Price Indices and the Value of Innovation with Heterogenous Patients*

Claudio Lucarelli†
Sean Nicholson‡
Nicholas Tilipman§

March 31, 2022

Abstract

Many countries use uniform cost-effectiveness criteria to determine whether to adopt a new medical technology for the entire population. This approach assumes homogeneous preferences for expected health benefits and side effects. We examine whether new prescription drugs generate welfare gains when accounting for heterogeneous preferences by constructing quality-adjusted price indices in the market for colorectal cancer drug treatments. We find that while the efficacy gains from newer drugs do not justify high prices for the population as a whole, innovation improves the welfare of sicker, late-stage cancer patients. A uniform evaluation criterion would not permit these innovations despite welfare gains to a subpopulation.

Keywords: Innovation, Cost-of-living, Healthcare cost, Heterogeneity
JEL Classification Codes: I11, I31, L00

*We thank Mark Duggan, David Molitor, Kate Bundorf, Adam Ullman, Ron Katznelson, and seminar participants at Hoover IP2 and Stanford University. Artem Gulish and Brigid Farrell provided excellent research assistance.

†Department of Health Care Management, The Wharton School, University of Pennsylvania. Contact: clucarel@wharton.upenn.edu

‡Department of Policy Analysis & Management, Cornell University. Contact: sean.nicholson@cornell.edu

§Division of Health Policy and Administration, School of Public Health, University of Illinois at Chicago. Contact: tilipman@uic.edu

1 Introduction

Persistent health care spending growth in the United States has been a policy concern for the last eight decades and the subject of substantial research and policy activity (Newhouse, 1992). Innovation of medical technology has been identified as the consensus catalyst for this growth, despite several policy efforts (Smith et al., 2009). Most policies and market evolutions have involved changes to health insurance, such as increased patient cost-sharing, managed care, prospective payment mechanisms, and, more recently, narrow provider networks and high-deductible health plans. None of these efforts have had a direct impact on innovation or the adoption of new technology, and therefore their effect has occasionally decreased the level of spending but not its growth rate (Chernew and Newhouse, 2011; Cutler et al., 2000; McWilliams et al., 2016; Manning et al., 1987).

Given the limited success of these interventions, the question of whether society obtains enough value for the price of medical care continues to be an important question, and controversial in situations where the stakes are high for patients, payers, providers, and manufacturers (Lakdawalla et al., 2012). Some countries in the Organization for Economic Cooperation and Development (OECD) have introduced centralized health technology assessment agencies to evaluate the value of new technology. The most prominent such institution is the United Kingdom's National Institute for Health and Clinical Excellence (NICE), which employs cost-effectiveness analysis (CEA) to assess a technology's cost per incremental unit of benefit.[1] Although the United States has no such system in place, periodically there are proposals to transition to a system of centralized medical technology value assessment that would allow public insurers such as Medicare to negotiate prices based on CEA evidence. (Frank and Nichols, 2019).

The relative simplicity of CEA as a tool for allocating scarce medical resources and its effectiveness at reducing spending explains its widespread use in European countries and why it is frequently considered for the U.S. However, using a single threshold to decide whether a particular technology is cost-effective is at odds with conventional economic models of demand, whose appeal is precisely to be able to capture the heterogeneity of different consumers' values for a good. In fact, the heterogeneity of willingness to pay that is central to a demand function highlights that a homogeneous valuation does not exist, unless preferences and outcomes are homogeneous in the population, which is very unlikely. The homogeneous cost effectiveness (CE) thresholds used to determine coverage or market entry, and the homogeneity of preferences assumed in CEA, are problematic for the efficient allocation of health care resources. Indeed, Garber and Phelps (1997) studied the economic foundations of CEA and concluded that it can identify egalitarian allocations, but applying a homogeneous rule to a heterogeneous population is not likely to yield a Pareto optimal allocation. Although the political equilibrium might generate a social marginal valuation that places uniform value on outcomes, we show that the resulting one-size-fits-all policy could be detrimental for some segments of the population based on their heterogeneous private value for new technology.[2]

[1] A widely used metric to measure the benefits of medical technology and interventions is a Quality Adjusted Life Year (QALY). This measure incorporates both the expected length and quality of life

[2] Through its recent action, the U.K. has also recognized that using a homogeneous CE threshold may produce suboptimal allocations. Specifically, the Cancer Drug Fund was established in the U.K. in 2010 to pay for oncology treatments that NICE did not recommend for coverage. Between 2010 and 2016, Lakdawalla et al. (2014) documents that one billion pounds were invested in the Fund to assuage public concerns regarding NICE's recommendations

In this paper we examine whether pharmaceutical innovation has delivered enough value to different groups of patients in the U.S. to justify high prices. We focus on the market for colorectal cancer treatments, a setting that has seen both substantial innovation as well as considerable price increases over the past several decades. We first rely on the discrete choice literature to estimate a model of oncologists' demand for colorectal cancer treatments (McFadden, 1974; Berry, 1994; Berry et al., 1995; Nevo, 2000; Berry and Pakes, 2007). We next compute the consumer surplus for different patient subpopulations resulting from new regimens entering the market. Finally, we construct a series of quality-adjusted price indices for each group by computing the change in drug prices consistent with the calculated welfare effect. This allows us to assess whether, for these patient subpopulations, quality-adjusted prices rose or fell once new goods entered the market.

The market for colorectal cancer treatments provides an ideal setting for estimating quality-adjusted price indices. From a policy perspective, the innovation-driven changes in drug attributes and prices of cancer treatments has led to considerably different approaches across countries regarding whether and how to regulate access to medical care. As an example, comparing the market shares of colorectal cancer regimens between the U.S. and Europe in 2005, it is clear that regimens with higher expected survival rates (measured in months within each segment, with more effective regimens appearing at the top of the figure) had a larger market share in the U.S. versus Europe, as shown in Figure 1. These differences may lead to significant differences in health outcomes. Indeed, Stevens et al. (2015) show that cancer mortality reductions have been larger in the countries that spend more on treatment. Considering its high prevalence in every country, cancer is therefore an area where the lack of a consensus between governments and their citizens regarding the value of new technology and how to measure that value could have a substantial impact on welfare.

In addition, the colorectal cancer market is attractive for purposes of estimation. It is a setting where *physicians* are sensitive to price. Because colorectal cancer treatments are usually administered in a physician's office, physicians take ownership of the drugs they infuse to their patients, are aware of prices, incur substantial carrying costs, and are subject to oversight from health insurers. As a result, physicians consider both clinical and financial considerations when choosing drug treatments for their patients, which leads to fewer instances of supplier-induced demand. This contrasts with most health care markets where physicians are aware of attributes but insulated from price, or where patients face only part of the price.

A key contribution of our paper is to account for physician and patient heterogeneity when estimating demand for drugs. Our model posits that physicians act as agents for their patients, but permits them to have idiosyncratic preferences for drug attributes. We specify a model where, in choosing a therapy for a patient, physicians take into account a patient's value for each efficacy measure (e.g., median life expectancy when the drug was tested in a randomized controlled trial), their tolerance for side effects, and the extent to which patients are willing to accept greater toxicity to achieve greater efficacy.[3] We also allow physicians' utility to depend on the regimen price. As mentioned above, physicians in this setting incur substantial carrying costs and, therefore, should

that blocked or delayed access to new therapies.

[3]These attributes, and the heterogeneity in preferences patients exhibit for them, have been recognized in the medical literature. Kravitz et al. (2004) state that the ability to predict how a patient will respond to therapy will depend on what they call response to treatment, vulnerability to side effects, and how those are traded with each other in terms of utility.

have aligned incentives with their patients. However, one possible concern is that physicians could potentially profit from more expensive drugs after being reimbursed by the patient's insurer.[4] To address this concern, in our preferred approach we specify a highly flexible model that allows physician preferences for price and non-price attributes to vary with their patients' characteristics. In addition, our model allows for *unobserved* heterogeneity in preferences for price across physicians through the inclusion of random coefficients.

We estimate our model using detailed and novel microdata on oncologists' prescription drug choices from IntrinsiQ, which we combine with drug regimen pricing and market share data from IMS Health and the Surveillance Epidemiology and End Results (SEER) datasets. We also obtain regimen efficacy and side effects measures directly from FDA-approved package inserts that describe the results of phase 3 clinical trials conducted for each drug. These data afford us several advantages. First, our pricing measure reflects the actual transaction price between customers and wholesalers, thereby allowing us to directly observe the costs physicians incur for each regimen. Second, our clinical trial data and market share data allow us to estimate physician treatment responses to various non-price characteristics of each regimen. Finally, our micro data allow us to follow oncologists' treatment decisions for individual patients over time, as new regimens enter the market and as the characteristics of their patient panels change. In effect, this allows us to separate the extent to which physician prescribing behavior is driven by changes in patient characteristics versus physicians' idiosyncratic preferences for specific drug regimens, such as price.

In our demand model, we identify physican's price sensitivity by relying on "Hausman-style" instruments (Nevo, 2000, 2001). Specifically, our instruments for a particular regimen in a particular time period are the average prices of all other regimens in prior time periods. The primary identifying assumption is that these average lagged prices are correlated with the price of a regimen in a particular period, but not with any period-specific shocks to regimen demand. Such instruments originally come from the observation that cost information is rarely observed, and more so at the product level. Therefore, prices in other markets could be good candidates for instruments given their correlation with marginal cost.[5] We provide the details of our instrumental variables in section 4.

Our approach allows us to address two important challenges that are relevant in markets with substantial innovation. The first challenge relates to correctly measuring the cost of living, which has been a concern since the introduction of the Consumer Price Index (CPI) more than 100 years ago. The CPI Commission (Boskin et al., 1998), appointed in 1995, concluded that most of cost of living mismeasurement comes from the introduction of new goods and changes in the quality of goods. Our model explicitly incorporates such changes in quality through our measure of consumer surplus. Alternative and simpler price aggregators such as Laysperes and Paasche, are known to

[4]Exploiting changes in reimbursement policy by Medicare, Jacobson et al. (2016) report that physicians' changes in treatments administered are small compared to the survival benefits for lung cancer patients, indicating that physicians behave altruistically. These results are also consistent with Kolstad (2013), who reports that physicians intrinsic motivation is about four times larger than the extrinsic motivation induced by profit incentives. Cockburn and Anis (2001) and Dunn (2012) make similar assumptions that physicians act in the best interest of patients when they construct medical price indices for rheumatoid arthritis and cholesterol drugs, respectively.

[5]Critically, markets can be defined geographically or over time, as in Nevo (2001) who uses the prices of all the quarters in his data to construct instruments for prices. We go a step further, and similarly to Berry et al. (1995), exploit the fact that cost shocks to competing products have an impact on own prices through competitive pressures, however, are unlikely to be correlated with the product-specific unobserved quality.

have limitations in terms of how they handle new goods, the substitution patterns they impose, and their resulting biases in estimating the true cost of living index. The Laspeyres index, for instance, measures the change in the cost of a fixed basket of goods from a base period, and therefore, assumes no substitution due to relative price changes, which usually overstates the true cost of living. The opposite is true for the Paasche index, which weighs by current consumption patterns. Our model allows us to obtain the substitution patterns directly from demand estimation.[6] The second challenge is to account for the heterogeneity of patients' willingness to pay due to different price sensitivity and clinical outcomes from treatment. Our approach derives values directly from the demand function, and therefore the heterogeneity in willingness to pay is built naturally into the price index.

We generate several indices for purposes of comparison for the period between 1993 and 2005. During this period, all new products remained on patent, which provides the most conservative scenario for a price index because we do not include decreases in prices due to eventual generic or biosimilar competition.[7] First, we construct a "naive" price index that simply reports the mean price of cancer regimens in each quarter, relative to our initial period (1993), without any adjustments for changes in regimen attributes but allowing the market shares (or bundle weights) to change. The naive index is based on the prices physicians pay to acquire each regimen and the market share of each regimen. This index shows a large increase in prices of about 29,000% (see Figure 6), which emphasizes the importance of quality-adjusting in a market with innovation and quality improvement.

We next construct several price indices from the discrete choice demand estimation methods described above. These methods vary in how much preference heterogeneity is allowed, starting from a simple case in which all heterogeneity is accounted for by the idiosyncratic logit error term (McFadden, 1974), and extending to a random coefficients model whose distribution depends on observed patient and unobserved physician characteristics (Berry et al., 1995, 2004). In models with idiosyncratic tastes for products, the addition of new products mechanically increases consumer surplus as the error term generates positive demand, and therefore every new product increases consumer surplus. For this reason, we also estimate a pure characteristics model of demand in order to test whether the price index changes when removing the idiosyncratic error term that is present in the logit and random-coefficients models (Berry and Pakes, 2007). The latter model includes only idiosyncratic tastes for characteristics and, to the best of our knowledge, has only been applied empirically in a few studies of computer markets (Song, 2007, 2008; Nosko, 2014). Following Nevo (2003)'s insight, we are explicit regarding our assumptions about the evolution of the outside option and the unobserved components of utility, and we report price indices for each case.

Our results show that accounting for patient heterogeneity matters, with the price index differing

[6] Since (Hicks, 1940) it has been widely accepted that a promising approach for constructing cost of living price indices is to measure the change in consumer surplus and welfare implications from the introduction of new goods using estimated demand systems (Trajtenberg, 1989; Pakes et al., 1993; Petrin, 2002; Nevo, 2003; Hausman, 1996). Related to this literature, but taking a more general view on innovation, Murphy and Topel (2006) also use consumer surplus to measure the value of improvements in health and longevity in the U.S., which they later relate to the value of investing in basic research to decrease cancer mortality.

[7] By July 2019, four of the five patent-protected drugs in our sample were competing with lower-priced generic or biosimilar versions of their molecules.

for distinct patient types. Our principal finding is that innovation in colorectal cancer treatments appear to have largely benefited patients with late-stage cancers undergoing multiple rounds of treatment, as compared to patients with early-stage cancers. Indeed, we find that patients with advanced cancers undergoing multiple treatment rounds experience a quality-adjusted price decrease of 44 percent, whereas patients with early-stage cancer receiving their first drug treatment experience an increase of 128 percent. In the first case, the value that patients obtain from the more expensive treatment greatly offsets its price, whereas the situation is reversed for patients with less advanced cancer. If one disregards heterogeneity and only uses an average value, such as with CEA, the price index increases by between five and 15 percent, and one would conclude that the innovation we observed in this period was not worth it. If a uniform rule prevented patients with advanced cancer from receiving their preferred treatment, these patients would have experienced large welfare losses. Our results are consistent with the conclusions reached by Lakdawalla et al. (2012), who found that patients at the end of life place large value on hope, and are willing to pay more for therapies that offer the possibility of a longer survival period. Our results are also consistent with survey data suggesting that people substantially value health investments for sick patients (Nord et al., 1995). Similarly, Bauer et al. (2021) develop a theoretical framework for valuing health improvements and find that the willingness to pay for medical treatment by a sick individual is several times greater than what a healthy person would be willing to pay for preventive care that improves longevity by the same amount. Ignoring heterogeneity in value assessment will lead to one-size-fits-all decision by payers that are disconnected from the preferences of the physician-patient dyads.

The heterogeneity in values we find, and the implied divergent evolution of the price index for different types of patients, is also informative for the efficient design of health insurance. A homogeneous rule based on thresholds determined by CEA would provide coverage for treatments below the CE threshold, and no coverage for treatments above it. Optimal health insurance should allow for the differences in value to express themselves in the market, by either offering a variety of plans that accommodate this heterogeneity in preferences with varying premia (Pauly, 2017), or by allowing patients to internalize treatment costs at the margin through the use of "top-up" insurance, where patients pay the incremental cost relative to a fully covered baseline treatment (Einav et al., 2016). Based on our results, a one-size-fits-all decision rule based on CEA or homogeneous criteria will not lead to optimal decisions for the adoption of technology to treat colorectal cancer.

In the next section we provide the institutional background that underlies our modeling assumptions; section 4 describes our model for estimating the demand system under various assumptions and the computation of the price indices; in section 3 we describe our data; section 5 presents our results on demand estimation and the evolution of the price indices; and in section 6 we conclude.

2 Background

2.1 Medical Price Indices

There are several important empirical challenges when measuring a medical price index (Berndt et al., 2000). One challenge is that quantities of the components in an index change as new goods

displace old goods in the market. Therefore, a price index must allow the bundle of goods to change over time as well as the weight on each component in the bundle. Another challenge is that often transaction prices are not observed in health care. Most publicly available data sets contain "list prices," i.e. prices that pharmaceutical firms, medical device firms, physicians, or hospitals charge for a product or service rather than what they actually receive. Health insurers usually receive substantial discounts off of list prices, but net prices are often unobserved.

In addition, due to health insurance, most consumers do not face the full price of medical care and may consume beyond the point where the marginal value of care equals the full price. Patients also rely on physicians to provide information regarding the value of medical goods and services. The implication is that consumer purchases in the medical market will not necessarily reveal their marginal valuation of a good or service. Finally, and most relevant for our paper, price increases are often accompanied by quality improvements. In our setting, we look at the decisions of informed physicians, who are aware of prices, and where rebates are unusual. Therefore, perhaps the greatest empirical challenge for constructing meaningful medical price indices is how to account for the changing quality of medical products and services. For example, life expectancy and mortality rates are often used as proxies for quality, but in some medical settings improvements in the quality of life are more important than the length of life. Information on such measures is often difficult to obtain.

There have been relatively few studies of whether medical prices are rising or falling once one takes into consideration the attributes of the new products and consumers' valuations of those attributes. Existing studies can be placed into three groups based on the method used. One method to account for changing quality is to measure actual changes or expected changes in health outcomes (e.g., life expectancy for a heart attack patient or remission of a depressive episode) due to the adoption of new medical technologies, monetizing these improvements based on separate estimates from the literature (e.g., $100,000 per year of life), and subtracting the value of the health gains from the rising costs. Cutler et al. (1998) show that the life expectancy of heart attack patients increased by eight months between 1984 and 1991. The value of per-patient expected longevity ($11,100) increased three times more than average treatment costs ($3,600) during this time period in real terms, which implies that the quality-adjusted price index fell by about one percent annually. Berndt et al. (2002) conclude that the real cost of treating major depression decreased by about two percent per year between 1991 and 1996 once one takes into account the probability that a patient's depression will go into remission. Eggleston et al. (2011) conclude that the cost of treating diabetes declined between 1999 and 2009 once one accounts for the value of reduced coronary heart disease mortality. Lakdawalla et al. (2015) find that the estimated value of health benefits produced by new colorectal cancer drugs were about equal to their higher prices, whereas prices fell for drug treatment of multiple myeloma patients once accounting for expected health outcomes.

A second method is to estimate hedonic price indices by regressing prices on objective measures of product attributes and time indicator variables. Cockburn and Anis (2001) find that rheumatoid arthritis prices rose over time, even after accounting for each drug's expected efficacy and toxicity. Dunn (2012) reports that the prices of cholesterol drugs fell by 30 percent between 1996 and 2007 once one accounts for each drug's expected change in a patient's bad and good cholesterol levels.

The two studies closest to our approach, in the third group, also construct price indices, or the components of an index, based on discrete choice demand models. Estimating a nested multinomial logit model, Trajtenberg (1989) shows that the social returns to CT scanner innovation exceeded the R&D costs, especially when the first products were introduced (1972-1978). Dunn (2012) uses a logit model to estimate patients' demand for cholesterol drugs and concludes that quality-adjusted prices fell by about 25 percent between 1996 and 2007, similar to the results from the hedonic model.

2.2 Medical Treatment of Colorectal Cancer

Colorectal cancer is an appropriate case study for examining the welfare effects of medical innovation because it is a common health condition, the majority of patients today are treated with drugs that did not exist 25 years ago, and treatment costs are rising rapidly. According to the National Cancer Institute (NCI), approximately 140,000 patients are diagnosed with colorectal cancer in the United States each year, resulting in about 50,000 deaths annually. This places colorectal cancer as the fourth most common cancer based on number of new patients, after breast, prostate, and lung. It is estimated that people born today will have a 4.5 percent chance of being diagnosed with colorectal cancer over their lifetime. According to the NCI, between 2005 and 2011 colorectal cancer patients had a 65 percent chance of surviving for five years. The probability a patient will survive for five years ranges from 90 percent for those diagnosed with Stage I cancer to 13 percent for those diagnosed with Stage IV (or metastatic) cancer.[8]

Colorectal cancer is treated differently depending on the stage of the disease at diagnosis. Most patients with a Stage I, II, or III disease at diagnosis will have the tumor removed surgically, or resected, in a hospital. The National Comprehensive Cancer Network (NCCN) recommends that patients with Stage III disease receive six months of drug treatment following the resection; they do not recommend drug treatment for Stage I patients; and they recommend that Stage II patients consider drug treatment after discussing the potential benefits and side effects with their physician. Most Stage IV patients are treated with drugs, either to shrink the tumor so that it can be resected, following a resection, or when the tumor is unresectable in order to lengthen and improve the quality of life. If a patient's cancer progresses after the initial drug treatment, patients often receive a second regimen (i.e., a second line of therapy) that might have been tested on similar patients and approved as a second-line therapy, or might be a regimen formally approved as a first-line treatment. The latter case would be an example of a physician prescribing a drug off label. In this paper we do not model the process by which a physician and her patient decide whether or not to receive drug treatment; we examine prescription drug choices conditional on the decision to receive drug treatment.[9]

[8] Cancers are classified into four stages, with higher numbers indicating that the cancer has spread to the lymph nodes (Stage III) or beyond its initial location (Stage IV).

[9] There is some evidence that the characteristics of cancer patients who receive prescription drug treatment has changed over time. Brouwer et al. (2018) find a small increase in the probability that patients in the Netherlands with Stage II, III, or IV colorectal cancer were treated with prescription drugs in 2000-2004 relative to 1995-1999. Likewise, Davidoff et al. (2013) find that cancer patients (of all types and stages) in the U.S. were more likely to receive drug treatment in 2000-2007 relative to 1995-1999. These results are consistent with healthier patients selecting into treatment over time. Because we include a patient's cancer stage in our micro empirical analysis, which should be the most important patient characteristic affecting a physician's treatment decision, we do not believe our parameter

Treating colorectal cancer in the US is expensive. Paramore et al. (2006) report that between 1998 and 2004, which is part of the time period we study in this paper, newly-diagnosed metastatic colorectal cancer patients in the United States received $97,000 of additional medical treatment in the 13 months following their diagnosis relative to a similar (based on age, gender, and geography) group of people (Paramore et al., 2006). Most of the spending occurred for hospital care ($37,400, on average), including the surgical resection, and physician visits ($34,600), including drugs and their administration.

In this paper we report a price index for drug treatment only, not for the total cost of treating colorectal cancer. We focus on drug costs because this is where the treatment innovation has occurred since 1996, and drug innovation has been the main driver of the subsequent increase in treatment costs and improvements in health. In our data set, the average cost of providing six-months of drug treatment to a colorectal cancer patient increased from about $127 in 1993 to over $36,000 in 2005 (see Figure 6). Hospital costs, on the other hand, did not increase substantially during this time period for colorectal cancer patients nor were there any surgical innovations that would have lengthened life or improved the quality of life substantially.

About 95 percent of colorectal cancer patients who were treated with drugs between 1993 and 1996 received a chemotherapy regimen called 5-FU/leucovorin. This regimen was inexpensive because the patents on the two component drugs (fluorouracil and leucovorin) had expired, so there were many companies offering low-cost generic versions. Between 1996 and 2004, five new drugs were approved to treat colorectal cancer. By 2005, these drugs collectively captured about 82 percent of the market. Most of the new drugs are biologics that target the inner workings of cancer cells, whereas 5-FU/leucovorin is a standard chemotherapy treatment that targets all fast-growing cells. Although the new drugs extended life in randomized clinical trials relative to chemotherapy and sometimes have less severe side effects due to their targeted nature, they are also priced much higher than standard chemotherapy drugs. Most colorectal cancer patients who are treated with pharmaceuticals receive multiple drugs in the form of a regimen rather than a single drug, similar to anti-retroviral "cocktail" treatments for AIDS patients. For example, the regimen with the greatest market share in 2005 contained four separate drugs: bevacizumab, oxaliplatin, fluorouracil, and leucovorin. Most of our analysis in this paper, therefore, is conducted at the level of a regimen rather than a drug. We describe the characteristics of the various drug regimens in greater detail in the section 3.

2.3 Medical Technology Adoption and Use: United States Versus Europe

In 2018, prescription drug spending per capita in the Unites States was $1,024 versus an average of $495 for the five most populous European countries.[10] Danzon and Furukawa (2008) show that US spending is higher due to greater use of newer (and more expensive) drugs and higher prices for the same drug, especially drugs still under patent protection.

Cost effectiveness analysis is a major reason European countries spend less on pharmaceutical

estimates are biased due to changes over time.

[10] Hartman et al. (2020) and OECD Health Statistics (2019): https://www.oecd.org/health/health-data.htm. Adjusted for exchange rates and purchasing power parity. Earlier years were used for some European countries when necessary.

drugs than the US. In the United Kingdom, for example, The National Institute for Care and Health Excellence (NICE) generally recommends rejecting new medical technologies with an incremental cost effectiveness ratio (relative to the standard of care) above $50,000 per quality-adjusted life year (QALY) for the population as a whole (Claxton et al., 2015). In 2007, just after our sample period, NICE recommended not covering bevacizumab and cetuximab, the two most expensive colorectal cancer drugs that were launched in the US in 2004.[11] As a result, most patients in the UK who want these drugs have to pay the full price themselves. The benefit of NICE, and similar organizations in other European countries, is that they often put pressure on pharmaceutical firms to cut their prices in order to get covered by European public payers. Cost effectiveness also restricts the use of therapies deemed to be priced above the average patient's valuation. The drawback of applying a uniform decision rule for the entire population is that some patients are likely to value the rejected therapies above the incremental price. Einav et al. (2016) propose a "top-up" health insurance policy where patients pay the incremental cost of the more expensive treatment to efficiently sort patients across treatment options.

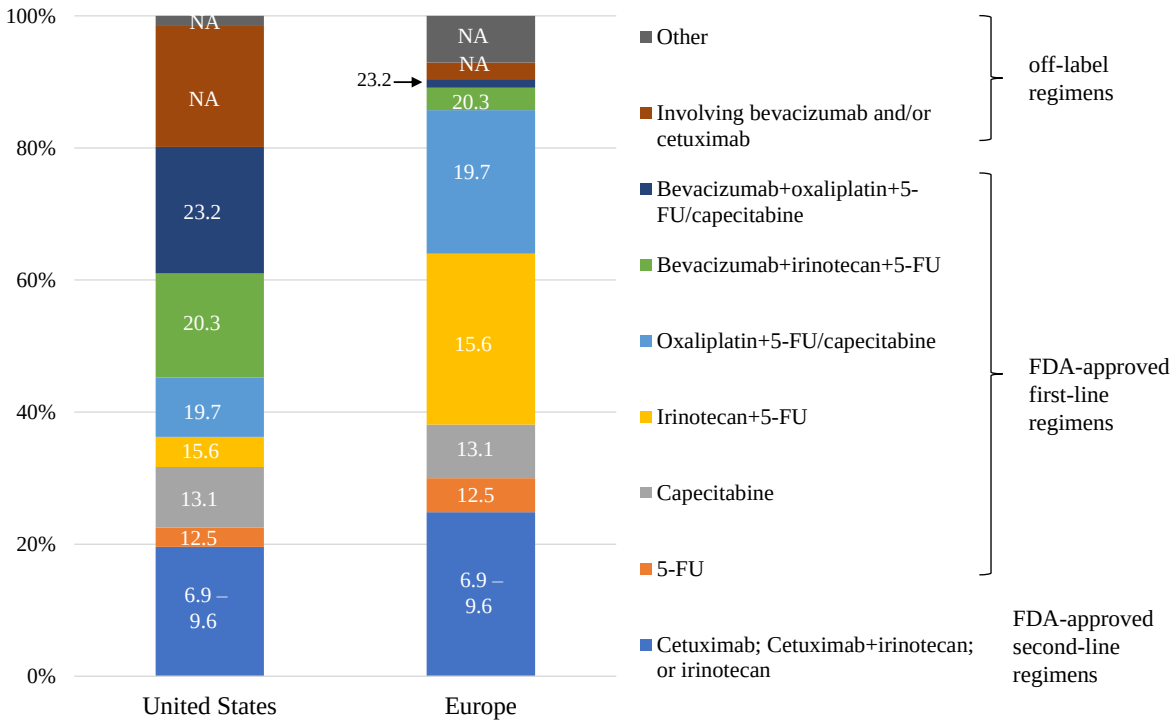
To depict these tradeoffs in our context, in Figure 1 we compare the colorectal cancer drug regimens received by patients in the US versus the five largest European countries for the 12-month period between April 2005 and March 2006, which is at the end of our sample period.[12] To highlight the important role that preferences play, we focus on patients who are likely to value the health benefits of prescription drugs highly: patients diagnosed with Stage IV colorectal cancer who are receiving their second line (or higher) of drug treatment because the cancer progressed during or after their initial treatment. As we show in section 5, these patients experienced a reduction in quality-adjusted prices in the US, whereas patients with less advanced cancer experience a price increase. The height of each bar in Figure 1 indicates a regimen's market share, and the regimens are ordered from the lowest median length of survival at the bottom to the highest at the top, with the median survival months depicted inside each bar. The survival data are based on randomized controlled trials, which avoids concerns about patient selection. Regimens that have not been formally approved by the Food and Drug Administration (FDA) (i.e., off-label uses) do not have reported survival data, thus the "N/A."

Beginning at the bottom of Figure 1, about 20 percent of second-line metastatic patients in the US receive a second-line regimen, with either irinotecan or cetuximab as a component in the regimen. The use of these regimens is about five percentage points higher in Europe. The median survival is low for the second-line therapies (6.9 to 9.6 months) in large part because the patients enrolled in the trials had already been treated for colorectal cancer. The next 61 and 66 percent of patients in the US and Europe, respectively, were treated with an approved first-line therapy. However, many more US patients received a first-line therapy that included the expensive drug bevacizumab (the green and dark blue bars), which both have median survival in excess of 20 months (when tested on metastatic patients receiving their initial drug treatment). Finally, US patients were much more likely to receive an off-label regimen involving either cetuximab or bevacizumab (e.g., either of these drugs combined with a drug approved for a type of cancer other

[11] https://www.nice.org.uk/guidance/ta118/chapter/4-Evidence-and-interpretation.

[12] US data are from IntrinsiQ and European data are from Synovate. France, Germany, Italy, Spain, and the United Kingdom are included in the European average.

Figure 1: Market Shares of Colorectal Cancer Regimens in the United States and Europe



Notes: Market share of drug regimens received by patients diagnosed with Stage IV colorectal cancer who are receiving their second line (or higher) of drug treatment for April 2005 through March 2006. European shares are averages for France, Germany, Italy, Spain, and the United Kingdom. US data are from IntrinsiQ and European data are from Synovate. The height of each bar indicates a regimen's market share and the median survival months from the regimen's randomized controlled trial is depicted inside each bar. The regimens are ordered from the lowest median length of survival at the bottom to the highest at the top. Regimens that have not been formally approved by the FDA (i.e., off-label uses) do not have reported survival data, thus have "N/A" for survival months.

than colorectal) relative to European patients. Cetuximab is another expensive drug. In summary, physicians in the US were much less constrained in how they treated patients, which contributed to much higher spending but also presumably better expected outcomes. However, there would have been little data at the time documenting the extent to which effective first-line therapies translated into effective second-line therapies.

2.4 Physician Behavior and Modeling Approach

In most markets one can use prices to infer consumers' willingness to pay for products and services because consumers are well informed and face the full price. Several features of the health care market create challenges for estimating and interpreting a quality-adjusted price index. Physicians (or sometimes hospitals on behalf of their physicians) purchase oncology drugs from wholesalers (who previously purchased them from drug manufacturers), administer them to patients in their offices, and then bill the patient's insurance company for the drugs.[13] Health insurers reimburse physicians for cancer drugs and the time required to administer the drugs to patients, and control moral hazard by requiring patients to pay a portion of the treatment cost and by restricting (expensive) treatments to patients who are not likely to benefit from them (e.g., requiring a physician to receive authorization from the insurer before treating a patient).[14] Finally, patients rely on physicians to recommend a treatment. Thus, colorectal cancer patients do not face the full price of drugs and rely on physicians to articulate both the choice set and the attributes of the drugs in the choice set. Thus, physicians have an opportunity to exploit their informational advantage and recommend particular drugs if the health insurer's reimbursement amount deviates substantially from physicians' drug acquisition costs.

We model the choice of a colorectal cancer drug regimen from a physician's perspective by positing that physicians choose treatments to maximize their own utility. We focus on physicians because they are well informed about attributes and prices, they care about their patients' health, and they are subject to oversight by profit-maximizing private health insurers. Physicians observe the full price of oncology drugs when they purchase them from wholesalers and store them in their offices. Physicians should be aware of the efficacy and safety of oncology drugs based on how the drugs performed in randomized controlled trials during the approval process (and recorded on the FDA-approved drug label), as well as by observing how their own patients respond to the drugs.

One criticism of our assumption is that physicians may, in fact, be insulated from the price of oncology drugs because health insurers ultimately reimburse them for drug acquisition costs. Although physicians are eventually reimbursed, they do take temporary ownership of oncology drugs. As such, physicians face the possible risk of not being reimbursed by health insurers and may incur substantial carrying costs. For example, a physician who pays $50,000 for the drugs in one patient's regimen and experiences a three-month delay between when she acquires the drugs and when her practice is reimbursed by a health insurer would incur an inventory carrying cost of $625 at an interest rate of five percent. Furthermore, if physicians care about patients' out-of-

[13] Based on data from IMS Health, 59 percent of colorectal cancer drugs in the third quarter of 2005 were purchased by physician offices/clinics and 28 percent by hospitals. The remainder was purchased by retail and mail order pharmacies, health maintenance organizations, and long-term care facilities.

[14] Medicare patients, for example, pay 20 percent of the price of drugs administered by physicians if they have Part B coverage and no Medigap supplemental insurance.

pocket costs they will internalize part of the price. Both carrying costs and patients' out-of-pocket costs, which are often a percentage of the price the insurer pays, increase with a drug's price.

Another criticism of our assumption is that if physicians earn profits on oncology drugs and profits influence prescribing decisions, this might bias the price and attribute estimates in our model. In the early 2000s the federal government concluded that oncologists were earning profits on most oncology drugs by acquiring them for less than the Medicare reimbursement amount (Government Accountability Office, 2001). At the time Medicare reimbursed oncologists 95 percent of a drug's listed average wholesale price (AWP), whereas physicians could acquire many drugs from wholesalers for substantially less than AWP. Most private health insurance companies reimbursed physicians using a similar formula, so these profits occurred for all patient types [15] Most of these profits were eliminated in 2005 when Medicare began reimbursing oncologists based on a drug's actual average selling price (ASP) rather than its list price (Medicare Payment Advisory Commission, 2006).

There is evidence that the profits oncologists earned on drugs did affect treatment decisions, especially prior to 2004. Several papers identify effects by examining variations in profits across oncology drug treatments due to geographic reimbursement rules (Jacobson and Newhouse, 2006) or variation within treatments over time following the 2005 Medicare policy change (Jacobson et al., 2010, 2016), or the entry of generic drugs (Conti et al., 2012). Our interpretation of these studies is that the magnitudes of the effects are small. In Jacobson and Newhouse (2006), for example, a one-standard deviation increase in reimbursement generosity is associated with an increase of about five percent in the cost of the drug treatments prescribed to colorectal cancer patients.

One factor supporting our assumption is that health insurers have the incentive to encourage physicians to tradeoff the full price of drugs against patients' health benefits, and the tools to do so. When physicians use expensive drugs it forces a health insurer to charge its customers higher premiums, but when patients receive treatments that improve their health, the insurance plan becomes more attractive to prospective enrollees. Therefore, insurers should set a patient's drug co-insurance rate, physicians' drug profits, and its oversight policies (e.g., the criteria for preauthorizing a proposed oncology treatment) to encourage optimal drug treatment decisions. Dunn (2012) and Cockburn and Anis (2001) make similar assumptions that physicians and insurers act in the best interest of patients when interpreting their medical price indices.

The heterogeneity in physicians' preferences we assume in the model below is driven primarily from physicians treating patients who have different survival goals and tolerance to side effects. Other sources of heterogeneity among physicians are documented by Chan et al. (2021) in relation to physicians' skill levels. Although important in specialties like radiology, where they report that variation in skill can explain a sizable part of the variation in diagnostic decisions, we do not have reason to believe that higher quality oncologists would have more homogeneous practice styles, or that improving skills of lower-skilled physicians would lead to more standardization. On the contrary, a recent panel of the National Comprehensive Cancer Network Policy Summit [16] defines higher quality oncologists as those able to correctly interpret patient priorities. Another

[15] In the IntrinsiQ data set we use, Medicare patients account for about one-half of all colorectal cancer patients who receive drug treatment.

[16] https://www.nccn.org/about/news/newsinfo.aspx?NewsID=1686

source of heterogeneity is knowledge, as reported in Agha and Molitor (2018), who find that the diffusion of oncology drugs is related to the geographic location of the physicians running clinical trials. However, the magnitude of these effects are not large enough to explain the treatment heterogeneity we observe, and the reported effects are also short-lived. In addition, among all the drugs these authors studied, colorectal cancer drugs were the ones with the largest number of trial sites, which suggests that for colon cancer this other source of heterogeneity is less of a concern.

3 Data

3.1 Prices and Market Shares

We use a number of different data sources to collect four types of information: drug prices, regimen market shares, the recommended quantity/dose of each drug used in each regimen, and regimen attributes from randomized controlled clinical trials. Wholesalers purchase drugs from manufacturers and then sell them to physician practices, hospitals, retail pharmacies, and other customers. IMS Health records transactions between wholesalers and its customers.[17] Specifically, IMS Health collects information on the sales in dollars and the quantity of drugs purchased by 10 different types of customers (e.g., hospitals, physician offices, retail pharmacies) from wholesalers in each quarter from 1993 through the third quarter of 2005. Prices and quantities are reported separately by National Drug Classification (NDC) code, which are unique for each firm-product-strength/dosage-package size. We calculate the average price paid per milligram of active ingredient of a drug by averaging across the different NDC codes for that particular drug. IMS Health reports the invoice price a customer actually pays to a wholesaler, not the average wholesale price (AWP), which often differs substantially from the true transaction price. We use nominal rather than real prices because any deflator would itself be a price index, and we do not want to build one index on top of another.

The price we calculate includes on-invoice discounts (e.g., for paying the wholesaler promptly) but does not include any discounts or rebates a customer may receive from a manufacturer after purchasing the product from the wholesaler. Based on interviews with oncologists and an analysis reported in Lucarelli et al. (2010), we do not believe that manufacturers offered substantial rebates during this period for the drugs in our sample.[18] A recent paper Howard et al. (2015) agrees that "we do not believe that rebates – refunds from manufacturers to hospitals, physicians, pharmacies and third party insurers – are large in the market for new anticancer drugs." Although we have information on 10 different types of customers, we focus on the prices paid by the two largest customers – hospitals and physician offices – because most oncology drugs are infused in a physician's office or hospital clinic.

[17] IMS Health is now part of a company called IQVIA

[18] For the five patent-protected colorectal cancer drugs in our study, these authors compared prices that include discounts and rebates to the IMS prices that we use in this paper. They found that prices from the two data sources were within two to four percent of one another, which is consistent with no or small rebates/discounts. Although pharmacy benefit managers, or PBMs, are able to negotiate price discounts for health insurers for many patient-administered oral drugs, PBMs are less effective at negotiating discounts on the physician-administered drugs that we examine in this paper. And discounts tend to be smaller for drugs treating life-threatening conditions such as cancer.

Most colon cancer patients who receive drug treatment are treated with a regimen that contains two or more component drugs. The IMS Health data contain information on market share by drug, but not market share for the combinations of drugs (regimens) actually used by patients. We rely, therefore, on two different sources for regimen-specific market shares. IntrinsiQ is a company that provides information systems to oncologists to help them determine the proper drug dosing for their cancer patients. As a result, IntrinsiQ collects monthly data from its oncology clients on the types of drugs used for patients. IntrinsiQ provided data on the proportion of US colorectal cancer patients (of all ages) treated with drugs who are treated with each regimen for each month between January 2002 and September 2005.[19]

We derive market shares for the 1993 to 2001 period from the Surveillance Epidemiology and End Results (SEER) data set, which tracks the health and treatment of cancer patients over the age of 64 in states and cities covering 26 percent of the United States population.[20] We calculate the proportion of colorectal cancer patients who are treated with each drug regimen in each quarter based on Medicare claims data available in SEER. In October 2003, approximately 48 percent of all colorectal cancer patients treated with cancer therapy drugs were 65 years or older.[21] In the 1993 to 2001 period, when there were relatively few treatment options for colorectal cancer, we include all regimens that contain drugs that were explicitly approved by the Food and Drug Administration (FDA) for colorectal cancer and had a market share greater than two percent. Market shares of all other drugs are combined into an outside option, which in this early period will consist primarily of off-label drugs - drugs approved for conditions other than colorectal cancer that are used on colorectal cancer patients.[22] In the 2002 to 2005 period, the outside option includes off-label drugs, regimens with less than one percent market share in the third quarter of 2005 (the end of the sample period), and regimens with missing attribute data.

Market shares for the 12 regimens in our sample and the outside option are plotted in Figure 2. The regimens are also described more fully in Table 1 and Table 2. Between 1993 and 1996, about 95 percent of colorectal cancer patients were treated with 5-FU/leucovorin, which at that time was generic.[23] Irinotecan (brand name Camptosar) was approved by the FDA for treating colorectal cancer in 1996, and over the next several years the market share of irinotecan (approved as a second-line treatment for metastatic colorectal cancer patients who had already been treated with a different drug regimen) and irinotecan combined with 5-FU/LV (approved as a first-line treatment) grew at the expense of 5-FU/LV.[24] Capecitabine (Xeloda), a tablet that produces the same chemical response as 5-FU/LV, was approved for treatment of colorectal cancer in April of

[19] Because we observe the market shares of regimens among patients with colorectal cancer, we do not need to worry about off-label use. Off-label use occurs when a physician treats a colorectal cancer patient with a drug that has not been approved by the FDA to treat colorectal cancer, or when a physician uses a drug approved for colorectal cancer on a patient with a different type of cancer. In October 2005, seventy-six percent of patients being treated with the four drugs approved solely for the treatment of colorectal cancer (irinotecan, oxaliplatin, cetuximab, and bevacizumab) actually had colorectal cancer. That is, off-label use accounted for approximately 24 percent of the quantities of these drugs.

[20] SEER contains data on the incidence rate of cancer among the non-elderly, but only has medical claims available for Medicare patients.

[21] Based on data from IntrinsiQ.

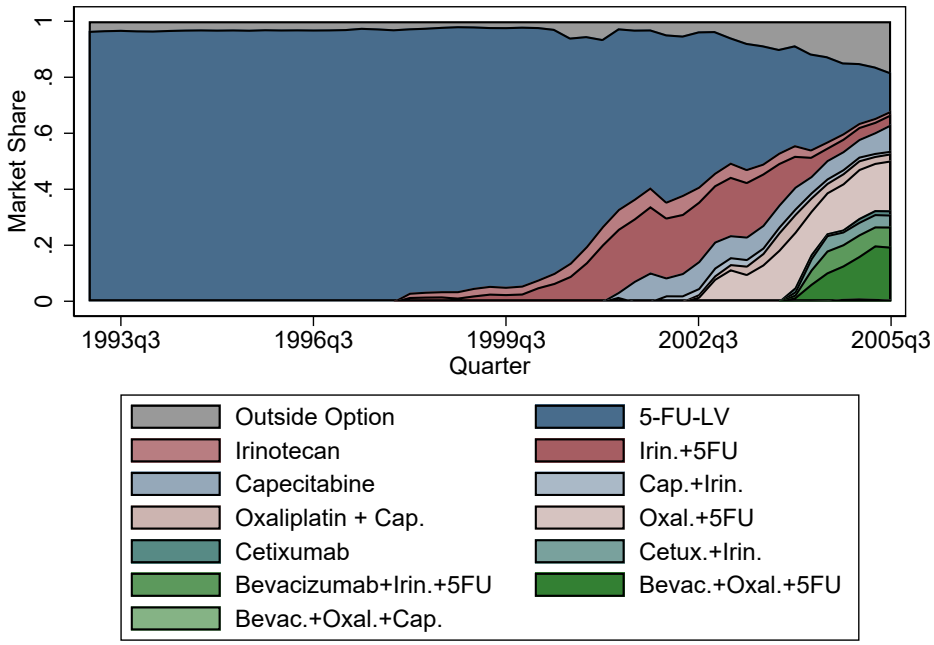
[22] Off-label use is more likely to occur if a patient's initial treatment has been unsuccessful.

[23] 5-FU contains the drug fluorororacil.

[24] Because it takes Medicare a while to code new drugs into their proper NDC code, for several quarters a new drug will appear in the outside option.

2001 and was administered as a standalone therapy or combined with irinotecan. All other drugs for treating colorectal cancer in our sample other than capecitabine are delivered intravenously under the supervision of a physician or nurse.

Figure 2: Regimen Market Shares



Notes: Market shares of colorectal cancer regimens between 1993 and 2005. Market share is measured as the percentage of colon cancer patients who are treated with drugs that are treated with a specific regimen. Data from IntrinsiQ and SEER.

Oxaliplatin (Eloxatin) was introduced in August of 2002, followed by cetuximab (Erbitux) and bevacizumab (Avastin) in February of 2004. By the third quarter of 2005, two of the regimens created by these three new drugs (oxaliplatin + 5-FU/LV; and bevacizumab + oxaliplatin + 5-FU/LV) surpassed the market share of 5-FU/LV, whose share had fallen to about 14 percent. The market shares of several regimens change sharply in the first quarter of 2002 when we use market share data from IntrinsiQ rather than SEER. One explanation for these changes is that Medicare patients may be treated with different regimens than non-Medicare patients. Another possible explanation is that the samples used by IntrinsiQ and/or SEER may not be consistent.[25] In order to smooth market shares between the pre- and post-2002 periods, we apply a regimen-specific factor to adjust the pre-2002 market shares based on the ratio of total (from IntrinsiQ) to Medicare-only (from SEER) market shares for the four quarters of 2002, when the two data sets overlap.

In order to calculate the price per regimen, we require information on the quantity of each drug in a regimen. The National Comprehensive Cancer Network (NCCN) reports the typical amount of active ingredient used by physicians for the major regimens. We supplement this where necessary with dosage information from drug package inserts, conference abstracts, and journal articles. Dosage information is reported in Appendix C. For example, the standard dosage schedule for the

[25] The SEER sample is drawn from locations representing 26 percent of the U.S. population.

regimen with the second largest market share in 2005 is 85 milligrams (mg) of oxaliplatin per meter squared of a patient's body surface area infused by IV on the first day of treatment, followed by a 1,000 mg infusion of 5-FU per meter squared of surface area on the first and second treatment days, and a 200 mg infusion of leucovorin per meter squared on the first and second treatment days. This process is repeated every two weeks. We price the regimens for a representative patient who has 1.7 meters squared of surface area (Jacobson and Newhouse, 2006) weighs 80 kilograms, and is treated for 24 weeks. Regimen prices are derived by multiplying the average price a customer paid per milligram of active ingredient in a quarter by the recommended dosage amounts for each drug in the regimen over a 24-week period.[26]

3.2 Regimen Attributes

We obtain most of the attribute information for each regimen from the FDA- approved package inserts that accompany each drug. These inserts describe the phase 3 randomized controlled clinical trials that were conducted, including the number and types of patients enrolled in the trials, the treatments administered to the treatment and control groups, the health outcomes for patients in both groups, and the side effects experienced by those patients. The benefit of data from clinical trials is that the regimen attributes are not affected by patient selection. Often there are multiple observations for a regimen, either because a manufacturer conducted separate trials of the same regimen, or because a regimen may have been the treatment group in one clinical trial and the control group in a subsequent trial run by a different firm. In these cases we calculate the mean attributes across the separate observations. Where necessary, we supplement the package insert information with abstracts presented at oncology conferences and in journal articles.

The attribute information is summarized in Table 1 and Table 2, organized according to the year when each regimen was introduced. Table 1 shows three measures of a regimen's efficacy: the median number of months patients survive after initiating therapy; the percentage of patients who experience a complete or partial reduction in the size of their tumor (i.e., the response rate); and the mean number of months (across patients in the trial) before their cancer advanced to a more serious state.[27] For all three of these measures, higher values are associated with superior health outcomes. We also record whether a regimen contains the capecitabine tablet, which should make the administration of the regimen more convenient for a patient, and whether the regimen is approved (and was tested) as a second-line treatment. Efficacy measures for second-line regimens will generally be worse than those for first-line regimens because the patients' cancer is likely to be more advanced at the beginning of the clinical trial and the first treatment was not completely successful.

We also collected data on the percentage of patients in phase 3 trials who experienced either a grade 3 or a grade 4 side effect for five separate conditions: abdominal pain, diarrhea, nausea, vomiting, and neutropenia. These are displayed in Table 2. Although many more side effects are recorded for most regimens, these five were consistently recorded across the 12 regimens in the sample. Side effects are classified on a standard one to four scale, with four being the most severe.

[26] The regimens are priced using data for the contemporaneous quarter only.

[27] Cancers are classified into four stages, with higher numbers indicating that the cancer has metastasized beyond its initial location.

Table 1: Drug Regimens and Efficacy Measures

Regimen	Launch Year	Price (2005)	Survival Months	Efficacy Measures	
				Response Rate	Time to Progression
5-FU + Leucovorin	1991	75	12.5	20.8	4.7
Irinotecan (Second Line)	1996	23,478	9.6	14.3	4.2
Irinotecan + 5-FU/LV	1996	20,124	15.6	35.4	6.7
Capecitabine	2001	9,223	13.1	21.0	4.4
Ironetecan + capecitabine	2001	21,385	15.6	35.4	6.7
Oxaliplatin + 5-FU/LV	2002	25,426	19.7	46.8	9.1
Oxaliplatin + capecitabine	2002	31,936	19.7	36.5	8.1
Cetuximab (Second Line)	2004	53,859	6.9	10.8	1.5
Cetuximab + irinotecan (Second Line)	2004	73,519	8.6	22.9	4.1
Bevacizumab + oxaliplatin + 5-FU/LV	2004	76,636	23.2	41.0	9.9
Bevacizumab + oxaliplatin + capecitabine	2004	57,541	23.2	41.0	9.9
Bevacizumab + irinotecan + 5-FU/LV	2004	46,991	20.3	45.0	10.6

Notes: All attribute information is based on results of patients in Phase 3 randomized controlled clinical trials. Median survival is given in months. Time to progression is measured as the mean number of months for a tumor to advance to a more severe stage.

Higher values for the side effect attributes are associated with worse health outcomes.

New colorectal cancer regimens tend to be more efficacious than the existing regimens, with side effect profiles that are sometimes more and sometimes less severe than earlier regimens. Consider the new entrant in 1996, irinotecan + 5-FU/LV (third row of Table 1). Relative to patients who received 5-FU/LV in a clinical trial (first row of Table 1), patients in clinical trials who received irinotecan + 5-FU/LV lived 3.1 months longer, on average, had a 14.6 percentage point higher probability of experiencing a reduction in the size of their tumor, and experienced a two month delay in the time it took for the cancer to advance to a more severe state. However, patients taking the new regimen were more likely to experience four of the five side effects listed in Table 2.

Oxaliplatin + 5-FU/LV, which was launched in 2002 (sixth row of Table 1 and Table 2), is more efficacious and has fewer severe side effects than irinotecan + 5-FU/LV. Patients in clinical trials of the former regimen lived an average of 3.8 months longer, had a 10.7 percentage point higher probability of experiencing a reduction in the size of their tumor, and experienced a 2.4 month delay in the time it took for the cancer to advance to a more severe stage relative to the latter regimen. Oxaliplatin + 5-FU/LV patients are also less likely to experience a grade 3 or 4 side effect for four of the five measures relative to irinotecan + 5-FU/LV. Finally, the arrival of bevacizumab + oxaliplatin + 5-FU in 2004 increased the median survival time by about four months relative to oxaliplatin + 5-FU/LV, with substantial improvements with one side effect measure and worse performance on the other four measures.

Two new second-line regimens entered the market in 2004 to compete against the first second-line regimen (irinotecan) that was launched in 1996.[28] cetuximab + irinotecan has a substantially better response rate than irinotecan administered by itself, although median survival is shorter. The new regimen also is superior to irinotecan on all five of the side effect measures.

3.3 Individual Choice Data

To estimate micro-moments with observed physician and patient heterogeneity, we rely on a disaggregated sample of the IntrinsiQ data from January 2003 through September 2005. As mentioned above, the data contain monthly reports from a sample of oncologists on the types of drugs used for patients. Our data contain identifiers for the provider, including site of service (i.e., physician practice), location (geographic region), and attending physician. In addition, it contains patient information including a patient identifier, age, cancer stage, weight, and the number of the course of treatment (e.g., first-round, second-round, etc.). We also observe which drug regimen any particular physician at any particular site prescribed to a particular patient for a particular round of treatment.

We aggregate the data to a panel of physician-patient-round treatment choices. We drop Stage I cancer patients as well as patients undergoing either third- or fourth-round treatments due to the small number of observations for which these occur. Summary statistics for the individual choice sample are reported in Table 3. Older patients appear to be disproportionately prescribed lower-priced regimens, whereas younger patients appear to be prescribed higher-priced regimens. For instance, the average age for 5-FU + LV, the lowest-priced regimen on the market ($75) is 69

[28] Regimens that include the tablet, capecitabine, are chemically equivalent to regimens that include 5- FU/LV.

Table 2: Drug Regimens and Side Effects Measures

			Side Effects Measures				
Regimen	Launch Year	Price (2005)	Abdominal Pain	Diarrhea	Nausea	Vomiting	Neutropenia
5-FU + Leucovorin	1991	75	5.5	10.4	4.8	4.4	33.7
Irinotecan (Second Line)	1996	23,478	16.0	31.0	17.0	12.0	26.0
Irinotecan + 5-FU/LV	1996	20,124	5.3	24.0	11.9	8.0	39.5
Capecitabine	2001	9,223	9.5	15.0	4.0	4.5	3.0
Ironetecan + capecitabine	2001	21,385	5.3	24.0	11.9	8.0	39.5
Oxaliplatin + 5-FU/LV	2002	25,426	6.0	15.4	4.4	5.5	38.8
Oxaliplatin + capecitabine	2002	31,936	6.0	21.9	15.6	11.3	3.7
Cetuximab (Second Line)	2004	53,859	9.0	2.0	2.0	3.0	5.0
Cetuximab + irinotecan (Second Line)	2004	73,519	8.0	22.0	6.0	7.0	5.0
Bevacizumab + oxaliplatin + 5-FU/LV	2004	76,636	8.0	23.1	7.9	8.6	12.2
Bevacizumab + oxaliplatin + capecitabine	2004	57,541	8.0	23.1	7.9	19.0	12.0
Bevacizumab + irinotecan + 5-FU/LV	2004	46,991	8.0	34.0	1.0	1.0	21.0

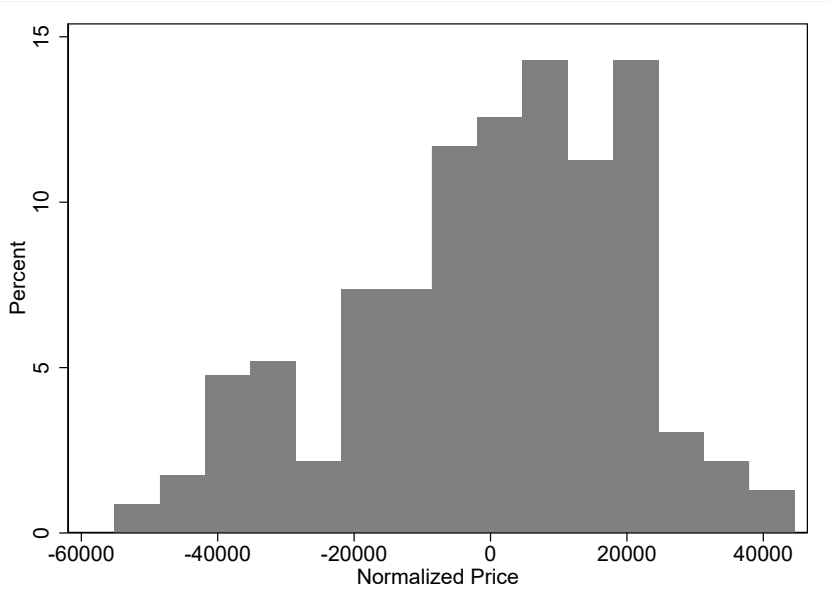
Notes: All attribute information is based on results of patients in Phase 3 randomized controlled clinical trials. Side Effects are measured as the percentage of patients who experienced a serious (i.e. grade 3 or 4) side effect of the given type in the randomized control trial"

years old whereas the average age of patients being prescribed capecitabine ($9,223) is 74 years old. Conversely, the average age for bevacizumab + oxaliplatin + 5-FU LV ($76,636) is only 63 years old versus the average age for cetuximab ($53,859) of 66.

In addition, patients with more advanced cancer tend to be prescribed either second-line treatments (e.g. 73 percent of patients on irinotecan are Stage IV patients, compared with only six percent of Stage II patients) or treatments that are otherwise higher-priced (e.g. 71 percent of patients on bevacizumab + oxaliplatin + capecitabine are Stage IV patients compared with only seven percent being Stage II patients). Patients whose cancer is early-stage tend to be prescribed lower-priced products (e.g., 34 percent of patients on 5-FU LV are Stage II cancer patients, and only 14 percent are Stage IV patients).

Apart from patient characteristics, attending physicians also appear to have preferences for specific regimen characteristics. To see this, we plot the distribution of physician fixed effects from a regression of regimen price on age, gender, round of treatment, cancer stage, quarter, region, and physician fixed effects. The results, depicted in Figure 3, show that there is considerable variation across physicians in the average price per chosen regimen even when netting out market, time, and patient characteristics. Controlling for patient observables, this suggests that individual physicians are willing to pay from $60,000 below to $40,000 above the average price of a regimen. This variation motivates our inclusion of unobserved random coefficients on regimen price and second-line regimens in some of the demand models that follow. Intuitively, these unobserved coefficients will be identified by correlation in drug regimen characteristics prescribed by individual physicians across patients over time.

Figure 3: Residual Average Regimen Price Across Physicians Net of Patient Characteristics



Notes: This figure plots the distribution of physician fixed effects (where the average is normalized to 0) from a regression of regimen price on patient characteristics as well as quarter, region, and physician dummies. Dependent variable is the price level of the regimen.

Table 3: Summary Statistics of Patient Characteristics by Drug Regimen

				Individual Attribute			
Regimen	Price (2005)	Age	Female	Second-Round	Stage 2	Stage 3	Stage 4
5-FU + Leucovorin	75	68.99	0.48	0.97	0.34	0.52	0.14
		(12.47)	(0.50)	(0.18)	(0.47)	(0.50)	(0.35)
Irinotecan (Second Line)	23,478	69.24	0.49	0.40	0.06	0.21	0.73
		(12.49)	(0.50)	(0.49)	(0.25)	(0.41)	(0.45)
Irinotecan + 5-FU/LV	20,124	65.63	0.47	0.81	0.04	0.21	0.76
		(11.60)	(0.50)	(0.40)	(0.18)	(0.41)	(0.43)
Capecitabine	9,223	73.68	0.51	0.80	0.24	0.39	0.37
		(13.07)	(0.51)	(0.40)	(0.43)	(0.49)	0.49)
Ironetecan + capecitabine	21,385	62.09	0.52	0.70	0.03	0.36	0.61
		(14.07)	(0.51)	(0.47)	(0.17)	(0.49)	(0.50)
Oxaliplatin + 5-FU/LV	25,426	63.66	0.46	0.83	0.12	0.46	0.41
		(12.12)	(0.50)	(0.38)	(0.33)	(0.50)	(0.49)
Capecitabine + Oxaliplatin	31,936	66.33	0.41	0.67	0.17	0.30	0.52
		(11.46)	(0.49)	(0.47)	(0.38)	(0.46)	(0.50)
Cetuximab (Second Line)	53,859	65.92	0.46	0.54	0.08	0.15	0.77
		(7.12)	(0.52)	(0.52)	(0.28)	(0.38)	(0.44)
Cetuximab + irinotecan (Second Line)	73,519	66.38	0.44	0.13	0.05	0.18	0.77
		(10.99)	(0.50)	(0.34)	(0.22)	(0.39)	(0.42)
Bevacizumab + oxaliplatin + 5-FU/LV	76,636	63.05	0.45	0.73	0.07	0.22	0.71
		(12.33)	(0.50)	(0.44)	(0.25)	(0.41)	(0.45)
Bevacizumab + oxaliplatin + capecitabine	57,541	67.50	0.36	0.71	0.07	0.14	0.79
		(15.68)	(0.50)	(0.47)	(0.27)	(0.36)	(0.43)
Bevacizumab + irinotecan + 5-FU/LV	46,991	62.92	0.49	0.50	0.04	0.17	0.79
		(12.33)	(0.50)	(0.50)	(0.19)	(0.37)	(0.41)
Obs: 4,155							

Notes: Standard Deviations in parentheses. Summary statistics by regimen using individual choice data from IntrinsiQ. All data are summaries from 2003-2005.

Since we only observe individual choices between 2003 and 2005, we estimate our heterogeneous demand parameters from Equation 2 and Equation 3 using the "micro-BLP" procedure outlined in section 4 for only these years. We next assume that the distribution of patient characteristics remains fixed for the preceding years, and use our aggregate market shares from 1993-2003, as well as our estimated heterogeneous parameters, to iterate for δ_{jt} as described in Equation 10.

4 Model

4.1 Physician Choice of Drug Regimen

We model a physician's utility when choosing a regimen to treat a patient as a random utility model, in which physicians act as partial agents, as discussed in section 2. In making these decisions, physicians take into account the efficacy and side effects that a regimen offers to their patients. They also take into account the regimen price they face. Our model allows for heterogeneity in physician's preferences, which vary both in terms of unobserved physician heterogeneity and patient observables. Our exposition is centered on the most general version we can estimate that exploits both observed and unobserved patient and physician characteristics in a random coefficients framework with micro data (Berry et al., 2004). However, we also estimate and present results of alternate model specifications for comparison. These include a logit model with no individual heterogeneity (Berry, 1994), a random coefficients model with aggregate data (hereafter referred to as "BLP") (Berry et al., 1995), and a pure characteristics model that omits the well-known idiosyncratic error term present in most discrete choice models (Berry and Pakes, 2007). Our intent with these special cases is to demonstrate how our estimated price indices and welfare calculations change with varying model assumptions.

Equation 1 shows attending physician's a utility when treating patient i with regimen j at time t. The price paid by the physician is denoted by p_{jt}, and we do not allow for the possibility of price discrimination or rebates from pharmaceutical companies.[29] The characteristics of a regimen in terms of both efficacy and side effects are denoted by x_{jt}. We allow for a static unobserved quality component ξ_j, and time-variant deviations from its mean $\Delta\xi_{jt}$. The model allows for random coefficients on price and other characteristics, and their distributions depend on both observed and unobserved characteristics of patients and physicians. We also include a taste shock at the product level, ε_{aijt}, which we assume is distributed Type I Extreme Value.

$$u_{aijt} = -\alpha_{ai} p_{jt} + \beta_{ai} x_{jt} + \xi_j + \Delta\xi_{jt} + \varepsilon_{aijt} \tag{1}$$

where:

$$\alpha_{ai} = \alpha + \theta_1^{\nu} \nu_a + \theta_1^{d} d_i \tag{2}$$

and

$$\beta_{ai} = \beta + \theta_2^{\nu} \nu_a + \theta_2^{d} d_i \tag{3}$$

[29] We discuss in section 3 that these rebates were either small or nonexistent in this market. We therefore do not subscript the price variable by a.

Equation 2 explicitly shows how preferences for a regimen might depend on characteristics of the physician that are unobserved to the econometrician ν_a, and characteristics that are observed in our patient-level data d_i. Here, ν_a are a series of unobserved *physician* preference shocks drawn from a normal distribution, and d_i are a series of observed characteristics of patients of type i. θ_1^ν and θ_1^d represent parameters to be estimated on unobserved (physician) and observed (patient) price heterogeneity, respectively. θ_2^ν and θ_2^d are parameters for unobserved and observed regimen characteristic heterogeneity. This flexible model allows for the possibility of imperfect agency by physicians, who might prefer higher priced regimens that are more profitable. Observing the choices of physicians for several patients over time, and controlling for observable differences across patients, we are able to identify the random coefficient on price and characterize the distribution of physicians' preferences over price.

Our outside option ($j = 0$) includes off-label colorectal cancer drug treatments, regimens with very small market shares, and regimens with missing efficacy or side effects attributes. For estimation purposes, we assume normalize the outside option to 0, i.e.

$$u_{ai0t} = 0$$

Following Berry (1994); Berry et al. (1995) and the substantial subsequent literature, we can group in the mean utility δ_{jt} all the terms that do not depend on individual heterogeneity and rewrite equation Equation 1 as:

$$u_{aijt} = \delta_{jt} - (\theta_1^\nu \nu_a + \theta_1^d d_i)p_{jt} + (\theta_2^\nu \nu_a + \theta_2^d d_i)x_{jt} + \varepsilon_{aijt} \tag{4}$$

where

$$\delta_{jt} = -\alpha p_{jt} + \beta x_{jt} + \xi_j + \Delta\xi_{jt} \tag{5}$$

Integrating over the distribution of the observed patient characteristics d_i, physician unobservables ν_a and the taste shock ε_{aijt}, we obtain the market shares:

$$s_{jt}(\theta) = \sum_{a,i} s_{aijt}(\theta) = \sum_{a,i} \frac{exp\left(\delta_{jt} - (\theta_1^v v_a + \theta_1^d d_i)p_{jt} + (\theta_2^v v_a + \theta_2^d d_i)x_{jt}\right)}{1 + \sum_k exp\left(\delta_{kt} - (\theta_1^v v_a + \theta_1^d d_i)p_{kt} + (\theta_2^v v_a + \theta_2^d d_i)x_{kt}\right)} \tag{6}$$

4.2 Estimation and Identification

We estimate the model by "concentrating out" the objective function; searching over the parameters that account for heterogeneity for a given mean utility δ_{jt}, and then finding the means for the attributes with a linear projection, similar to Berry et al. (1995) and Nevo (2000). We construct two sets of micro-moments following Berry et al. (2004). The first set of moments is constructed by matching the observed and predicted covariances between regimen characteristics and physician/patient attributes. Specifically, for characteristic x^k interacted with patient characteristic d^r, the moment is:

$$G_{k,r}^1(\theta) = \frac{1}{N} \sum_{a,i,j,t} x_{jt}^k d_{ait}^r \left(\mathbb{1}\{y_{ait} = j\} - s_{aijt}(\theta)\right) \tag{7}$$

where N is the total number of physician-patient combinations in the data and $\mathbb{1}\{y_{ait} = j\}$ is an indicator for whether individual i seeing physician a at time t choose regimen j.

The second set of moments matches observed and predicted covariances of the regimen characteristics, x_j^k chosen for patient i by attending physician a, and the regimen characteristics of all *other* regimens chosen by attending physician a for patients other than i throughout the sample. Specifically, the moment is:

$$G_k^2(\theta) = \frac{1}{N} \sum_{a,i,j,t} x_{aijt}^k \frac{1}{AI - 1} \sum_{-i,t} x_{a,-i,t}^k \left(\mathbb{1}\{y_{ait} = j\} - s_{aijt}(\theta) \right) \tag{8}$$

where x_{aijt}^k represents characteristic k chosen by physician a for patient i in time t, $x_{a,-i,t}^k$ is the characteristic k of regimen j chosen by physician a at time t for any patient other than i, and AI is the total number of patients treated by physician a throughout the sample. In other words, this moment takes as given the observed *average characteristic* of regimens chosen by physician a for patients other than i and matches the observed and predicted covariance between this statistic and the characteristic chosen by physician a for patient i at time t.

We stack these moments together and search for the values of θ that minimizes the sum of square moments.[30]

Equation 7 helps to identify all of the observed heterogeneity parameters (θ_2), while Equation 8 pins down the *unobserved* heterogeneity parameters (θ_1). In practice, we estimate two unobserved heterogeneity parameters: price and second-line regimens.

$$G(\theta)' G(\theta) \tag{9}$$

To identify the random coefficients in Equation 2 and Equation 3, we take advantage of the fact that we observe many physicians making treatment decisions for many patients as the choice set changes. Suppose, for example, that within a particular physician's patient panel, controlling for patient characteristics and choice set availability, he tends to prescribe low-priced regimens only. Another physician, on the other hand, tends to prescribe high-priced regimens only, controlling for patient characteristics. This variation identifies the unobserved price taste parameter for each physician. If a new regimen enters the market that is similar to an existing regimen, and we observe a physician substitute from the old to the new regimen, the covariance between the regimens' attributes helps to further pin down the random coefficient.

To estimate the parameters of attributes in the mean utility function in Equation 5, in theory we could also have constructed an additional series of moments to match the predicted aggregate market shares from Equation 6 to the actual shares in the individual data. This would produce an additional JxT moment equations, with only $J + K + 1$ parameters to estimate (the ξ_j for each regimen j, α, and the β parameters for each characteristic k). However, as with Berry et al. (2004), this would have required a simultaneous search over not only the θ "heterogeneous" parameters, but also the $J + K + 1$ "mean" parameters as well. To expedite the search process, we instead implement a two-step procedure. First, we employ the contraction mapping from Berry et al. (1995)

[30] Because this system is exactly identified (one moment inequality for each parameter to be estimated), it is invariant to the choice of weighting matrix. We therefore choose the identity matrix.

to recover δ_{jt}:

$$\delta_{jt}^{h+1}(\alpha, \beta, \theta_1) = \delta_{jt}^{h}(\alpha, \beta, \theta) + (ln(s_{jt}) - ln(S_{jt}(\delta_{jt}^{h}, \theta))) \tag{10}$$

Second, following Berry (1994), we estimate Equation 5 using two-stage least squares to recover the mean regimen parameters. Because the unobserved drug regimen characteristics are likely to be correlated with price, estimating Equation 5 requires instrumental variables. We obtain our instruments by using the supply-side market equilibrium conditions. As price is a function of marginal costs and markups, any exogenous variable that shifts marginal costs or markups should be a valid instrument. Bresnahan et al. (1997) use as instruments the plausibly exogenous number of products in the market in period t and the sum of the observed characteristics of the competitors, which measure how crowded and competitive the product space is. These sets of instruments should affect markups via changes in the competitive environment, and therefore should be correlated with price but uncorrelated with a regimen's unobserved characteristics.

However, product attributes do not vary much over time in our sample due to infrequent product entry and exit. Therefore, our first-stage regressions using these "BLP-style" instruments are generally weak, and the estimation results do not differ measurably from standard OLS models. Therefore, we follow Nevo (2000) and Nevo (2001) by using "Hausman-style" instruments, which use the prices in other markets as instruments given their correlation with marginal cost. We note here that other markets can be defined geographically or over time. For instance, Nevo (2001) uses all 20 quarters in his data to construct his instruments. We go a step further and, inspired by Berry et al. (1995), construct two instruments with the lagged prices of other regimens. For the price of regimen j in period t, one instrument is the average price in period $t-1$ of all other regimens other than j. A second instrument is the average price in period $t-1$ of regimens produced by firms whose drugs are not used in regimen j. These types of instruments have been considered valid by Gandhi and Nevo (2021). The key identifying assumption is that these instruments are uncorrelated with the current-period demand shock but are correlated with price of regimen j. For this to hold, we require that the demand shock for regimen j in period t is uncorrelated with the demand shock for regimen k in period $t-1$, a condition likely to hold true unless there exists a time-persistent market-level demand shock.

One potential concern with such instruments is the possibility of a persistent informational shock that changes physicians' prescribing behavior, e.g. learning that a particular regimen has some unobserved side effects or lower efficacy than reported. These types of shocks, however, would likely results in significant changes to the perception of the quality of the twelve regimens in our sample relative to the outside option (i.e. the unobserved mean utility, ξ_t.) Indeed, as we show in Appendix A, there are considerable fluctuations in the size of the outside option over time. As such, these types of shocks would be sufficiently captured by our inclusion of time dummies in Equation 5. We also construct instruments using t-2 and t-3 lags in case there is still some serial correlation left; our results are robust to the use of these alternative instruments.[31]

[31] We thank a referee for suggesting these instruments.

4.3 Constructing the Price Indices

We build a series of price indices for different demand specifications. Following our model description, we focus the exposition in this section on the most general model that has observed and unobserved heterogeneity, and summarize the other models as special cases. The price index is built based on changes in the compensating variation derived from the above model. The compensating variation provides a measure of how much income could be taken away from (or given to) a physician to leave him indifferent between facing the old choice set and the new improved (inferior) choice set. Given the Type I Extreme value assumption on the taste shock, and the presence of observed and unobserved heterogeneity, the compensating variation is calculated as:

$$CV_{ait} = \frac{u_{ait} - u_{ait-1}}{\alpha_{ai}}$$

where u_{ait} is the unconditional indirect utility $u_{ait} = \max_j u_{aijt}$ and α_{ai} is the marginal utility of income of physician a when treating patient i. Small and Rosen (1981) show that u_{ait} can be computed as:

$$u_t = ln \sum_j exp(\delta_{jt} - (\theta_1^{\nu}\nu_a + \theta_1^d d_i)p_{jt} + (\theta_2^{\nu}\nu_a + \theta_2^d d_i)x_{jt})$$

Trajtenberg (1990) shows that if the price change takes the form of a shift by a factor of $(1-\mu_t)$ in the distribution of prices but the variance remains the same, then the price index can be obtained as:

$$PI_t = (1 - \mu_t)PI_{t-1}$$

where:

$$\mu_t = \frac{\bar{CV}_t}{\bar{CV}_t + \bar{p}_t}$$

Here $\bar{p}_t$ is the average price in period t and:

$$\bar{CV}_t = \frac{1}{ns}\sum_{a,i}^{ns} CV_{ait}$$

and ns refers to the number of patientXprovider observations. Similar logic can be used to construct price indices for distinct groups of physician and patient populations, by constructing the compensating variation just for those groups (e.g. $\bar{CV}_{ait}$).

We construct price indices for several special cases of this model. If all heterogeneity is captured by the regimen-level taste shock, this results in a logit model with no random coefficients. Introducing random coefficients with aggregate data permits only the unobserved heterogeneity and presumes patient observables do not vary across physician. For each of these we construct price indices, keeping the objects that enter the the compensating variation consistent with the underlying demand specification.

4.3.1 Assumptions on the Unobserved Components of Demand

As noted by (Nevo, 2003), the welfare implications of new product introductions are sensitive to how one interprets the estimated coefficients of the time variables and unobserved product characteristics. Therefore, we plot a separate index associated with two different scenarios. In the first scenario, we assume that the period-specific unobserved quality of the inside goods (i.e., ξ_t) are fixed, but that the value of the outside option (i.e. u_{0t}) changes over time. We refer to this scenario as "u_{0t} changes." In the second scenario, we assume that the value of the outside option does not change over time and that, therefore, the average period-specific unobserved value of the inside goods does change over time ("u_{0t} fixed"). Under this scenario, the estimated coefficients on the time indicators measure changes in the mean quality of the inside goods over time, whereas in the former scenario they measure changes in the value of the outside option.

The interpretation of these time dummies significantly affects the indices. Suppose, for instance, that the market share of the outside option increases substantially between period $t-1$ and period t. It is impossible to identify whether, net of observable differences in observed product characteristics between period $t-1$ and period t, this decrease in demand reflects decreases in unobserved quality of the inside goods in period t (the second scenario, "u_{0t} fixed"), or whether it reflects an increase in the unobserved quality of the outside option in period t (the first scenario, "u_{0t} changes"). If it is the former, this represents a welfare loss for consumers. If it is the latter, this represents a welfare gain.

We believe the data in our application are most consistent with the first scenario described above. In particular, between 2000 and 2005, the market share of the outside option increased significantly by about 16 percentage points. If this increase were driven by a deterioration in quality of "inside goods" then we would expect to see physicians treating patients largely with component drugs whose active ingredients are not approved by the FDA for colorectal cancer in any regimen, but rather were approved for different types of cancer altogether. However, the converse is true: a large share of the outside option consists of *component drugs* whose active ingredients are approved by the FDA for colorectal cancer (similar to the "inside goods"), but *regimen cocktails* that were not approved (see Appendix A for more details). As such, in our preferred specification, we feature price indices where the period-specific mean utility of the inside goods are assumed to be fixed, while the value of the outside option is assumed to change over time. In Appendix A, we also present price indices that assume the opposite situation ("u_{0t} fixed"). We do this for two reasons. First, the actual situation is likely to be some combination of the two scenarios above, even though we believe it is much closer to "u_{0t} changes." Second, in other applications the data might be more consistent with "u_{0t} changes." Therefore, our approach allows us to depict how different the colorectal cancer quality-adjusted price indices are when one adopts polar-opposite assumptions about changes in unobserved quality.

We also plot indices for the above two scenarios under two different assumptions about the unobserved time-and-product-specific deviation in quality from the mean (i.e. $\Delta\xi_{jt}$). In estimating the demand models, $\Delta\xi_{jt}$ are the equivalent of residuals that can be backed out given the estimated parameters. There is, therefore, one residual for every product-time combination throughout the sample period. Practically, these residuals capture two things: a change in unobserved quality of

products between periods, or a change in consumer tastes for products between periods. However, whether we view changes in $\Delta\xi_{jt}$ as changes in quality or tastes affects the implied price indices.

Suppose, for instance, that we observe that $\Delta\xi_{jt} < \Delta\xi_{jt-1}$, i.e., the unobserved deviation of product j from the time-specific mean went down between periods $t-1$ and t. One interpretation is that this reflects a decrease in quality of product j. Another interpretation is that, for unobserved reasons, consumers' tastes changed such that they needed fewer purchases of product j to achieve the same utility. In the former case, this reflects a loss in welfare and a rise in the price index for the product; the product became less effective at the same price, and therefore the consumer purchases less of it. In such a case, we would want $\Delta\xi_{jt}$ to vary over time to reflect this changing quality. In the latter case, the price index need not increase, and we would therefore want to fix $\Delta\xi_{jt}$. We therefore report both sets of indices–one where we assume "$\Delta\xi_{jt}$ changes" and one where we fix it to product j's initial value ("$\Delta\xi_{jt}$ fixed").[32]

One final concern with the model outlined above is that welfare gains generated from product introductions may be upward biased due to the presence of the idiosyncratic error term (Berry and Pakes, 2007). As a result, price indices may be artificially low following a product entering the market, particularly if that product has close substitutes with existing regimens. This is a particularly salient issue in our context as we are explicitly concern with the value of permitting such new products. To address these concerns, we present a special case of our demand model: a pure characteristics model, which assumes all the heterogeneity derives from different values for characteristics, but that no welfare is generated from a purely "product-level" shock. In this case the estimation and computation is substantially different from the other cases that we describe above. We outline the estimation of the pure characteristics model below.

4.4 Pure Characteristics Model

While the random-coefficients model mitigates some of the substitution problems inherent in a logit model, it still relies on the assumption of the Type I Extreme Value idiosyncratic error term, ε_{ijt}. The presence of this term allows the model to be solved and the mean utility levels, δ_{jt}, to be derived as a function of observed market shares. However, the error term is problematic when calculating welfare because it guarantees that regardless of product characteristics, there are certain consumers who would purchase *any* product (i.e., market shares cannot be zero for any product regardless of its price and attributes). This implies that product entry necessarily increases the utility of some consumer, and therefore consumer utility increases as more products are added to the product space.

Berry and Pakes (2007) show how this shortcoming can lead to biased welfare estimates. They propose eliminating the idiosyncratic error term from the model altogether, which alters the model so that it becomes purely a model of "tastes for characteristics." The randomness of choices comes from different consumer sensitivity to price and heterogeneous product attribute preferences, as opposed to heterogeneity in the "tastes for products" models described above. The benefit of estimating this model is that it generates more plausible substitution patterns by eliminating the Type I error assumption.

[32] See (Nevo, 2003) for a much more detailed explanation on this distinction, as well as for an additional motivation for treating $\Delta\xi_{jt}$ as changing over time.

We define physician utility in the pure characteristics model to be

$$u_{ijt} = \delta_{jt} - \alpha_i p_{jt} + \beta_i x_{jt} \tag{11}$$

where δ_{jt} is defined as in Equation 5; the combination of all the attributes for which the parameters are non-random (i.e., physicians agree on the value of these attributes). Note that all the heterogeneity now comes from the α_i term, the price sensitivity of physician i, and β_i, the physician taste for attributes. For tractability, we allow physicians to differ in their sensitivity to price and to have heterogeneous preferences for just one non-price attribute: whether a regimen has been formally approved by the FDA as a second-line treatment.

Following Berry and Pakes (2007) and Song (2008), we derive the market share equation first by ordering the products such that product 1 has the lowest price and product J has the highest price. That is:

$$p_1 < p_2 < ... < p_J$$

Given this price ordering, consumer i will buy product j if and only if:

$$\begin{aligned} u_{ijt} &> u_{ikt} \\ \delta_{ijt} - \alpha_i p_{jt} &> \delta_{ikt} - \alpha_i p_{kt} \end{aligned}$$

where $\delta_{ijt} = \bar{\delta_{jt}} + \beta_i x_{jt} + \xi_j + \Delta\xi_j$. Thus we have the cutoff points:

$$\alpha_i < \frac{\delta_{ijt} - \delta_{ikt}}{p_{jt} - p_{kt}} \quad if \quad p_{jt} > p_{kt}$$

$$\alpha_i > \frac{\delta_{ikt} - \delta_{ij}}{p_{kt} - p_{jt}} \quad if \quad p_{kt} > p_{jt}$$

Therefore consumer i will buy good j if:

$$\alpha_i < \min_{k<j} \frac{\delta_{ijt} - \delta_{ikt}}{p_{jt} - p_{kt}} = \bar{\Delta}_{ijt}$$

and

$$\alpha_i > \max_{k>j} \frac{\delta_{ikt} - \delta_{ijt}}{p_{kt} - p_{jt}} = \underline{\Delta}_{ijt}$$

and so the share for good j becomes:

$$s_j = \int F(\bar{\Delta}_{ijt}|\beta_i) - F(\underline{\Delta}_{ijt}|\beta_i) 1[\bar{\Delta}_{ijt} > \underline{\Delta}_{ijt}] dG(\beta_i)$$

where $F(.|\beta_i)$ is the cumulative distribution function (CDF) of α_i conditional on β_i, and $dG(\beta_i)$ is the CDF of β_i. For our model, we assume a lognormal distribution for α_i, so that $log(\alpha_i) \sim N(0, \theta_1)$ and a normal distribution for β_i, so that $\beta_i \sim N(\beta_2, \theta_2)$.

The parameters of the model to be estimated are $\theta = (\beta, \beta_2, \theta_1, \theta_2)$. Note that if there were only one random coefficient, for instance on price, then this model would collapse to a vertical model

(see Bresnahan et al. (1997) and Lucarelli and Nicholson (2009)). In a vertical model, the δ_{jt} can be derived analytically, and thus estimating the model is relatively straightforward. However, with the presence of two random coefficients, one can no longer back out the mean utilities.

In addition, this model imposes a computational burden; dropping the idiosyncratic error term implies that a contraction mapping that can be used to estimate the full random-coefficients model may not exist, as demonstrated by Berry et al. (1995). We therefore follow Berry and Pakes (2007) and Song (2008) and use a three-step procedure to estimate the model. We first re-estimate the full random-coefficients model and rescale the logit error term by a factor that drives it toward zero. This brings the mean utility closer to its true value in the absence of an error term. We then use a combination of a fixed-point homotopy method and a Newton-Raphson search. We elaborate on details of the search procedure in Appendix B.

Constructing the price indices is similar to the methods described above. The only difference is that without the idiosyncratic logit error, u_{it} is no longer expressed as the traditional "log-sum" formula, but instead: $u_{it} = \max_j u_{ijt}$

5 Results

5.1 Parameter Estimates

We first report the parameter estimates from each of the models using aggregate data only. In the first column of Table 4 we report estimated coefficients from the OLS logit model without instrumenting for price; in the second column we report results of the IV logit that addresses the endogeneity of prices; and in the third column we report coefficients from the random coefficients model (BLP) that both instruments for price and incorporates physician heterogeneity in drug preferences. The final column reports the parameter results from the pure characteristics model that incorporates physician heterogeneity for price as well as second-line regimens, while omitting from the model the idiosyncratic error term, ε. Because many of the attributes are highly correlated, we include in all models a reduced set of attributes that proxy for relevant quality indicators of the regimen. To proxy for efficacy, we include the median survival months from the randomized clinical trial and to proxy for potential side effects, we include the share of trial patients experiencing vomiting symptoms. We also include an indicator for whether the regimen was a tablet–a measure of convenience–an indicator for whether the regimen was approved by the FDA as a second-line regimen, and an indicator for whether any drug in the regimen was produced by the company Pfizer. The latter is meant to capture brand preferences for one of the major innovators of such second-line therapies. Specifically, it allows us to rationalize the fact that certain Pfizer regimens (e.g., Irinotecan) manage to gain significant market share despite having fairly similar observed attributes as other regimens in the data.[33] In all of the models, we rescale prices to log points rather than levels.[34]

[33] For instance, the regimen Irinotecan + 5-FU-LV maintains more than a 20% market share, compared with only an approximate 6% market share for Capecitabine, despite the fact that they have relatively similar survival, side effects, and the latter is considerably less costly. See Figure 2.

[34] Average regimen prices rise very substantially from about $127 in 1993 to $36,000 in 2005. As such, estimating the models with log price rather than the price level increases the amount of price variation early in the sample period for purposes of identifying the coefficient on price, and prevents the index from swinging wildly. For this reason, we

Table 4: Parameter Estimates from Quality-Adjusted Models with Aggregate Data

	(1) OLS	(2) IV Logit	(3) BLP	(4) Pure Char.
(Log) Price	-0.278***	-1.280***	-1.372***	
	(0.0559)	(0.304)	(0.341)	
Survival Months	0.0127	0.570***	0.599***	0.506*
	(0.0405)	(0.175)	(0.179)	(0.262)
Vomiting	-0.0614**	-0.243***	-0.261***	-0.185***
	(0.0249)	(0.0683)	(0.0861)	(0.0512)
Tablet	-1.209***	0.798	0.946	1.919***
	(0.196)	(0.672)	(0.719)	(0.477)
Second Line	-0.630	5.324***	3.648	4.283
	(0.442)	(1.877)	(3.551)	(3.651)
Pfizer	-0.288*	1.600**	1.699***	1.498***
	(0.166)	(0.616)	(0.621)	(0.556)
Constant	2.165***	1.824**	2.792	2.956
	(0.449)	(0.799)	(2.485)	(4.654)
θ_1			-0.180	0.348**
			(0.233)	(0.172)
θ_2			2.458	1.008
			(3.399)	(1.008)
Observations	208	208	208	208

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Parameter estimates from various demand models. Column 1 reports results from an OLS demand model. Column 2 reports results from the logit model with Hausman instrumental variables. Column 3 reports results from the full (BLP) model. Column 4 reports results from the pure characteristics model. θ_1 refers to the random coefficient on price, θ_2 refers to the random coefficient on second-line regimens.. In the BLP model, $\alpha_i = N(\alpha, \theta_1)$. In the pure characteristics model, $log(\alpha_i) = N(0, \theta_1)$.

The logit and BLP results align closely with Song et al. (2017). Comparing the (log) price coefficients between column 1 and columns 2 and 3 confirms that there is a positive correlation between a drug's price and demand shocks, and that the instrumental variables mitigate the problem of price endogeneity. The price coefficient is -0.278 in the OLS logit, versus -1.280 and -1.372 in the in the IV logit and BLP models, respectively. The price coefficient is significantly different from zero at the one-percent level in all three models. In Table D.1, we present robustness checks on our IV estimates where we lag our Hausman instruments by two and three periods instead of the one-period lag in our main specification. Encouragingly, the parameter estimates on price as well as the efficacy attributes are all quite similar.

The sign on the median months that patients survived in the treatment's randomized controlled trial is positive across all the models with exception of the OLS model. For the IV logit model, our estimates imply that physicians, on average, are willing to pay an additional 44 percent for an additional month of patient survival (where the mean survival months among these regimens is 13.2). Evaluated at the mean regimen price in our sample ($21,113), this implies physicians are willing to pay $9,400 per additional month. When accounting for expected side effects associated with treatment, the average expected *quality-adjusted* life years (QALYs) across the regimens (based on market share weights) is 53 percent of the expected (unadjusted) life years.[35] This implies that physicians value a year of perfect health at $211,000, which is higher than the most common estimates for the United States ($50,000 - $150,000) (Neumann et al., 2014), but lower than the preferred estimate of $373,000 used by Murphy and Topel (2006). Meanwhile, the coefficient on the percentage of patients in the randomized controlled trial who experienced a severe vomiting reaction to the treatment is negative across all four models, suggesting that physicians value drug regimen efficacy and dislike side effects. Specifically, the IV logit estimates suggest that physicians are willing to pay an additional 19 percent to reduce the expected share of patients experiencing vomiting by one percentage point (or an additional $4,000 evaluated at the mean price).

The coefficient on "tablet" is positive, but not significant in the IV logit model. The magnitude of the estimate suggests that consumers are, on average, willing to pay about 62 percent more for a regimen that has the convenience of a tablet component. This translates to approximately $13,000 for this convenience at the mean regimen price. Meanwhile, the coefficient on the second-line indicator is positive and significant. This positive coefficient captures the fact that second-line regimens have low efficacy measures (e.g., survival months) because they are tested on the sickest patients and are often quite expensive. Nevertheless, second-line regimens are often prescribed to patients who failed on a first-line regimen (i.e., the cancer progressed) and are seeking a viable clinical alternative or complement. Our estimates suggest physicians are, on average, willing to pay about 400 percent more for second-line regimens. As such, this coefficient plays a large role in the implied welfare gains from the launch of second-line treatments, as discussed in subsection 5.2.

also plot all of the indices below in logs.

[35] We estimated each regimen's expected QALY by calculating the average length of time clinical trial patients spent in five health states, and applying previously-published adjustments for the quality of life associated with each of these states (Brown and Hutton, 1998; Brown et al., 2001; Lloyd et al., 2006): stable metastatic disease without experiencing a side effect from chemotherapy; tumor is responding to treatment (partial or complete response rate) with no side effect; responding to treatment with a side effect; not responding to treatment with a side effect; and the disease has progressed to a more advanced state. For side effects, we used the probability a colorectal clinical trial patient experienced any grade 3 or grade 4 side effect.

Finally, the coefficient on the Pfizer regimen indicator is positive in the final three models. In Appendix Table D.2, we present alternate estimates where we remove the Pfizer indicator and replace it with additional efficacy measures. The results are similar to our main specification.

Turning to the random coefficients, the BLP and pure-characteristics models incorporate random coefficients on a regimen's price and whether it is FDA approved as a second-line regimen. Neither of the estimated standard deviations are significantly different from zero in the BLP model. The coefficient on the standard deviation of price is positive and significant in the pure-characteristics model. This indicates that colorectal cancer treatments that are similarly priced are perceived by physicians to be closer substitutes than treatments that are differently priced.

The distributions of the estimated α and β coefficients for the demand models are displayed in Figure 4. Panel (a) depicts α_i for the BLP and pure characteristics models.[36] Although the estimated mean price sensitivity is lower in the pure characteristics model (approximately 1.0 versus the estimated 1.5 in the BLP model), it is clear from the figure that consumers have more heterogeneous price preferences in the pure characteristics model, as the distribution of α_i is considerably more disperse about the mean. The implication is that some of the heterogeneity that the logit and BLP models previously placed on the idiosyncratic error term is now being loaded onto α_i. Looking at one standard deviation below and above the estimated mean price sensitivity coefficient, our pure characteristics model implies that physicians are willing to pay between 37 percent less ($8,000) and 77 percent more ($16,385) than the mean for an additional month of survival, suggesting considerable heterogeneity. Conversely, moving to Panel (b), the β_i coefficient (valuation of second-line treatments) exhibits more variation among physicians in the BLP relative to the pure characteristics model. Nevertheless, although the coefficient is estimated imprecisely, our pure characteristics model implies physicians are willing to pay between an additional 240 percent and 800 percent for second-line regimens when evaluated at plus or minus one standard deviation of the mean. Overall, these distributions indicate that allowing for unobserved random preferences is an important driver of demand: in particular, models with more flexibility (i.e., pure characteristics) characterize physicians as having wider variation in price sensitivity, but tighter deviations in their predisposition towards recommending second-line treatments.

We now turn to parameter estimates of the demand model with both observed and unobserved heterogeneity across physicians *and* patients from the individual choice data. The results are displayed in Table 5. The coefficient estimates for mean regimen attributes are largely consistent with those from the estimation of the macro moments from Table 4. Physicians dislike higher-priced regimens and those with side effects, but value regimens with higher efficacy. The coefficient on the second-line indicator is positive suggesting that, controlling for other regimen characteristics, physicians prefer second-line treatments. Recall that second-line regimens have relatively short survival months because they are tested on patients with more advanced cancer.

The estimated coefficients on the interactions between drug regimen characteristics and observed patient attributes, however, highlight the importance of allowing for heterogeneity when analyzing the welfare impact of innovation. Specifically, we allow for interactions between regimen price and

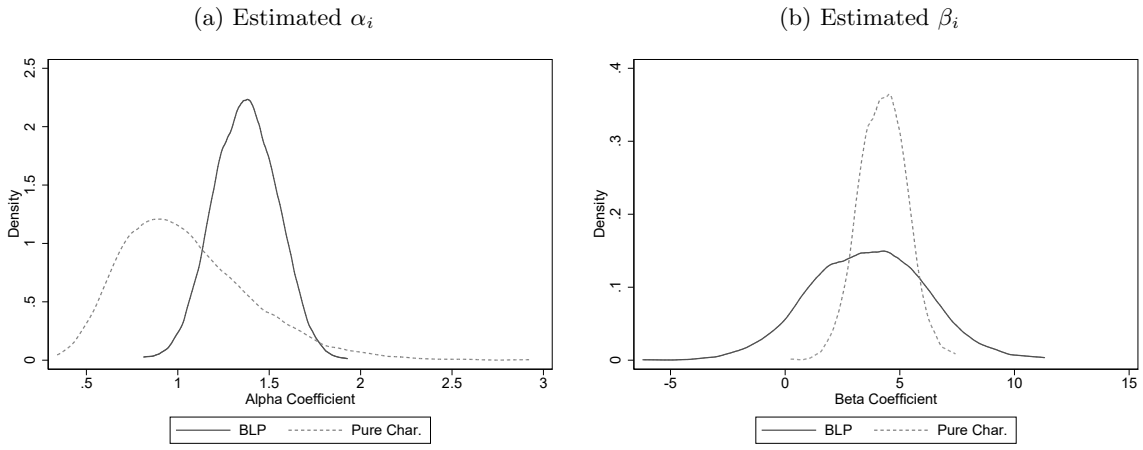
[36] Note that α_i here is depicted as being positive to be consistent with the specification of our utility model in Equation 1, in which the price variable, p_{jt}, is preceded with a negative sign. In this depiction, therefore, a positive α_i implies that consumers are sensitive to price, as expected.

Table 5: Parameter Estimates Allowing for Patient Heterogeneity

	(1)
Log Price	-2.000***
	(0.497)
Survival Months	0.979***
	(0.286)
Vomiting	-0.634***
	(0.112)
Tablet	2.481**
	(1.100)
Second-Line Indicator	8.657***
	(3.070)
Pfizer	2.946***
	(1.007)
Constant	6.238***
	(1.307)
AgexPrice	-0.757***
	(0.045)
AgexVomiting	0.538***
	(0.026)
AgexSecondLine	0.0642
	(0.484)
Round2xPrice	0.145***
	(0.004)
Round2xVomiting	-0.122***
	(0.001)
Round2xSecondLine	2.731***
	(0.246)
StageIIIxPrice	0.301***
	(0.030)
StageIIIxVomiting	-0.100***
	(0.008)
StageIVxPrice	0.591***
	(0.062)
StageIVxVomiting	-0.133***
	(0.017)
θ_1^v	-0.502***
	(0.039)
θ_2^v	0.014
	(3.313)
Observations	208

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Parameter estimates from micro-BLP demand models with observed and unobserved heterogeneity. θ_1^v refers to the random unobserved coefficient on price. θ_2^v refers to the random coefficient on second-line treatments. "Round2" refers to whether a patient is undergoing their second-or-higher round of treatment. "Vomiting" refers to share of patients experiencing vomiting symptoms after taking the regimen.

Figure 4: Estimated Price Sensitivity and Second-Line Valuation Coefficients



Notes: This figure plots estimated distributions of price sensitivity (α_i, Panel (a)) and valuation for second-line treatments (β_i, Panel (b)) across sampled physicians for the BLP and pure characteristics models. α_i represents the mean estimated price sensitivity of physicians in the sample, along with the variance estimated from the random coefficients. Note that the positive sign of α_i is consistent with consumers being price sensitive, given our specification in Equation 1. β_i represents the mean estimated coefficient on the utility of second-line treatments, along with the variance estimated from the random coefficients.

side effects with a patient's age (divided by 100), whether the patient is being treated for Stage III cancer, and whether the patient is being treated for highly advanced (Stage IV) cancer. In addition, we allow for interactions between patients undergoing their second-or-higher round of therapy with an indicator for whether the regimen was approved by the FDA as a second-line treatment. The cancer patients with the shortest expected life expectancy will be those whose tumor is already advanced when diagnosed (i.e., Stage III or Stage IV), and patients where the disease continues to progress and advance even after receiving an initial drug treatment (i.e., second line). These patients might be more willing to pay more for expensive therapies that are effective, have limited side effects, and are specifically designed for their clinical situation (i.e., treatments approved by the FDA as second-line therapies for patients who have already received an initial drug treatment).

Indeed, based on the results from Table 5, physicians treating patients who are undergoing a second-round of treatment as well as those with advanced cancer (i.e., Stage III or Stage IV, where Stage II is omitted from the regression and Stage I patients were omitted from the sample altogether), place greater value on regimens with minimal side effects and are less price sensitive relative to physicians treating patients with better treatment prospects and longer expected survival. Our model implies that physicians treating patients for Stage IV cancer are willing to pay about 70 percent more for an additional month of survival. Evaluated at the mean regimen price, this implies that Stage IV patients are willing to pay about $177,000 for an additional year of survival, about an extra $51,000 relative to Stage II cancer patients. Stage IV cancer patients are willing to pay about 54 percent more for a one percentage point lower likelihood of vomiting, amounting to an additional $11,000. For Stage II cancer patients, this amounts to about $7,000.

Moreover, physicians treating patients with a second-round of treatment place significantly extra value on second-line regimens. These results provide the intuition for why the price index that allows for heterogeneity falls over time for patients with a Stage IV cancer receiving a second round of treatment when expensive second-line regiments arrive in the market, as we discuss below.

The coefficients on the age interactions in Table 5 indicate that physicians treating older patients have relatively strong preferences for lower-priced regimens and are somewhat less sensitive to side effects relative to physicians treating younger patients. The former results would be consistent with Medicare enrollees being required to pay a greater percentage of the drug price relative to younger patients covered by private health insurance. Finally, the unobserved coefficient on price, θ_1^v, is negative and insignificant unlike the results from the pure-characteristics model, but similar to the results of the macro-BLP model.

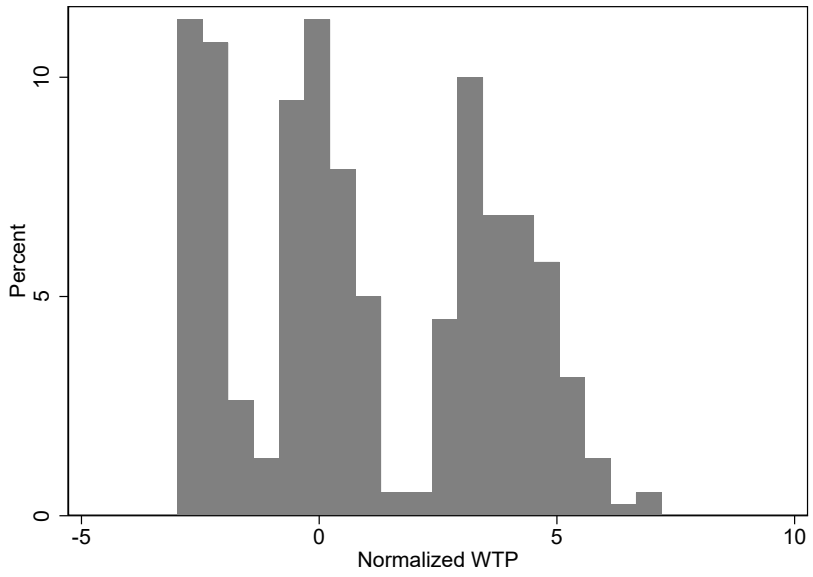
Broadly, these results indicate that there is significant heterogeneity in physicians' regimen preferences across *patients* in addition to idiosyncratic preferences of physicians towards certain regimens. To further highlight this, Figure 5 plots the distribution of estimated WTP for one of the notable entrants towards the end of the sample: bevacizumab, or Avastin. Specifically, the figure shows the variation in WTP for the Bevacizumab+Oxaliplatin+5-FU-LV regimen across patients in 2004, holding fixed the variation in estimated unobserved physician preference heterogeneity. The mean WTP across patients is normalized to 1 so that each point can be interpreted as how much a particular physician is WTP relative to the average patient characteristic. Indeed, the figure shows substantial heterogeneity in physician preferences across patient observables, certain patient types yielding a WTP of more than 7 times the average for Avastin relative to the outside option and other types only willing to pay 3 times *less* the average.

5.2 What are the Gains from New Drug Regimens?

We now use the estimated models to produce quality-adjusted price indices for colorectal cancer regimens between 1993 and 2005. The naive price index (in log points) is plotted in Figure 6. Note that prices are relatively stable in the pre-1998 period where 5-FU/LV dominated the market and all other products were included in the outside option. The share-weighted prices began to rise in 1998 after irinotecan was introduced, and then continued to increase until the 2005 period. This is further documented in Table 1, where it can be seen that the 2005 price of irinotecan was considerably higher than that of 5-FU/LV, and that this continued to be the case in 2001 when capacitabine was introduced, in 2002 when oxaliplatin was launched, and in 2004 when bevacizumab entered the market.

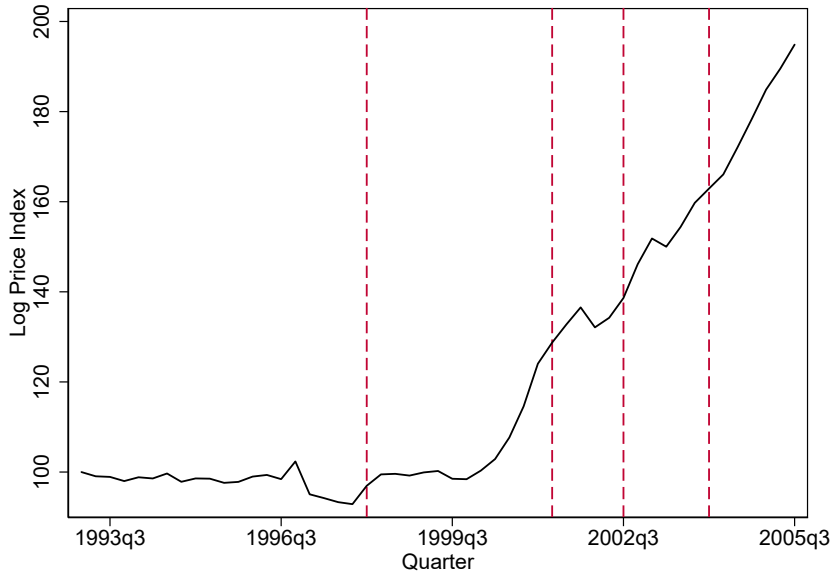
We plot the log price indices for the models using aggregate data (logit, BLP, and pure characteristics) in Figure 7. The figure presents our preferred index where the quality of the inside goods is assumed to be fixed at zero (i.e., the coefficients on the time indicators capture changes in the mean quality of the outside goods), and $\Delta\xi_{jt}$ is assumed to reflect changes in unobserved product quality over time. In Appendix A, we present indices separately for each of the four assumption combinations regarding the outside option, u_{0t}, and unobserved product quality, $\Delta\xi_{jt}$, as described in detail in subsubsection 4.3.1. We also present evidence justifying our preferred interpretation of the outside option in the appendix.

Figure 5: Distribution of WTP for Bevacizumab+Oxaliplatin+5-FU-LV in 2004



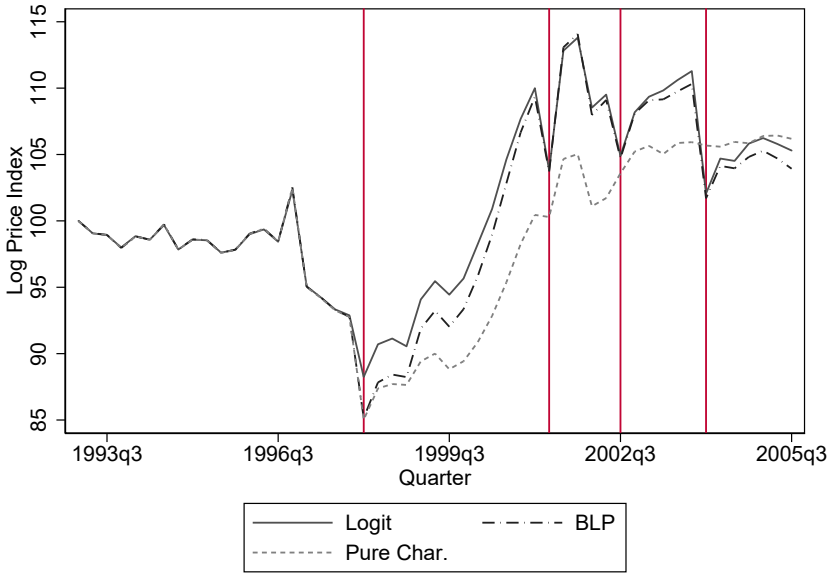
Notes: This figure plots the distribution of WTP for Bevacizumab+Oxaliplatin+5-FU-LV across patients in the sample in 2004, holding physician preferences for regimens fixed. Mean WTP for the regimen is normalized to 1. These values can, therefore, be interpreted as how much more certain patients are willing-to-pay for the drug relative to how much the average patient is willing-to-pay for the drug relative to the outside option.

Figure 6: Naive Price Index



Notes: This figure plots the price of regimens weighted by market share between 1993 and 2005. Prices are for a 24-week treatment cycle. Dashed red lines reflect the arrival of four new regimens to the market: irinotecan, capecitabine, oxaliplatin, and bevacizumab, in chronological order.

Figure 7: Price Indices with No Patient Heterogeneity



Notes: This figure plots the *aggregate* price indices from the estimated logit, BLP, and pure characteristics models. All indices are on a log scale (i.e., the dependent variable is the "log" price). This graph plots the index from our preferred specification, which assumes that the outside option (μ_{0t}) changes in quality each period and that the period-specific deviations from the mean in estimated regimen quality ($\Delta\xi_{jt}$) change each period. Dashed red lines reflect the arrival of four new regimens to the market: irinotecan, capecitabine, oxaliplatin, and bevacizumab, in chronological order.

The first notable result from these indices is that, regardless of modeling assumptions used, adjusting for quality through our discrete choice framework significantly reduced the extent to which the indices rose relative to the naive index. Whereas the naive index rose by about 94 log points between 1993 and 2005, the quality-adjusted indices each rose by six log points. This implies the survival and side effects gains brought on by these entrants were highly valued by physicians. This can further be seen through the notable drop in the index surrounding the introduction of each new regimen (depicted by the vertical red lines).

During the initial period before irinotecan was introduced, all three quality-adjusted price indices estimated from demand models in Panel (a) are relatively flat and similar to the naive index, as would be expected. They each fall slightly throughout this period, reflecting the fact that only one product was available before 1996 (5-FU + LV), and its product characteristics did not change other than experiencing minor price decreases (producing a small decrease in the index). When irinotecan was introduced in 1996 there was a noticeable reduction in all of the price indices. This reflects the fact that physicians valued the substantial improvement in expected longevity by more than the price increase. Although the price of irinotecan ($20,000 in 2005 dollars) was well above 5-FU + LV ($75), irinotecan improved survival from 12.5 months to 15.6 months. Beyond this, irinotecan was the first second-line regimen approved by the FDA (irinotecan with 5-FU + LV is a first-line regimen, whereas irinotecan by itself is a second-line regimen). The substantial valuation on second-line treatments in our demand models served to reinforce a declining index during this period.

All of the price indices rise substantially after the introduction of irinotecan until 2001 when the only products on the market continued to be 5-FU + LV and irinotecan. While their quality attributes were consistent, their prices rose substantially over this time period. Between 1996 and 2001, both irinotecan regimens experienced price increases of about 25 percent, translating into an increase of more than $3,600. As a result, the logit and BLP indices spiked by about 25 log points during this time period. The pure-characteristics index rose only about 15 log points, driven by the fact that the average mean price sensitivity in the pure characteristics model is estimated to be lower than that of the BLP and logit models (and, as a result, physicians are estimated to lose less utility from irinotecan price increases).

The index then falls sharply when capecitabine and oxaliplatin were introduced in 2001 and 2002, respectively. Oxaliplatin further increased the median survival to 19.4 months, but with a higher price ($25,000 for a six-month treatment). Capecitabine is an oral tablet that provides time savings by allowing patients to avoid time consuming infusions, but it is much more expensive than 5-FU-LV, which it is chemically equivalent to. In spite of these price increases, however, the value of the survival gains and tablet convenience led to a decrease in the logit and BLP indices by a little over five log points following the introduction of capecitabine and a little under 10 log points following the introduction of oxaliplatin. The pure characteristics index, however, remained relatively flat during this time period. This suggests that most of the welfare gains associated with the introduction of these new regimens were driven by the idiosyncratic error term. When the term is removed by the pure characteristics model, the welfare gains associated with higher survival are muted.

After this point, the indices stabilize or fall for the rest of the sample period when bevacizumab

and cetuximab launch in 2004. Part of the observable drop visible in 2004 for the logit and BLP indices are attributed to the survival gains exhibited by bevacizumab. However, recall that between 2000 and 2005 the market share of the outside option increased substantially, so the stability or drop in the indices is due in large part to our interpretation that the quality of the outside option was increasing over time (a point we return to below).

Overall, towards the end of the sample period all three indices are 21 percent to 35 percent higher than they were in 1993, implying a slight reduction in welfare, despite a substantial increase in price.[37] As a point of comparison, the CPI rose 35 percent during this time period. Although pharmaceutical firms offered superior products over time, they set prices that extracted most of the welfare gains, which should provide strong incentives to invest substantially in R&D.

The ordering of the three demand-based indices seem intuitive. The BLP and pure characteristics indices drop about five log points more than the logit index in 1997 when irinotecan is launched. The larger drop for the former two indices is driven by the fact that these models allow consumers to have heterogenous preferences for a treatment's price and whether it is formally approved as second-line treatment, whereas the logit model assumes homogenous preferences. Irinotecan was included in both first- and second-line regimens in 1997. Physicians who placed a high value on second-line regimens would therefore benefit from irinotecan's launch, but that would not be captured in the logit model. This is particularly true since prior to irinotecan, there were no FDA-approved second-line regimens. As a result, irinotecan had no close substitutes in the product space and, as such, produced substantial welfare gains. After 1998, the BLP and logit indices closely track one another.

The pure characteristics index does not rise as substantially as the other two indices after the introduction of irinotecan. As mentioned previously, this reflects the fact that between 1996 and 2001, no new products were introduced into the product space. In addition, among the existing products, the only changing attribute was a substantial increase in the price of irinotecan. Given that the pure characteristics model estimated a lower mean price elasticity among physicians, the price index did not rise during this time period as much as the other indices.

Conversely, the pure characteristics index also did not *fall* by as much as the logit and BLP indices upon introduction of new regimens. For instance, when bevacizumab was launched in 2004, both the BLP and logit indices fell by about 10 log points with the introduction of this flagship product. The pure characteristics index, however, remained stable. As mentioned, a key weakness of the logit and BLP models is that because of the idiosyncratic error term, ε, each new product introduction, irrespective of where it is in the quality space, yields welfare gains for consumers. This pushes the implied quality-adjusted price index downward. The pure characteristics model corrects this flaw by removing the error term, thereby allowing consumers to value newly introduced products exclusively by their characteristics.

The implication is that if products are introduced that do not offer substantial improvements

[37] Using the pure characteristics price index as an example, the index is 6.18 log points higher in the third quarter of 2005 relative to the first quarter of 1993 (i.e., log(pt)/log(p0) = 1.0618). Because prices in the first quarter of 1993 were \$127, this implies that in levels prices rose to \$171 ($127^{1.0618}$), or 35 percent. The log transformation method is more accurate for small versus large price changes. For example, prices increased by 29,000 percent between 1993 and 2005, whereas the log-price method depicted in Figure 5 implies an increase of only 12,000 percent. The log-price method therefore understates increases and decreases in quality-adjusted prices in our application, but captures the directional and qualitative effects.

in efficacy, side effects, or substantial decreases in price, the pure characteristics index will rise relative to the logit index (or not fall as much). This is indeed what we see towards the end of the sample period. Consider, for instance, cetuximab, a second-line treatment introduced in 2004. This treatment is quite expensive and offers substantially fewer survival months relative to the existing second-line treatment (irinotecan). The logit model, however, predicts that despite the fact that cetuximab appears to be dominated by irinotecan, consumers still benefit from its introduction. The pure characteristics model, on the other hand, is flat when cetuximab is launched.

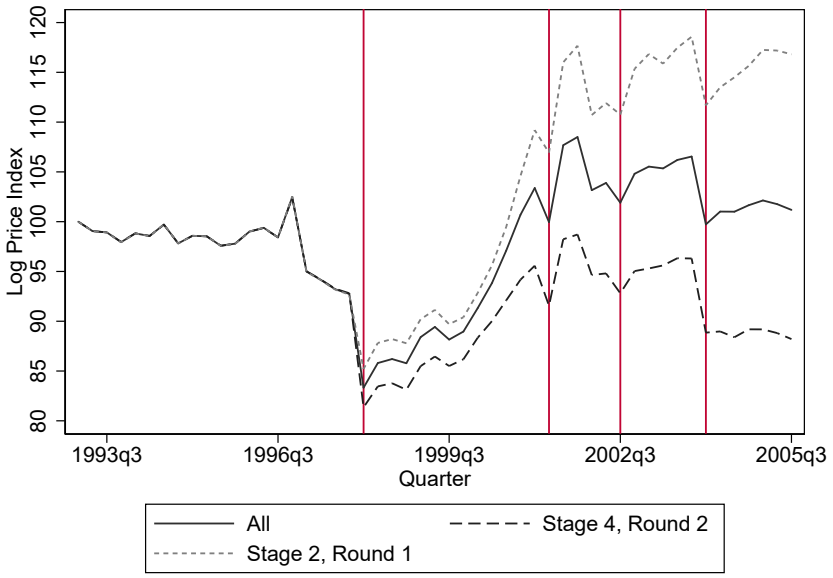
The three price indices depicted Figure 7 rely on physicians' valuation of how drug attributes will affect their patients' health, and the value of health gains. If, however, health insurers fail to create the optimal incentives for patients and physicians, or if physicians exploit their information advantage to capture profits on drug treatments at the expense of patients' health outcomes, these indices may not accurately measure changes in social welfare.

In a separate paper, Lakdawalla et al. (2015) use the same data to present a price index for colorectal cancer drug treatment that does not rely on physicians' valuations of drug attributes, and they arrive at a similar conclusion regarding flat quality-adjusted prices. Specifically, they use the performance of the drugs in randomized clinical trials, reported here in Table 1 and Table 2, and the observed market shares to estimate a patient's expected quality adjusted life years (QALYs) between 1998 and 2005. Applying a value of $100,000 per QALY from the economics literature, they then subtract the change in monetized expected health gains from the change in the drug treatment costs, which is similar to the method used by Cutler et al. (1998) for heart attack treatment and Eggleston et al. (2011) for diabetes treatment. Lakdawalla et al. (2015) find that this quality-adjusted price index, which imposes a value per QALY and forecasts QALY changes based on changes in market share and the survival and side effects of each regimen, increased by $1,400 over this time period, versus the unadjusted increase of $36,000 in the per-patient average drug treatment cost. That is, a price index that does not rely on physicians' subjective valuations is essentially flat, similar to the three indices depicted Figure 7.

Although the aggregate indices shed light on the overall welfare effects of new colorectal cancer regimens, as well as how those effects vary by modeling assumptions, the indices above do not capture the considerable heterogeneity that may exist among patients and physicians who may place different value on different attributes of particular treatments. In Figure 8 we depict the indices constructed from the Micro-BLP model, our preferred specification, for two subgroups. Alternate indices are presented in Appendix A separately for the four scenarios regarding whether the quality of the outside option changes over time and how to interpret the period-specific unobserved value of the inside products. The subgroups we highlight are relatively healthy patients diagnosed with Stage II cancer who are receiving their first-round of drug treatment, versus relatively sick patients diagnosed with Stage IV cancer who are receiving their second-round of drug treatment.

Figure 8 plots the price indices for the two subgroups under our preferred assumptions where the outside option quality changes over time and $\Delta\xi_{jt}$ changes, reflecting changes in unobserved product quality over time. The index plotted for the entire population looks fairly similar to the BLP index plotted in Figure 7. In particular, the index begins fairly flat then falls with each new product entrant, and rises in between. The one notable exception is the *extent* to which quality-adjusted prices rise over time and the extent to which the introduction of new regimens increase

Figure 8: Estimated Price Indices with Physician and Patient Heterogeneity



Notes: This figure plots the price indices from the estimated micro-BLP models, incorporating both observed and unobserved heterogeneity. These are indices for the entire population, as well as for the population of patients with Stage II cancer diagnoses undergoing their first-round of treatment and patients with Stage IV diagnoses undergoing their second-or-higher round of treatment. All indices are on a log scale (i.e. the dependent variable is the "log" price). All indices assume our preferred specification: assuming that the outside option (μ_{0t}) changes in quality each period and that the period-specific deviations from the mean in estimated regimen quality ($\Delta\xi_{jt}$) change each period. Dashed red lines reflect the arrival of four new regimens to the market: irinotecan, capecitabine, oxaliplatin, and bevacizumab, in chronological order.

welfare. In general, price increases are more muted when patient heterogeneity is introduced to the model. By the end of the sample period, the index ends up rising just five percent, implying a slight reduction in welfare (in nominal prices). Moreover, the introduction of new regimens also caused lower fluctuations in the index. For instance, in Figure 7, the introduction of bevacizumab in 2004 led to a decline of about 10 log points for the BLP index. However, once patient heterogeneity is incorporated, the introduction of this flagship product decreases the index by about half that amount. This, in effect, mimics the behavior of the pure characteristics model: the introduction of heterogeneity in patient preferences "soak up" effects that would have otherwise been attributed to idiosyncratic preferences for new products.

While the index implies that the gains from new products were not quite worth the price increases, the behavior of the index for the two subgroups highlights the stark contrast in patient valuations of these cancer therapies. In particular, the quality-adjusted price indices for the two subgroups diverge substantially with the introduction of capecitabine (in 2001) and oxaliplatin (in 2002). By the end of the sample period the quality-adjusted price index for the healthy patients is 128 percent higher than it was in 1993, whereas it is 44 percent *lower* for the sicker patients. Notably, even after the introduction of bevacizumab and cetuximab in 2004, the price index for relatively healthier patients continued to increase through the end of the sample period, suggesting that the survival gains from this regimen did not bring enough value to early-stage cancer patients. However, the price index for sicker patients continue to mildly *decrease*, even post-launch.

Overall, the results indicate that the gains from expensive colorectal cancer regimens accrued mostly to relatively sick patients with repeated interactions with oncologists. These populations experienced a substantial improvement in welfare due predominantly to the development of both (a) regimens that promised significantly higher survival and low side-effects and (b) valuable second-line therapies. Conversely, the high prices brought by these regimens caused welfare to fall for the relatively healthy patients being treated for the first (and probably last) time. Given that the latter group represents a larger share of the population, the overall index for the entire population exhibited a slight increase, suggesting overall welfare declines from product introductions in aggregate.

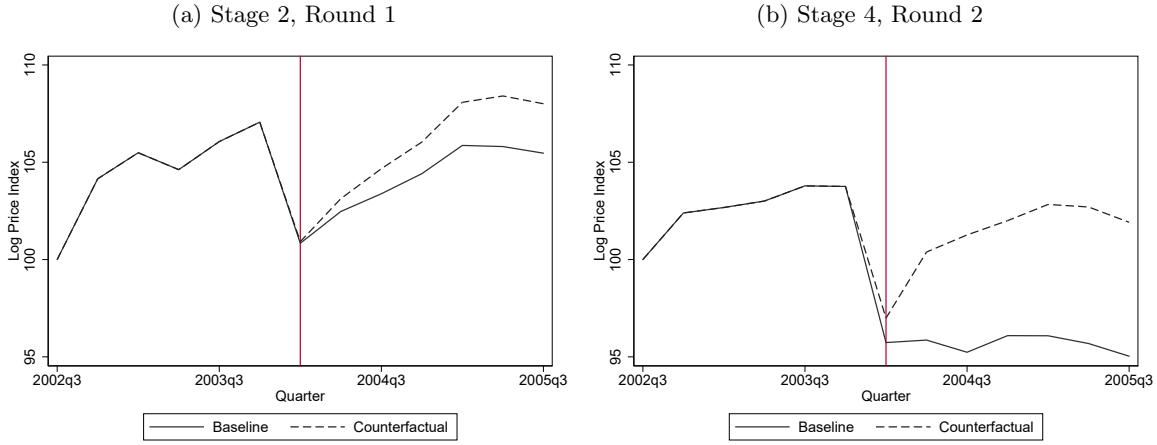
5.3 Welfare Gains Relative to CEA Methods

We next turn to presenting a counterfactual in which the United States adopted CEA criteria to evaluate whether a regimen was approvied for market distribution. Rather than specify a specific QALY threshold for which regimens in our sample would be approved , we opt for a simpler approach to simulate the welfare effects of these policies: dropping regimens from our sample that were not approved in the UK under the CEA criteria. Specifically, using our estimated parameters from the micro-BLP model with observed and unobserved heterogeneity, we recompute the price indices after eliminating all regimens with Bevacizumab and Cetuximab from the physicians' choice sets (thereby removing all market entrants in 2004). We choose these regimens for two reasons. First, in 2007, NICE explicitly recommended against covering these two drugs for all patients, including those requiring second-line treatments. Second, between 2005 and 2006, 56.9% of metastatic patients receiving a second (or higher) line treatment in the US were prescribed a regimen that contained

Bevacizumab or Cetuximab. This is compared with only 1.4% in the UK, suggesting that, even prior NICE's decision, providers in the UK were rarely adminstering these drugs. As such, these regimens would have almost certainly not been approved in the US under CEA criteria of evaluation.

Figure 9 shows the result of these counterfactuals for two patient populations: those with earlier-stage cancers undergoing their first round of treatment and those with later-stage cancers undergoing their second (or higher) round of treatment. To highlight the specific welfare impacts of removing these two regimens in 2004, we only show the price index starting from the third quarter of 2002, when Oxaliplatin (the drug that just entered prior to Bevacizumab and Cetuximab) was introduced. For the Stage II population, the removal of these products, in the observed choice set, the log index rose about 5 points from 2002, whereas we project it would have risen about 8 points for this population had CEA policies been implemented. Evaluated at the mean price of regimens in the third quarter of 2002 (about $20,000), this represents a welfare loss of approximately an additional $11,000 relative to the baseline choice set.

Figure 9: Welfare Gains Relative to Simulating UK CEA Policies



Notes: This figure plots the price indices estimated from the micro-BLP models, incorporating both observed and unobserved heterogeneity. Panel (a) reports indices for patients in earlier stage cancers undergoing their first round of treatment, whereas panel (b) reports indices for later-stage patients undergoing their second-or-higher round of treatment. Each figure plots the baseline index (normalized to 100 in the third quarter of 2002) as well as a counterfactual index where regimens containing Bevacizumab or Cetuximab are removed. All indices are on a log scale (i.e. the dependent variable is the "log" price). All indices assume our preferred specification: assuming that the outside option (μ_{0t}) changes in quality each period and that the period-specific deviations from the mean in estimated regimen quality ($\Delta\xi_{jt}$) change each period.

For the population Stage IV cancer patients, the welfare effects are fairly similar, though somewhat more pronounced. For these relatively sicker patients, the log price index declines by about 5 points from 2002 at baseline (consistent with the decline seen in Figure 8). If the market entrants in 2004 were removed, however, the index *rises* by approximately 2 points by the end of the sample. Overall, the net welfare loss of product removal for this population is approximately $12,000. While this net welfare loss is similar to the magnitudes borne by earlier-stage patients, it is notable that

at baseline, this relatively sicker population sees net welfare gains of about $8,000 per patient (as the price index declines), whereas the relatively healthier population sees no such gains from these products. We project these gains to be completely eliminated if CEA criteria were implemented.[38]

6 Conclusion

In this paper, we use price and market share data from 1993 to 2005, clinical trial data, and patient and physician characteristics to estimate a series of quality-adjusted price indices for colorectal cancer treatments. In particular, we compute a naive price index that does not adjust for improving attributes, and compare it with indices based on logit, full random-coefficients BLP, and pure characteristics demand models.

We find that accounting for patient heterogeneity is important. Newer, more effective colorectal cancer treatments benefited patients with late-stage cancers whose initial treatment was not successful relative to patients with early stage cancers. The former group of patients experienced a quality-adjusted price decrease of 44 percent, whereas the latter group of patients experienced an increase of 128 percent. When we ignore heterogeneity and only use an average value, such as is commonly done in European countries, the quality-adjusted price index increases by between five and 15 percent, implying that the innovation was not worth it. A uniform rule preventing patients with advanced cancer from receiving newer treatments would reduce welfare for these patients. During our sample period, five of the six drugs were patent-protected and able to mark-up prices well above production costs. By July 2019, however, generic or biosimilar versions of all but one of the drugs had entered the market and driven down prices due to greater competition.

Our results show that adjusting for quality and accounting for patient heterogeneity matter. The naive price index greatly overestimates the price increase. We find that the pure characteristics quality-adjusted (log) price index rises by five to 15 percent for the entire patient population over a 13-year period compared with a 29,000% percent increase exhibited in a model without quality adjustments. Moreover, allowing for flexible substitution patterns matters for generating accurate welfare estimates. By shedding the idiosyncratic logit error term that is a workhorse of traditional demand models, the pure characteristics model produces more realistic substitution patterns as new colorectal cancer treatments enter the market and replace existing regimens. Quality adjustment methods that rely on traditional logit models overestimate welfare gains of product introductions by up to 10 log points.

Our calculations regarding the value of new technology do not include the insurance value of innovation (Lakdawalla et al., 2017): how healthy individuals value reducing the variance of outcomes by converting previously uninsurable physical risk because of the lack of treatment into insurable financial risk. The value measured in our setting is for patients that are already sick, for whom treatment innovation improves utility in the bad state of the world, making illness less

[38] It is also likely that these figures are underestimated. Like in our baseline specification, we assume that the outside option's value changes over time as the quality of "outside goods" increases. If CEA policies were adopted, however, it is very likely that use of these off-label products would be highly limited, as is the case in Europe. Some of the welfare gains of outside option changes would, therefore, be muted. If we assume the scenario where the outside option remains fixed, we predict the net welfare loss associated with the removal of Bevacizumab and Cetuximab would be closer to $25,000 per patient.

unpleasant through either increased efficacy or lower side effects.

References

Agha, L. and Molitor, D. (2018). The local influence of pioneer investigators on technology adoption: Evidence from new cancer drugs. *Review of Economics and Statistics*, 100(1). 2.4

Bauer, D., Lakdawalla, D., and Reif, J. (2021). Health risk and the value of life. *NBER Working Paper*. 1

Berndt, E. R., Bir, A., Busch, S. H., and Frank, R. G. (2002). The medical treatment of depression, 1991-1996: Productive inefficiency, expected outcome variations, and price indices. *Journal of Health Economics*, 3(21):373–396. 2.1

Berndt, E. R., Cutler, D., Frank, R. G., Griliches, Z., Newhouse, J., and Triplett, J. E. (2000). *Handbook of Health Economics*. Elsevier Science. 2.1

Berry, S., Levinsohn, J., and Pakes, A. (2004). Differentiated product demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*. 1, 4.1, 4.2, 4.2

Berry, S. T. (1994). Estimating discrete choice models of product differentiation. *Rand Journal of Economics*, 25(2):242–262. 1, 4.1, 4.1, 4.2

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890. 1, 5, 4.1, 4.1, 4.2, 4.2, 4.2, 4.4

Berry, S. T. and Pakes, A. (2007). The pure characteristics demand model. *International Economic Review*, 48(4):1993–1225. 1, 4.1, 4.3.1, 4.4, 4.4, B, B

Boskin, M. J., Dulberger, E. R., Gordon, R. J., and an Dale Jorgenson, Z. G. (1998). Consumer prices, the consumer price index, and the cost of living. *Journal of Economic Perspectives*, 12(1). 1

Bresnahan, T. F., Stern, S., and Trajtenberg, M. (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics*, 28(0):S17–S44. 4.2, 4.4

Brouwer, N. P., Bos, A. C., Lemmens, V. E., Tanis, P. J., Hugen, N., Nagtegaal, I. D., de Wilt, J. H., and Verhoeven, R. H. (2018). An overview of 25 years of incidence, treatment and outcome of colorectal cancer patients. *International Journal of Cancer*, 143. 9

Brown, R. E. and Hutton, J. (1998). Cost-utility model comparing docetaxel and paclitaxel in advanced breast cancer patients. *Anticancer Drugs*, 9(10). 35

Brown, R. E., Hutton, J., and Burrell, A. (2001). Cost effectiveness of treatment options in advanced breast cancer in the uk. *Pharmacoeconomics*, 19(11). 35

Chan, D. C., Gentzkow, M., and Yu, C. (2021). Selection with skills:evidence from radiologists. *(Forthcoming) Quarterly Journal of Economics*. 2.4

Chernew, M. E. and Newhouse, J. (2011). Health care spending growth. In Pauly, M. V., Mcguire, T. G., and Barros, P., editors, *Handbook of Health Economics, Volume 2*, pages 1–43. Elsevier. 1

Claxton, K., Sculpher, M., Palmer, S., and Culyer, A. J. (2015). Causes for concern: Is nice failing to uphold its responsibilities to all nhs patients? *Health Economics*, 24. 2.3

Cockburn, I. M. and Anis, A. H. (2001). Hedonic analysis of arthritis drugs. In Cutler, D. M. and Berndt, E. R., editors, *Medical Care Output and Productivity*, pages 439–462. University of Chicago Press, Chicago. 4, 2.1, 2.4

Conti, R. M., Rosenthal, M. B., Polite, B. N., Bach, P. B., and Shih, Y.-C. T. (2012). Infused chemotherapy use in the elderly after patent expiration. *American Journal of Managed Care*, 18(5). 2.4

Cutler, D., McLellan, M., Newouse, J., and Remler, D. (1998). Are medical prices declining? evidence from heart attack treatments. *Quarterly Journal of Economics*, 63(4):991–1024. 2.1, 5.2

Cutler, D. M., Mclellan, M., and Newhouse, J. P. (2000). How does managed care do it? *RAND Journal of Economics*, 31(3). 1

Danzon, P. M. and Furukawa, M. F. (2008). International prices and availability of pharmaceuticals in 2005. *Health Affairs*, 27(1). 2.3

Davidoff, A. J., Shaffer, T., Erten, M. Z., Shoemaker, S., Zuckerman, I. H., Pandya, N., Stuart, B. C., Bryant-Comstock, L. G., and Shenolikarl, R. (2013). Use and spending on antineoplastic therapy for medicare beneficiaries with cancer. *Medical Care*, 51(4). 9

Dunn, A. (2012). Drug innovations and welfare measures computed from market demand: The case for anti-cholesterol drugs. *American Economic Journal: Applied Economics*, 4(3):167–189. 4, 2.1, 2.4

Eggleston, K. N., Shah, N. D., Smith, S. A., and Berndt, E. R. (2011). Quality adjustment for health care spending on chronic disease: Evidence from diabetes treatment, 1999-2009. *American Economic Review: Papers and Proceedings*, 101(3). 2.1, 5.2

Einav, L., Finkelstein, A., and Williams, H. (2016). Paying on the margin for medical care: Evidence from breast cancer treatments. *Quarterly Journal of Economics*, 8(1). 1, 2.3

Frank, R. G. and Nichols, L. M. (2019). Medicare drug-price negotiation—why now...and how. *New England Journal of Medicine*, 381. 1

Gandhi, A. and Nevo, A. (2021). *Handbook of Industrial Organization, Volume 4*, chapter Empirical Models of Demand and Supply in Differentiated Products Industries. Elsevier. 4.2

Garber, A. M. and Phelps, C. E. (1997). Economic foundations of cost-effectiveness analysis. *Journal of Health Economics*, 16. 1

Government Accountability Office (2001). Medicare: Payments for covered outpatient drugs exceed providers' cost. Technical report, Government Accountability Office. 2.4

Hartman, M., Martin, A. B., Benson, J., and Catlin, A. (2020). National health care spending in 2018: Growth driven by accelerations in medicare and private insurance spending. *Health Affairs*, 39(1). 10

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In Bresnahan, T. F. and Gordon, R. J., editors, *The Economics of New Goods*, pages 207–248. University of Chicago Press, Chicago. 6

Hicks, J. R. (1940). The valuation of the social income. *Economica*, 7(26). 6

Howard, D. H., Bach, P. B., Berndt, E. R., and Conti, R. M. (2015). Pricing in the market for anticancer drugs. *Journal of Economic Perspectives*, 29(1):139–162. 3.1

Jacobson, M., Chang, T. Y., Newhouse, J. P., and Earle, C. C. (2016). Physician agency and competition: Evidence from a major change to medicare chemotherapy reimbursement policy. *Journal of Economic Behavior and Organization*, 134. 4, 2.4

Jacobson, M., Earle, C. C., Price, M., and Newhouse, J. P. (2010). How medicare's payment cuts for cancer chemotherapy drugs changed patterns of treatment. *Health Affairs*, 29(7). 2.4

Jacobson, M. and Newhouse, J. P. (2006). Does reimbursement influence chemotherapy treatment for cancer patients? *Health Affairs*, 25. 2.4, 3.1

Kolstad, J. T. (2013). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7). 4

Kravitz, R. L., Duan, N., and Braslow, J. (2004). Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Quarterly*, 82(4). 3

Lakdawalla, D., Shafrin, J., Lucarelli, C., Nicholson, S., Khan, Z. M., and Philipson, T. J. (2015). Quality-adjusted cost of care: A meaningful way to measure growth in innovation versus the value of health gains. *Health Affairs*, 34(4):555–561. 2.1, 5.2

Lakdawalla, D. N., Jena, A. B., and Doctor, J. N. (2014). Careful use of science to advance the debate on united kingdomâs cancer drugs fund. *Journal of the American Medical Association*, 311(1). 2

Lakdawalla, D. N., Malani, A., and Reif, J. (2017). The insurance value of medical innovation. *Journal of Public Economics*, 145. 6

Lakdawalla, D. N., Romley, J. A., Sanchez, Y., Maclean, J. R., Penrod, J. R., and Philipson, T. (2012). How cancer patients value hope and the implications for cost-effectiveness assessments of high-cost cancer therapies. *Health Affairs*, 31(4). 1

Lloyd, A., Nafees, B., Narewska, J., Dwilde, S., and Watkins, J. (2006). Health state utilities for metastatic breast cancer. *British Journal of Cancer*, 95(6). 35

Lucarelli, C. and Nicholson, S. (2009). A quality-adjusted price index for colorectal cancer drugs. *NBER Working Paper*. 4.4

Lucarelli, C., Nicholson, S., and Town, R. J. (2010). The effect of physician reimbursement on chemotherapy treatment decisions and patient outcomes. *Unpublished Manuscript, Cornell University*. 3.1

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., and Leibowitz, A. (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *American Economic Review*, 77(3). 1

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*. Academic Press, New York. 1

McWilliams, J. M., Hartfield, L. A., Chernew, M. E., Landon, B. E., and Schwartz, A. L. (2016). Early performance of accountable care organizations in medicine. *New England Journal of Medicine*, 374. 1

Medicare Payment Advisory Commission (2006). Report to the congress: Medicare payment policy. Technical report, Medicare Payment Advisory Commission. 2.4

Murphy, K. M. and Topel, R. H. (2006). The value of health and longevity. *Journal of Political Economy*, 114(5):871–904. 6, 5.1

Neumann, P. J., Cohen, J. T., and Weinstein, M. C. (2014). Updating cost-effectiveness–the curious resilience of the $50,000-per-qaly treshold. *New England Journal of Medicine*, 371(9). 5.1

Nevo, A. (2000). A practitioner's guide to estimation of random coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548. 1, 4.2, 4.2

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342. 1, 5, 4.2

Nevo, A. (2003). New products, quality changes and welfare meausres computed from estimated demand systems. *The Review of Economics and Statistics*, 85(2):266–275. 1, 6, 4.3.1, 32, A

Newhouse, J. (1992). Medical care costs: How much welfare loss? *Journal of Economic Perspectives*, 6(3):3–21. 1

Nord, E., Richardson, J., Street, A., Kuhse, H., and Singer, P. (1995). Maximizing health benefits vs. egalitarianism: An australian survey of health issues. *Social Science and Medicine*, 41(10). 1

Nosko, C. (2014). Competition and quality choice in the cpu market. *Working Paper*, Forthcoming. 1

Pakes, A., Berry, S. T., and Levinsohn, J. H. (1993). Applications and limitations of some recent advances in empirical industrial organization: Price indices and the analysis of environmental change. *American Economic Review: Papers and Proceedings*, 83. 6

Paramore, C. L., Thomas, S. K., Knopf, K. B., .s Cragin, L., and Fraeman, K. H. (2006). Estimating cost of care for patients with newly diagnosed metastatic colorectal cancer. *Clinical Colorectal Cancer*, 6(1). 2.2

Pauly, M. V. (2017). The questionable economic case for value-based drug pricing in market health systems. *Value Health*, 20(2). 1

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729. 6

Small, K. and Rosen, H. (1981). Applied welfare economics with discrete choice models. *Econometrica*, 49(1):105–130. 4.3

Smith, S., Newhouse, J., and Freeland, M. S. (2009). Income, insurance, and technology: Why does health spending outpace economic growth? *Health Affairs*, 28(5):1276–1284. 1

Song, M. (2007). Measuring consumer welfare in the cpu market: An application of the pure characteristics demand model. *Rand Journal of Economics*, 38(2):429–446. 1, B

Song, M. (2008). The quality-adjusted price index in the pure characteristics demand model. *Journal of Business and Economic Statistics*, 28(1):190–199. 1, 4.4, B

Song, M., Lucarelli, C., and Nicholson, S. (2017). Merger with inter-firm bundling: A case of pharmaceutical cocktails. *RAND Journal of Economics*, 48(3). 5.1

Stevens, W., Philipson, T. J., Khan, Z. M., MacEwan, J. P., Lintchicum, M. T., and Goldman, D. P. (2015). Cancer mortality reductions were greatest among countries where cancer care spending rose the most, 1995â2007. *Health Affairs*, 34(4). 1

Trajtenberg, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *Journal of Political Economy*, 97(2). 6, 2.1

Trajtenberg, M. (1990). Product innovations, price indices, and the (mis)measurement of economic performance. *NBER Working Paper No. 3261*. 4.3

A Alternate Assumptions on Unobserved Components

In Figure A.1, we present alternate indices with varying assumptions on unobserved product quality and outside options. Here we explain why our preferred price indices assume that the unobserved quality of the outside option is changing over time. In the first quarter of 2000, only 2.1 percent of the patients in our sample were treated with a drug treatment that is included in our outside option – a drug regimen that was approved for colorectal cancer but had a very small market share throughout the sample period, where the efficacy and side effect data were unavailable, or where the regimen was not approved by the FDA for colorectal cancer (i.e., an off-label treatment). By the fourth quarter of 2003, on the other hand, the market share of the outside option had increased to 8.4 percent, with 2.8 percent receiving a regimen involving a drug that was approved for colorectal cancer (e.g., oxaliplatin, irinotecan, or capecitabine plus other component drugs), and the remaining 5.6 percent involving drugs approved for other types of cancer. Two years later, in the third quarter of 2005, the cumulative market share of the outside option had more than doubled to 18.3 percent, with almost 10 percent of this accounted for by a regimen involving bevacizumab (Avastin) plus other component drugs. Bevacizumab is expensive and is included in regimens with the longest median survival (see Figure 1). Physicians appear, therefore, to be shifting over time to regimens that use drugs approved for colorectal cancer combined with drugs that were not included in the trials that were submitted for FDA approval. This suggests that physicians attached substantial value to the treatments included in the outside option, which they relied on increasingly over time.

Panel (a) of Figure A.1 depicts our preferred specification, as in Figure 7. Panel (c) depicts the three indices assuming that the quality of the outside option changes over time but $\Delta\xi_{jt}$ remains fixed at the first-period values (i.e. assuming $\Delta\xi_{jt}$ reflects a change in tastes, rather than product quality). Here, nearly all variation between the indices have been eliminated. All of the indices appear identical to the indices in Panel (a) pre-1997 (when there is only one available product), but thereafter exhibit a steady decline. By the end of the sample period, consumers experienced welfare gains of about 10 log points (or about a 40 percent decline in quality-adjusted prices). The implication is that the combination of time dummies and residuals explain a lot of the variation across indices. At first glance this might be seem counterintuitive, but it conforms to expectations when considering the implications of (Nevo, 2003). In particular, large swings both in aggregate and product-specific market shares in the sample imply that there are certain behaviors that are difficult to explain through observable characteristics alone, and they are captured by $\Delta\xi_{jt}$ and ξ_t.

Both plots above are constructed based on a scenario where the coefficients on the time indicators are interpreted as measuring how the quality of the outside option changes over time. We believe, as mentioned above, that given the prevalence of off-label use of bevacizumab in the later years of the sample, this interpretation is the most consistent with the data. However, in Panel (b) of Figure 7 we plot the log price indices for the three price indices under a scenario where the quality of the outside option is assumed to be fixed over time at zero (i.e., the coefficients on the time indicators capture changes in the mean quality of the inside goods), and $\Delta\xi_{jt}$ is assumed to reflect changes in unobserved product quality over time. As market share shifts to the outside option after 2000, all three price indices increase by more than those in Panel (a) and Panel (c).

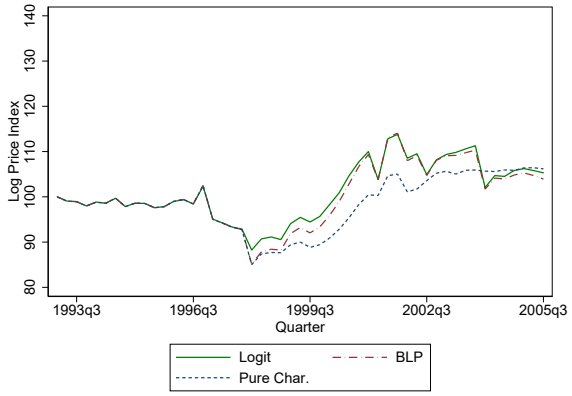
Assuming time dummies reflect changing product quality appears to result in a significant deterioration of the mean quality of inside goods towards the latter part of the sample (when a larger share of patients switch to the outside option), driving all the indices upward. By the end of the sample period, the price indices are 18 to 33 log points (139 percent to 395 percent) higher than in 1993, implying substantial reductions in welfare.

Finally, in Panel (d) of Figure 7 we depict the three indices under the assumption that the quality of the outside option remains fixed (as in Panel (b)), but that $\Delta\xi_{jt}$ also remains fixed as product j's first period value. The pattern of these indices is similar to those in Panel (b), although the three indices do not rise as much by end of the sample period, implying greater welfare gains (or smaller losses).

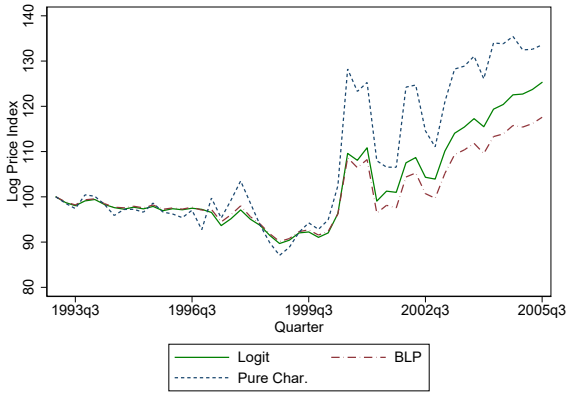
In Figure A.2, we plot the indices from the models with physician and patient heterogeneity under each of these assumptions. Again, panel (a) reflects our preferred specification, as in Figure 8. The indices in the other three panels have a similar pattern where the sickest patients experience welfare improvements due to innovation (and the prices of those products), and healthier patients never fare as well as the sicker patients. In Panel (c) and Panel (d), the welfare differences are fairly small, whereas they are larger in Panel (a) and Panel (b). Looking across the four panels, it is apparent that consumer heterogeneity is critically important when estimating welfare gains from introductions of new regimens, and therefore the implied quality-adjusted prices due to innovation. The two subpopulations experience very different welfare effects from the innovation, such that an aggregate index would incorrectly conclude that all consumers experience the same outcomes.

Figure A.1: Estimated Price Indices Without Patient Heterogeneity

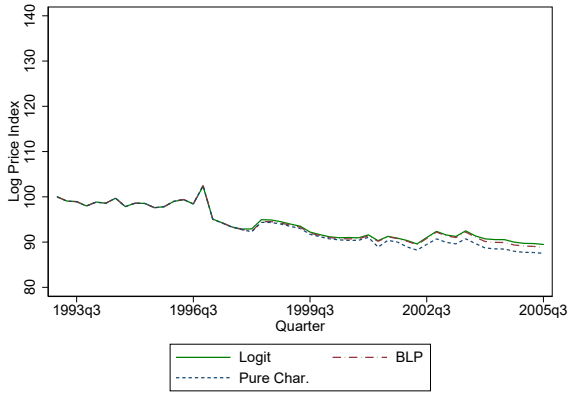
(a) u_{0t} change, $\Delta\xi_{jt}$ changes



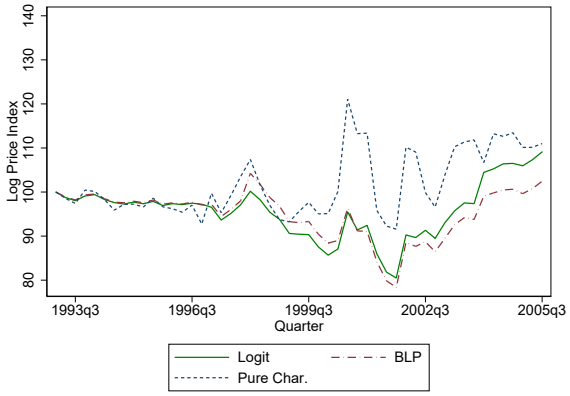
(b) u_{0t} fixed, $\Delta\xi_{jt}$ changes



(c) u_{0t} change, $\Delta\xi_{jt}$ fixed



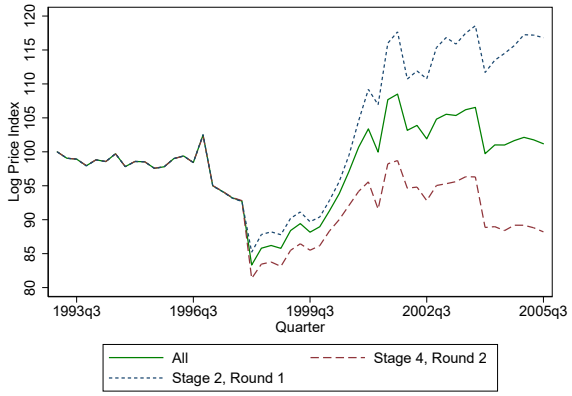
(d) u_{0t} fixed, $\Delta\xi_{jt}$ fixed



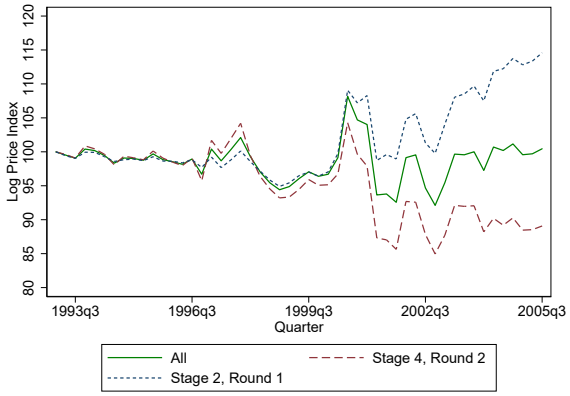
Notes: This figure plots the *aggregate* price indices from the estimated logit, BLP, and pure characteristics models. All indices are on a log scale (i.e. the dependent variable is the "log" price). Panel (a) plots the index from our preferred specification: assuming that the outside option (μ_{0t}) changes in quality each period and that the period-specific deviations from the mean in estimated regimen quality ($\Delta\xi_{jt}$) change each period. Panel (b) plots the indices assuming that μ_{0t} is normalized at 0 each period (and therefore the period-specific mean unobserved utility, ξ_t, changes), while $\Delta\xi_{jt}$ changes. Panel (c) plots indices where μ_{0t} changes each period, but $\Delta\xi_{jt}$ remains fixed at the regimens first-period value (its estimated value in the quarter the regimen first is introduced). Panel (d) plots indices assuming μ_{0t} is fixed at 0 each period and that $\Delta\xi_{jt}$ is also fixed at each regimen's first-period value.

Figure A.2: Estimated Price Indices with Patient and Physician Heterogeneity

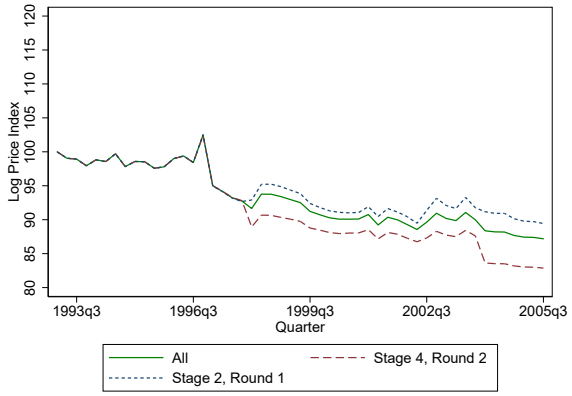
(a) u_{0t} changes, $\Delta\xi_{jt}$ changes



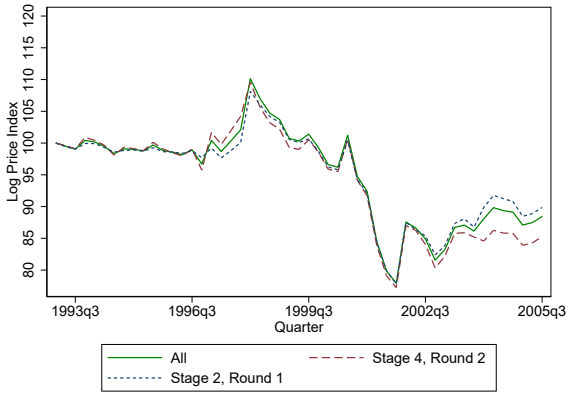
(b) u_{0t} fixed, $\Delta\xi_{jt}$ changes



(c) u_{0t} changes, $\Delta\xi_{jt}$ fixed



(d) u_{0t} fixed, $\Delta\xi_{jt}$ fixed



Notes: This figure plots the price indices from the estimated micro-BLP models, incorporating both observed and unobserved heterogeneity. Each panel presents indices for the entire population, as well as for the population of patients with Stage II cancer diagnoses undergoing their first-round of treatment and patients with Stage IV diagnoses undergoing their second-or-higher round of treatment. All indices are on a log scale (i.e. the dependent variable is the "log" price). Panel (a) plots the index from our preferred specification: assuming that the outside option (μ_{0t}) changes in quality each period and that the period-specific deviations from the mean in estimated regimen quality ($\Delta\xi_{jt}$) change each period. Panel (b) plots the indices assuming that μ_{0t} is normalized at 0 each period (and therefore the period-specific mean unobserved utility, ξ_t, changes), while $\Delta\xi_{jt}$ changes. Panel (c) plots indices where μ_{0t} changes each period, but $\Delta\xi_{jt}$ remains fixed at the regimens first-period value (its estimated value in the quarter the regimen first is introduced). Panel (d) plots indices assuming μ_{0t} is fixed at 0 each period and that $\Delta\xi_{jt}$ is also fixed at each regimen's first-period value.

B Estimating the Pure Characteristics Model

The pure characteristics model described in subsection 4.4 is notoriously difficult to estimate. The challenge stems from the absence of the idiosyncratic error term, ϵ_{ijt}, present in both the logit and random-coefficients models. In the previous models, the error term ensured that the market share function was a smooth function of the regimen characteristics, which allowed for easy integration. Without the error term, there is no longer a guaranteed contraction mapping that can be used to equate the predicted and observed market shares, and hence back out the mean utility, δ_j. We, therefore, resort to alternate methods of estimating the mean product quality as described by Berry and Pakes (2007). Our methods also follow closely that of Minjae Song (Song, 2007, 2008).

As in Berry and Pakes (2007) and Song (2008), the estimation procedure proceeds in three steps. The first step is to bring the mean product quality closer to the true product quality by use of a contraction mapping, as in BLP. Recall the utility function in BLP is:

$$u_{ijt} = -\alpha p_{jt} + \beta x_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt} + \xi_j + \Delta\xi_j + \epsilon_{ijt} \tag{12}$$

In principle, then, the utility function from the pure characteristics model is a limiting case of the utility function from the random coefficients model as ϵ_{ijt} approaches zero. We thus begin by implementing the BLP contracting mapping with a scaling factor applied to the error term that gradually proceeds to zero. The market share in this model collapses to:

$$S_{jt}(\delta_{jt}, \alpha^u, \beta^u) = \frac{1}{ns}\sum_{i=1}^{ns} \frac{exp[(\delta_{jt} - \alpha^u v_i p_{jt} + \beta^u v_i x_{jt})\mu]}{1 + \sum_{k=1}^{J} exp[(\delta_{kt} - \alpha^u v_i p_{kt} + \beta^u v_i x_{kt})\mu]} \tag{13}$$

where μ is the scaling factor and gradually grows larger in the estimation routine. In practice, as μ grows larger, the exponential function rapidly blows up, resulting in incalculable or missing mean values. We iterate on μ and proceed slowly until the point that we can no longer compute a δ_j. The closes δ_j we obtain using this scaled share function brings us closer to the true mean value from the pure characteristics model, and we hence use this value as the starting point for the next part of the procedure.

The second step is an element-by-element fixed-point homotopy method. This goal of this inversion is to find a mean utility value, δ_j, to satisfy the following equation:

$$|S_{jt}(\delta_j, \delta_{-j}, \theta) - s_{jt}| < tol \tag{14}$$

The difficult is that this inversion is not guaranteed to be a weak contraction mapping. Therefore, both Berry and Pakes (2007), and Song, combine this element-by-element inversion with a homotopy method using the following:

$$\delta'_j(t) = (1-t)\delta_{0j} + t\delta_j, j = 1, ..., J \tag{15}$$

where δ_{0j} is an initial guess for the mean value, δ_j is the current iteration of the mean value, and t is between 0 and 1. As Song points out, when $t = 1$, this collapses back to the strict element-

by-element inverse, which is not guaranteed to contract. However, when $t < 1$, there is a strict contraction mapping guaranteed, albeit the fixed point the model converges to may not necessarily be the true value of the pure characteristics model, δ_j. Therefore, much like the first step of the procedure in which the BLP contraction is implemented with a scaling factor to bring the value of δ_j closer to the true value, the homotopy method is implemented with a value of t that is strictly less than 1, but repeated while approaching one very slowly. Following Berry & Pakes, and Song, we begin with $t = 0.99$ and increase by 0.0025, with the element-by-element mapping repeating at least 50 times before altering the value of t.

The final step is to use the Newton-Rhapson search method. This occurs when the element-by-element inversion in Equation 15 is not satisfied after an iteration of the homotopy search, yet all the predicted market shares are non-zero. This implies the true value of δ_j has not been found yet, but since the current predicted market shares are non-zero, an alternate, more rapid search method for smooth functions can be implemented. While both Berry & Pakes, and Song had to rely on the Newton method in their simulations of the pure characteristics model, our implementation rarely relied on this method, with the homotopy method approaching the true δ_j and satisfying Equation 15 most of the time.

C Composition and Dosages

Table C.1: Composition and Dosages of Chemotherapy Drugs

Regimen	1^{st} Drug	2^{nd} Drug	3^{rd} Drug	4^{th} Drug
5-FU + LV	425 mg of 5-Fu/m^2/day for days 1-5, every 4 weeks	20 mg of LV/m^2/day for days 1-5, every 4 weeks		
Irinotecan (Pfizer)	125 mg of irinotecan per week/m^2 for 4 weeks, every 6 weeks			
Irinotecan + 5-FU/LV	180 mg of irinotecan/m^2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and 2, every 2 weeks	200 mg of LV/m^2 on day 1 and 2, every 2 weeks	
Capecitabine (Roche)	2,500 mg of capecitabine per m2/day for days 1-14, every 3 weeks			
Capecitabine + Irinotecan	70 mg of irinotecan/m^2/week, every 6 weeks	2,000 mg of capecitabine per m^2/day for days 1-14, every 3 weeks		
Oxaliplatin (Sanofi) + 5-FU/LV	85 mg of oxaliplatin per m^2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and day 2, every 2 weeks	200 mg of LV/m^2 on day 1 and day 2, every 2 weeks	
Oxaliplatin + Capecitabine	130 mg of oxaliplatin per m^2 on day 1, every 3 weeks	1,700 mg of capecitabine per m^2/day for days 1-14, every 3 weeks		
Cetuximab (ImClone)	400 mg of cetuximab per m^2 on day 1; then 250 mg/m^2 once a week, every 6 weeks			
Cetuximab + Irinotecan	400 mg of cetuximab per m^2 on day 1; then 250 mg/m^2 once a week, every 6 weeks	125 mg of irinotecan per week/m^2 for 4 weeks, every 6 weeks		
Bevacizumab (Genentech) + Oxaliplatin + 5-FU/LV	5 mg of bevacizumab per kg, every 2 weeks	85 mg of oxaliplatin per m2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and day 2, every 2 weeks	200 mg of LV/m^2 on day 1 and day 2, every 2 weeks
Bevacizumab + Irinotecan + 5-FU/LV	5 mg of bevacizumab per kg, every 2 weeks	180 mg of irinotecan per m^2 on day 1, every 2 weeks	1,000 mg of 5-FU/m^2 on day 1 and day 2, every 2 weeks	200 mg of LV/m^2 on day 1 and day 2, every 2 weeks
Bevacizumab + Oxaliplatin + Capecitabine	7.5 mg of bevacizumab per kg, every 3 weeks	130 mg of oxaliplatin per m^2 on day 1, every 3 weeks	1,700 mg of capecitabine/m^2/day for days 1-14, every 3 weeks	

mg=miligram of active ingredient; m^2=meter squared of a patient's surface area; kg=kilogram of a patient's weight.
Source: National Comprehensive Cancer Network, Colon Cancer, Version 2.2006; package inserts.

D Additional Demand Tables

Table D.1 presents the results of our baseline IV logit specification, as well as additional specifications that lag our Hausman instruments by two periods and three periods (thereby generating more limited serial correlation). The results are extremely similar to the baseline specification, with the price coefficients as well as those of all efficacy measures remaining virtually unchanged.

Table D.1: Parameter Estimates from Models with Alternate Instruments

	(1) IV Logit	(2) IV Lag 2	(3) IV Lag 3
(Log) Price	-1.280***	-1.405***	-1.389***
	(0.304)	(0.383)	(0.413)
Survival Months	0.570***	0.639***	0.631***
	(0.175)	(0.219)	(0.235)
Vomiting	-0.243***	-0.266***	-0.263***
	(0.0683)	(0.0821)	(0.0866)
Tablet	0.798	1.047	1.017
	(0.672)	(0.827)	(0.882)
Second Line	5.324***	6.065**	5.974**
	(1.877)	(2.342)	(2.515)
Pfizer	1.600**	1.835**	1.806**
	(0.616)	(0.762)	(0.815)
Constant	1.824**	1.782**	1.787**
	(0.799)	(0.871)	(0.864)
Observations	208	208	208

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Parameter estimates from various demand models. Column 1 reports results from the logit model with Hausman instrumental variables. Column 2 reports estimates where the instruments are lagged by two periods. Column 3 reports estimates where the instruments are lagged by three periods.

In Table D.2, we present our baseline specification, as well as alternate models that remove the firm indicator for Pfizer. The results in Column 2 are fairly similar to those in Column 1. There are two notable exceptions. First, the price sensitivity coefficient becomes somewhat more muted. Second, the coefficient on "tablet" flips signs (although it remains insignificant). Although seemingly counterintuitive, the reason this occurs is that, as shown in Figure 2, most Pfizer regimens retain fairly high market shares upon launch despite having largely similar efficacy and side effects measures as other regimens. Notably, Irinotecan + 5-FU-LV retains an approximate 25 percent market share compared with about six percent for Capecitabine. This is in spite of the fact that Capecitabine is considerably less expensive, is a tablet regimen, has approximately the same survival rate, and has a *lower* incidence of vomiting, as shown in Table 1 and Table 2. Without an indicator for the flagship Pfizer regimens, the model therefore rationalizes this behavior by muting the utility gains from tablet convenience.

There is, however, one alternative explanation for this behavior: Pfizer regimens tend to have somewhat higher incidence of tumor reduction (i.e., response rate). To test this, Column 3 presents the results of a specification where the firm indicators are dropped but the "response rate" efficacy

measure is included in addition to survival months. Here, the results return to being nearly identical to the baseline specification. The price coefficient, in particular, is quite similar at -1.360 (vs. -1.280). The utility gains from regimens with higher response rates are positive and significant, as expected. However, the utility on survival months becomes somewhat more muted and loses significance. This latter phenomenon occurs simply due to the fact that survival months, response rates, and price are highly correlated.[39] In any case, we take this as evidence that our baseline specification captures the utility gains from Irinotecan introduction quite well.

Table D.2: Parameter Estimates from Additional IV Logit Specifications

	(1)	(2)	(3)
(Log) Price	-1.280***	-0.983***	-1.360***
	(0.304)	(0.184)	(0.346)
Survival Months	0.570***	0.398***	0.117
	(0.175)	(0.108)	(0.0779)
Response Rate			0.197**
			(0.0759)
Vomiting	-0.243***	-0.156***	-0.133***
	(0.0683)	(0.0416)	(0.0450)
Second Line	5.324***	3.815***	6.078***
	(1.877)	(1.260)	(2.223)
Tablet	0.798	-0.0132	1.061
	(0.672)	(0.387)	(0.785)
Pfizer	1.600**		
	(0.616)		
Constant	1.824**	2.223***	3.182***
	(0.799)	(0.704)	(0.910)
Observations	208	208	208

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Parameter estimates from various IV logit demand models. Column 1 reports the baseline specification. Column 2 reports results from a model with no firm indicators. Column 3 reports results from no firm indicators and an additional efficacy variable: response rate.

[39] Survival months and response rates have a 91% correlation.