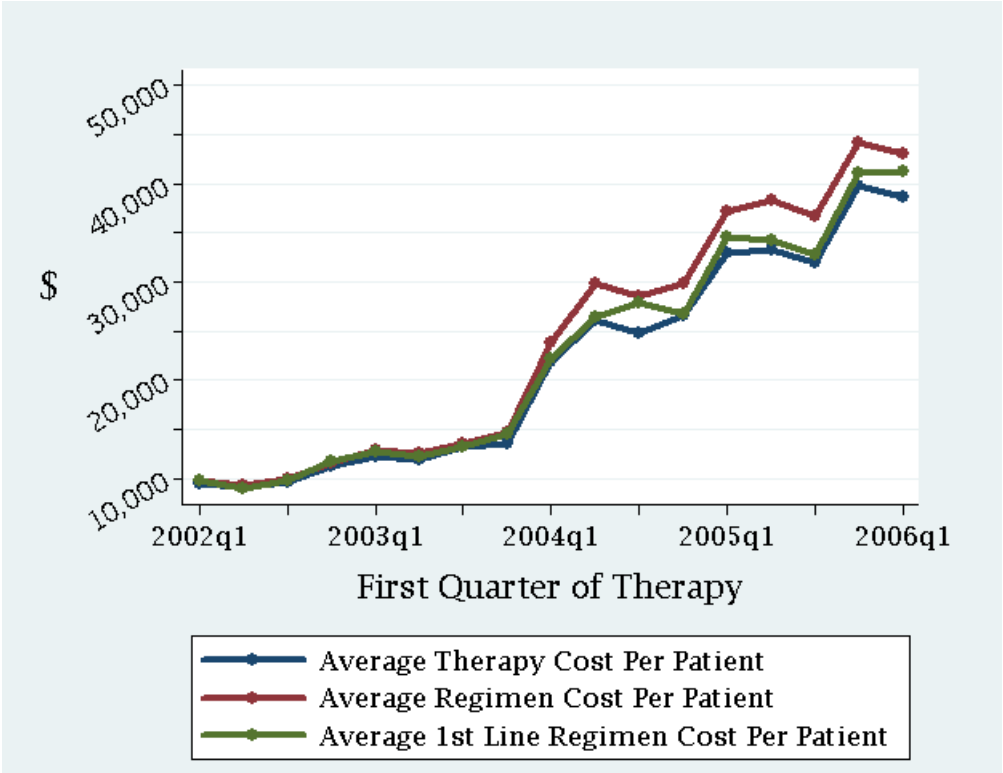
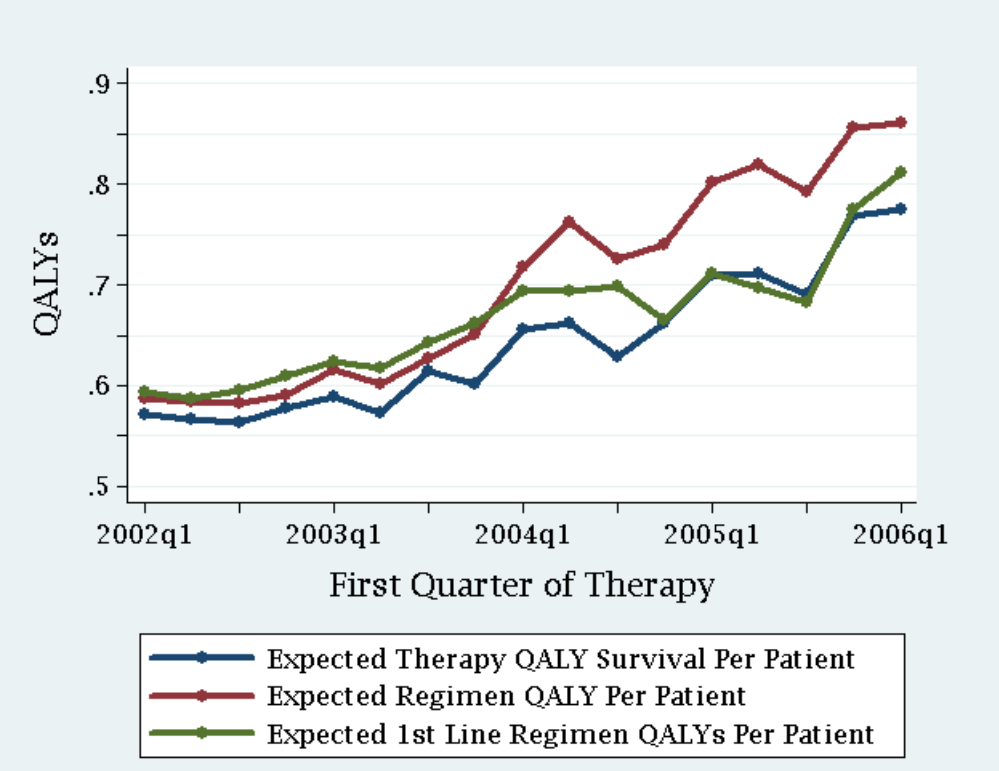


Figure 3.5: Expected QALYs of Survival from CRC Chemotherapy, 2003-2006



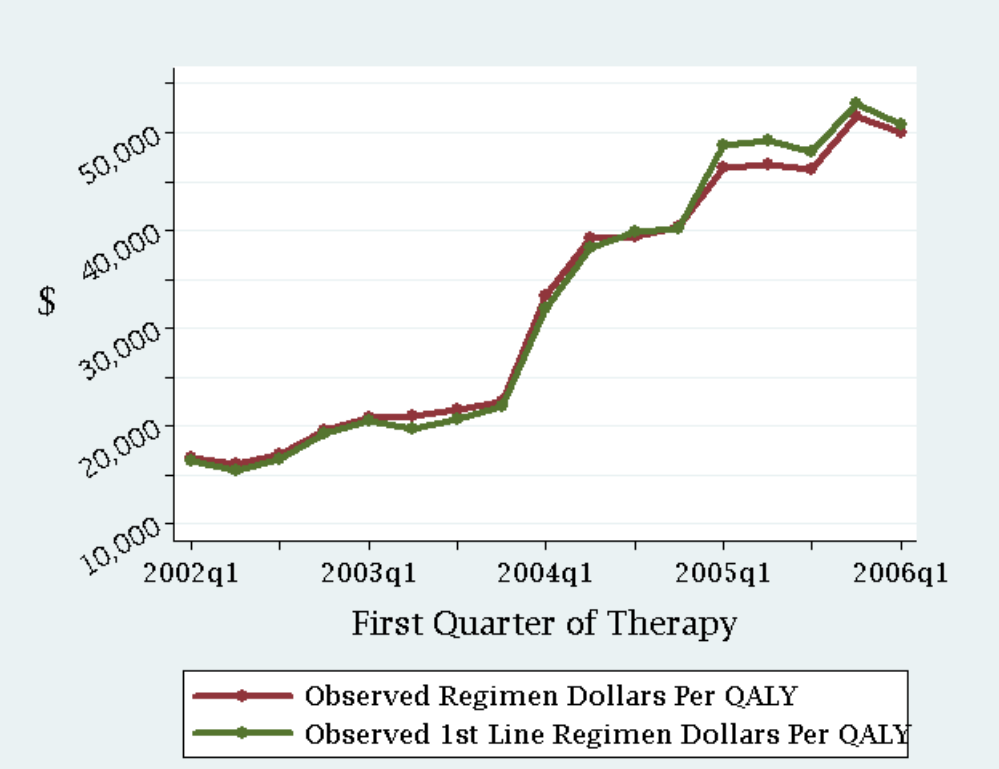
increasingly used the new agent oxaliplatin in combinations, the largest increase is due to the introduction of cancer therapies called monoclonal antibodies, wherein the immune system is used to inhibit tumors.[3] Bevacizumab and cetuximab are very expensive treatments but provide statistically significant survival benefits.

Across regimens, the expected QALYs of survival are presented in Figure 3.5 and show a marked increase over time as well. In general, metastatic colorectal cancer patients experience little over one year of overall survival with considerable side effects from chemotherapy, so that the adjustment for quality results in a very low QALY survival. Still, over the period QALY of survival increases by roughly 40%.

Combining the two trends to examine the cost per QALY of survival, Figure 3.6 shows

[3]Drugs of this type are identifiable by their "-mab" suffix.

Figure 3.6: Dollars Per QALY of Survival from CRC Chemotherapy, 2003-2006



that spending per QALY more than tripled over the four years, reaching a high of more than $50,000. Although the Independent Payment Advisory Board is prohibited from using cost effectiveness measures, specifically any dollar amount per QALY, to determine Medicare coverage, what the US is getting from its spending on metastatic colorectal chemotherapy has dramatically fallen over time.

3.3.2 Non-Small Cell Lung Cancer

Because most chemotherapy regimens for metastatic NSC lung cancer were introduced before the middle of the 1990s, the market shares show much less variability by 2003 than those for colorectal cancer. The age of the drugs also corresponds to the much lower level of overall spending since off-patent molecules are much less expensive than those with patent protection. Still, Figure 3.7 shows a dramatic increase in spending over 1995. Though it represents only a small fraction of the market, the bevacizumab-based regimen is far more costly than other therapy regimens and its use rises during that time period.

The use of bevacizumab to treat lung cancer between 2004 and 2005 is a testament to physicians' willingness to provide care to patients. Bevacizumab was first approved to treat colorectal cancer in February 2002, when it was still in trial to test efficacy for lung cancer. The data from that trial was released in March 2006 and the FDA granted approval in October of the same year, but long after physicians began prescribing the treatment as a first-line therapy to patients. The decline in cost at the end of 2004 may actually be an artifact of the data rather than decreased usage of the bevacizumab regimen. If physicians try to gain reimbursement from Medicare for bevacizumab, they may be using a non-specific drug code rather than that supplied by the Centers for Medicare Studies through the HCPCS coding system. This would understate the usage of BevPacCar to treat NSC lung cancer and in turn the average patient cost of chemotherapy.

Figure 3.7: Average Patient NSCLC Chemotherapy Costs, 2003-2006

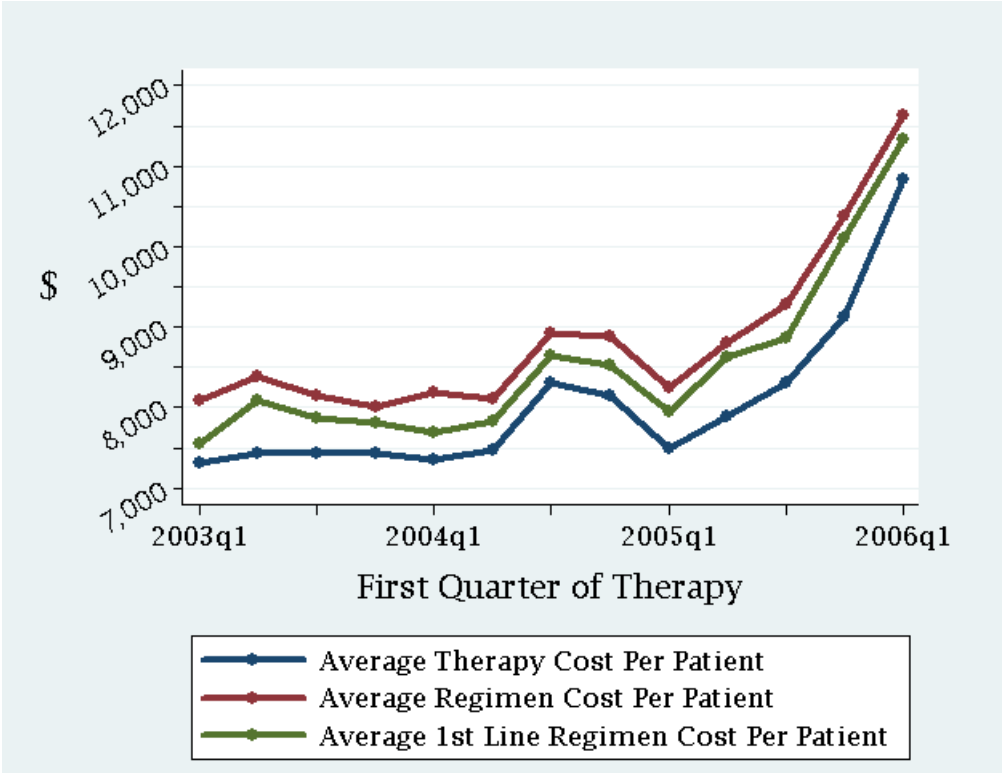


Figure 3.8: Expected QALYs of Survival from NSCLC Chemotherapy, 2003-2006

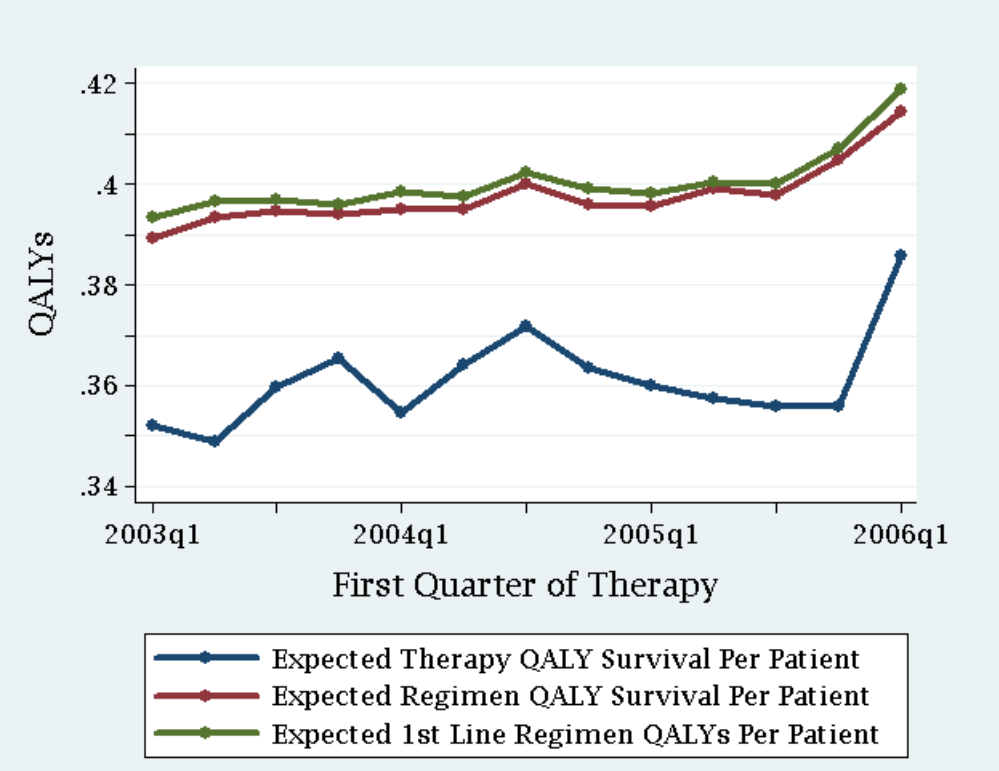
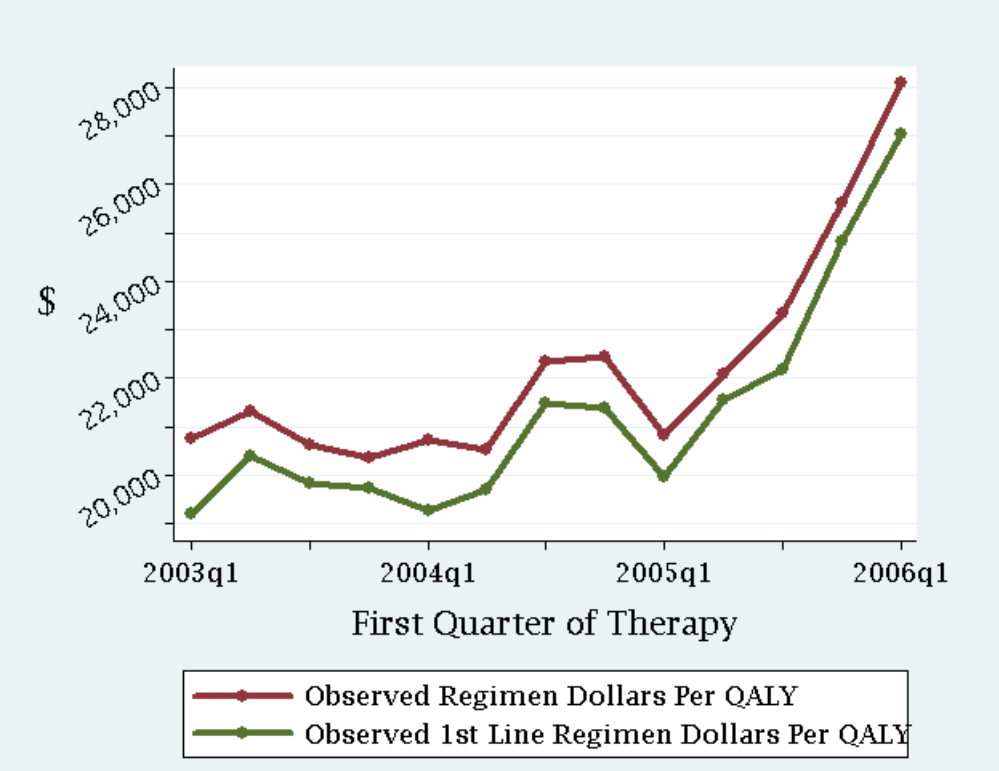


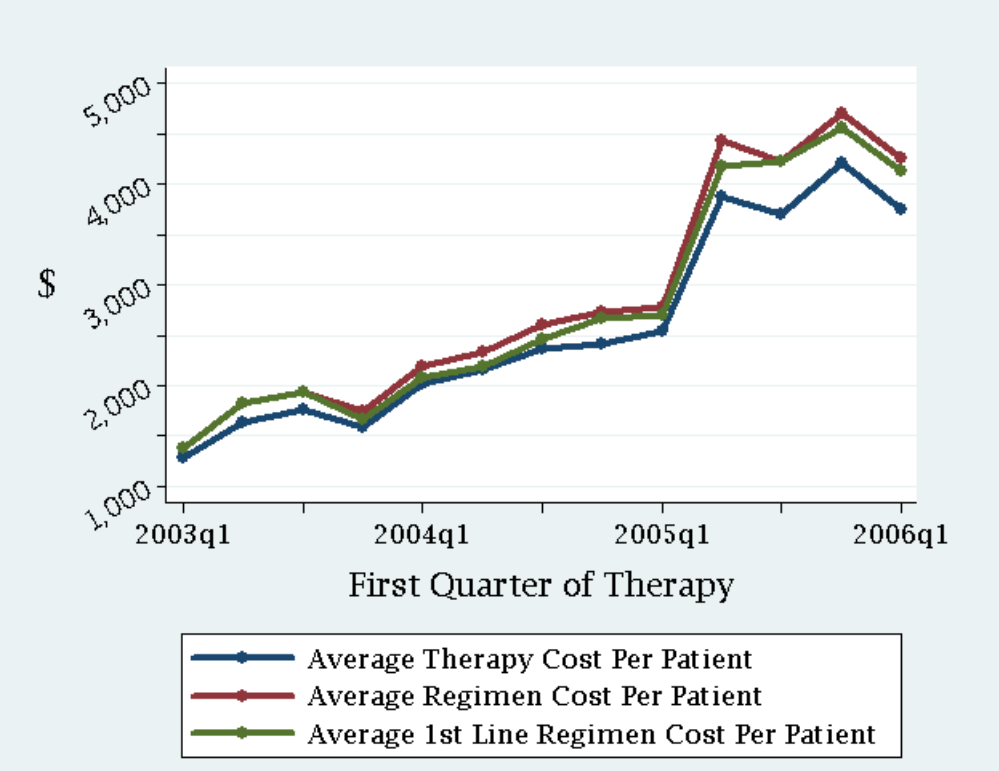
Figure 3.9: Dollars Per QALY of Survival from NSCLC Chemotherapy, 2003-2006



Although the drugs approved to treat NSC lung cancer over the period were constant, Figure 3.8 shows a moderate but continuous increase in expected QALYs of survival over time. In this figure and the corresponding graph of QALYs of survival for breast cancer therapy, the unadjusted expected QALYs of survival per patient is significantly different from that of the adjusted regimens. This difference is the result of a higher proportion of patients receiving the "other" regimen, stable usage shares of actual regimens, and the assumed zero QALY survival of "other" in the naïve estimation.

Despite the potential under reporting of BevPacCar in the data, the cost per QALY of survival for NSC lung cancer also increased over the period. Figure 3.9 shows the 38 % increase in the cost of one QALY of survival from chemotherapy over just three years. Although the increase is not as large as that for colorectal cancer, the annual growth rate of

Figure 3.10: Average Patient BRC Chemotherapy Costs, 2003-2006

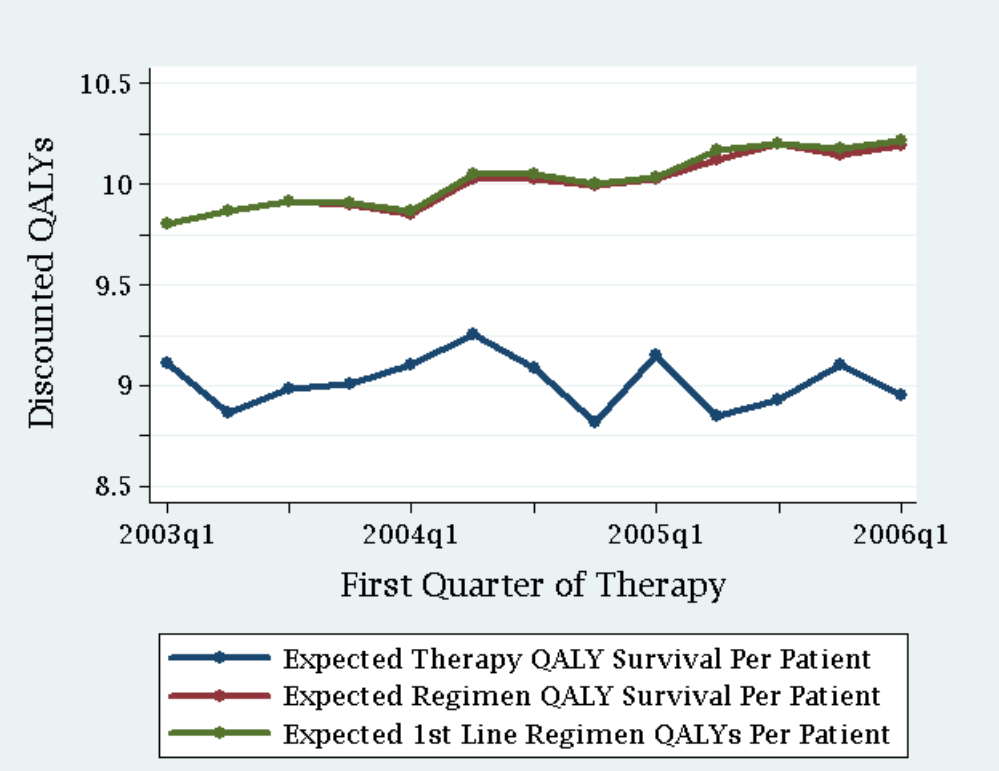


29.1% far outpaces the growth in overall health care spending.

3.3.3 Breast Cancer

Between 2003 and 2006, the combination of doxorubicin and cyclophosphamide remained the dominant adjuvant treatment of breast cancer, though overall treatment trended away from older fluorouracil based regimens toward newer paclitaxel and trastuzumab based regimens. Historically adjuvant treatment for breast cancer was inexpensive. The confirmation in 2005 that trastuzumab, another monoclonal antibody, was effective at treating localized HER2-positive cancer, that is, tumors over-expressing an influential hormone gene, dramatically increased the average cost of care since trastuzumab-based regimens were more than three

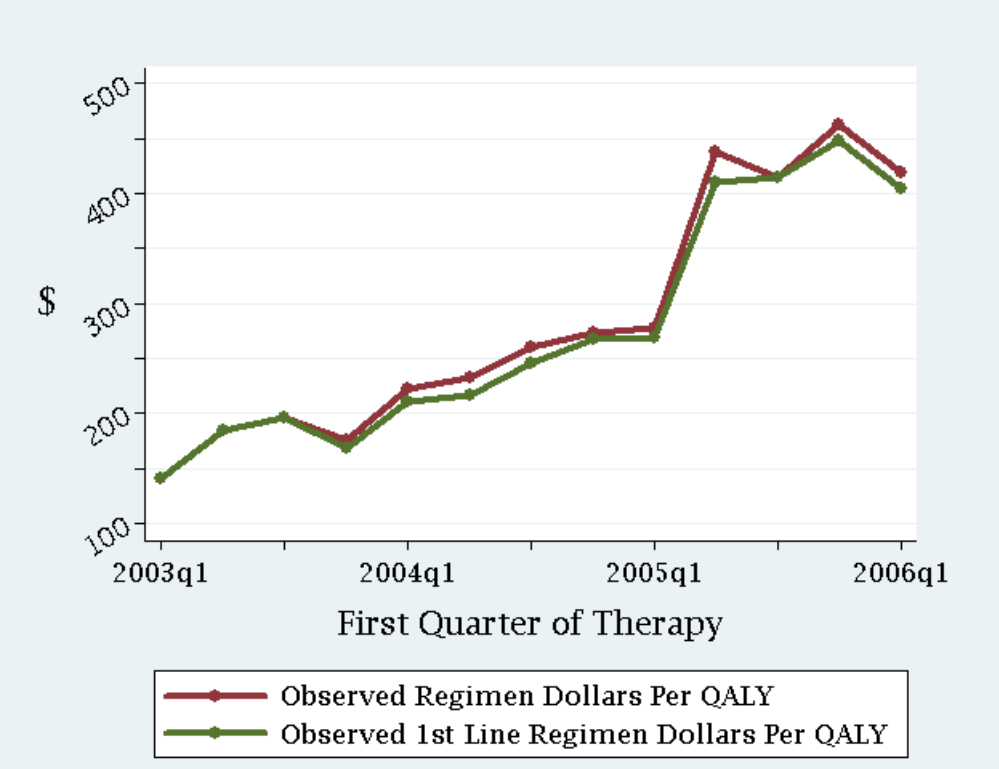
Figure 3.11: Expected QALYs of Survival from BRC Chemotherapy, 2003-2006



times as expensive as the non-trastuzumab-based regimens. Thus, despite trastuzumab's small share of the market, Figure 3.10 clearly depicts its inclusion in therapy regimens as having a dramatic increase in the cost of care.

Since adjuvant chemotherapy is used for smaller, localized tumors, survival from breast cancer chemotherapy is much longer than for either of colon or lung cancer. Figure 3.11 bears out this difference and shows that QALYs of survival from chemotherapy regimens steadily increased between 2003 and 2006. That steady increase stands in sharp contrast to the abrupt increase in overall cost of adjuvant chemotherapy. Figure 3.12 shows that the cost per QALY of survival jumped with the increased use of expensive therapies in 2005. Like the cost of one QALY of survival from colorectal cancer, the three year increase here equates to a roughly 200% increase in the cost of each QALY.

Figure 3.12: Dollars Per QALY of Survival from BRC Chemotherapy, 2003-2006



3.4 Discussion and Conclusions

Unsurprisingly, the cost of chemotherapy, like health care spending overall, increased between 2002 and 2006. In the cases presented here, the increase ranges from 38% for metastatic NSC lung cancer, to over 300% for metastatic colorectal cancer. Though that cost increase has been met with increases in quality-adjusted survival, the growth rate for the latter is dwarfed by the former, so that the cost per QALY has increased significantly across each scenario. Shifts in the market composition toward newer, more expensive though marginally more efficacious regimens prompted these changes, with the rise of monoclonal antibody therapies being a primary example of that shift.

Although this study sets out to study the changes in average benefits from chemotherapy use, the question directly posed to payers is the value derived from any new particular therapy. Incremental cost effectiveness ratios (ICERs) are widely used, wherein a new regimen's cost is compared to a consistent reference drugs cost and the difference is scaled by their difference in quality adjusted life years saved or lost. This ratio is the basis for recommendation by the United Kingdom's National Health Service's National Institute for Health and Clinical Excellence (NICE). Although the Patient Protection and Affordable Care Act expressly prohibited the used of such measures, NICE, for example, has recommended against the use of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer on the basis of its cost and incremental effectiveness (National Institute for Health and Clinical Excellence, 2007). Identifying consistent reference treatments and employing ICERs throughout would enable an analysis such as this to be more in line with a policymaker or private payer's question. Nevertheless, identifying and quantifying at even a broad level as here can make clear how spending per patient translates into outcomes across time.

REFERENCES

Abrevaya, Jason and Jerry A. Hausman (1999). "Semiparametric Estimation with Mismeasured Dependent Variables: an Application to Duration Models for Unemployment Spells." *Annales d'conomie et de Statistique/Annals of Economics and Statistics*, 55/56, 243–275.

Ackerberg, Daniel (2003). "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination." *International Economic Review*, 44(3), 1007–1040.

Allison, Paul D. (1982). "Discrete-time methods for the analysis of event histories." *Sociological Methodology*, 13, 61–98.

Allison, Paul D. (1984). *Event History Analysis: Regression for Longitudinal Event Data*, vol. 46 of *Quantitative Applications in the Social Sciences*. Sage Publications, Beverly Hills, CA.

Amemiya, Takeshi (1985). *Advanced Econometrics*. Harvard University Press, Cambridge, MA.

Baker, Sharyn, Jaap Verweij, Eric Rowinsky, Ross Donehower, Jan Schellens, Louise Grochow, and Alex Sparreboom (2002). "Role of Body Surface Area in Dosing of Investigational Anticancer Agents in Adults, 1991–2001." *Journal of the National Cancer Institute*, 94(24), 1883–1888.

Barreca, Alan, Melanie Guldi, Jason M. Lindo, and Glen Waddell (2010). "Running and Jumping Variables in RD Designs: evidence Based on Race, Socioeconomic Status, and Birth Weights." IZA Discussion Papers 5106, Institute for the Study of Labor (IZA).

Becker, G. S. and K. M. Murphy (1988). "A theory of rational addition." *The Journal of Political Economy*, 96(4), 675–700.

Burstein, Harold J. (2012). "If the Choices Are All the Same, Why Not Prefer the Less-Expensive Option?" *Journal of the National Comprehensive Cancer Network*, 10(4), 425–426.

Camacho, Nuno, Bas Donkers, and Stefan Stremersch (2011). "Predictably non-Bayesian: Quantifying Salience Effects in Physician Learning About Drug Quality." *Marketing Science*, 30(2), 305–320.

Caplan, Arthur (2011). "Will Evidence Ever Be Sufficient to Resolve the Challenge of Cost Containment?" *Journal of Clinical Oncology*, 29(15), 1946–1948.

Centers for Disease Control and Prevention (2011). "Vital Signs: Colorectal Cancer Screening, Indcidence, and Mortality – United States, 2002-2010." *Morbidity and Mortality Weekly Report*, 60, 884–889.

Chaloupka, Frank J. and Kenneth E. Warner (2000). "The Economics of Smoking." In Culyer, Anthony J. and Joseph P. Newhouse (Eds.) "Handbook of Health Economics," Elsevier, Amsterdam.

Ching, Andrew (2010). "Consumer Learning and Heterogeneity: Dynamics of Demand for Prescription Drugs After Patent Expiration." *International Journal of Industrial Organization*, 28(6), 619–638.

Ching, Andrew and Masakazu Ishihara (2010). "The Effects of Detailing on Prescribing Decisions Under Quality Uncertainty." *Quantitative Marketing and Economics*, 8(2), 123–165.

Chintagunta, Pradeep, Renna Jiang, and Ginger Jin (2009). "Information, Learning, and Drug Diffusion: The Case of Cox-2 Inhibitors." *Quantitative Marketing and Economics*, 7(4), 399–443.

Coscelli, Andrea and Matthew Shum (2004). "An Empirical Model of Learning and Patient Spillovers in New Drug Entry." *Journal of Econometrics*, 122(2), 213 – 246.

Cosslett, Stephen R. (1993). "Estimation from Endogenously Stratified Samples." In Maddala, Gangadharrao Soundalyarao, Calyampudi Radhakrishna Rao, and Hrishikesh D. Vinod (Eds.) "Handbook of Statistics," Elsevier, Amsterdam.

Crawford, Gregory and Matthew Shum (2005). "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica*, 73(4), 1137–1173.

DeCicca, Philip, Donald Kenkel, and Alan Mathios (2008). "Cigarette taxes and the transition from youth to adult smoking: smoking initiation, cessation, and participation." *Journal of Health Economics*, 27(4), 904 – 917.

Dickstein, Michael (2011). "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice." Job Market Paper, Harvard Business School, Boston, MA 02163.

Erdem, Tülin (1998). "An Empirical Analysis of Umbrella Branding." *Journal of Marketing Research*, 35(3), 339–351.

Erdem, Tülin and Michael Keane (1996). "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science*, 15(1), 1–20.

Erdem, Tülin and Baohong Sun (2002). "An Empirical Investigation of the Spillover Effects of Advertising and Sales Promotions in Umbrella Branding." *Journal of Marketing Research*, 39(4), 408–420.

Forster, Martin and Andrew M. Jones (2001). "The Role of Tobacco Taxes in Starting and Quitting Smoking: duration Analysis of British Data." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 164(3), 517–547.

Greenlee, Robert, Mary Beth Hill-Harmon, Taylor Murray, and Michael Thun (2001). "Cancer Statistics, 2001." *CA: A Cancer Journal for Clinicians*, 51(1), 15–36.

Gurney, Howard (1996). "Dose Calculation of Anticancer Drugs: A Review of the Current Practice and Introduction of an Alternative." *Journal of Clinical Oncology*, 14(9), 2590–2611.

Hausman, J. A., Jason Abrevaya, and F. M. Scott-Morton (1998). "Misclassification of the dependent variable in a discrete-response setting." *Journal of Econometrics*, 87(2), 239–269.

Hausman, Jerry A. (2001). "Mismeasured variables in econometric analysis: problems from the right and problems from the left." *The Journal of Economic Perspectives*, 15(4), 57–67.

Heitjan, Daniel F. and Donald B. Rubin (1990). "Inference from Coarse Data Via Multiple Imputation with Application to Age Heaping." *Journal of the American Statistical Association*, 85(410), 304–314.

Ibrahim, Amna (2003). "Past FDA Approvals in Colorectal Cancer." In "Clinical Trial Endpoints in Colorectal Cancer," Center for Drug Evaluation and Research, Food and Drug Administration, Washington, DC.

Janakiraman, Ramkumar, Shantanu Dutta, Catarina Sismeiro, and Philip Stern (2008). "Physicians' Persistence and Its Implications for Their Response to Promotion of Prescription Drugs." *Management Science*, 54(6), 1080–1093.

Janakiraman, Ramkumar, Catarina Sismeiro, and Shantanu Dutta (2009). "Perception Spillovers Across Competing Brands: A Disaggregate Model of How and When." *Journal of Marketing Research*, 46(4), 467–481.

Jena, Anupam, Seth Seabury, Darius Lakdawalla, and Amitabh Chandra (2011). "Malpractice Risk According to Physician Specialty." *New England Journal of Medicine*, 365(7), 629–636.

Jenkins, Stephen P. (1995). "Easy estimation methods for discrete-time duration models." *Oxford Bulletin of Economics and Statistics*, 57(1), 129–136.

Jones, Andrew M. (1999). "Adjustment costs, withdrawal effects, and cigarette addiction." *Journal of Health Economics*, 18(1), 125–137.

Jönsson, Bengt and Nils Wilking (2007). "The Burden and Cost of Cancer." *Annals of Oncology*, 18(Supplement 3), iii8–iii22.

Keane, Michael P. and Robert M. Sauer (2009). "Classification error in dynamic discrete choice models: implications for female labor supply behavior." *Econometrica*, 77(3), 975–991.

Kenkel, Donald S., Dean R. Lillard, and Alan D. Mathios (2003). "Smoke or fog? the usefulness of retrospectively reported information about smoking." *Addiction*, 98(9), 1307–1313.

Kenkel, Donald S., Dean R. Lillard, and Alan D. Mathios (2004). "Accounting for misclassification error in retrospective smoking data." *Health Economics*, 13(10), 1031–1044.

Kessler, Daniel and Mark McClellan (1996). "Do Doctors Practice Defensive Medicine?" *Quarterly Journal of Economics*, 111(2), 353–390.

Klabunde, Carrie, Arnold Potosky, Julie Legler, and Joan Warren (2000). "Development of a Comorbidity Index Using Physician Claims Data." *Journal of Clinical Epidemiology*, 53(12), 1258–1267.

Lee, Donghoon and Kyungchul Song (2011). "Simulated MLE for Discrete Choices Using Transformed Simulated Frequencies." Working paper, University of British Columbia.

Lovett, Mitchell (2008). *Unstable Consumer Learning Models: Structural Models and Experimental Investigation*. Ph.D. thesis, Duke University.

Lucarelli, Claudio, Sean Nicholson, and Minjae Song (2010). "Bundling Among Rivals: A Case of Pharmaceutical Cocktails." Working Paper 16321, National Bureau of Economic Research.

Machlin, S. R., J. C. Kleinman, and J. H. Madans (1989). "Validity of mortality analysis based on retrospective smoking information." *Statistics in Medicine*, 8(8), 997–1009.

Manski, Charles F. and Francesca Molinari (2010). "Rounding Probabilistic Expectations in Surveys." *Journal of Business and Economic Statistics*, 28(2), 219–231.

Murphy, Sherry L., Jiaquan Xu, Kenneth D. Kochanek, and Division of Vital Statistics (2012). "Deaths: Preliminary Data for 2010." *National Vital Statistics Reports*, 60(4), 1–60.

Narayanan, Sridhar and Puneet Manchanda (2009). "Heterogeneous Learning and the Targeting of Marketing Communication for New Products." *Marketing Science*, 28(3), 424–441.

Narayanan, Sridhar, Puneet Manchanda, and Pradeep Chintagunta (2005). "Temporal Differences in the Role of Marketing Communication in New Product Categories." *Journal of Marketing Research*, 42(3), 278–290.

National Comprehensive Cancer Network (1996). "Colorectal Cancer Practice Guidelines." *Oncology*, 10, 140–175.

National Comprehensive Cancer Network (2004). "Colon Cancer." In "Clinical Practice Guidelines in Oncology," NCCN, Fort Washington, PA.

National Institute for Health and Clinical Excellence (2007). "Bevacizumab and Cetuximab for the Treatment of Metastatic Colorectal Cancer." In "NICE Technology Appraisal Guidelines," National Health Service.

Nicholson, Sean, Claudio Lucarelli, and David A. Asch (2011). "What Does the US Buy and What Does It Get from Its Cancer Chemotherapy Drugs Compared to European Nations?" Working paper, Cornell University.

NIH Consensus Conference (1990). "Adjuvant Therapy for Patients with Colon and Rectal Cancer." *Journal of the American Medical Association*, 264(11), 1444–1450.

Osborne, Matthew (2011). "Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination." *Quantitative Marketing and Economics*, 9(1), 25–70.

Patient Protection and Affordable Care Act (2010). *Public Law 111-148*. 124 Stat. 119.

Patrick, D. L., A. Cheadle, D. C. Thompson, P. Diehr, T. Koepsell, and S. Kinne (1994). "The validity of self-reported smoking: a review and meta-analysis." *American Journal of Public Health*, 84(7), 1086–1093.

Poterba, J. M. and L. H. Summers (1995). "Unemployment benefits and labor market transitions: a multinomial logit model with errors in classification." *Review of Economics and Statistics*, 77(2), 207–216.

Ries, L.A.G., M.P. Eisner, C.L. Kosary, B.F. Hankey, B.A. Miller, L. Clegg, A. Mariotto, E.J. Feuer, and B.K. Edwards (Eds.) (2004). *SEER Cancer Statistics Review, 1975–2001*. National Cancer Institute, Bethesda, MD.

Schrag, Deborah, Laura Cramer, Peter Bach, and Colin Begg (2001). "Age and Adjuvant Chemotherapy Use After Surgery for Stage III Colon Cancer." *Journal of the National Cancer Institute*, 93(11), 850–857.

Soares, Michael (2005). "Off-Label Indications for Oncology Drug Use and Drug Compendia: History and Current Status." *Journal of Oncology Practice*, 1(3), 102–105.

Suranovic, Steven M., Robert S. Goldfarb, and Thomas C. Leonard (1999). "An economic theory of cigarette addiction." *Journal of Health Economics*, 18(1), 1–29.

Torelli, Nicola and Ugo Trivellato (1993). "Modelling inaccuracies in job-search duration data." *Journal of Econometrics*, 59(1-2), 187–211.

Train, Kenneth (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, 2nd edn.

Vohra, Sunita, Larissa Shamseer, and Margaret Sampson (2011). "Efficacy Research and Unanswered Clinical Questions." *Journal of the American Medical Association*, 306(7), 709.

Warren, Joan, Linda Harlan, Angela Fahey, Beth Virnig, Jean Freeman, Carrie Klabunde, Gregory Cooper, and Kevin Knopf (2002). "Utility of the SEER-Medicare Data to Identify Chemotherapy Use." *Medical Care*, 40(8), IV55–IV61.