CHAPTER

16

The industrial organization of health care markets☆

Ben Handel[a,b,*] **and Kate Ho**[c,b,d]
[a]*UC Berkeley, Berkeley, CA, United States*
[b]*NBER, Cambridge, MA, United States*
[c]*Princeton University, Princeton, NJ, United States*
[d]*CEPR, London, United Kingdom*
*Corresponding author: e-mail address: *handel@berkeley.edu*

CONTENTS

☆ We are grateful to Jason Abaluck, Zarek Brot-Goldberg, David Dranove, Martin Gaynor, Soheil Ghili, Gautam Gowrisankaran, Michael Geruso, Jonathan Holmes, Robin Lee, Mark Shepard, Amanda Starc, and Robert J Town, for very helpful comments on earlier drafts of this chapter. We thank Catherine Che, Yuki Ito, Yikun Jiang, and Matteo Saccarola for excellent research assistance.

Handbook of Industrial Organization, Volume 5, ISSN 1573-448X. https://doi.org/10.1016/bs.hesind.2021.11.016

Abstract

In this chapter we outline the tools that have been developed to model and analyze competition and regulation in health care markets, and describe particular papers that apply them to policy-relevant questions. We focus particularly on the I.O. models and empirical methods and analyses that researchers have formulated to address policy-relevant questions, although we also provide an overview of the institutional facts and findings that inform them. We divide the chapter into two broad sections: (i) papers considering competition and price-setting among insurers and providers and (ii) papers focused specifically on insurance and market design. The former set of papers is largely concerned with models of oligopolistic competition; it is often focused on the U.S. commercial insurance market where prices are market-determined rather than being set administratively. The latter focuses on insurance market design with an emphasis on issues raised by asymmetric information, leading to adverse selection and moral hazard. In addition, we discuss the literature on consumer choice frictions in this market and the significant implications of those frictions for I.O. questions.

Keywords

Health care markets, Health insurance, Market structure and firm performance, Firm behavior

1 Introduction

The health care sector worldwide is large and growing at a fast pace. In 2018, health care spending in the U.S. amounted to $3.6 trillion, or 17.7% of GDP, an increase of

4.6% compared to the previous year. Each of the major components of the sector—hospital services; physician services; and health insurance—is larger than most other industries in the U.S. economy. Private health insurance spending grew 5.8% in 2018 to $1.24 trillion; hospital expenditures grew 4.5% to $1.19 trillion; while physician and clinical services expenditures grew 4.1% to $726 billion.[1]

The last decade has seen an explosion of papers considering industrial organization questions in health care markets. Rather than importing methods from previous studies of other industries, researchers have developed models that have later been exported to consider similar questions outside health care. The sheer size of the industry helps explain this level of interest by I.O. researchers. Beyond this, the health care market has several distinguishing features which present new challenges for economists and policymakers and which are the subject of the substantive conceptual and empirical literature reviewed in this chapter.

Perhaps the most important feature that separates health care markets from typical product markets is that health care as a product is often viewed as a "right" or moral imperative. This facilitates a large role for the public sector in regulating health care markets, borne out primarily through policies to promote access to care. Examples include the provision of subsidies for insurance, subsidies for care, and market regulations that seek to limit inequities (e.g. limiting price discrimination based on health status). More broadly, the notion that consumers have at least a partial "right" to high-quality health care implies that unfettered free markets are unlikely to succeed in achieving social goals regarding health care provision. Other tools to deal with care rationing through regulation are likely to be needed and, in practice, lead to myriad regulatory institutions that are important to understand when assessing the industrial organization implications of different policies.

There are also several direct market-related issues that arise in health care markets but not in typical product markets. First, health insurance markets must contend with adverse selection: the situation where consumers who are less healthy select into more generous insurance, driving up the price of that insurance and rendering it less accessible for others. Health insurance markets are among the primary markets where adverse selection is an empirical concern and the health literature leverages insights on selection markets more broadly, dating back to Akerlof (1970) and Rothschild and Stiglitz (1976) and summarized in a companion chapter in this handbook by Einav et al. (2021). Adverse selection may be especially acute in health insurance settings relative to other selection markets (e.g. car insurance, credit markets) because limits on price discrimination, imposed due to concerns about equity, enhance the likelihood of such selection.

Second, the health care sector is a classic example of a vertical market structure, where various segments of suppliers operate in the supply chain to deliver and pay for consumer care. Health care providers source medical devices and pharmaceutical products from upstream providers and work across specialties to provide care for

[1] Source: Centers for Medicare and Medicaid Services, https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.

consumers with heterogeneous conditions and needs. While the interaction between health care providers is complex in itself, providers also negotiate with private insurers to determine payments for their services under the threat of exclusion from the insurer's network. Consumers typically only pay a small portion of these negotiated prices out of pocket, and consequently, do not substantially discipline markups via demand responses as they would in a typical market.

In this chapter we outline the tools that have been developed to model and analyze competition and regulation in health care markets, and describe particular papers that apply them to policy-relevant questions. We focus particularly on the IO models and empirical methods and analyses that researchers have formulated to address policy-relevant questions, although we also provide an overview of the institutional facts and findings that inform them.[2] We also assess key opportunities for future research.

We divide the chapter into two broad sections: (i) papers considering competition and price-setting among insurers and providers and (ii) papers focused specifically on insurance and market design. The former set of papers is largely concerned with models of oligopolistic competition; it is often focused on the US commercial insurance market where prices are market-determined rather than being set administratively. We provide an outline of recent papers in this literature that is centered around a model of price negotiations between insurers and providers. For earlier surveys that are less tightly focused on a single framework, see Gaynor and Town (2011) and Gaynor et al. (2015).[3] In addition, in this section we study issues related to (i) price regulation and gaming by medical providers and (ii) vertical integration across different parts of this supply chain.

The second component of the literature focuses on insurance market design with an emphasis on issues raised by asymmetric information (Arrow (1963), Pauly (1968)) leading to adverse selection and moral hazard. In addition, we discuss the literature on consumer choice frictions in this market and the significant implications of those frictions for IO questions. These factors are relevant to almost all health insurance markets: the recent literature often uses data from the health insurance exchanges set up by the Affordable Care Act (see, e.g., Kaiser Family Foundation (2010)) but also studies other regulated exchanges such as Medicare prescription drug markets and Medicare Advantage privatized health plans. In addition, we discuss the industrial organization of employer-sponsored insurance markets, which cover over half of the U.S. population. We also discuss (i) international contexts that rely on regulated markets to provide health insurance and (ii) the interplay between the time-horizon of insurance contracts and efficient provision of coverage.

[2] Readers who are interested in learning more about institutional details and data are directed to other survey papers such as Gaynor and Town (2011) and Gaynor et al. (2015).

[3] There is also a literature considering quality competition when prices are set by regulators, for example in the US fee-for-service Medicare or in the UK National Health Service. A different model is needed here: e.g., the extent to which increasing competition leads to quality improvements depends partly on whether provider fees are greater than their costs. We consider these papers briefly in Section 2 but leave detailed discussion to other authors.

These two broad sections follow the evolution of the literature closely. Few existing papers combine insights and models across the two broad areas we consider, due to the significant complexity of simultaneously considering the key elements across both. We follow the structure of these literatures in this chapter with the understanding that, in many cases, it makes sense to focus on methods and ideas from one part of the literature or the other. With that in mind, we do discuss connections between these two components of the literature throughout and note opportunities for future work that brings them together.

Of course, given the complexity and breadth of the health care sector, we omit some topics that are relevant and interesting. We cover the extensive literature on moral hazard and consumer price sensitivity only to the extent that it is relevant for market competition and regulation. We include a focused and limited discussion of issues related to pharmaceutical markets though, clearly, there are many important and rich IO questions related to advertising, patents, and the complex supply chain in these markets. These topics are covered in more depth in a companion chapter on innovation and I.O. by Bryan and Williams (2021) as well as in a recent handbook chapter by Scott Morton and Kyle (2011). The literature on productivity is covered in a separate chapter by De Loecker and Syverson (2021), while that on matching markets (including kidney exchanges) is in a chapter by Agarwal and Budish (2021). While we discuss recent developments in competition policy in the UK and some aspects of market design in other countries, we largely focus on U.S. health care markets. Finally, the chapter focuses on the formal health care sector and does not discuss other issues that relate to health, such as nutrition, exercise, and addiction, all of which are linked to important I.O. questions.

The rest of this chapter proceeds as follows. Section 2 discusses competition and market function for health care provider and health care service markets, including how those markets relate to and interact with health insurance. Section 3 discusses consumer demand in health insurance markets while Section 4 covers market design and regulation of these markets. Section 5 concludes with a discussion of new promising directions for research.

2 Insurers and providers: competition and antitrust

The recent literature covering provider and commercial insurer competition and price-setting behavior is largely focused on modeling price negotiations in the context of a vertical market. Our objective in this section is to outline these papers and provide a common framework to help readers understand the way the literature has developed over time and the opportunities for future research.

One might reasonably ask why IO economists have focused so heavily on developing empirically relevant models of bilateral bargaining to apply to the U.S. commercial health care market. The reason is clear when we view the market through the lens of a multi-stage model of firm and consumer behavior. It is characterized by bilateral oligopoly: providers are the upstream firms; insurers are downstream; and (with the exception of small physician groups) provider payments are often determined through bilateral negotiations. The outcomes of these price negotiations are critically important for the level of spending and spending growth, which as discussed are very high in the U.S., and for provider incentives to invest in technology or quality. They are also directly relevant for premium setting, and hence also for the demand system that determines the match of consumers to plans. It is difficult to make statements about optimal policy or market design without a good understanding of how prices are determined. A bargaining model must therefore necessarily form the core of any model of the market. We provide the "Nash-in-Nash" model that has been adopted by most papers in this literature and then discuss its implications and limitations in health care and the steps recent authors have taken to extend it.

The timing of the multi-stage model is as follows.[4] In the first stage providers make investments that determine their quality. Second, when the outcomes of these investments have been realized, they negotiate with insurers to determine insurers' provider networks and the prices paid to providers. Third, insurers choose their premiums to maximize their objective functions, taking into account their expected payments to providers, their own characteristics and those of competing insurers. The final two stages account for consumer behavior. Consumers observe each insurer's provider network and other characteristics, including premiums, and choose their plans. Finally, when the enrollment process is complete, some consumers become sick and utilize providers either from within their insurers' networks or (facing a higher out-of-pocket price) from outside.

We discuss the stages of the model in reverse order. Each stage has an impact on the overall equilibrium outcome, and therefore on welfare. Clearly every stage is related to the others: optimal choices in one stage are functions of expectations regarding the rest. The functioning of the market in each stage also has independent policy relevance. For each we provide a model that has been estimated empirically. We adopt the notation from one paper, Ho and Lee (2017), so that the model follows smoothly from demand through price negotiations and premium setting to implications for quality investment incentives. We discuss the different modeling assumptions made by different authors, which are often constrained by data availability, and their relative advantages and disadvantages.[5] We also review issues that have not been factored into the overall framework but where (often model-based) empirical evidence exists that could usefully be accounted for in future iterations.

[4] Our discussion focuses on the commercial health care market. A variant on the model would apply to other US health care markets where private insurers compete for enrollees, such as Medicare Advantage and Medicare Part D. Papers considering these markets are discussed in Section 5.

[5] Different authors have had access to different types of data, and this has directly affected their modeling choices. For example, early papers (Capps et al. (2003), Ho (2006)) utilized hospital discharge data together with (at best) insurance plan market share data covering different markets and time periods. In contrast, Ho and Lee (2017) observed individual household-level choices of health insurance plans, hospital admissions for individuals in those households, and actual prices paid to hospitals. This enabled them to make fewer assumptions when developing and estimating their model.

A note on the structure-conduct-performance approach

This chapter does not discuss in detail the large literature that uses the "structure-conduct-performance" paradigm to study how competition affects health care market outcomes. This approach originated in the broader empirical industrial organization literature in the 1950s through 1970s. The researcher typically conducts a regression analysis where the dependent variable is a market outcome like profit or price and the primary explanatory variable is a measure intended to capture the structure of the market, often the Herfindahl-Hirschman Index (HHI), defined as the sum of squared market shares. Other explanatory variables are also included to account for other (exogenous) reasons for variation in the outcome variable. The objective is to relate market structure to firm performance, with conduct (which is not observed) generating the estimated relationship between these two variables.

Berry et al. (2019) explain that the field of industrial organization moved away from the structure-conduct-performance approach in the late 1980s for several reasons (Bresnahan (1989), Schmalensee (1989)), although it has persisted longer in health economics than elsewhere. We summarize their explanation here. The first issue is measurement: measuring concentration is inherently difficult because, as discussed below in the health care context, markets are not observed directly in the data. Economic profits are difficult to measure accurately, and in health care, even prices should presumably be adjusted to account for differences in health status severity across consumers. The second issue is endogeneity of the concentration variable, which (as a function of market shares) is likely to be correlated with unobserved demand shifters that also affect price. Instruments are difficult to find because, by definition, the regression of prices on concentration depends on all elements of demand and marginal costs. The most fundamental problem, however, is that different changes in primitives can produce the same observed correlations between concentration and prices, with very different implications for welfare. That is, there is no causal relationship between the two variables. A positive observed correlation between market concentration and prices, for example, could be caused by firm mergers that jointly generate higher concentration, higher markups, higher prices and lower consumer surplus. However, the correlation could also be observed if large firms have low marginal costs but high fixed costs, since in that case firm size (and industry concentration) may be associated with high markups, in part in order to recover the fixed costs (Demsetz (1973)). A correlation is also possible, in a differentiated products setting, if a reduction in search costs shifts market share towards firms with high quality products, increasing concentration and also consumer welfare (Autor et al. (2020)). Thus the results of these types of regression analyses are at best suggestive. As Bresnahan (1989) argued, we cannot clearly consider the impact of concentration unless we focus on three objects: a demand system; the marginal costs of firms; and the nature of oligopolistic competition. The following sections investigate these objects, and their combination, in the setting of commercial health insurance.

2.1 Facts about insurer and provider prices and competition

We provide some brief facts about the industry to motivate our discussion of the literature on consumer choices, price and network negotiations, and the impacts of consolidation in both upstream and downstream markets.

2.1.1 Insurer concentration and premium variation

US commercial health insurance markets are concentrated and becoming more concentrated over time. Dafny (2015) uses public data sources to construct estimates of the national market share of the four largest insurers in the commercial market over the period 2006-2014.[6] The four-firm concentration ratio increased from 74% to 83% over that time. By comparison, Dafny cites the equivalent figure for the airline industry as 62%.

Dafny (2010) accesses new data from a benefits consulting firm on the plans purchased and premiums paid by 776 large employers over the period 1998-2005. These data are not necessarily complete by market or nationally representative, but they represent one of the most complete and extensive datasets containing prices and quantities for the insurance market. Dafny examines the effect of shocks to employer profitability on changes in the insurance premiums they pay. The idea is that if insurers possess no market power then the premiums they charge will not vary with employer profitability. Only if insurers have market power will they be able to price discriminate on this basis. Dafny finds strong evidence that premiums increase with the buyer's profitability. This effect decreases in magnitude with the number of insurers in the market, consistent with insurer market power falling with the number of firms.

2.1.2 Insurer network differentiation

Insurers differentiate themselves on a number of dimensions other than price. These include wellness programs; their patient-facing portals; and the extent to which they share data with physicians.[7] The breadth of plans' provider networks is arguably one of the most important sources of differentiation because it determines consumer access to hospitals and physicians in their local markets and also has a potentially meaningful effect on health care spending. Restricted network plans have been used as a means to control health insurer costs since the growth of managed care plans in the 1980s. The degree of selectiveness differs over time and across markets. In a sample of 43 major US markets in 2003, Ho (2006) found that 85% of hospital-HMO pairs in the commercial market agreed on a contract (i.e., on average, only 15% of hospitals were excluded). In contrast, Dafny et al. (2015) document that only 57% of potential links were formed by HMO plans on the Texas exchange. The 2017

[6] These data include privately insured lives across individual, small employer group, and large employer group markets. Market conditions (including barriers to entry and concentration) differ across these markets.

[7] Personal relationships among insurance sales representatives, brokers, and employer benefits managers may also be important in determining consumers' plan choices.

Employer Health Benefits Survey, released by the Kaiser Family Foundation, suggests that networks are increasingly narrow in the employer-sponsored market too. For 15% of employers offering health insurance, the largest health plan offered had high performance or tiered networks that provided financial or other incentives for enrollees to use particular hospitals. Nine percent offered a plan considered to be a narrow network plan.

2.1.3 Provider price dispersion and provider consolidation

Cooper et al. (2019) uses a large dataset of medical claims, covering 28% of individuals with employer-sponsored insurance in 2008-11, to document the substantial variation in spending and prices across the US. Health spending per privately insured consumer differs by a factor of three across geographic areas, with half of the variation coming from provider price variation.[8] Prices for any particular service vary substantially at every level: across regions; across hospitals within a region; and across patients within a hospital. Market structure is highly correlated with prices: monopoly hospitals' prices are 12% higher than those in markets with at least 4 hospitals, and prices increase after mergers of nearby hospitals. This is consistent with a fairly substantial literature which finds that commercial prices tend to increase following mergers of hospitals in the same geographic and product market without a significant quality improvement.[9]

Craig et al. (2021) consider a different cut of the data. They quantify the extent of variation in negotiated hospital prices between insurers for a given hospital in a dataset covering medical claims in Massachusetts. They show that between-payer price variation is similar in magnitude to between-provider variation. Prices are higher for insurance plans that are not fully-insured (i.e., for Administrative-Service Only, or self-insured, plans).

Clemens and Gottlieb (2017) look at physician (rather than hospital) payment structures in commercial health care markets. Given the small size of many physician practices, one might expect that commercial prices might be determined at least partly based on federally administered Medicare rates (which formally apply only to Medicare-eligible patients aged over 65). The authors use a novel identification strategy to investigate this idea. They construct a link between Medicare and private sector claims data and exploit a one-time sharp reduction of Medicare's payments for surgical procedures relative to non-surgical procedures. They find that a \$1 decrease in Medicare's payment for a surgical procedure results in a \$1.16 decline in commercial payments for that service, within a year of the Medicare change. Medicare's influence is particularly strong in areas with concentrated insurance markets and competitive provider groups. Note that these findings do not rule out bilateral

[8] This mirrors findings for Medicare spending from the Dartmouth Atlas, but that variation is due almost entirely to quantities because prices are set centrally.

[9] See Dranove and White (1994), Dafny (2009), Haas-Wilson and Garmon (2011), Farrell et al. (2011), Gaynor and Town (2011), among others.

price negotiations between insurers and physician groups. Some of the estimated relationship may be mechanical, as many contracts explicitly define prices as a percent of Medicare rates. When contracts are renegotiated, the relation may change, but analyzing this requires examining prices over multiple years, when other factors may be at play. The authors interpret their findings through the lens of a bargaining model where Medicare payment rates are an important part of physicians' outside options, and an increase in these rates enables them to negotiate higher commercial prices. A second paper, Clemens et al. (2017), uses a different dataset to consider these ideas further. The authors ask whether private insurers' payments to physicians mirror Medicare rates precisely, or whether they depart from this benchmark. They find that prices for 25% of physician services, representing almost half of spending, differ from Medicare rates, and that departures are particularly likely for large physician groups and capital-intensive care where the gains at stake are relatively high.

Overall, this literature provides evidence of substantial dispersion in the prices paid by commercial insurers to both physician groups and hospitals, for a given service, across insurers, regions, and providers. A model of consumer demand and price negotiations is needed to predict the likely effects of market regulation and antitrust policy for both types of providers.

2.2 Demand: consumer choice of providers

2.2.1 Discrete choice models: consumer demand for hospitals

Three early papers use estimated discrete choice models of consumer demand for hospitals to assess the likely price effects of horizontal mergers between hospitals: Town and Vistnes (2001), Capps et al. (2003), and Gaynor and Vogt (2003). We describe Capps et al. (2003) in detail and then return to summarize the others. At the time these papers were written, the market definition used by the courts—a key input into merger analysis—was determined by the Elzinga-Hogarty approach. The idea was to use aggregate patient inflows and outflows to define market boundaries, expanding the market size until two criteria were satisfied: little (few patients traveling) out from the inside and little in from the outside. The authors argued that, while this approach made sense for the homogeneous product markets for which it was designed, it tended to over-estimate the size of the market for hospitals and hence under-estimate the price that a merged entity was likely to support.

There are two reasons why the method is problematic for use in hospital markets. First, hospitals provide differentiated products to heterogeneous consumers whose needs differ substantially in ways that affect market size calculations. For example, some patients with certain severe conditions are willing to travel large distances for care: if we assume that this is true for all patients, we will define large geographic markets for hospitals, implying permissive merger rulings. However, the willingness of *some* patients to travel does not eliminate the market power hospitals have in their local neighborhoods, because many patients with less serious conditions have a strong preference for access to nearby providers. The second issue is the need to account for the vertical nature of the health care market: the fact that commercial insurance carriers act as intermediaries between hospitals and their patients. Since insurers bargain with providers over reimbursement rates, and then bundle these providers' services and market their plans to consumers before diagnoses are known, the data on ex post hospital utilization may not accurately reflect insurer pricing incentives. In particular, we need to account for the fact that consumers might place a high ex ante value on access to particular hospitals when choosing their plans—implying an incentive for the insurer to include those hospitals in the network, even at a high negotiated rate—even if they turn out not to utilize these hospitals in the following year. This ex ante nature of pricing, together with the heterogeneity of patient diagnoses and preferences, needs to be accounted for in any model that predicts the price effects of hospital mergers.

Capps et al. (2003) use a discrete choice model of consumer demand for hospitals that conditions on diagnosis and other observables to predict ex ante consumer willingness-to-pay for the hospital to be included in the network. This willingness-to-pay is envisioned as a measure of market power that, unlike patient flows, accounts for the differentiated goods market and is directly related to the prices that hospitals are able to negotiate. They use a multinomial logit demand framework for estimation.

Suppose consumers are grouped into age-gender categories (or "types") κ; each type requires admission to a hospital with probability γ_κ^a. Conditional on admission, the individual receives one of six diagnoses $l \in \mathcal{L} \equiv$ {cardiac, cancer, neurological, digestive, labor, other} with probability $\gamma_{\kappa,l}$. Assume that individuals can only visit a hospital in their market m and insurer's network, and individual k of type $\kappa(k)$ with diagnosis l derives the following utility from hospital i:

$$u_{k,i,l,m}^H = \delta_i + z_i \upsilon_{k,l} \beta^z + d_{i,k} \beta_m^d + \varepsilon_{k,i,l,m}^H \tag{1}$$

where δ_i are hospital fixed effects, z_i are observed hospital characteristics (such as indicators for teaching hospitals and for-profit hospitals, the number of beds, and indicators for hospitals providing particular services), $\upsilon_{k,l}$ are characteristics of the consumer (such as diagnosis or income), $d_{i,k}$ represents the distance between hospital i and individual k's zip code of residence (and has a market-specific coefficient), and $\varepsilon_{k,i,l,m}^H$ is an idiosyncratic error term assumed to be i.i.d. Type 1 extreme value. There is no outside option since the data include only patients who are sick enough to go to a hospital for a particular diagnosis. The authors estimate the parameters of this equation via maximum likelihood using their hospital discharge data.

Specification differences across papers

Eq. (1) follows the specification in Ho and Lee (2017). Capps et al. (2003) use a slightly different specification. For example, they include observed hospital characteristics rather than hospital fixed effects, and travel time (and its square) rather than distance. Further, while Ho and Lee (2017) observe the network of each insurer and can therefore accurately specify the choice set of each patient, assuming that the enrollee can choose any hospital in his local market that is included in his insurer's network, Capps et al. (2003) do not observe networks and therefore have to make a different assumption. They estimate the utility equation for patients enrolled in Medicare or indemnity insurance, who have unrestricted hospital choice sets, and proceed by assuming that managed care enrollees (i.e., the consumers of interest for their merger analysis) have the same preferences conditional on observable characteristics.

The out-of-pocket price is not included in Eq. (1). Capps et al. (2003) include price in their initial utility equation but later note that prices are unobserved and (at least for consumers enrolled in managed care plans) likely to differ little across hospitals, and hence exclude it. The assumption that negotiated hospital prices do not influence individuals' allocations to hospitals could also potentially be justified by arguing that negotiated prices are difficult to observe, even for patients who face coinsurance payments, and that insurers have few other levers to steer patients to cheaper hospitals. However, as discussed in Ho and Pakes (2014) and Gowrisankaran et al. (2015), this abstracts away from both patient and referring physician price sensitivity, both of which may be important, and the implications of the model for prices would change if it was relaxed. We return to these issues below.

Willingness-to-pay (WTP)

The model predicts the probability that an individual k—who lives in market m, is enrolled in insurance plan j, and has diagnosis l—visits hospital i. In addition, the estimates are used to construct a measure of consumers' ex-ante expected willingness-to-pay (WTP) for the insurer's network. This object, which varies by age and gender, is the focus of Capps et al. (2003); it will also be included as a plan characteristic in the model of consumer demand for insurance plans below. Given the assumption on the distribution of $\varepsilon^H_{k,i,l,m}$, individual k's WTP for the hospital network offered by plan j is

$$WTP_{k,j,m}(\mathcal{G}_{j,m}) = \gamma^a_{\kappa(k)} \sum_{l \in \mathcal{L}} \gamma_{\kappa(k),l} \underbrace{\log\left(\sum_{h \in \mathcal{G}_{j,m}} \exp(\hat{\delta}_h + z_h \upsilon_{k,l} \hat{\beta}^z + d_{h,k} \hat{\beta}^d_m) \right)}_{EU_{k,j,l,m}(G_{j,m})},$$

where the expression is a weighted sum across diagnoses of the expected utility of a hospital network conditional on a given diagnosis ($EU_{k,j,l,m}(G_{j,m})$), scaled by the probability of admission to any hospital.[10]

Capps et al. (2003) use the estimated parameters of their utility equation to predict the change in $WTP_{k,j,m}$ when hospital i is added to the insurer's network. Integrating over the distribution of consumer locations, demographics, and likely clinical indications, summing over consumers in the market, and renormalizing by an (assumed

[10] $EU_{k,j,l,m}(G_{j,m})$ is the expected value of the maximum of $\{u^H_{k,i,l,m}\}$ across all hospitals in $G_{j,m}$ before the realization of the (demeaned) error terms $\{\varepsilon^H_{k,i,d}\}$.

constant) price coefficient, they obtain a \$-valued variable $\Delta WTP_{i,j,m}$ that summarizes the population's ex ante willingness-to-pay to include hospital i in the network offered by insurer j.[11]

Finally, the authors consider a simple myopic version of a Nash bargaining model. It suggests that the insurer should include hospital i in its network whenever the population $\Delta WTP_{i,j,m}$ exceeds the cost of including it $\Delta C_{i,j,m}$. Further, if every hospital captures a proportion α of the surplus it generates, then the contribution the hospital earns above its variable costs from the interaction with managed care insurer j is:

$$\pi_{i,j,m} = \alpha(\Delta WTP_{i,j,m} - \Delta C_{i,j,m}) + u_{i,j,m}$$

The authors point out that, while this simple framework ignores potentially interesting issues (discussed below) that are raised by more detailed models of bargaining in vertical oligopoly, it has the advantage of suggesting a simple regression framework for estimating α. They make some further modifications to account for the lack of data on hospital costs before estimating a positive, statistically significant relationship between observed hospital profits and their ΔWTP measure. They use the estimates to demonstrate the model's prediction that hospital profits, and hence (under assumptions about price-cost margins) prices could change substantially following hypothetical mergers between particular pairs of hospitals in San Diego, even in cases where the Elzinga-Hogarty approach would imply very large hospital markets and little or no merger effect.

Related papers

Recall that two other early papers address a similar question to Capps et al. (2003) in the context of an estimated model. Town and Vistnes (2001) is an earlier paper that develops a model of a "switching regime" where the insurer has two options if it fails to reach agreement with a given hospital: either drop it without replacement, or replace it with the next-best alternative. The authors predict enrollees' valuations for having access to different networks of hospitals using an estimated utility equation for inpatient hospitals. They then regress the log price on the log of the expected utilities from the two alternative networks. The estimates indicate that hospital prices decrease when the alternative networks available to the insurer are attractive. Hospital mergers, which reduce the number of alternative networks, can have substantial effects on prices. Gaynor and Vogt (2003) focus on the effect of ownership type (non-profit versus for-profit) on firm behavior. They estimate a model of demand and pricing in the hospital industry, allowing the behavior for for-profit and not-for-profit firms to differ, and find that not-for-profit hospitals face less elastic demand; have lower marginal costs; and higher markups (but lower prices) than for-profit hospi-

[11] Since the authors do not observe the network offered by each insurer, they define $\Delta WTP_{i,j,m}$ as the population willingness-to-pay to add the hospital to the otherwise-complete network (i.e., this does not differ across insurers).

tals. Simulated mergers in relatively concentrated markets lead to substantial price increases for both non-profit and for-profit hospitals.

Impact on antitrust policy

Overall, the models developed in these papers generate the empirically relevant prediction that prices may increase in response to horizontal mergers between hospitals. Merger price effects are predicted when the merging hospitals are substitutes at the point of service (for a particular patient). Together with a hospital merger retrospectives project launched by the Federal Trade Commission in 2002, which generated evidence that past mergers had resulted in hospital price increases, these papers have had a substantial effect on antitrust merger policy. The FTC now relies on a methodology very similar to those in Town and Vistnes (2001) and Capps et al. (2003) to predict the "direct effects" of mergers. The proportion of proposed horizontal mergers that were abandoned, or prevented after FTC investigation, has increased dramatically since this policy change.[12]

Given the influence of this literature on antitrust policy, it is important to assess the predictive relevance of discrete choice models in the hospital setting and their usefulness in predicting price effects of mergers. A few recent papers contribute to this process of assessment. Garmon (2017) compares the results of several potential merger screening methods to the actual post-merger price changes after 28 hospital mergers measured relative to controls. He finds that willingness-to-pay based methods more accurately flag potentially anticompetitive mergers for further investigation than previous methods based on Hirschman-Herfindahl Indices, but there is heterogeneity in their predictive ability. Raval et al. (2020) use a series of natural disasters that unexpectedly removed hospitals from patients' choice sets to evaluate the ability of discrete choice models to predict patient substitution to other hospitals post-disaster. Most models considered in the paper under-predict large substitutions, but there is a significant improvement using models that include a random coefficient on distance. This line of research is clearly very useful for practitioners as well as researchers.

2.2.2 Physician - patient interactions

The recent I.O. literature on health care markets has focused on hospital rather than physician services, abstracting away from patients' choices of primary care providers or specialists and treating physician and patient as a single "aggregate" decision-maker rather than modeling physicians' hospital referral decisions. In reality, physicians may exploit their informational advantage to refer patients to a specific hospital, thereby influencing choice. In that case, a model which accounted for patient choices of physicians, and for physician incentives, would have different implications for the impact of policy on patient choice and hence incentives to change prices or quality.

[12] However, the Capps et al. (2003) simple model predicts that there is no price effect of a merger between hospitals that are not direct substitutes at the point of care. We return to this issue below.

Ho and Pakes (2014) investigate the response of physicians' hospital referrals to the cost-control incentives imposed by managed care insurers. Commercial health insurers in California often pay large physician groups through global capitation contracts where the physician group is paid a fixed amount per patient to cover all treatment costs (including hospital costs). An alternative is professional services capitation, usually including a "shared risk" arrangement where the physician group receives a share of the savings made relative to some pre-agreed benchmark, again including hospital costs. The authors consider birth episodes. They estimate a utility equation for each insurer that summarizes the preferences implied by hospital referrals. They use a partially identified framework that addresses both measurement error in expected price and the fact that the price for a particular patient is likely to be correlated with the unobserved hospital quality for that severity. The estimates indicate that physicians' price responsiveness, when referring patients to hospitals, increases significantly with the capitation rate of insurer payments. The price coefficients are also much more negative than those from a more standard multinomial logit analysis which does not address either of the two estimation issues. Finally the authors show that, while the trade-off between distance and price differs substantially across insurers, the trade-off between distance and quality does not. That is, the estimates suggest that highly capitated, more price-sensitive plans tend to send their patients longer distances to access similar-quality services at a lower price. Consistent with this, severity-adjusted outcomes also do not differ significantly across plans. These findings are in line with Song et al. (2011) which studies the Alternative Quality Contract implemented by Blue Cross Blue Shield of Massachusetts in 2009 and finds that physicians responded to five-year global budget contracts by referring patients to lower-priced outpatient providers with no reduction in quality of care.

2.2.3 Other related issues

Several recent papers investigate other policy-relevant issues that affect consumer demand for providers and that the supply side (pricing) may respond to. While not always embedded in a model of provider pricing, they may be important inputs into future iterations of such a model.[13]

Price transparency

The wide variation in the price of health care procedures across providers is often difficult for consumers to observe. Various recent price transparency initiatives at the employer, insurer and state level have attempted to improve consumer price-shopping, steer patients to lower-priced options, and thereby exert downwards pressure on provider prices. The evidence on the efficacy of these initiatives is somewhat mixed. Several papers (Lieber (2017), Whaley (2015), Desai et al. (2016)) consider the effects of initiatives by individual employers or insurers, finding small or no

[13] Narrow and tiered networks are also relevant here; see Section 2.4 for an outline of these papers and the issues they raise.

effects on shopping behavior. The first papers to assess the impact of a price transparency initiative that was broadly available to all individuals in a state are Brown (2019) and Brown (2020). In March 2007 the state of New Hampshire introduced a website that allowed individuals to access information on their insurer-specific out-of-pocket prices for a subset of medical procedures. Since this intervention was market-wide, it had the potential to generate substantial supply-side effects on pricing in response to any demand-side effects from changes in shopping behavior. Brown (2019) uses difference-in-differences analyses that utilize two sources of variation: the timing of introduction of the website and variation in procedures included. Results for outpatient medical imaging services indicate that patient demand responded to price information, particularly for patients who had not yet met their deductible, and that prices fell by a modest amount (2% on average) in response to this. Brown (2020) develops an empirical model of demand and supply for medical imaging services that explicitly accounts for consumer uncertainty about prices. Consumers with rational expectations receive noisy price signals, form beliefs based on these signals, and then make decisions consistent with these beliefs. The introduction of the website is assumed to make a subset of consumers fully informed about prices; the proportion of such consumers is estimated as part of the model. Counterfactual simulations suggest that price transparency makes residual demand more price elastic and this can decrease the incentive for providers to negotiate high prices. (There is also an offsetting effect because price transparency makes consumers less likely to choose high-priced providers, reducing the incentive for insurers to refuse to include such providers in their networks, but this is dominated by the first effect in the estimated model). Overall, if all consumers had price information, prices could fall in equilibrium by as much as 20%.[14] In related ongoing work, Alcott et al. (2021) study a randomized-controlled trial in New York that provides information about negotiated prices to both consumers and providers. They find limited evidence that consumers attend to the information and discipline prices, but some evidence that providers use this information to benchmark their own negotiated rates, in some cases leading to increased prices.

Surprise billing

Cooper and Scott-Morton (2016) documents the prevalence of "surprise" out-of-network billing, under which consumers receiving care at an in-network hospital are nevertheless treated by, and receive unexpected bills from, out-of-network physicians. Over one in five patients who went to in-network emergency rooms in the authors' data were treated by out-of-network emergency physicians. This is possible because physicians in the US often negotiate contracts with insurers independently from the hospitals in which they practice. That is, particular institutional features

[14] There has been some debate among researchers about the possibility that price transparency could also affect firm price negotiations directly, e.g. by facilitating collusion (Cutler and Dafny, 2011), or alternatively, resolving uncertainty about other firm bargaining weights or costs. This is discussed further in the section on negotiations for medical devices below.

in this market generate a profitable loophole that some providers choose to exploit. Cooper et al. argue that, in addition to harming individual patients who receive unexpectedly large bills for their care in the short run, surprise billing also creates frictions that impede the long-run functioning of price competition in the commercial market. The ability to charge patients on an individual (unexpected) basis enables some physicians to choose to go out-of-network, thereby bypassing the downwards pressure on prices that is generated through negotiations with insurers. This is particularly feasible for physicians in emergency departments or other specialties where their services are part of a wider bundle of hospital care and cannot be avoided once the hospital choice is made. Alternatively, such physicians may use the threat of going out-of-network to negotiate increased in-network payment rates. The paper assesses the impact of a New York State law that introduced binding arbitration between emergency physicians and insurers, hence reducing physician threat points, and shows that it significantly reduced both the practice of out-of-network billing and the level of in-network emergency physician payments.

2.3 Demand: consumer choices in insurance markets

The next component of the model considers consumer choice of health insurance plans. We follow the literature by specifying the indirect utility from an insurance plan as a function of consumer and plan attributes, as in a traditional discrete choice analysis, rather than using the theory of expected utility under uncertainty to build a demand model based on individual risk preferences and exposure to risk. Each approach has advantages relative to the other. The framework considered here requires fewer assumptions but prevents the researcher from estimating risk aversion and other parameters that may be useful to answer some questions. See Section 4 for the literature that considers those issues.[15]

Ho (2006) was the first paper to model consumer demand for insurers in a way that explicitly accounted for the value of the hospital network offered by each plan. Ho and Lee (2017)'s more comprehensive data allow them to estimate a more detailed utility equation. Suppose each household chooses among the insurance plans in the choice set, taking all household members' hospital preferences into account. We can write the utility a household or family f receives from choosing insurance plan j in market m as

$$u^M_{f,j,m} = \delta_{j,m} + \alpha^{\phi}_f \phi_j + \sum_{\forall \kappa} \alpha^W_\kappa \sum_{k \in f, \kappa(k)=\kappa} WTP_{k,j,m} + \varepsilon^M_{f,j,m}\,, \tag{2}$$

where $\delta_{j,m}$ is an insurer-market fixed effect that controls for physician networks, brand effects, and other insurer characteristics and ϕ_j is the premium faced by the household (in reality this may vary across family sizes and is scaled to reflect

[15] Einav et al. (2010a) provide further discussion of the relative advantages and disadvantages of these two approaches to modeling demand for insurance.

the fact that the employer may make a contribution towards premiums). The term $\sum_{\forall\kappa}\alpha_\kappa^W\sum_{k\in f,\kappa(k)=\kappa}WTP_{k,j,m}(\cdot)$ accounts for the household's WTP for the insurer's hospital network by summing over the value of $WTP_{k,j,m}$ for each member of the household multiplied by an age-sex category specific coefficient, α_κ^W. Finally ε_{fjm}^M is a Type 1 extreme value error term. The parameters are estimated under the assumption that each household chooses the insurance plan that maximizes its expected utility among those available to it. This specification is consistent with households choosing an insurance product prior to the realization of their health shocks and aggregating the preferences of members when making the plan decision.

The estimated coefficients on WTP in the health plan utility equation are positive and significant in both Ho (2006) and Ho and Lee (2017). Ho (2006) uses the estimates to quantify the welfare loss from restricted hospital choice due to exclusion from insurers' networks. The loss to society is approximately \$16 per consumer, or \$1 billion per year over 43 major US markets studied, holding prices fixed. The author notes that this may be substantially outweighed by the price and premium reductions resulting from insurers' ability to threaten hospital exclusion during price negotiations with providers.

Specification differences across papers

The detailed data available to Ho and Lee (2017) enable them to make relatively weak assumptions for estimation. In particular, they have access to household level plan choice data rather than aggregate market shares, and thus can model plan choices at the household level rather than the individual level and permit household characteristics (such as the number of children and household income) to affect choices. In addition, they observe—and condition on—the identities of the insurers in each household's choice set. In contrast, Ho (2006) uses different datasets to estimate hospital versus insurer demand, and does not see family composition. She therefore models both insurer and hospital demand at the individual consumer (rather than household) level. The insurance plan utility equation is estimated from aggregate market share data. Individuals' employer-based choice sets are not observed, so the estimates are interpreted as reflecting the joint preferences of the employers choosing the choice sets and the employees choosing their plans.

Ho and Lee (2017) do not explicitly model household responsiveness to deductibles, copays, or coinsurance rates. Accounting for them would add complexity to the pricing and premium-setting equations set out below. They note that, as long as the financial generosity of plans (outside of premiums) does not vary when an insurer is added to or removed from a market, the impact of deductibles and copays will be absorbed into plan-market fixed effects and not affect the analysis. However, this simplification abstracts away from possible consumer sensitivity to hospital prices through out-of-pocket payments, an issue assessed in Gowrisankaran et al. (2015) (Section 2.4 below) among others.

2.4 Supply: insurer-hospital price negotiations and premium setting

Provider payments from commercial health insurers are determined by bilateral bargaining between insurers and providers. We now outline a model of these negotiations, developed in Ho and Lee (2017), that considers pricing at two levels of a vertical market.[16] A demand system like the one outlined above is an important model input. While the model is necessarily somewhat involved, it (or something similar) is an essential tool to consider a number of policy-relevant questions. Ho and Lee (2017) use their estimated model to examine the impact of insurer competition on premiums, prices, and welfare in the U.S. commercial health care industry. The model captures the standard intuition that reduced insurer competition may increase health expenditures by raising insurer premiums and, with them, the payments made to medical providers. Offsetting this, a more concentrated insurance sector can also strengthen insurers' bargaining leverage when negotiating with medical providers. This additional effect—a variant of countervailing power (Galbraith, 1952)—can mean that greater downstream concentration can reduce total hospital payments. The overall effect is theoretically ambiguous. The model makes clear that the effects of reduced competition on premiums and payments are closely linked; any overall assessment of the impact of insurer competition needs to consider them together as part of the same model.

2.4.1 A model of price and premium negotiations

Ho and Lee (2017) assume that commercial insurers engage in simultaneous bilateral Nash bargaining ("Nash-in-Nash" bargaining) over premiums with a large employer, and (at the same time) in simultaneous bilateral Nash bargaining with hospitals over prices in each market. This bargaining protocol, which assumes that each bilateral negotiation maximizes bilateral Nash products (taking the outcomes of all other bargains as given), was proposed in Horn and Wolinsky (1988). It has been used in applied work to model upstream-downstream negotiations over input prices in oligopolistic vertical markets (e.g., Crawford and Yurukoglu (2012), Grennan (2013), Gowrisankaran et al. (2015)). See the companion chapter on vertical markets by Lee, Whinston and Yurukoglu for further details on the model.

More formally: consider the set of insurers (or MCOs) $\mathcal{M}$ that are offered by an employer, and a group of markets that contain a set of hospitals $\mathcal{H}$. Let the network of hospitals and MCOs be represented by $\mathcal{G} \subseteq \{0,1\}^{|\mathcal{H}|\times|\mathcal{M}|}$, where $ij \in \mathcal{G}$ denotes that hospital i is present in MCO j's network. Assume that an enrollee in MCO $j \in \mathcal{M}$ can only visit hospitals in j's network, $\mathcal{G}^M_j$; similarly, $\mathcal{G}^H_i$ denotes the set of insurers that have contracted with (and are allowed to send patients to) hospital i.

[16] Papers which consider other health insurance contexts, like Medicare where prices are set administratively, necessarily rely on different models of supply that fit their particular settings.

Take the network $\mathcal{G}$ as given, and assume the following timing, which follows the overall framework described in this section:

1a. The employer and the set of MCOs bargain over premiums $\boldsymbol{\phi} \equiv \{\phi_j\}_{j \in \mathcal{M}}$, where ϕ_j represents the per-household premium charged by MCO j.

1b. Simultaneously with premium bargaining, all MCOs and hospitals $ij \in \mathcal{G}$ bargain to determine hospital prices $\boldsymbol{p} \equiv \{p_{ij}\}$, where p_{ij} denotes the price paid to hospital i by MCO j for treating one of j's patients.

2. Given hospital networks and premiums, households choose to enroll in an MCO, determining household demand for MCO j in market m, denoted by $D_{j,m}(\mathcal{G}, \boldsymbol{\phi})$, and the number of enrollees $D^E_{j,m}(\mathcal{G}, \boldsymbol{\phi})$ (Section 2.3 above).

3. After enrolling in a plan, each individual becomes sick with some probability; those that are sick visit some hospital in their network. This determines $D^H_{ijm}(\mathcal{G}, \boldsymbol{\phi})$, the number of individuals who visit each hospital i through each MCO j in market m (as in Section 2.2).

The assumption that premiums and hospital prices are simultaneously determined, rather than sequential, simplifies computation and estimation of the model because prices can be treated as fixed when evaluating payoffs from premium bargaining and premiums can be held fixed when considering both agreement and disagreement payoffs from hospital price bargaining (see the equations below). While the assumption implies some theoretical restrictions on the relationship between prices and premiums, it does not mean that premiums are independent of negotiated hospital prices in equilibrium. There is a relationship between the two, which is influenced by consumer demand and other factors, because the two variables are constrained to be "optimal" (i.e., maximize their bilateral Nash products) with respect to each other in equilibrium. Thus, for example, an increase in premiums due to insurer mergers can have a positive effect on hospital prices in equilibrium (intuitively, as the "larger pie" generated by increased concentration is shared between insurers and hospitals) and, since hospital prices are assumed to be linear, double marginalization will also be present.[17] Note also that, while not entirely realistic, the simultaneity assumption may not be unreasonable in this setting. Since prices and premiums are set at staggered intervals and are fixed for different period lengths in reality, the obvious alternative timing assumption—that premiums immediately adjust to changes in negotiated prices—may not be more accurate.

[17] However, the simultaneity assumption does rule out some vertical incentives. For example, when considering raising prices to an insurer, a hospital does not internalize the effect of its price increase on downstream premiums. Further, the assumption implies that premiums do not change upon disagreement (off the equilibrium path). The effect of these issues may be limited in Ho and Lee (2017) because hospitals are small contributors to overall premiums that are set at the state level. They may, however, be important in other settings.

Profit equations

MCOs and hospitals seek to maximize profits when bargaining over negotiated prices and premiums. Assume that the profits for MCO j are given by

$$\pi_j^M(\mathcal{G}, \boldsymbol{p}, \phi_j) = \sum_m \left(\phi_j \boldsymbol{D}_{j,m}(\cdot) - D_{j,m}^E(\cdot)\eta_j - \sum_{h \in \mathcal{G}_{j,m}^M} D_{h,j,m}^H(\cdot) p_{h,j} \right) . \tag{3}$$

The first term on the right-hand side of (3) represents total premium revenues obtained by MCO j. The second term is non-inpatient hospital costs, such as payments to primary care physicians, while the third represents expected payments made to hospitals in MCO j's network for inpatient services.

The corresponding expression for the profits for a hospital i is:

$$\pi_i^H(\mathcal{G}, \boldsymbol{p}, \boldsymbol{\phi}) = \sum_{n \in \mathcal{G}_i^H} D_{in}^H(\cdot)(p_{in} - c_i) , \tag{4}$$

which sums, over all MCOs n with which hospital i contracts, the number of patients it receives multiplied by an average margin per admission (where c_i is hospital i's average cost per admission for a patient).

All demand-related terms in either MCO or hospital profits, ($D_{j,m}(\cdot)$, $D_{j,m}^E(\cdot)$, $D_{ijm}^H(\cdot)$), are determined by the two-stage demand model described in the previous subsection. Any components of MCOs' or hospitals' profits that do not vary with the network, including fixed costs, do not affect the subsequent analysis and are omitted.

Employer-insurer bargaining over premiums

Assume that premiums for each MCO are negotiated with the employer via simultaneous bilateral Nash bargaining, where the employer maximizes its employees' welfare minus its total premium payments. This assumption nests the standard Nash-Bertrand model of premium setting. Negotiated premiums ϕ_j for each MCO j thus satisfy

$$\begin{aligned} \phi_j = \arg\max_{\phi_j} [\underbrace{\pi_j^M(\mathcal{G}, \boldsymbol{p}, \{\phi, \boldsymbol{\phi}_{-j}\})}_{GFT_j^M}]^{\tau^\phi} \\ \times [\underbrace{W(\mathcal{M}, \{\phi, \boldsymbol{\phi}_{-j}\}) - W(\mathcal{M} \setminus j, \boldsymbol{\phi}_{-j})}_{GFT_j^E}]^{(1-\tau^\phi)} \forall j \in \mathcal{M} , \end{aligned} \tag{5}$$

(where $\boldsymbol{\phi}_{-j} \equiv \boldsymbol{\phi} \setminus \phi_j$) subject to the constraints that the terms $GFT_j^M \geq 0$ and $GFT_j^E \geq 0$. These terms represent MCO j's and the employer's "gains-from-trade" (GFT) from coming to agreement and having MCO j in the employer's choice set. The MCO's gains-from-trade are its profits from being offered by the employer; its disagreement outcome is assumed to be 0. The employer's gains-from-trade are the difference between its objective $W(\cdot)$ when MCO j is and is not offered. Ho and Lee model $W(\cdot)$ as the employer's total employee welfare net of its premium payments to insurers.

Given the timing assumptions, outside options from disagreement are determined by removing MCO j from the employer's choice set, holding fixed premiums and negotiated hospital prices for other MCOs, but allowing employees to choose new insurance plans (but not switch employers).

The Nash bargaining parameter for this negotiation is $\tau^{\phi} \in [0, 1]$, where $\tau^{\phi} = 1$ implies that MCOs simultaneously set profit-maximizing premiums (i.e., compete via Nash-Bertrand), and $\tau^{\phi} = 0$ implies that the employer pays each MCO only enough to cover its costs.

Insurer-hospital bargaining over hospital prices

As with premiums, hospital prices $\boldsymbol{p}$ are determined via simultaneous bilateral Nash bargaining.[18] Each negotiated price per-admission $p_{ij} \in \boldsymbol{p}$ between hospital $i \in \mathcal{H}$ and MCO $j \in \mathcal{M}$ (for all $ij \in \mathcal{G}$) maximizes the pair's bilateral Nash product:

$$p_{ij} = \arg\max_{p_{ij}} \left[\pi_j^M(\mathcal{G}, \boldsymbol{p}, \boldsymbol{\phi}) - \pi_j^M(\mathcal{G} \setminus ij, \boldsymbol{p}_{-ij}, \boldsymbol{\phi})\right]^{\tau_j} \times \left[\pi_i^H(\mathcal{G}, \boldsymbol{p}, \boldsymbol{\phi}) - \pi_i^H(\mathcal{G} \setminus ij, \boldsymbol{p}_{-ij}, \boldsymbol{\phi})\right]^{(1-\tau_j)}, \tag{6}$$

where the Nash bargaining parameter for MCO j is represented by $\tau_j \in [0, 1]$. That is, each price p_{ij} maximizes the product of MCO j and hospital i gains-from-trade, holding fixed all other negotiated prices $\boldsymbol{p}_{-ij} \equiv \boldsymbol{p} \setminus p_{ij}$, where $\pi_j^M(\mathcal{G} \setminus ij, \boldsymbol{p}_{-ij}, \boldsymbol{\phi})$ and $\pi_i^H(\mathcal{G} \setminus ij, \boldsymbol{p}_{-ij}, \boldsymbol{\phi})$ are MCO j and hospital i's disagreement payoffs. Since each bilateral bargain occurs concurrently with premium setting, both parties to a disagreement believe that the new network will be $\mathcal{G} \setminus ij$, and all other prices $\boldsymbol{p}_{-ij}$ and premiums $\boldsymbol{\phi}$ remain fixed.

2.4.2 Equilibrium negotiated premiums and hospital prices

Next the authors derive the first-order conditions for the premium and hospital price bargaining equations in (5) and (6) to examine how insurer competition affects their determination in equilibrium.

Insurer premiums

Setting the first-order conditions of (5) equal to 0 (for a given network and set of premiums $\boldsymbol{\phi}_{-j}$, and set of negotiated prices $\boldsymbol{p}^*$) generates a set of conditions that correspond to the standard Nash-Bertrand first-order conditions when $\tau^{\phi} = 1$:

[18] The authors' empirical application actually allows for insurers to bargain simultaneously with hospital systems.

$\partial\pi_j^M(\cdot)/\partial\phi_j = 0$ for all j. However, holding fixed $\boldsymbol{\phi}_{-j}$, if $\tau^\phi < 1$, then the equilibrium premium for j will likely be lower than that predicted under Nash Bertrand premium setting. It will be particularly low, for example, if the MCOs' Nash bargaining parameter (τ^ϕ) is low; the profits that MCO j receives from the employer ($\pi^M(\cdot)$) are high; the employer's gains-from-trade with the MCO (GFT_j^E) are low; or the harm to the employer from higher premiums ($-\partial GFT_j^E(\cdot)/\partial\phi_j$) is large.

Hospital prices

Turning now to hospital prices, the first-order conditions of (6) (for a given network, set of premiums $\boldsymbol{\phi}$, and set of negotiated prices $\boldsymbol{p}^*_{-ij}$) are

$$\begin{aligned}\underbrace{p^*_{ij} D^H_{ij}}_{\text{total hospital payments}} =& (1-\tau_j)\left[\underbrace{[\Delta_{ij} D_j](\phi_j - \eta_j)}_{\text{(i) premium and enrollment effect}} - \underbrace{\left(\sum_{h\in\mathcal{G}^M_j\setminus ij} p^*_{hj}[\Delta_{ij} D^H_{hj}]\right)}_{\text{(ii) price reinforcement effect}}\right] \\ &+ \tau_j\left[\underbrace{c_i D^H_{ij}}_{\text{(iii) hospital cost effect}} - \underbrace{\sum_{n\in\mathcal{G}^H_i\setminus ij}[\Delta_{ij} D^H_{in}](p^*_{in} - c_i)}_{\text{(iv) recapture effect}}\right] \quad \forall ij \in \mathcal{G},\end{aligned} \tag{7}$$

where we have dropped the arguments of all demand functions for expositional convenience, and $[\Delta_{ij} D_j] \equiv D_j(\mathcal{G},\cdot) - D_j(\mathcal{G}\setminus ij,\cdot)$, and $[\Delta_{ij} D^H_{hj}] \equiv D^H_{hj}(\mathcal{G},\cdot) - D^H_{hj}(\mathcal{G}\setminus ij,\cdot)$ represent the adjustments in demand when hospital i and MCO j come to a disagreement.

Eq. (7) decomposes the determinants of negotiated payments when MCO j and hospital i bargain over the gains-from-trade created when that hospital is included in j's network. These gains are primarily obtained by MCOs through higher premiums and additional enrollees. The first line of the equation, representing MCO j's gains from having hospital i on its network, comprises two terms. The first is the premium and enrollment effect: the effect of hospital i's inclusion in MCO j's network on the MCO's premium revenues. Second is the price reinforcement effect: the adjustment in payments per enrollee that j makes to other hospitals in its existing network upon dropping i. The second line of (7), representing hospital i's gains from being included in MCO j's network, also comprises two terms: the hospital cost effect indicates that every unit increase in hospital i's costs results in a τ_j unit increase payments. Finally, the recapture effect is the adjustment in hospital i's reimbursements from other MCOs when i is removed from MCO j's network.

These terms have intuitive effects on equilibrium negotiated prices. The premium and enrollment effect indicates that the greater is the loss in an MCO's premium revenues (net of non-hospital costs) from losing access to a hospital, the more that hospital is paid. The price reinforcement effect indicates that, if hospital i's patients on MCO j substitute to cheaper hospitals when i is dropped, hospital i is paid less than if its patients switch to more expensive hospitals. The Nash bargaining parameter determines how completely the hospital is able to pass through cost increases to the MCO. And finally, the recapture effect represents hospital i's "opportunity cost" from being in MCO j's network: the more hospital i would be paid by other MCOs if i dropped MCO j, the more MCO j pays i in equilibrium.

Effects of reducing insurer competition

The authors use the bargaining first-order conditions to assess the impact of a reduction in insurer competition due to the removal of an insurer from the employer's plan menu. They note first that the logic from Nash-Bertrand premium setting is still present: the removal of an insurer tends to reduce "competitive pressures" (i.e., the elasticity of demand with respect to premiums for each MCO), and thus increase premiums. However, if $\tau^{\phi} < 1$, there are additional effects. The term $GFT_j^E(.)$ may either increase or decrease for an MCO j when a rival insurer is removed, depending on the relative cost-effectiveness of the different plans and other attributes that affect their value to the employer. Any net increase in $GFT_j^E(.)$ would reduce the extent to which premium-setting departed from Nash-Bertrand behavior and lead to additional upward pressure in j's premiums, but might also be offset by an increase in $\pi_j^M(.)$. Overall, in this model, it is possible for premiums to either increase or decrease as a result of removing an MCO from the choice set, even holding hospital payments fixed.

The effects on hospital prices are also non-trivial. The removal of a rival MCO affects every term in Eq. (7). The effect on the first term is theoretically ambiguous. On one hand, when a rival insurer is no longer competing for the same enrollees, the loss of hospital i typically results in a smaller adjustment to MCO j's enrollment; this is a primary source of MCOs' additional bargaining leverage when negotiating with hospitals in less competitive markets, and can lead to lower negotiated prices. On the other hand, a less competitive insurance market may generate higher premiums, which will tend to increase negotiated prices. The effects on terms (ii) and (iv) of the equation are even more difficult to sign because they are affected by changes in demand for all MCOs n and are also a function of the equilibrium prices paid to all hospitals. Finally, the impact of insurer competition on term (iii) will be limited in this model because hospital costs per admission are assumed not to be a function of realized hospital demand.

Overall, the equilibrium effects of insurer competition on negotiated prices and premiums (and consequently on consumer welfare and industry profits) depend on underlying demand primitives, firm heterogeneity, and institutional details. Ho and Lee estimate the components of the model using detailed household and individual-level data from the California Public Employees' Retirement System (CalPERS), a benefits manager that provides pension and health benefits to California state and public employees. The details of the demand system are described in previous sections of this chapter. The remaining components of the model—insurer non-inpatient hospital costs $\boldsymbol{\eta}$ and Nash bargaining parameters $(\tau^{\phi}, \boldsymbol{\tau})$—are estimated jointly using two-step GMM based on three sets of moments: two derived from the premium and price first-order conditions and a third "margin moment", which is particularly helpful in pinning down insurers' non-inpatient hospital costs, that fits observed to predicted insurer margins defined as the MLR obtained from financial reports from the California Department of Managed Health Care.

The authors' simulations demonstrate that neither premiums nor hospital reimbursements need necessarily increase upon the removal of an insurer. In their setting, two particular factors help determine the premium changes for one insurer (in the application this is Blue Shield) when a rival insurer is removed: (i) the size and attractiveness to enrollees of the insurer that is removed; and (ii) the presence of effective premium setting constraints. If premiums are not constrained by the employer—modeled by assuming Nash Bertrand premium-setting—premiums for Blue Shield are predicted to rise when its rival is removed. This is consistent with other results in the literature (e.g., Dafny, 2010; Dafny et al., 2012; Trish and Herring, 2015). Predicted increases are larger on removal of Kaiser, which has approximately a 40% statewide market share, than Blue Cross, with a 16% share. However, when premium setting is constrained—modeled by assuming Nash bargaining as laid out above—premium increases are predicted to be smaller, and even negative when Blue Cross is removed. Countervailing power effects are also empirically relevant: they may lead to lower negotiated prices and hence limited premium increases. Finally, in every counterfactual, there is substantial heterogeneity in predicted price changes across providers and markets.

2.4.3 Other related papers

Several other papers also use models of bargaining between hospitals and health insurers to consider policy-relevant questions, often related to hospital mergers. Capps et al. (2003) and Town and Vistnes (2001), both discussed above, are early examples that use a simple bargaining concept to motivate a regression equation linking willingness-to-pay to observed hospital profits or prices.

Gowrisankaran et al. (2015) develops and estimates a model of hospital-insurer bargaining which is used to predict the impact of hospital mergers on negotiated prices. In contrast to Ho and Lee (2017), the main analysis assumes that premiums are fixed and enrollees do not choose health plans; that is, insurers do not compete with one another in the downstream market for enrollees. The insurer objective function is not its profit; this would be a function of enrollment which the authors do not model. Instead insurers maximize a weighted sum of (negative) expected payments to in-network hospitals and the WTP variable derived in Capps et al. (2003).[19] This

[19] The authors also compare their results to a calibrated version of their model in which insurers engage in Nash-Bertrand premium setting.

simplification comes at the cost of ignoring the impact of networks on enrollment and hence insurer profit, and the equilibrium relationship between premiums and hospital prices, that are captured in Ho and Lee (2017). Instead the authors add flexibility on other dimensions. Most notably, they include an out-of-pocket price due to coinsurance in the hospital utility equation. Under the assumption that hospital prices are known to consumers, they use the Nash-in-Nash bargaining framework to derive a first-order condition for prices, conditional on hospital demand parameters and observables, which indicates that patients who face coinsurance payments are sensitive to hospital prices at the point of care. This adds a new lever—consumer price sensitivity—that affects negotiated price levels in equilibrium. It can also affect price dispersion: since MCOs use prices to steer patients towards cheaper hospitals, they have less incentive to exclude high-priced providers from their networks, so higher-cost hospitals may have relatively high negotiated prices in equilibrium.[20]

Prager and Tilipman (2020) also consider a Nash-in-Nash model of hospital-insurer bargaining, focusing their attention on modeling hospital disagreement points. Rather than assuming the hospital's share of patients from a particular plan falls to zero on disagreement, they allow patients to access out-of-network hospitals at potentially higher out-of-pocket prices. The authors show how this feature can affect negotiated prices for in-network hospitals and explore its implications for the impact of proposed policies to impose caps on out-of-network prices.

2.4.4 Revisiting provider consolidation

The model developed in Ho and Lee (2017) has implications for hospital merger price effects that are broader than those from the prior literature. Dafny et al. (2019) note that the simplified theoretical framework in Capps et al. (2003), which (along with the similar model in Town and Vistnes (2001)) is used as an input to assess the anti-competitive threat from hospital mergers, makes the implicit assumption that there can be no increase in bargaining leverage unless the merging parties are competing to provide the same set of services to the same set of patients. Since insurance enrollment is not modeled, insurance markets are assumed independent of patient markets and irrelevant for hospital negotiations, and patient demand is not permitted to be linked across markets. Thus, while the model generates the empirically relevant prediction that prices may increase in response to mergers between same-market hospitals, there is no predicted price effect of cross-market mergers.

Dafny et al. (2019) argue that the fact that insurers bundle together a network of hospitals' services before selling the resulting insurance plans to their customers can generate linkages across markets that may lead to cross-market merger price effects. For example, since insurers sell their plans not directly to individual consumers but instead to employers and households, there may be "common customers" who value the services offered by merging parties that operate in distinct patient markets. These common customers could be large employers that demand insurance products covering hospital services in multiple distinct geographic markets, i.e. areas where their

[20] Brown (2020) makes a similar point in the context of price transparency; see Section 2.2.

employees live and work. A merged cross-market hospital system that covers those regions may be able to demand higher reimbursement rates from insurers. Hence common customers can create links across markets which in turn generate price effects of cross-market hospital mergers.[21]

Two recent papers provide early empirical evidence of price effects of cross-market hospital mergers. Lewis and Pflum (2017) use a difference-in-differences analysis to assess the impact of cross-market hospital mergers on prices. They find that independent hospitals acquired by out of market systems raise price by 17-18 percent, and the effects are larger when the acquiring system is larger or when the acquired hospital is smaller (by number of beds). They argue that greater post-merger bargaining skill is the most credible explanation for the results. Dafny et al. (2019) examine two distinct samples of acute-care hospital mergers, addressing concerns about the exogeneity of which hospitals are parties to transactions by focusing on hospitals that are likely to be "bystanders" rather than the drivers of transactions. They estimate that prices for hospitals acquiring a new system member in the same state but not the same narrow geographic market ("adjacent treatment hospitals") increase by 7-10 percent relative to control hospitals, while non-adjacent treatment hospitals have small, statistically insignificant relative price changes. Acquirers raise their own prices, not just those of targets, suggesting that significant quality or bargaining improvements (such as might arise for targets following a takeover) are unlikely to be the source of price increases. Price effects are largest when the merging parties have hospitals in closer geographic proximity.

Schmitt (2018) has a related but distinct finding: mergers that increase multimarket contact between hospital systems raise prices by 6% in markets where there is no change in market structure from the merger. Brand and Rosenbaum (2019) review this small literature and conclude that there is reasonable empirical evidence that cross-market mergers lead to higher prices. However, they argue that more work is needed to understand the mechanisms behind the estimated effects and their implications for antitrust policy.

2.4.5 Network formation

The literature discussed above focuses on predicting hospital prices for a given hospital network. As noted, however, insurance plans are increasingly differentiated in terms of network breadth. This raises interesting questions about the determinants of hospital networks and the impact of narrow networks on prices and other outcomes. A small literature (Gruber and McKnight, 2016; Atwood and LoSasso, 2016; Wallace, 2019) suggests that employers realize spending reductions with no adverse effect on clinical quality when they offer narrow network plans.

Ho (2009) takes a first step towards modeling hospital network formation and contracting, assuming a take-it-or-leave-it offers model to determine hospital prices.

[21] Common customers can also be households that demand services of hospitals in the same geographic market but different product markets, e.g. pediatric and cardiac hospitals. See also Vistnes and Sarafidis (2013) for a further discussion of these issues.

She uses the estimated model of consumer demand for hospitals and insurers given the network offered from Ho (2006) as an input. In a partially identified framework, Nash conditions on observed networks (no plan could improve its profits by reversing a decision with any hospital; no hospital could do better by offering a null contract to any plan) are sufficient to place bounds on the determinants of equilibrium hospital profits despite the fact that the data do not include information on hospital prices. The estimates indicate that hospitals in systems, and those that are particularly attractive to patients, capture high markups while those with higher costs per patient receive lower markups than others. However, this paper is limited to using the observed variation in network breadth to make inferences about hospital profits, in a setting where provider prices and profits are not observed. It does not take the next step of endogenizing network formation.

Shepard (2020) picks up the thread of this literature. He studies network variation across plans and within-plan over time on the Massachusetts Health Connector insurance exchange that was established in 2006. He demonstrates substantial adverse selection against plans that offer the star academic hospitals (most obviously the hospitals in the Partners system in Boston). The sickest consumers, and also those who may be less sick but who have a strong preference for Partners hospitals, tend to enroll in these plans. Thus insurers have an incentive to exclude Partners hospitals in order to cream-skim enrollees, avoiding those who are costly to insure (either because of their preferences or their medical risk). Shepard demonstrates the magnitude of these incentives using a two-stage model of insurer and provider choice combined with a cost model that is estimated from claims data. He notes that a large insurer dropped the Partners hospitals in 2012. This led to a substantial shift of enrollees away from the plan, but it improved the plan's bottom-line profitability while reducing profits for its peers that continued to offer Partners access. By the end of 2014, only one plan covered Partners, and it had been purchased by Partners in 2013. The cream-skimming incentive for narrower networks is most relevant in individual health insurance markets as opposed to markets where employers select plans for their employees. This may help explain why narrow network plans have proliferated in exchanges while remaining uncommon in employer-based insurance.

Tilipman (2021) takes the ideas in Ho (2009) further, adding switching costs to the insurer choice equation and estimating a model of consumer demand for physicians (as well as hospitals) that also allows for inertia. He uses a partially identified framework, under simple assumptions regarding price and premium adjustments, to bound the fixed costs faced by employers choosing their menus of plans in order to maximize an objective function that weights employee surplus and (negative) premiums and fixed costs. Offered plans differ in their hospital and physician networks as well as their premiums. Tilipman uses the estimates to explore the influence of switching costs on employers' health plan offerings, and particularly the networks of the plans they offer, and finds it is substantial. We discuss this paper in more detail in Section 4 below.

A model of price negotiation and network formation: Nash-in-Nash with threat of replacement

Ho and Lee (2019) investigate the issues raised in the previous literature by developing a model that allows them to endogenize insurer network formation. They study the same CalPERS setting that was analyzed in Ho and Lee (2017), noting that in the year following their data, one of the large insurers (Blue Shield HMO) had requested permission to exclude a substantial number of hospitals from its network. Networks were previously constrained by CalPERS to be essentially complete. The authors argue that one possible reason why an insurer might choose to exclude a hospital is related to price negotiation. By excluding some hospitals from its network, an insurer may be able to negotiate lower prices with those that remain.

Ho and Lee develop the Nash-in-Nash price negotiation model from their previous paper on two dimensions. First, they introduce an initial stage where a single insurer commits to a particular network prior to bargaining. Second, they modify the bargaining framework to allow out-of-network hospitals to influence negotiated payments. Intuitively, we can think of each bilateral insurer-hospital pair as engaging in simultaneous Nash bargaining over their gains-from-trade (as in Nash-in-Nash bargaining), with the important addition that each insurer is permitted not just to drop its bargaining partner but also to replace it with any alternative hospital that is not already on its network. In the hospital-insurer setting, this may be intuitively more reasonable than the Nash-in-Nash assumption that an insurer's only alternative to reaching agreement with a particular hospital is to exclude it without replacement, even if other alternative hospitals exist outside the network. The extension is also useful in rationalizing exclusion, because it effectively generates an endogenous cap on Nash-in-Nash prices, where the effectiveness of the cap depends on the existence of credible alternative (currently excluded) negotiating partners. The new bargaining solution, Nash-in-Nash with Threat of Replacement, assigns each included hospital a payment equal to the minimum of (i) what would be paid under Nash bargaining, holding fixed other hospital prices at NNTR prices and (ii) the price that would make the insurer indifferent between keeping the hospital in its network, and replacing it with the "best" alternative at the lowest price that the alternative provider would be willing to accept.[22]

One further component of the model is important for understanding its implications. The NNTR solution is defined to be admissible only for networks that are stable, i.e. that satisfy conditions implying that no party has a unilateral incentive to terminate a contract based on negotiated prices. The authors show this is equivalent to assuming that any hospital included in the network must generate higher joint

[22] The theory builds on the literature on bargaining with outside options, particularly the discussion of "deal me out" versus "split the difference" models in Binmore et al. (1989) and the model of bargaining with endogenous outside options in Manea (2018). Insights from these papers are adapted to a bilateral contracting environment where firms can negotiate with multiple partners and there are externalities between bargains.

surplus with the insurer than any excluded hospital. Overall, in order to affect equilibrium bargaining outcomes, excluded hospitals must be sufficiently attractive—i.e., must have the potential to generate enough surplus—to provide the insurer with additional leverage in bargaining, but not so much that excluding them makes the network unstable. Thus the NNTR solution concept provides an opportunity for the detailed data and demand model to inform which hospitals are most likely to be excluded, and why. Note that, if excluded hospitals are sufficiently unattractive that they are not credible outside options, then the NNTR solution coincides with the Nash-in-Nash outcome. This would also be true if the network was complete.

The paper then proceeds to simulations, based on the estimated model in Ho and Lee (2017) combined with the new bargaining concept, which demonstrate that NNTR provides the insurer with a new empirically-relevant incentive to exclude hospitals: while the Nash-in-Nash solution has difficulty rationalizing any exclusion by Blue Shield in the relevant year, the NNTR solution does not. The model can also explain why an insurer might choose to exclude more than one hospital from its network in some markets. Since hospitals are differentiated on multiple (vertical and horizontal) dimensions, the most effective threat for replacing one hospital that remains in the network may be different from the most effective threat for another.

Note that the Ho and Lee (2019) model endogenizes network formation only in settings with a single strategic insurer. Their application is particularly suitable for this assumption because it is plausible to assume that only a single insurer chooses its network strategically.[23] However, there remains a question of how to expand the model to account for multiple strategic insurers. Ho and Lee (2019) suggest that the most direct extension would assume that each insurer simultaneously announced the set of hospitals it was willing to negotiate with, and that prices were then determined by the NNTR bargaining solution. If appropriate extensions were made to the concept of stability, this would allow authors to begin to analyze insurer competition with endogenous networks and prices. However, the extension would face at least two challenges. The first is computational, since the number of potential combinations of networks to be analyzed would increase rapidly with the number of insurers. The second is conceptual: the NNTR solution only allows one side of the market (insurers) to threaten replacement of the other (hospitals). In a setting with multiple strategic insurers, it seems unreasonable to ignore the possibility that hospitals could do the same, leveraging excluded insurers in negotiations. This introduces issues related to endogenizing each agent's outside options given the contracting externalities embedded in the model. For example, if one hospital threatens to replace insurer j with insurer j', does it predict that j might also replace it with an alternative? Does this affect the identity of the hospital's optimal replacement for j? Further, are the information and timing assumptions defined such that some alternative (excluded) bargaining partners

[23] The other two insurers are Kaiser Permanente, a vertically integrated insurer that owns its hospitals and does not contract with other providers, and Blue Cross PPO, a broad network plan which included essentially every hospital in the markets it covered.

maintain some bargaining leverage rather than being forced down to their reservation prices? Many of these questions raise dynamic considerations; they suggest an alternative approach of specifying a fully dynamic network formation game. One promising route forward is provided in Lee and Fong (2013), which proposes a dynamic model of network formation and bargaining in a bilateral oligopoly between multiple upstream and downstream firms. Their model also endogenizes networks and outside options, in a setting where agents anticipate future (not contemporaneous) adjustments to the network while bargaining.

Ghili (2020) and Liebman (2020) are recent papers that are complementary to Ho and Lee (2019). Each develops a model that also allows excluded hospitals to affect insurers' negotiated rates with included hospitals in a static framework. These models of negotiation are embedded into a demand model similar to that in Ho (2006) and Ho and Lee (2017). Details of the bargaining frameworks differ across papers. Ghili (2020) is based on pairwise stability conditions (as in Jackson and Wolinsky (1996)) which imply that any hospital-insurer pair that is excluded in equilibrium must be unable to generate positive gains from trade conditional on the prices negotiated by included hospitals. This is a stronger assumption than the stability condition in Ho and Lee (2019), making it more difficult for the model to predict exclusion. The author rationalizes the networks observed in the data by estimating a recurring fixed cost of contracting. He allows multiple insurers to choose networks simultaneously, although he does not address the issues relating to contracting externalities outlined above. Liebman (2020) adapts a bargaining protocol from Collard-Wexler et al. (2019). He allows an insurer to commit to a maximum number of hospitals that it will contract with; upon disagreement, a replacement hospital is perceived to be randomly chosen from an exogenously determined set of potential replacements. This is an interesting modification to the model that takes seriously the idea that as the insurer excludes more hospitals, the continuation value of those that remain falls, leading them to accept lower prices. However, it does not allow the estimated demand model, and detailed data, to help inform which hospitals are good substitutes for others and hence might profitably be excluded.

Other authors have considered yet more complex questions related to network formation. In recent years insurers have begun to offer tiered hospital networks, in which some hospitals are made available at lower coinsurance rates than others. Intuitively, this should have the effect of constraining consumer choice—and making consumers aware of price differences between hospitals—without fully removing providers from the choice set. Prager (2020) considers consumer demand in the presence of tiered networks. She shows that consumers substitute towards hospitals on more preferred tiers, presumably generating incentives for insurers to establish tiered networks and also affecting price negotiations in the presence of such tiers. Consistent with this intuition, Starc and Swanson (2021) provide evidence from Medicare Part D data that drug plans with more restrictive preferred pharmacy networks pay lower retail drug prices. To our knowledge, no paper yet estimates a model that accounts for these supply side effects.

Distinguishing features of NNTR versus Nash-in-Nash bargaining

We know of no simple way to test whether the observed variation in a particular dataset is consistent with the Nash-in-Nash model, the NNTR concept, or some alternative bargaining model. Ho and Lee (2019) note that, under their assumptions, the Nash-in-Nash model does not rationalize exclusion by Blue Shield in the year in which it in fact proposed to exclude a number of hospitals, while their NNTR solution predicts this outcome quite naturally. This is helpful evidence to support the NNTR concept but it is not definitive. Other modifications to the Nash-in-Nash framework such as fixed costs of contracting (as in Ghili (2020)), or increasing marginal costs at the insurer-hospital level, could also fit this dimension of the data.

Similarly, it is tempting to try to distinguish between models by asking whether the breadth of networks (the proportion of hospitals in the market that are included) is positively correlated with prices. Unfortunately, the intuition that suggests such a test is incomplete. There are two offsetting effects when the network is expanded under NNTR: (a) the standard effect, also relevant for Nash-in-Nash bargaining, under which adding substitutes to the network reduces each hospital's marginal contribution to the joint surplus and hence reduces payments; (b) the additional "threat of replacement" effect that reducing the number of excluded providers potentially reduces the insurer's ability to play included hospitals against excluded ones. The relative importance of these effects depends on many factors including the characteristics of the relevant market and the providers in that market. Particular examples considered in Ho and Lee (2019) show that hospital payments may not decrease monotonically under NNTR as hospitals are excluded. In fact it is straightforward to construct examples where the first (Nash-in-Nash) effect dominates, even under NNTR bargaining, and prices actually increase as the network is narrowed. For these reasons we caution against using correlations in the data, either in the cross-section or over time, to attempt to rule out any particular bargaining framework.

2.5 Provider markets and quality

In stage one of the five-stage model, providers make investment decisions that help determine their quality. Investments are influenced by provider expectations regarding their impact on consumer decisions to enroll in an insurance plan that offers the hospital, and to choose it when they require care, and on the hospital's leverage when negotiating with insurers over inclusion in the network and over prices. This full model is complex; to our knowledge no authors have yet attempted to estimate it in full. The predicted impact of any policy to encourage investment is likely to be complicated by multiple equilibria and substantially affected by frictions such as asymmetric information regarding provider quality, and (for hospital quality) problems generated when the referring physician has more information about quality than the patient.

Quality when prices are administered

Questions regarding provider quality investment are simplest in markets where prices are determined administratively. The fee-for-service US Medicare market and the UK National Health Service (NHS) are good examples. Two fairly recent studies (Cooper et al. (2011), Gaynor et al. (2013)) examine the impact of a 2006 NHS reform that was intended to promote competition on hospital quality. Prices were administratively determined based on patient diagnoses, so that when patient choice of hospital was introduced, hospitals were forced to compete solely on non-price dimensions. These studies use difference in differences estimation (where the differences are before and after the reform and across more and less concentrated markets). Although they differ in the precise methods employed, both Cooper et al. (2011) and Gaynor et al. (2013) find that, following the reform, risk-adjusted mortality from acute myocardial infarction (AMI) fell more at hospitals in less concentrated markets than at hospitals in more concentrated markets.

Gaynor et al. (2016) take the analysis further, estimating a discrete choice demand model for patients requiring coronary artery bypass graft (CABG) surgery using a method that explicitly accounts for the 2006 reform's mandate that patients had to be offered a choice of five hospitals. No such requirement was in place before the reform. The authors assume that the choices made by the referring physician post-reform fully reflect patient preferences. To estimate preferences pre-reform, they use an approach which draws from the consideration set literature (Goeree (2008), Mehta et al. (2003)), assuming the patient choice set contains a subset of the full list of options which is determined by the physician's preferences based on local administrative boundaries. This allows them to estimate the extent to which patient choices were constrained in the period before the reform. They find that the constraints were considerable; there is clear evidence of improved sorting of patients to higher-quality hospitals after the reform; and that an increase in mortality for the average hospital led to a five-times-larger reduction in market share post-reform than pre-reform. Hospitals which experienced the largest increase in elasticity also had the largest reduction in mortality rates after the reform. Overall, the paper provides an important model-based decomposition of the mechanisms behind the impact of hospital competition on quality in markets where prices are fixed.

Moscelli et al. (2021) notes that theory predicts an improvement in quality due to competition only if price is greater than marginal cost. In that case firms facing competition have an incentive to increase their quality in order to attract patients. If price is below marginal cost, increases in competition can actually lead to reductions in quality. The authors investigate whether the 2006 NHS reform affected the rate of emergency hospital readmission after surgery for elective treatments—hip and knee replacements—where regulated prices are unlikely to be sufficient to cover hospital costs. The authors show that after the reform, hospitals facing more rivals reduced quality, increased waiting times and reduced length of stay for these two procedures. They found no effect on outcomes (mortality rates) for CABG treatments.

Quality with negotiated prices

The recent literature that considers provider quality in markets where prices are negotiated is small and not model-based. We omit studies that follow a structure-conduct-performance approach (see Gaynor et al. (2015) for a discussion of these papers). A few papers use variation caused by hospital mergers or entry to address the issue. Ho and Hamilton (2000) and Capps (2005) both examine the impact of hospital mergers on quality of care. Ho and Hamilton (2000) study 130 hospital mergers of various types over the period 1992 to 1995 using hospital specific fixed effects to control for time invariant hospital characteristics that may be related to merger. Capps (2005) compares merging to non-merging hospitals in New York state during 1995-2000. Neither study finds significant effects of hospital mergers on the quality of care. Romano and Balan (2011) study the impact of a particular consummated merger between Evanston Northwestern Hospital and Highland Park Hospital which was the subject of an antitrust suit by the Federal Trade Commission. The authors again use a difference-in-differences approach. They find no significant impact of the merger on many quality measures, but there is a significant negative impact on some and a few have positive impacts.

Cutler et al. (2010) utilizes the repeal of entry-restricting regulation (certificate of need regulations) in Pennsylvania to examine the effect of entry of hospitals into the market for Coronary Artery Bypass Graft (CABG) surgery. The authors argue that overall production is capacity constrained—cardiac surgeons are a scarce input and their supply cannot easily be altered—so that hospital entry cannot lead to increased quantities of CABG surgery but may increase the market shares of high quality surgeons. This is confirmed in the empirical analysis. Estimates suggest that in markets where entrants had 11-20 percent market shares of CABG surgeries, high quality surgeons' market shares increased 2.1 percentage points more than for standard quality surgeons. Overall, the authors conclude that entry led to increased quality, but that there was no net effect of entry on social welfare.

Several papers assess the impact of widely-used quality "report cards" on consumer choices of physicians and also of health plans. Studies analyzing physician choice tend to find positive but small effects of reported quality on provider market shares (Cutler et al., 2004; Dranove and Sfekas, 2008). Other papers investigate the possibility that quality reporting could lead to selection by providers against sicker patients. Dranove et al. (2003) compares outcomes for cardiac patients in the Medicare population in New York and Pennsylvania located in areas with and without report cards. They find that report cards improved the match of patients to hospitals—a gain from the provision of information—but also find evidence of selection against sicker patients. Dafny and Dranove (2008) considers the effect of Medicare HMO report cards, distributed to 40 million Medicare enrollees, in 1999-2000, on subsequent health plan choices. The authors find that, while the public report cards generated changes in enrollment, these effects are smaller than market-based learning as measured by the trend towards higher-quality plans in the years prior to the intervention.

Kolstad (2013) considers the impact of quality report cards on CABG surgeon quality in the same Pennsylvania setting as Cutler et al. (2010). Kolstad notes that report cards may affect physician behavior and hence physician quality, both because information on quality affects consumer demand and hence physician profits (an "extrinsic" motivation) and because of an "intrinsic" motivation to observably perform well relative to a reference group of peers for reasons unrelated to profit. He estimates a model of consumer demand for surgeons and shows that variation in demand for quality, and the competitive structure of markets, leads to large differences in extrinsic incentives between surgeons. Those facing stronger profit incentives due to report cards show greater quality improvements, but the response to intrinsic incentives is four times larger, leading to significant declines in risk-adjusted mortality.

A useful next step in this literature would be to specify and estimate an equilibrium model of provider investment in quality. To our knowledge this has not yet been attempted for health care providers (although essentially the same question has been modeled in education; see Neilson (2020), Allende et al. (2020), Allende (2020)). The small number of papers already discussed that use partially identified models to consider insurer network breadth—a measure of quality for *insurers* rather than providers—may suggest an approach that makes estimation feasible. We view these issues as an opportunity for future research.

2.6 Medical organization responses to price regulation

The impact of prices on medical organization care delivery is central to the ongoing debate on how to control health care spending. While there is much ongoing work in health economics on how to best regulate prices to ensure efficient health care utilization, a full review of this work is beyond the scope of this paper. See, e.g., Agarwal et al. (2020) for a discussion of bundled payment models and, e.g., Frech et al. (2015) for a discussion of accountable care organizations (ACOs), two policy initiatives that focus on pricing and provider behavior. See, e.g., Clemens and Gottlieb (2014), for a broader analysis of how health care supply, technology diffusion, and patient outcomes respond to regulated price changes.

There are, however, a few recent notable health and IO papers that use sophisticated models of price gaming and productivity to study contract design questions. Most notably, two recent papers (Eliason et al. (2018) and Einav et al. (2018)) have both studied (i) how long-term care hospitals (LTCHs) strongly game Medicare price regulation and (ii) how pricing regulation for LTCH care could be changed to better facilitate efficient utilization.

LTCHs care for patients who have been discharged from an acute care hospital but who still need meaningful care post-discharge. The Medicare payments incentives studied in Eliason et al. (2018) and Einav et al. (2018) rise linearly up until a "magic day" when the LTCH receives a large lump sum additional payment. After that point, the LTCH receives no further payment. Formally, the marginal prices p_t for keeping a patient in the LTCH an extra day at day t are:

$$\begin{aligned} p \text{ if } t < t^m \\ P - (t^m - 1)p \text{ if } t = t^m \end{aligned}$$

$$0 \text{ if } t > t^m$$

Here, P is a total lump sum payment the LTCH receives if a patient stays longer than t^m days (typically 25 days). This total amount is larger than the sum of the linear fee leading up to day t^m, such that the marginal payment on $t = t^m$ is large and positive, equal to approximately $13,500 on average (Einav et al. (2018)).

Both papers document very strong LTCH responses to these incentives, with huge spikes in patient discharges immediately after the magic day. The papers investigate a range of alternative explanations but conclude that the LTCH behavior is a response to these clear pricing incentives. Eliason et al. (2018) document that for-profit LTCHs are more likely to game the price regulation relative to not-for-profit LTCHs while Einav et al. (2018) show that there are no detectable mortality changes resulting from this price gaming.

Both paper estimate dynamic structural models in order to study counterfactual contract design for LTCHs. In Eliason et al. (2018), the LTCH has a daily flow utility per patient equal to:

$$u_t = \lambda_t + \alpha p_t$$

Here, λ_t represents the non-financial benefits and costs of keeping the patient for one more day and p_t is the marginal price for keeping the patient an extra day, as discussed above. The LTCH solves a dynamic problem when determining whether or not to keep the patient another day:

$$V_t(\epsilon_t) = u_t + \max[\epsilon_{kt} + \delta E V_{t+1}, \epsilon_{dt}]$$

The authors assume that the idiosyncratic shocks for keeping the patient ϵ_{kt} and discharging the patient ϵ_{dt} are Type I extreme value distributed such that the probability the LTCH discharges the patient on day t, conditional on not being discharged earlier, is:

$$\frac{1}{1 + e^{\delta E V_{t+1}}}$$

The authors solve the model using backwards induction starting from a terminal date T to determination continuation values and discharge probabilities as a function of candidate parameters. The model presented in Einav et al. (2018) is similar in spirit, but differs in several key ways including that it (i) assumes stationarity in patients' health processes and (ii) allows for a richer discharge set to upstream acute care hospitals as well as downstream to a less intensive facility (the only option allowed in Eliason et al. (2018)). Ultimately, the modeling approaches are complementary and allow for slightly different counterfactuals and differential emphasis on patient / LTCH heterogeneity.

Eliason et al. (2018) find that (i) LTCHs would discharge patients a week earlier on average without the lump sum "magic day" payment and that (ii) cost-plus reimbursement policies would eliminate the discharge spike after the magic day but would actually lead to longer stays on average because patients remain marginally profitable for the entire stay, not just through day 25. Einav et al. (2018) focus on alternative contract designs that hold the LTCH harmless and highlight a contractual design that leads to $2,100 (5%) savings per patient for Medicare without leading to lower LTCH profits. They also look at non-Pareto contracts that generate substantially greater savings to Medicare but lead to lower LTCH profits, with ambiguous implications for patient well-being.

In addition to these papers on LTCH incentives, there are several recent papers on dialysis clinic incentives. Grieco and McDevitt (2016) model the quantity-quality tradeoff for these clinics accounting for unobservable and endogenous choice of treatment quality and, separately, for differential productivity across clinics conditional on treatment quality. They find that reduced quality leading to a one percentage point increase in septic infections allows for the clinic to see an additional 1.6% patients, holding other factors fixed. Eliason et al. (2020) study the implications of mergers and acquisitions for health spending and outcomes at dialysis clinics. They study the acquisitions of over 1,200 independent clinics by several large organizations and document the transfer of several key strategies employed by the large clinics. These include (i) moving patients to more highly reimbursed drugs (ii) replacing higher skilled nurses with less-skilled technicians, and (iii) waitlisting fewer patients for kidney transplants. The authors show that patients fare worse as a result of these changes with key implications for worsened hospitalization and mortality rates. These results also relate to Dafny (2005), who finds greater price responsiveness by for-profit firms when faced with varying Medicare prices.

It is important to note that there are also many important papers on mechanism design and kidney exchange, which happens outside the price system. We do not discuss these papers here since they are discussed in depth in the companion chapter in this handbook by Agarwal and Budish (2021).

2.7 Vertical consolidation between providers

The model outlined so far in this chapter abstracts away from an important component of health care market structure: the extent of 'vertical' consolidation between physician groups and secondary care providers such as specialists and hospitals. Such consolidation, which is perhaps more accurately viewed as consolidation of complements, has increased substantially over the last 10-20 years. There is a growing literature considering the implications of this change. Authors emphasize the agency problems raised by the patient-physician relationship, as patients rely on their physician for information about which services are needed as well as for provision of the services themselves. Vertical consolidation could increase the incentive for a primary care physician to refer a patient to the newly-owning hospital, perhaps for treatments that would not otherwise be recommended; alternatively, it could improve coordination across care settings (physician's office and hospital) and reduce duplication of care. Physician or hospital prices could also be affected (although most papers do not specify a mechanism for such price effects).

Recent papers considering the impact of physician-hospital mergers include Baker et al. (2016), which uses detailed data on hospital admissions of Medicare beneficiaries, matched to data on physician practice ownership, to investigate whether integration of referring physicians with hospitals affects hospital choice. The authors estimate multinomial logit discrete choice demand models that account for hospital and physician characteristics (including distance to the patient's home) as well as ownership details. They find that patients are significantly more likely to be admitted to hospitals that own physicians, and substantially (33.4 percentage points) more likely if the patient's physician is owned by the particular hospital. Brot-Goldberg and de Vaan (2020) also investigate the impact of provider consolidation on referrals, holding prices fixed. They use detailed data from Massachusetts to study the impact of integration that generates large systems of health providers, assessing the trade-off between productive efficiency of care and potential inefficient allocation of patients across specialists in the context of PCP referrals to orthopedic joint surgeons. The authors estimate a model of PCP referrals that allows for measured heterogeneity in cost outcomes across specialists. They conclude that consolidation generates productive efficiencies, reducing expected costs per patient, but that incentives to refer patients internally for non-efficiency reasons are also substantial.

Capps et al. (2018) go a step further, investigating the effect of hospital acquisitions of physician practices on prices as well as spending. They use detailed proprietary data (covering several geographically dispersed states) to identify hospital-physician integration at the individual physician level and study physician transaction prices before and after integration. They document substantial changes in market structure: the share of spending by vertically integrated doctors increased in their sample from 17.7% to 27.2% between 2007 and 2013. Difference-in-differences regressions indicate that this change in market structure led to an average physician price increase of 14.1%. Approximately 45% of the price change was due to integrated organizations exploiting a reimbursement rule that allowed hospitals to charge "facility fees" for procedures by hospital-owned physicians. The authors also consider the impact of consolidation on total spending, focusing on hospital integration with primary care physicians, and find that any decrease in utilization from integration was insufficient to outweigh the increase in prices.

Dranove and Ody (2019) is a related paper that investigates the causes (rather than consequences) of the increase in hospital-physician integration in the authors' data. The core idea is that at least some portion of the observed change in market structure was a reaction to the reimbursement rules established by Medicare, and often followed by private insurers. In 2010, the Centers for Medicare and Medicaid Services replaced the physician survey used to estimate practice costs, leading to procedure-level shocks to Medicare prices. This tended to reduce the aggregate Medicare reimbursement for office-based care for physicians who were not integrated with a hospital, increasing the incentive to vertically integrate. A previous paper, Song et al. (2015), shows that three cardiology procedures shifted from physician offices to hospital outpatient offices in response to this change. This paper extends the analysis by examining a broader range of procedures; providing evidence that the 2010 price shock increased hospital employment of physicians and that this explains much of the shift to hospital outpatient offices; and showing that this single shock was sufficient to explain approximately 20% of the overall increase in physician employment by hospitals between 2009-13. That is, the authors argue, organizational structure responded to profit incentives introduced by changes to Medicare reimbursement rates.

Cuesta et al. (2020) considers the scenario of integration between hospitals and insurers (rather than hospitals and physicians). The authors adapt the model of vertical integration in the cable television market developed in Crawford et al. (2018) to fit their health care setting, accounting for the effects of insurer-hospital integration on negotiated hospital prices, consumer choices of insurance plan and their hospital choices. Vertical integration reduces inefficiencies due to double marginalization but also introduces incentives for integrated firms to increase prices to rivals in order to steer consumers to their own integrated partners. The authors estimate this model using detailed administrative data from Chile, finding evidence that vertical integration has led to increased negotiated prices and welfare reductions in their setting.

Overall, this small but growing literature promises to provide important insights for policy-makers and researchers on the effects of, and incentives for, integration across various components of the health care market. Further research would be very valuable in all these areas.

2.8 Medical devices: prices and bargaining

We conclude this section by outlining a small literature considering price negotiations for medical devices. These begin with Grennan (2013), an innovative paper which was among the first in the literature to use a Nash-in-Nash bargaining model to analyze price negotiations between upstream and downstream firms. The author uses a new panel dataset containing prices and quantities for coronary stents that are purchased by hospitals from medical device manufacturers. He begins by noting the substantial cross-sectional variation in prices, across different hospitals, for the same device being sold by the same manufacturer. He argues, in the context of the Nash-in-Nash model, that such dispersion could be caused by differences across hospitals in the gains from trade provided by a particular device, or alternatively by variation in Nash bargaining weights (which he interprets as a measure of bargaining ability). The model is estimated in two stages: first, doctor demand for stents is estimated using observed price and quantity data; then the bargaining framework, together with inputs from demand estimation, are used to estimate costs and relative bargaining abilities. Identifying assumptions are required to separate costs from bargaining weights: the author assumes that costs are determined entirely by observed type of stent, while (relative) hospital bargaining weights are permitted to vary more flexibly across hospitals, manufacturers, and time. Counterfactual simulations suggest that policies to limit suppliers' ability to price discriminate might actually work against hospitals, leading to a softening of competition and an increase in average price. Unsurprisingly, this finding is sensitive to assumptions about the impact of a move to uniform pricing on hospital bargaining weights.

In Grennan (2014), the author looks more closely at the bargaining weights estimated in the previous paper. He shows that variation in bargaining weights is four times larger than variation in demand for a given stent; that is, bargaining weights explain most of the observed variation in prices when viewed through the lens of the model. He then explores the variation in estimated bargaining weights, using the panel structure of the data to run a regression that projects their log onto manufacturer-hospital-pair dummy variables. Firm (manufacturer and hospital) fixed effects explain almost 30% of the variation in relative bargaining weights: that is, some firms consistently negotiate more favorable prices (conditional on estimated costs, willingness-to-pay and the extent of competition) than others. A further 36% of the variation in relative bargaining weights is pair-specific.

A very recent addition to the literature is Grennan and Swanson (2020). This paper investigates an important and policy-relevant question: does price transparency lead to price changes in a context with negotiations between upstream and downstream firms, and if so, in what direction? There are concerns among policy-makers that while (as already discussed) providing price information to consumers might make residual demand more price elastic and provide incentives for firms to negotiate low prices, there are potential offsetting effects if price information also affects firm price negotiations directly. The authors investigate components of this question empirically, using new data containing all purchase orders issued by a sample of hospitals that purchased a subscription to a web-based price benchmarking database. Focusing on 508 facilities negotiating prices for coronary stents, the authors consider two potential mechanisms through which benchmarking data could affect price negotiations. The first is based on an asymmetric information model where hospitals face uncertainty about suppliers' costs or bargaining weights: transparency reduces uncertainty, and with it, the dispersion in negotiated prices. The second is an agency model in which transparency allows hospital managers to better observe purchasing agents' effort and provide more effective incentives to reduce prices. Two sources of data variation are used to investigate these mechanisms: variation in hospitals' timing of joining the database; and the fact that new brands entered the market during the time period of the data. Estimated average price effects are small and noisy, but the estimates suggest that high-price hospital-brand combinations have significant price reductions when database information becomes available. Price declines are larger in cases with high purchase volumes. Various additional analyses, including tests based on timing of price changes relative to brand entry, are consistent with benchmarking solving an asymmetric information problem—helping hospitals learn about suppliers—rather than an agency problem. The authors note that their findings have potentially important implications for the empirical literature on bargaining. Further, while they do not directly consider the possibility that price transparency could help facilitate collusion—a concern raised by researchers assessing potential overall effects of price transparency on equilibrium prices—their broad findings that prices fall when benchmarking data is available is suggestive that, at least in this market, collusive behavior is not the dominant response.

3 Consumer choice in health insurance markets

A key premise underlying much of health insurance market design and regulation is that consumer choice can help (i) facilitate efficient matching between consumers and the available plans and (ii) encourage insurers to offer higher quality products and compete to lower prices. Discussed in depth in Enthoven et al. (2001), the notion of managed competition in insurance markets has made strong inroads into many health insurance provision contexts, especially in the United States. Regulated insurance exchanges that are founded on the notion of active consumers include Medicare Part D prescription drug insurance for seniors, Medicare Advantage (privatized Medicare), the state-by-state exchanges set up to target the uninsured in the Affordable Care Act, Medigap supplemental insurance for seniors, and privatized Medicaid markets for low income consumers. In addition, large group markets (e.g. large employers) in the United States often offer multiple plan options with the same underlying logic as discussed for these regulated exchanges. Finally, many international contexts also operate insurance markets following the managed competition paradigm, including, e.g., the Netherlands, Switzerland, Chile, and Australia.

Since choice is such a central input into insurance market regulation, modeling demand for insurance—in a way that accounts for issues raised by asymmetric information—has been a vital contribution of the health and industrial organization literature on insurance markets. Modeling demand for health insurance has much in common conceptually with modeling demand for other insurance markets (e.g. car or life insurance) though health markets have specific features related to the scale of the risks and the nature of the risks that need to be addressed in demand modeling. Some of the key features that relate directly to the insurance nature of health insurance are risk preferences, beliefs about health risk, and moral hazard (i.e. price sensitivity) in health care utilization. Here, we discuss these demand foundations in depth with a specific eye towards their use in health insurance markets: see Einav et al. (2021) for a broader discussion of demand estimation in insurance markets.

In addition to these insurance-specific dimensions, health insurance demand has two other key micro-foundations we will discuss in this chapter. One is choice frictions / behavioral choice foundations that the literature has studied extensively. These frictions have been studied in different forms but ultimately work against the premise in managed competition that consumers will help discipline the market. Such frictions have also been studied extensively in markets for other financial products (e.g. retirement savings plans), where consumers often have substantive difficulties choosing among products with important implications. We cover this work in some depth after going through a baseline model for insurance demand.

A second key foundation is non-financial preferences for insurance plans. The earlier portion of this chapter covers in depth how important provider networks are for insurance plan demand and digs into how consumers value those networks in detail. In addition to these specific preferences for networks, consumers may have brand preferences that reflect a broad range of factors specific to a given insurance carrier (some of which may be tangible factors and some of which may be intangible).

We discuss these non-financial preferences in the context of the specific papers we cover.

3.1 Health insurance demand: baseline model

Cardon and Hendel (2001) is one of the earliest papers that jointly models health plan choice and health care utilization in large group employer markets. The main contribution of this paper is the two-stage model of plan choice and utilization. This model nicely embeds the central welfare tradeoff between risk protection and moral hazard in insurance markets (see, e.g., Zeckhauser (1970)). If an insurer offers greater financial coverage, consumer surplus from risk protection increases but, potentially, the negative welfare impact of moral hazard (partially or fully wasteful care) increases. While Cardon and Hendel (2001) focuses on testing for asymmetric information, rather than welfare analysis of a specific policy or phenomenon, the framework is well set up for welfare analysis. Many papers, including many of those discussed for the remainder of this chapter, build on this framework by extending the micro-foundations and then applying the resulting model to study the welfare implications of specific policies.[24]

3.1.1 Health insurance demand: moral hazard

Cardon and Hendel (2001) set up their model with two stages. Working backwards, they first describe the second stage that models how consumers use health care once they have enrolled in a given health plan. While demand for health care is realized ex post to the purchase of insurance, consumer projections of (i) how much their out-of-pocket expenses will be during the year and (ii) how much health care they will consume given the plan design are important factors into ex ante demand for insurance.

In Cardon and Hendel (2001), consumer preferences for utilization are described as:

$$U(m_i, h_i) \tag{8}$$

$$h_i = x_i + s_i \tag{9}$$

Here, h_i is a consumer's utility from health and m_i is that numeraire good representing utility from additional money. x_i is their health care consumption, and s_i is

[24] For brevity, we focus this demand discussion on structural papers that dive into the micro-foundations of demand. There are many excellent papers that study health insurance demand without explicitly modeling these micro-foundations, e.g. Einav et al. (2010b). There are many reasons why frameworks that model demand with fewer assumptions, in a way that is sufficient to answer specific policy questions, can be preferable to a more structural approach. We discuss relevant health and industrial organization papers that estimate both types of demand models throughout this chapter. For a more detailed discussion of Einav et al. (2010b) and other related demand approaches, see the companion chapter by Einav et al. (2021) in this volume.

their realized health state in a given year. s_i is unknown at the time of insurance purchase and affects marginal utility for health care. Once s_i is realized, consumers can spend x_i to improve their utility from health. So, e.g., if one's health state is particularly poor, marginal utility at spending level x_i will be higher than the same marginal utility for someone in a good health state.

$$\begin{aligned} U^*_{ij}(s_i) = U^*(y_i - p_j, s_i, C_j) = \quad & \max_{x_i} U(m_i, h_i) \\ \text{s.t.} \quad & m_i + C_j(x_i) = y_i - p_j \end{aligned}$$

That is, a consumer chooses a total amount of health care spending x_i conditional on their income net of the plan premium $y_i - p_j$, plan cost-sharing characteristics C_j and their realized health state s_i. U^*_{ij} is the indirect utility of individual i with policy j. Here, C_j reflects the typical non-linear insurance contract that maps plan cost-sharing features to consumer medical spending, for a given total medical spending level. Cardon and Hendel (2001) specifically model plan deductibles and coinsurance rates for a wide range of large group plans and develop a methodology to solve for optimal usage x_i given the peculiarities of dealing with the non-convex budget set induced by most health insurance contracts.

There are a large number of papers that provide insight into different impacts of health care demand, conditional on a given set of insurance characteristics. Surveying this whole literature is well outside the scope of this chapter. However, there are several papers that help to summarize key insights from the literature.

Brot-Goldberg et al. (2017) studies demand for health care using a natural experiment at a large employer that required all employees to switch from free health care on the margin to a non-linear contract with a deductible, coinsurance rate, and out-of-pocket maximum. This paper shows that consumers are price sensitive on average, reducing spending by 15% as a result of the increase in marginal health care prices. The paper finds that these spending reductions are not due to price shopping for cheaper providers but rather due to straight quantity reductions. Consumers reduce quantities of "wasteful" care as well as quantities of "high-value" care, where these categories are defined on a medical basis. These findings, both in direction and magnitude, are consistent with other key papers in the literature, including, e.g., the RAND Health Insurance Experiment discussed in Newhouse (1993).[25]

In addition, following on earlier work in the literature, this paper shows that consumers have unusual responses to the true marginal cost of care during the insurance plan year. During a plan year, under a typical non-linear insurance contract, consumers face a "spot" price, which is the price of care they actually pay in that moment. However, the true price of care at a given point of time is the "expected marginal price" that consumers face at the end of the insurance contract. Given the way typical health insurance contracts are shaped, the spot price is almost always weakly higher

25 See also, e.g., recent papers by Baicker et al. (2015), Hackmann et al. (2015), and Kowalski (2016).

than the expected marginal price (the Medicare Part D donut hole is a key exception). Brot-Goldberg et al. (2017) find that consumers respond heavily to spot prices, even when they diverge sharply from the expected marginal price (the price that a homo economicus would use). The paper shows that this has important implications. Spot price responses lead to spending reductions for otherwise sick consumers who have zero expected marginal prices of care (because they are very likely to reach the out-of-pocket maximums in their plans). These findings are highly consistent with the findings of Aron-Dine et al. (2015), Einav et al. (2010), Abaluck et al. (2018), and Dalton et al. (2019), which primarily focus on demand responses to non-linear contracts in Medicare Part D.

3.1.2 Health insurance demand: plan choice

Regardless of the specifics of the second stage model for health care demand given contract design, when considering health insurance plan choice consumers have expectations over these second stage health and financial outcomes and use those expectations as an ingredient into their choice. Cardon and Hendel (2001) set up a discrete choice model of plan choice for the first stage. In the first stage, consumers choose a plan j from a set of J_i plans offered by their employer and can also choose to be uninsured. The indirect utility from plan choice incorporates second period optimal behavior for each policy j:

$$V_{ij}(\omega_i, a_{ij}) = E(U^*_{ij}(s_i)|\omega_i) + a_{ij} \tag{10}$$

$$= \int U^*(y_i, z_i, Z_j)\pi_i(dz_i|\omega_i, D_i) + a_{ij} \tag{11}$$

Here, π_i is the distribution of health state s_i at the time of plan choice. This distribution is conditional on a consumer's private signal about their health state distribution, ω_i, and also on observable demographics D_i. a_{ij} represents policy-specific random tastes. The consumer's first stage problem is:

$$\max_{j \in J} V_{ij}(\omega_i, a_{ij}) \; \forall j \tag{12}$$

In this setup, the authors model self-selection into policies j based on consumers' signals ω_i and their demographics D_i.

Similarly to the moral hazard literature, the literature on plan choices, risk preferences, and information about health risk is now extensive and beyond the scope of this chapter to fully survey (see, e.g., Einav et al. (2021) for a broader discussion). We summarize key insights on each of these dimensions here.

Quite a few papers have focused on estimating risk preferences and their implications for insurance choice and welfare. While Cardon and Hendel (2001) estimate a homogeneous risk aversion parameter for their population, most subsequent papers estimate heterogeneity in risk aversion to account for different preferences in the population. Einav et al. (2013), Handel (2013), and Carlin and Town (2009) estimate risk preferences for consumers choosing from plan menus that are only financially differentiated (which helps to eliminate confounds from non-financial plan attributes).

Abaluck and Gruber (2011), Ketcham et al. (2012), and Abaluck and Gruber (2016) are examples of papers that estimate risk preferences from larger plan menus in Medicare Part D where there is plan heterogeneity in non-financial characteristics.[26] These papers, as Cardon and Hendel (2001) do, typically assume constant absolute risk-aversion (CARA), which means that consumers' preferences for risky gambles don't change with their underlying wealth levels. This is a convenient assumption in the literature, so that wealth doesn't have to be modeled in a sophisticated way empirically.

The papers estimate a range of values for risk protection that are context-specific. The papers in employer markets typically find positive but moderate benefits from risk protection (in a welfare sense) and meaningful heterogeneity in those benefits. Generally speaking, the specific risk aversion coefficients estimated are likely not to be that useful out of context, because risk preference estimates are known to be quite sensitive to scaling gamble sizes (see, e.g., Rabin (1998)). However, these estimates should be quite relevant for thinking about the welfare benefits from risk protection in the given empirical context they are estimated, as long as the scenarios studied do not deviate too far from the scale of the gambles observed empirically. Another useful paper in this space is Aron-Dine et al. (2012), which doesn't structurally estimate risk preferences but does study how the ranking of choice riskiness is correlated within-individual across different financial products. They find that there is a high correlation between the riskiness of choices across contexts, suggesting that, even if these choices do not necessarily lead to a consistent structural risk aversion parameter, they are consistent with a heuristic model of risk preferences.

A second key element of structural insurance choice models is consumer beliefs about their distribution of out-of-pocket expenditures for plans in their choice sets. Most recent papers leverage detailed individual-level claims data sets to predict spending for the upcoming health plan year using (i) information about plan cost-sharing designs and (ii) claims diagnoses and costs. Claims diagnoses at the individual level have been especially helpful in projecting individual-specific risks going forward, since different diagnoses have different degrees of persistence and cost. Typically, papers model underlying health risk and then project it onto different plan designs, potentially allowing for moral hazard as in stage two of the Cardon and Hendel (2001) model described above.

Due to regulation that restricts pricing on observable variables as well as the granularity of individual-level data, most papers focus on using observable information to model risk. They typically perform a *correlation test*, or some variant of one, to help show that concerns of selection on private information are not particularly important to model for the questions the papers seek to answer. The correlation test estimates a plan choice regression with rich observables, a plan cost regression with similar observables, and then asks whether the residuals of the two regressions are correlated in the sense that those whose residuals favor more generous coverage are

[26] There are also quite a few influential papers in this literature in non-health insurance contexts. See, e.g., Cohen and Einav (2007) and Barseghyan et al. (2013).

also those whose residuals favor higher costs. See Chiappori and Salanie (2000) for an early reference in this literature and see Einav et al. (2010a) for a more extensive discussion of this literature. We note that some papers in the literature specifically seek to identify private information about health risk separately form moral hazard. For example, Cardon and Hendel (2001) uses the model described above, together with National Medical Expenditure Survey data, to test for asymmetric information in consumers' plan choices across a range of large employers. The authors separately identify adverse selection (via the signal ω) and moral hazard (via how x_i depends on C_j) with the assumption that choice of employer does not depend on the plans that employer offers.[27] In their data, the authors can't reject a null hypothesis of no private information used in plan choice, they find a high level of risk aversion, and they find a price elasticity of care similar in magnitude to that found in prior work, e.g. in the RAND Health Insurance Experiment (Newhouse (1993)).

Risk aversion and health risk estimates are key foundations of essentially all health insurance choice models. In addition, when plan networks vary this is an important potential source of value that consumers consider. See the earlier part of this chapter for a detailed discussion of consumer demand for medical providers and insurance plan networks. For non-financial, non-provider plan features, such as, e.g., brand equity and care rationing, we don't present a general discussion though we do address these important aspects in relevant papers throughout the rest of this chapter. Now, we turn to a detailed discussion of one other important demand dimension in insurance markets: choice frictions and behavioral foundations.

3.2 Choice frictions: active choice issues

Quite a few papers have ventured beyond the classical frameworks for insurance demand to study choice frictions and behavioral choice factors in insurance markets. These issues have been shown to be important in insurance markets, which are notoriously difficult for consumers to make choices in. Research in this area has been useful for studying health insurance market design and has also been useful for broader insights into the demand for complex financial products.

Typically, papers modeling these kinds of frictions start from the classical framework just described and modify that framework to account for choice problems. The literature naturally separates into two branches. The first studies active choice issues, where inertia / adherence to a default option is not the focus. The second studies inertia and passive choice issues, which are especially important in markets like health insurance where consumers have default options and the products are complex to evaluate.

[27] This assumption, which has been used quite frequently in the literature, implies that insurance choice sets are exogenous at the individual level. As a result, the authors can compare similar consumers facing different plan choice sets, and thus similar consumers who end up in different plans. This provides an instrument to identify consumer responses to cost sharing separately from their ex ante unobserved information about their health.

Handel and Kolstad (2015b) is one example of a paper modifying the classical expected utility setup to account for choice frictions. By sticking closely to the classical expected utility model, the paper shows how bringing additional data to bear on consumers' lack of knowledge (interpreted as the result of information frictions) impacts the conclusions that are drawn, relative to assuming biases and frictions away in a classical expected-utility framework.[28]

In this paper, where consumers choose from a set of plans offered by a large employer, the consumer's problem is to choose a plan j from set $\mathcal{J}$.

Consumer i's utility in health plan j, ex post to the realization of their health state, is:

$$u(W_i - P_{ij} + \pi_j(\psi_j, \mu_i) - s, \gamma_i). \tag{13}$$

u is assumed to be a concave utility function, implying that consumers have diminishing marginal utility for wealth and are risk averse. Following earlier work, such as, e.g., Cardon and Hendel (2001), consumers are assumed to have constant absolute risk aversion (CARA), meaning that the curvature of the utility function doesn't depend on baseline wealth.

This ex-post utility includes several components, some of which are the same regardless of health during the year. W_i is consumer wealth and P_{ij} is the premium contribution an individual i pays in plan j. π_j reflects the consumer's value for non-financial plan characteristics, such as provider networks or tax-advantaged health-savings accounts: this depends on plan characteristics ψ_j and a consumer's health type μ_i. In this formulation, each of these components is assumed to be independent of the health-risk realization.[29] Finally, the payment s is the consumer's out-of-pocket payment for health care, given an ex post realization of their health risk. This approach for modeling ex post utility is similar to the second-stage model in Cardon and Hendel (2001) discussed above, with the obvious omission here that consumers are assumed to have no moral hazard.

We now turn to *ex ante* consumer utility, similar to stage one in the Cardon and Hendel (2001) framework. Assume that a consumer faces uncertainty about their out-of-pocket spending in a given plan j, following the probability distribution $f_{ij}(s|\psi_j, \mu_i)$. The distribution of payments depends on the plan design and the consumer's health-risk type. Given this uncertainty, a consumer's expected utility for plan j is:

$$U_{ij} = \int_0^{\infty} f_{ij}(s|\psi_j, \mu_i) u(W_i - P_{ij} + \pi_j(\psi_j, \mu_i) - s, \gamma_i) ds. \tag{14}$$

Within this setup, the consumer will choose the plan j that maximizes her expected utility U_{ij}. Handel and Kolstad (2015b) depart from this baseline expected utility

[28] One potential downside of sticking closely to the classical expected utility framework is that the paper may mis-specify the underlying model for why mistakes occur.

[29] In certain settings, one may want to model π as a function of the ex post risk realization as well, since provider networks and health risk interact. We don't do so here for simplicity.

model by allowing for the consumer's beliefs (notated with "hats") to deviate from what they would be under full information and rational expectations:

$$\widehat{U_{ij}} \quad = \int_0^\infty f_{ij}(s|\widehat{\psi_{i,j}}, \widehat{\mu_i})u(W_i - P_{ij} + \widehat{\pi_{i,j}}(\widehat{\psi_{i,j}}, \widehat{\mu_i}) - s, \gamma_i)ds \tag{15}$$

Here, beliefs about plan characteristics, health risk, and health benefits are modeled allowing for both population-level and individual-level departures from the rational-model values.

Empirically, this framework allows for departures from baseline beliefs and information due to information frictions or biases more broadly. These frictions and biases may result from consumers not having easy access to key information; consumers not attending to readily available information; or consumers having difficulty integrating certain types of information into decisions. Handel and Kolstad (2015b) consider data from a large firm with over 50,000 employees where employees choose between two plans: a broad network PPO plan with no premium and no (in network) cost sharing, and a high-deductible plan with the same network and a linked health savings account subsidy (essentially a reverse premium). The paper presents descriptive evidence showing that consumers seem to substantially under-purchase the high-deductible plan (HDHP) based on its financial value relative to the simpler PPO option. The standard non-behavioral explanation is that these purchasing patterns reflect consumer risk aversion—but the degree of risk aversion necessary to rationalize these choices is very high. Given this backdrop, the authors implement a comprehensive survey to measure consumer information sets shortly after they make plan choices during open enrollment. The survey asks multiple choice questions to consumers about all aspects of plan choice, including perceptions about the health savings account subsidy, provider networks, and financial characteristics such as deductibles or coinsurance. In addition, the survey asks about perceived hassle costs of enrolling in a high-deductible plan where medical bills and health savings accounts may involve time and hassle costs relative to the hassle-free PPO option. The survey is linked to enrollment and detailed claims data at the individual-level, allowing the authors to study how individual choices relate to limited information. The authors show that consumers who lack knowledge about the high-deductible plan relative to the PPO plan are more likely to leave substantial sums of money on the table in their plan choices. The key point is that this money left on the table is not due to risk aversion, but to frictions or biases that result in limited knowledge.

The primary structural model the authors estimate is a baseline expected utility model with shifters that reflect changes in willingness-to-pay for the high-deductible plan as a function of limited information about that plan (as measured in the survey). This is very similar to the theoretical model in Eq. (15) but incorporates measures of limited information in a specific way. The main specification is:

$$U_{ij} \quad = \int_0^\infty f_{ij}(s)u_i(x_{ij})ds \tag{16}$$

$$u_i(x) = -\frac{1}{\gamma_i(\mathbf{D}_i)} e^{-\gamma_i(\mathbf{D}_i)x} \tag{17}$$
$$x_{ij} = W_i - P_{ij} - s + \eta(\mathbf{D}_i)\mathbf{1}_{j_t=j_{t-1}} + \mathbf{Z'}_i\beta\mathbf{1}_{HDHP} + \epsilon_{ij}. \tag{18}$$

Here, U_{ij} is an expected utility function for a risk averse consumer, following the model just discussed. Eq. (17) describes the functional form used to implement the constant absolute risk aversion model. x_{ij} measures the outcome (translated into monetary units) for each consumer during the year, given a realization of their health uncertainty. η is a term that addresses consumer inertia, modeled as an implied switching cost. Risk aversion γ and inertia η both vary with observable demographics D_i.

The authors include indicator variables related to consumers' information sets in the vector **Z**. For each question, they construct indicator variables for 'informed', 'uninformed', or 'not sure' answers as well as variables derived from answers to questions about hassle costs and knowledge of own health expenditures. $Z = 0$ indicates that a consumer is perfectly informed, while $Z = 1$ indicates that a consumer lacks information on a certain dimension. The coefficient β then measures the impact of that lack of information on willingness-to-pay for the high-deductible plan relative to the less complex PPO option.

This empirical approach to studying the impact of consumer frictions and biases has several advantages and disadvantages. One advantage is that measuring effective consumer information sets with surveys is often feasible. Another advantage is that the approach is simple, in the sense that the estimates tell us about the impact of survey-measured limited information on willingness-to-pay for different options. One disadvantage is that it doesn't posit a specific structural mechanism for how limited information impacts choices: a more structured version would allow for answers to survey questions to imply something specific about the precise nature of beliefs. But it is also difficult to link the responses directly to belief objects. This disadvantage makes it difficult to assess whether specific policy interventions to improve consumers' choices would be successful. Another potential disadvantage is that the baseline model used is a specific expected utility model that does not capture behavioral notions of how consumers respond to risk and uncertainty, which is an important topic.[30]

Handel and Kolstad (2015b) offer several results on the knowledge consumers lack and the resulting amount of money they leave on the table. The most influential gaps in knowledge are about available providers and treatments, and the perceived time and hassle costs for the high-deductible plan. For example, a consumer who

[30] While we are unaware of empirical papers studying non-standard consumer responses to risk and uncertainty in health insurance, Barseghyan et al. (2013) study non-linear probability weighting for consumers choosing car and property insurance policies and Grubb and Osborne (2015) studies overconfidence and myopia in cellular phone markets. These projects structurally identify alternative choice models, but typically assume full consumer information to do so. It should be possible to combine the Handel and Kolstad (2015b) approach with these others.

incorrectly believes that the PPO option grants greater medical access than the high-deductible plan (they grant the same access in reality) is willing to pay $2,267 more on average for the PPO over the one-year period of the insurance contract than a correctly informed consumer. Aggregating across all included measures for incomplete knowledge, the average consumer is willing to pay $1,694 more for the PPO relative to a fully informed consumer with zero perceived hassle costs. Consumer perceptions of relative hassle costs, which likely overstate true hassle costs, have a major impact, equaling approximately $100 per perceived extra hour of time spent on plan hassle.

Next, they find that including measures of consumer information into the model together with risk aversion significantly changes estimates of risk aversion. Framed in terms of a simple hypothetical gamble, a consumer with baseline model risk aversion (where information frictions are not taken into account) would be indifferent between taking on a gamble in which he gains $1000 with a 50 percent chance and loses $367 with a 50 percent chance. In other words, he would have to be paid a risk premium of roughly $633 in expectation to take this risky bet. In the primary model with survey variables included, the consumer is instead found to be indifferent between taking on a gamble with a $1000 gain and $913 loss (with 50% chance of each). This has meaningful implications for policy, for example altering conclusions of the benefits of forcing consumers into high-deductible plans.

It is important to note that this is just one example of quite a few papers that document consumer difficulties in making plan choices. Since our focus on this chapter is on the industrial organization implications of these choice issues, we describe these papers briefly (see Chandra et al. (2019) for a more extended treatment of these topics). Bhargava et al. (2017) document mistakes in active insurance purchases at a large firm where (i) many of the choices are dominated by other choices and (ii) many consumers choose those dominated choices, leaving meaningful sums of money on the table. Abaluck and Gruber (2011) show that consumers forego substantial savings in Medicare Part D choices, controlling for spending risk, risk preferences, and average brand preferences. Abaluck and Gruber (2011) find that a key reason consumers lose money on their plan choices is that they overweight premiums by a factor of 5 to 1 relative to expected out-of-pocket spending.[31] Heiss et al. (2010) also study consumer choice quality in Medicare Part D and find results that are consistent with those from Abaluck and Gruber (2011). Ketcham et al. (2012) show similar patterns in Part D plan choices and also study whether consumers learn to make better choices over time. They find evidence of poor consumer choices but, leveraging panel data,

[31] Despite the fact the consumers systematically seem to overweight premiums relative to their potential out-of-pocket spending, throughout many papers in the literature, it is not obvious that this means consumers have "overly generous" coverage relative to what a social planner would want. There are several reasons including (i) consumers may also over-weight salient cost-sharing features like deductible level relative to predicted out-of-pocket spending and (ii) Brot-Goldberg et al. (2017) and Abaluck et al. (2020) show that less generous coverage can, respectively, lead to consumers under-consuming high value care and lead to higher mortality rates downstream. Thus, it is still an open question whether or not the choice frictions documented in insurance choice lead to overly generous coverage, especially when viewed in light of choice frictions related to care consumption.

find that consumers may make better choices over time as they gain experience in the market. Specifically, they find that consumer overspending is reduced, on average, by $298 in their second year in the Part D market relative to their first. Some of this may be due to plan switching and some to plans delivering better value over time. Gruber et al. (2020) show that even trained intermediaries (brokers) struggle to make high-value choices on behalf of the consumers they serve, though brokers' choices improve when given access to sophisticated decision support.

3.3 Choice frictions: inertia and passive choice issues

In the last section we discussed choice frictions in insurance markets when consumers are actively considering insurance choices, i.e. they don't have a default option or they are actively considering different options despite having a default option. There has been as much, if not more, research on inertia in health insurance markets, which causes consumers to leave substantial surplus on the table. Inertia is especially large in typical insurance markets because consumers typically have a default option, usually their previously chosen plan. As shown in many contexts, defaults can be quite powerful and substantially lower active shopping and price sensitivity. Consumer inertia reduces the quality of consumer choices in such settings, as products evolve over time and consumers do not adjust accordingly.

Handel (2013) studies inertia using data from a large employer that spans six years (2004-2009). The employer changed the menu of options employees could choose from during the middle of this time frame and forced all employees to make active (non-default) choices from this new menu of options. Following that forced active choice, consumers had a default option of their previously chosen plan, despite the fact that the plan premiums and features changed significantly over time. The paper presents several pieces of descriptive evidence suggesting that inertia causes consumers to leave meaningful sums of money on the table. First, one product changed over time such that it became dominated by other options and, despite losing over a thousand dollars for sure, consumers continued to enroll in the newly dominated plan when it was their default option. Second, the active choices that new enrollees made were significantly better (in terms of money left on the table) than the choices of similar incumbent employees who had a default option. While active choices are far from perfect, choices become worse in an environment with a sub-optimal default option.

The paper estimates a structural model of consumer inertia, modeled as a switching or adjustment cost that could result from consumers having research / paperwork costs of switching or learning costs of using a new plan. The expected utility framework is similar to that in Cardon and Hendel (2001) as described in Eqs. (10). In Handel (2013), the money at stake for consumers in each ex post realized health state is;

$$x_{ij} = W_i - P_{ij} - s + \eta(\mathbf{X}_i^B)\mathbf{1}_{j_t=j_{t-1}} + \epsilon_{ij}. \tag{19}$$

Here, W_i is baseline consumer wealth, P_{ij} is the premium paid up front, and s represents out-of-pocket spending in plan j under the realized health state. Inertia is quantified by the amount of money consumers are willing to leave on the table to stick with their incumbent plan. In effect, the premium for the incumbent plan is lowered by η for consumers in this model. η is allowed to depend on observable characteristics, X, including other benefits choices consumers make (such as flexible spending account choices that must be actively made every year).

Inertia in this environment (and most health insurance environments) could result from any of the following micro-foundations:

1. **Switching costs:** Consumers could incur paperwork or hassle costs of switching plans. Consumers may also incur adjustment costs to learn how to use their new plan, or costs associated with needing to switch care providers. While this last cost (of switching providers) is not an issue in the Handel (2013) analysis, such costs will be relevant in many settings.
2. **Search costs:** Consumers could incur costs of searching through the different available plan options to determine if they want to switch. Typically, this would be modeled as a two stage model (as in Ho et al. (2017) described below) where consumers first decide whether to search and then decide whether to switch after searching.
3. **Inattention:** Consumers could be inattentive. They could rationally decide not to engage in the search process because search is too costly relative to expected benefits. Or they could less rationally neglect potential benefits of carefully considering plans and plan options.
4. **Naive present bias:** Consumers could believe that they will conduct research and make a new choice right before the choice deadline, but then when the time arrives not be willing or able to invest the time and effort to do so.

Handel (2013) does not distinguish between these micro-foundations, but shows how welfare conclusions are sensitive to the micro-foundations. In particular, his welfare analysis allows for a range of results that depend on whether or not inertia primarily results from a rational response to costs (e.g., of search) or a less rational response to perceived benefits and/or perceived costs.

The paper finds that consumers exhibit significant inertia: on average, consumers with a default option are estimated to leave \$2,032 on the table annually to stay with their default. Consumers who also make active flexible spending account elections leave an average of \$551 less on the table. Families, who have more money at stake, leave \$751 more on the table than single employees. There is no evidence that recent health shocks lead to active choices. The paper studies counterfactual policies where the extent of inertia is reduced by some proportion and consumers re-choose plans in the market. In the partial equilibrium analysis where plan prices do not adjust from re-sorting, a 75% reduction in the magnitude of inertia improves consumer welfare by 5.2% of paid premiums. Later in this chapter, when we discuss market design, we will discuss the part of this paper where prices are allowed to re-adjust as consumers make better choices due to reduced inertia.

Quite a few other papers study inertia in health insurance markets and show that it causes meaningful financial losses for consumers. Ho et al. (2017) study inertia in

Medicare Part D with a model of inattention. They model consumers with a default option making choices in two stages. First, they decide whether or not to engage with the market. This decision is influenced by a series of shocks (e.g., changes to the premium of their current plan) related to the market and their default option. Second, consumers who decide to engage in the market choose a plan following a standard active discrete choice model, where consumer i's utility for option j is denoted by $u_{i,j}$.

As the market evolves over time, consumers costlessly learn about how their current plan changes but have to pay a cost ε to learn about how the characteristics of other plans change. Consumers choose to pay this cost if the expected benefit of doing so outweighs the cost:

$$\mathbb{E}\left[\max_{j=1....J} u_{i,j,t+1}|\bar{X}_{i,k,t+1}\right] - u_{i,k,t+1} > \varepsilon_{i,t}.$$

Here, plan k is the choice a consumer is currently enrolled in and $\bar{X}_{i,k,t+1}$ includes the known characteristics for that plan. The expectation is taken over the characteristics of other plans that the consumer discovers if she pays the cost to search through the set of available plans. If the consumer pays the cost to search then she learns the characteristics of all plans in the market. The consumer is more likely to search if (i) she has a health shock that changes the value she receives from different plans; (ii) the characteristics of her current plan change; and/or (iii) she receives a signal that the market significantly evolved to make search valuable.

Empirically, the authors estimate this model without fully specifying consumers' beliefs about other options in the market prior to search. They model consumer attention as being a function of whether they experience shocks (v) that cause them to pay attention:

$$v_{i,t} = v_{i,p,t}\beta_1 + v_{i,c,t}\beta_2 + v_{i,h,t}\beta_3 + v_{i,e,t}.$$

Here, $v_{i,p}$ equals 1 if there is a premium increase for a consumer's own plan that exceeds the median weighted increase in the market; v_c equals 1 if there is a meaningful change to the out-of-pocket coverage characteristics for a consumer's own plan; v_h equals 1 if the consumer experienced an acute health shock in the past year, e.g., a significant increase in drug spending; and v_e is a random shock that spurs consumer search. With this framework, a consumer searches if her composite shock v is greater than some threshold value (related to $\varepsilon_{i,t}$ above). Then, if the consumer searches, she picks the plan that maximizes her expected utility, with full updated knowledge of all plan characteristics. If the consumer does not search then she remains in the plan that she is already enrolled in.

Ho et al. (2017) find substantial inertia in the Medicare Part D context: only approximately 10% of consumers switch plans each year and enrollees leave a lot of money on the table by not switching. Consistent with the model of inattention, consumers are more likely to switch when their own plan features (e.g., premium or cost-sharing) change but are less likely to search when alternative plan features change by similar amounts. The paper then studies how insurers price given the degree of inertia in the market, which we discuss later in this chapter. It is interesting to note that Handel (2013) and Ho et al. (2017) use similar data and identification strategies to study inertia, but assume different micro-foundations. In ongoing work, Brot-Goldberg et al. (2021) do distinguish between the potential micro-foundations underlying default effects, using the fact that Medicare Part D randomized default options for low-income individuals entering the market. They show that consumers almost always stick with the random default option and that, when switched out of that option, they stick with a new plan despite only having to pay a very small amount to switch back to their old plan. Consumers stick with default options even when the plans don't include drugs on the formulary that they clearly need. This paper strongly suggests that inattention is the primary driver for inertia and default effects in their context, rather than switching costs. Future work that continues to empirically distinguish between mechanisms for inertia in different populations and contexts will be valuable in this literature (Handel and Schwartzstein (2018)).

A range of other papers also document inertia in Medicare Part D. These papers include, but are not limited to, Ericson (2014), Polyakova (2016), Heiss et al. (2016), and Abaluck and Gruber (2016), with each approaching the inertia question from a distinct angle. In addition, Abaluck and Gruber (2016) find limited evidence that consumers learn to shop effectively for plans over time, contrary to the findings in Ketcham et al. (2012). Abaluck and Adams-Prassl (2021) and Coughlin (2021) model consumer consideration sets, where consumers only consider a subset of plan options at a given point in time, and apply this to studying inertia in plan choice in Medicare Part D. Finally, in the large employer and Medicaid managed care contexts, respectively, Strombom et al. (2002) and Marton et al. (2015) both show significant value left on the table due to consumer inertia.

4 Health insurance market design and regulation

4.1 Frameworks for studying insurance market design

A key premise underlying many health care markets around the world is that competition between payers (insurers) will reduce costs and lead to higher quality care. This paradigm, typically referred to as "managed competition," acknowledges that, though such competition can be valuable, it needs to be more heavily regulated than typical product markets (see, e.g., Enthoven et al. (2001)). There are a number of key reasons for why such regulation is important, including that (i) health care is often viewed as a "right" so that public subsidies often play a role in provision and that (ii) adverse selection is a potential concern. Adverse selection can occur both on the extensive margin (into or out of a market) or on the intensive margin (across plans in a given market) and manifests when consumer costs are correlated positively with choosing generous plans. Government regulation can help mitigate both dimensions of selection in a variety of ways, as we discuss in more detail below.

To study the design of insurance markets, it is useful to have in mind a benchmark model capturing firm behavior in a typical "managed competition" style market, often also referred to as an "exchange." There are two useful workhorse models in the literature, both of which effectively assume perfect competition between firms. The first is Einav et al. (2010b), who set up a model with the following key assumptions:

1. Competing insurers offer one specific regulated insurance product
2. Insurers cannot price discriminate between consumers
3. All consumers are allowed to purchase insurance from any insurer that they want to (guaranteed issue)
4. If you don't purchase insurance in the market, you get a baseline outside option that is publicly provided (e.g. traditional Medicare for seniors over 65 in the U.S.)

Regulated insurance products, no price discrimination, and guaranteed issue are all hallmarks of managed competition-style markets. The second useful workhorse model, Handel et al. (2015), makes similar assumptions to Einav et al. (2010b), with two key differences:

1. Competing insurers can offer multiple regulated classes of insurance options
2. There is a fully enforced mandate such that consumers must buy one option offered in the market

While the Einav et al. (2010b) assumptions may better reflect privatized add-ons to public insurance (e.g. Medicare Advantage HMOs), the Handel et al. (2015) assumptions may better reflect regulated exchanges with multiple types of plans and an insurance mandate, similar to the markets set up under the Affordable Care Act (ACA) in the United States.

The details of these studies are discussed in more depth in the handbook chapter on selection markets included in this volume [Einav et al. (2021)]. Here, we discuss the key implications of these two studies for studying the industrial organization of health care markets.

We present a model that incorporates both frameworks, following the comparison between these two papers presented in Weyl and Veiga (2017). Assume that a unit mass of individuals must purchase insurance that has either low baseline generosity L or higher generosity H. The prices of these options are (p_L, p_H) and $\Delta p = p_H - p_L$. We will discuss momentarily how these prices are determined. An individual's utility for H as opposed to L, not factoring in price, is denoted u. An individual buys plan H if their incremental utility u is greater than Δp.

Define $c_H(u)$ as the average cost of those enrolling in plan type H, i.e. $c_H(u)$ is $E[c_H|u \geq \Delta p]$ and $c_L(u)$ as $E[c|u < \Delta p]$.

Consider two different institutional setups for pricing. The model in Handel et al. (2015), which Weyl and Veiga (2017) refer to as "total pricing" (TP), each type of plan offered is responsible for its own costs, there is a fully enforced mandate, and no outside option. Insurers in the competitive market break even, such that, in any equilibrium:

$$p_H = E[c_H | u \geq \Delta p]$$
$$p_L = E[c_L | u < \Delta p]$$

Henceforth we will use the shorthand $\Delta P = \Delta AC$ to describe these equilibrium restrictions. Note that Handel et al. (2015) also discuss what happens in boundary cases where all consumers shift to one plan or the other. More broadly, in that paper the authors present an in depth discussion of the game theoretic assumptions required to (i) ensure that an equilibrium exists and (ii) determine which potential candidate equilibrium survives as the unique equilibrium. For a more in depth discussion of these points see both Handel et al. (2015) and the Handbook chapter in this volume by Einav et al. (2021). For the rest of this discussion, when discussing TP, we will operate under the assumptions Handel et al. (2015) maintain to ensure that an equilibrium exists and is unique.

Alternatively, Weyl and Veiga (2017) discuss the framework set up in Einav et al. (2010b) as one of "incremental pricing" (IP). Under IP, c_L is covered by the baseline provider, e.g. the federal U.S. government in the case of traditional Medicare. The high provider, who is now providing top up coverage relative to L, has to break even, but only on the incremental costs relative to c_L, rather than on the spread in average costs between those enrolling in L and those enrolling in H. Specifically, average cost for the competitive provided top up plan is:

$$AC_H = E[c_H - c_L | u \geq \Delta P] \tag{20}$$

The equilibrium in an IP market occurs where p_H is set such that $AC_H = \Delta P$.

Weyl and Veiga (2017) show a number of key relative properties for TP markets relative to IP markets. First, they show that, under an assumption of global adverse selection (expected costs are increasing in u), the average cost wedge relevant to equilibrium pricing is always going to be larger in TP than in IP. The difference in this wedge across these two setups is:

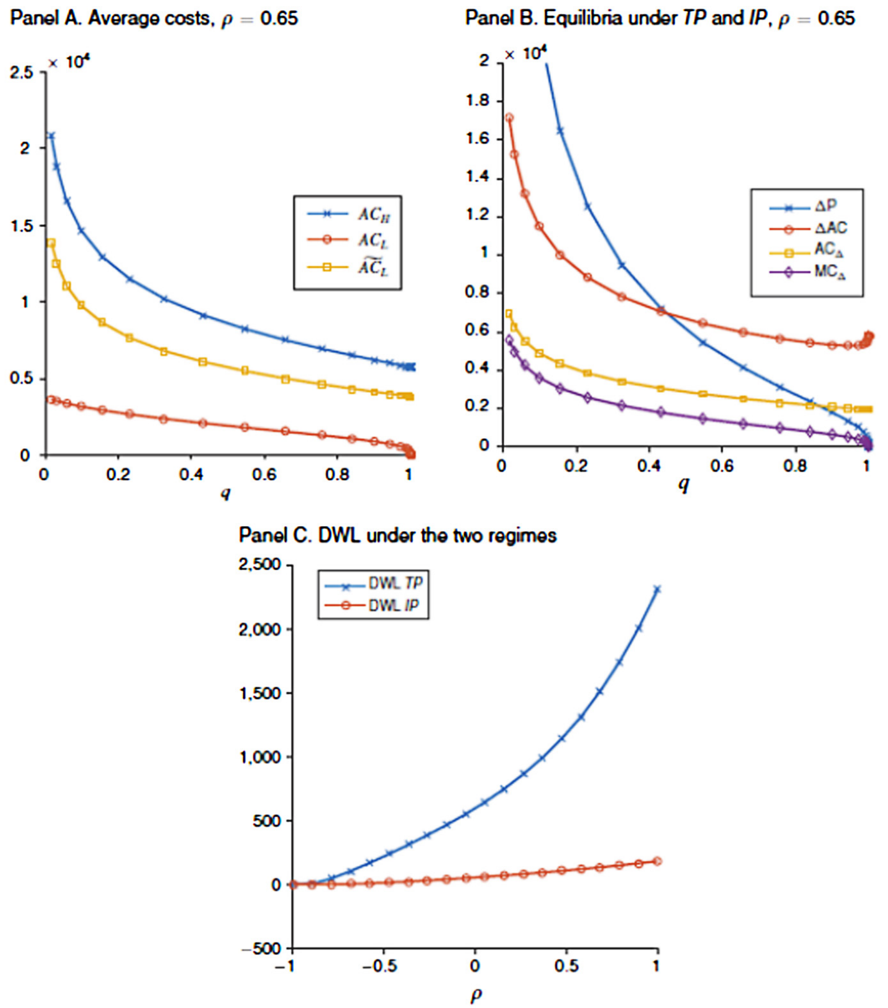
$$E[c_L | u \geq \Delta P] - E[c_L | u < \Delta P] \tag{21}$$

Weyl and Veiga (2017) show that, under certain reasonable assumptions, this implies that *for any* ΔP, ΔAC in TP is going to be bigger than AC_H in IP. This, in turn, implies that:

$$\Delta P{*}_{TP} > P{*}_{IP} > P * q^H_{TP} < q^H_{IP} < q^H_{*} \tag{22}$$

Here, P* is the socially optimal price differential, while the other starred prices are the equilibrium price differentials under each of the institutional setups. The key finding is that under TP, adverse selection will be stronger, market unraveling higher, and the deadweight loss from adverse selection higher, relative to IP.

Weyl and Veiga (2017) show that the assumed institutions have quantitatively important implications for results. Fig. 1, reproduced from Weyl and Veiga (2017),



Notes: In panel A, the difference between the top and middle curves is the relevant average cost under *IP*, while that between the top and bottom curves is relevant under *TP*. In panel B, the curves determining equilibrium for the baseline calibration, which also show the magnitudes of the local distortions characterized by Proposition 1. In panel C, DWL under the two institutions for alternative values of ρ.

FIGURE 1

This figure is reproduced from Weyl and Veiga (2017). Copyright American Economic Association; reproduced with permission of the AEJ: Microeconomics.

illustrates the differences between these two institutional setups, using parameter estimates from Handel et al. (2015).

Panel A in the figure plots three average cost lines as a function of ΔP. The difference between the top and the middle line reflects AC_H under IP, while the difference between the top and the bottom line reflects ΔAC under TP. Quantitatively, the simulations the authors run show that ΔAC under TP is approximately 3 times larger than the average cost of the high plan under IP. Panel B shows the equilibrium implications of this. Under TP, the market share enrolled in more generous coverage in equilibrium is about 0.4. Under IP, it is about 0.9. The social optimum has everyone enrolling in more generous coverage, since u > MC for all consumers by assumption. Thus, the differences in adverse selection and coverage generosity are meaningfully different under these two sets of institutional assumptions. Panel C shows the deadweight loss from adverse selection under each setup as a function of the underlying correlation between health risk and insurance value. For the value of this correlation estimated based on Handel et al. (2015), and in general as this correlation becomes larger, the deadweight loss from selection under TP institutions is much larger than that under IP institutions.

For further discussion on the specifics of these papers and models, see Einav et al. (2021). For our purposes here, there are several key takeaways as we move towards our discussion of insurance market function and regulation:

1. The assumed institutional setup (TP vs. IP) can have big implications for positive and normative estimates, including the degree of adverse selection
2. In empirical work, this distinction generally matters more when conducting counterfactuals (which use the model assumptions to project new scenarios) rather than in-sample analysis, where the observed pricing and shares tell us a lot about these key quantities, irregardless of the underlying model assumptions
3. Taken literally, there are quite a few policy environments where each framework applies well. For example, Handel et al. (2015), with the TP model, likely applies more naturally to ACA-like exchanges, especially when an effective insurance mandate is in place, while Einav et al. (2010b), with the IP model, likely applies better to a top up market like Medigap supplemental coverage (prevalent in many systems around the world).

4.1.1 Modeling imperfect competition

Quite a few empirical papers in the literature focus on imperfect competition in insurance markets. It is more challenging to set up a general framework for imperfect competition in selection markets, in part because firms with market power also consider risk selection as a key dimension of pricing decisions (i.e. they care about who buys at a given price, as well as the number of buyers). Since these empirical papers, which we discuss later, do not typically discuss overarching concepts related to imperfect competition and selection markets, we outline some general insights from Mahoney and Weyl (2017) on this topic here.

Mahoney and Weyl (2017) set up a symmetric model of imperfect competition that parameterizes the extent of market power and the extent of selection based on risk. They study the interaction between these two phenomena and find that policies that enhance efficiency by combating adverse selection in competitive markets may actually be welfare reducing when the market is imperfectly competitive. Here, the key insight is that if the equilibrium quantity of insurance provided is relatively high, and the slope of marginal revenue is relatively flat, then flattening the average cost curve through risk-adjustment transfers decreases quantity, though the social planner wants to increase quantity. If equilibrium quantity is low and marginal revenue is steep, risk-adjustment transfers still increase quantity and reduce selection. The authors show the converse is true for markets with advantageous selection.

The key conceptual insight, which is relevant to the empirical papers discussed later in this chapter, is that the conceptual comparative statics typical of competitive markets with adverse selection may not hold in imperfectly competitive selection markets. Though, in practice, many of the empirical papers in this area assume perfect risk-adjustment (i.e. assume away adverse selection) it is important to keep the rich potential interactions between imperfect competition and selection markets in mind when thinking about these papers.

4.2 Managed competition exchanges: examples

There are a few key examples of managed competition exchanges that are oft-studied in the health and industrial organization literature. We briefly describe these markets here and describe further details in the next section when discussing the specific papers.

1. **Medicare Part D:** Medicare Part D, which began selling prescription drug insurance to seniors in 2006, is probably the most studied health insurance exchange market. This program, which serves over 50 million seniors, is one of the leading examples of privatized insurance benefits being sold through a regulated, competitive market. The Part D market sells voluntary "stand-alone" drug plans direct to consumers, though consumers can also obtain Medicare drug benefits if they elect a Medicare Advantage HMO plan that includes those benefits.
   Medicare Part D markets, which are typically organized at the state level, have had on average about 30 insurance plan choices per market. Choices are differentiated based on (i) insurer brand (ii) insurer drug formulary (iii) insurer administration, and (iv) plan cost-sharing characteristics that determine the out-of-pocket payments consumers make. Medicare determines premium subsidies given to consumers on a sliding scale that relates to income and, for the lowest income, also provides cost-sharing subsidies. The specific design of the premium subsidy scheme has been the point of much study and is something we will discuss in more depth later in this section.
   While there have been many papers focused on consumer demand for Part D plans (see some discussion earlier in this article) in this section we will focus on papers that study specific aspects of regulation, competition and supply in the Part D market.
2. **Medicare Advantage:** Medicare Advantage, formerly known as Medicare Choice Plus or Medicare Part C, is a program that allows seniors to opt out of traditional Medicare benefits and receive benefits instead from a private insurer.

Insurers operating under Medicare Advantage operate under a range of regulations that define what services and what providers they have to provide access to. Medicare Advantage plans are typically thought of as HMO plans that have low cost-sharing, but ration care through other means, e.g. through access to a specific network of providers. As noted above, Medicare Advantage plans can also offer drug benefits and thus substitute for standalone Part D plans.

Medicare Advantage beneficiaries receive subsidies from the federal government and pay premiums, similarly to Part D. Insurers generally cannot price discriminate and, though risk-adjustment transfers are in place to mitigate adverse selection, it is unclear if these transfers are particularly effective (Brown et al. (2014). Medicare Advantage has grown in popularity over time, enrolling 36% (24.1 million) of Medicare-eligible enrollees in 2020 (up from 24% in 2010. Imperfect competition has been an important concern in Medicare Advantage markets, where several carriers dominate the landscape. The nationwide C3 index is 59%, with United Health Care, Humana, and BCBS making up over half the nationwide market share. In addition, since plans are offered on a county-specific basis, there are many counties where a low number of plans are offered, increasing concerns about market power. We discuss research related to imperfect competition in Medicare Advantage in more depth later in this chapter. See Kaiser Family Foundation (2020) for more details on the Medicare Advantage program.

3. **State exchanges:** The Affordable Care Act, passed in the United States in 2010, created the template for state-by-state insurance exchanges to be set up for consumers in the United States. The goal of the exchanges is to fill in the gaps and insure consumers who are not covered by an employer, Medicare, or Medicaid. Exchanges are either operated independently by states or run by the federal government (federally facilitated exchange).

   Plans typically are offered by county, and, unlike Medicare Part D, consumers typically have between 2 and 10 options to choose from. The lower number of plans has raised meaningful concerns about access and imperfect competition, especially in markets with a low number of plans.

   State exchanges follow federal regulations for managing competition and increasing access. Subsidies are given by the federal government according to a sliding scale between 133-400% of the federal poverty line (FPL). In addition to these subsidies, which increase access, price discrimination in essentially prohibited in the exchanges, increasing access for the predictably unhealthy. The Affordable Care Act also included a broad mandate to purchase insurance, though consumers can get insurance from any feasible source. One key difference between the exchanges and other managed competition setups (such as Part D) is that the majority of consumers have insurance through another source (employer sponsored coverage) so exchange designers need to be heavily concerned about substitution in and out of the market, as well as substitution across plans with the market.

   In addition to the state exchanges set up under the Affordable Care Act, several exchanges operated prior to the onset of this law, including the Massachusetts Connector Exchange, which has been the focus of much research. See Kaiser

Family Foundation (2011) for further details on the initial regulation of exchanges set up under the Affordable Care Act.

Of course, in addition to these key examples, there are many other examples of important regulated insurance markets both within and outside the United States. We discuss further examples outside of the United States later in this chapter. With these market examples in mind, we now discuss these markets and related regulation in more detail, following the conceptual and empirical literatures.

4.3 Regulation of competitive insurance markets

In the managed competition health insurance paradigm, there are a number of policy decisions that are made repeatedly, over time and across markets. Since these policy decisions are the focus of much conceptual and empirical work in the literature, we lay out some of the key policy decisions here before getting to the specific empirical papers that address them later on in this chapter.

1. **Subsidies:** health insurance markets almost always provide subsidies to consumers, due to the high cost of insurance and the fact that health care, in contrast to most other products, is often viewed as a human right. Both exchange markets and employers provide subsidies for insurance purchase and these subsidies typically cover well over half of plan premiums.
   This has several implications. First, in a short-run sense, employers and the government are responsible for much of the cost of health care consumed by enrollees. Second, even for a given average subsidy level, the way that subsidies are set can have meaningful implications for plan pricing and plan design. Enthoven et al. (2001) advocates for lump-sum subsidies where consumers have to pay fully on the margin if they want more expensive insurance coverage. This has the benefit of making consumers more price sensitive and having them bear the cost, e.g., of higher priced providers or very generous risk-protection. In turn, this should help stiffen competition, both in premiums and insurance product design.
   However, in practice lump-sum subsidies have several drawbacks. First, for lower-income consumers, they can make more expensive plans completely unaffordable, raising significant equity concerns. Second, Jaffe and Shepard (2020) note that subsidies that are linked to plan premiums, i.e. that rise with premiums to some degree, can help exchange designers mitigate uncertainty about premium changes over time. The central tradeoff discussed in this paper is an important one for subsidy policy design. On the one hand, price-linked subsidies can be helpful in preventing large fluctuations in the plan premiums seen by consumers over time. On the other, linking subsidies to prices softens price competition, since a plan price increase is automatically met by increased public payments, dampening any demand response.
   The authors theoretically model this tradeoff and then investigate it empirically using data from the Massachusetts health insurance Connector exchange (similar in spirit to the state-by-state exchanges set up under the ACA). They find that

price-linked subsidies in the Massachusetts exchange lead to price increases of between 1-6 percent and that the negative impact of these price increases meaningfully outweighs the expected gain from reducing insurance plan price uncertainty, though they acknowledge that quantifying such uncertainty may be difficult in many contexts. Pinning this uncertainty down is potentially quite important for several reasons. First, because of consumer inertia, lump-sum subsidies may necessitate greater churn but consumers may not actually respond to value creation and destruction fluidly in the market. Second, price-linked subsidies can reduce adverse selection death spirals relative to lump-sum subsidies (as discussed in Cutler and Reber (1998)), which is beneficial if market stability is valued in practice.

Related work by Tebaldi (2020) studies insurance plan competition and subsidy design in the California health insurance exchange set up under the Affordable Care Act. This paper makes the important point that, because younger households are both healthier and more price elastic, providing more generous subsidies to that group leads to equilibria where all buyers are better off and per-person public spending is lower. The paper implements a sophisticated discrete choice demand model that accounts for heterogeneity, and integrates this with a supply-side model with imperfect competition. It accounts for extensive margin adverse selection and assumes perfect risk-adjustment on the intensive margin.

In addition to studying an important policy question related to subsidy design, Tebaldi (2020) is one of the first papers to model imperfect competition in the exchanges set up under the ACA. He finds a meaningful degree of product differentiation, primarily due to carrier-specific differences and notes the important implications this has for subsidy design. One downside to studying this nascent market at its inception is that the model does not account for carrier uncertainty and learning when setting prices and establishing products in the market. This is a difficult but important area for future research. In addition, as other papers show, inertia may lead to meaningful reductions in price sensitivity and insurers may use invest-then-harvest policies to attract consumers at the market's inception (see the upcoming discussion on imperfect competition for more details).

There are also a number of papers that have studied subsidy design in Medicare Part D prescription drug insurance with an industrial organization focus. Decarolis (2015) studies competition in Part D markets with a specific focus on how low-income subsidies are delivered to consumers. The paper notes a number of reasons why firms price their plans with a specific focus on gaming the low-income subsidy system. First, about two-thirds of the 9 million LIS enrollees do not actively select an insurance plan. They are allocated by the Center for Medicare and Medicaid Services (CMS) to plans with a premium not greater than the LIS itself. Conditional on an insurer having at least one plan with its premium at or below the subsidy, the allocation rule keeps the LIS enrollees within the same insurer from year to year and, otherwise, allocates them at random across the insurers offering plans with premiums at or below the subsidy. Second, CMS pays the premiums for LIS enrollees in full. Third, the amount of the subsidy is an average of plan premiums. Fourth, all major insurers offer multiple plans that enter into the calculation of the subsidy.

Decarolis (2015) notes that, at the most basic level, this means that a firm offering multiple plans can maintain just one plan with a premium equal to the low income subsidy and set high premiums for all its other plans to inflate the subsidy. The paper sets up an index to show how manipulable the low income subsidy is in a given market for a given insurer. Then, it shows that this index is associated with most of the substantial rise in premium increases in Part D between 2006 and 2011. This suggests that, from a policy standpoint, redesigning the subsidy scheme could lower costs without reducing consumer welfare.

In related work, Decarolis et al. (2020) extend the Decarolis (2015) analysis and estimate a general model of imperfect competition in Part D with the goal of assessing desirable properties of subsidy design. The paper assumes that risk-adjustment transfers (discussed soon) perfectly mitigate adverse selection. Insurers have market power because they are differentiated by their brand equity / non-financial plan features. The authors find the observed Medicare Part D subsidy mechanism is close to optimal. Across a range of counterfactual analyses, they find that more efficient subsidy mechanisms in Medicare Part D share three features: (i) they continue to support demand elasticity through making prices matter; (ii) they limit firms' ability to exercise market power; and (iii) they continue to link the prices firms set with their marginal costs.

Miller et al. (2019) study optimal subsidy regulation on a different dimension, asking how a planner should allocate subsidies across different geographic regions when they have a fixed subsidy budget. Their framework allows for insurers to adjust both pricing and product characteristics in response to alternative subsidy designs, while accounting for consumer preference heterogeneity, consumer inertia heterogeneity, and firm cost heterogeneity. They study Medicare Advantage subsidies and find that the optimal subsidy scheme they solve for increases consumer surplus by 30 percent over the current approach and is reasonably easy to implement with a linear rule based on market-level observables.

2. **Risk-Adjustment transfers:** Risk-Adjustment transfers are another oft-discussed health insurance market policy. Risk-adjustment transfers use an algorithm to transfer money from insurers who enroll healthy consumers directly to insurers who enroll sicker consumers. These transfers combat adverse selection by making sicker consumers cheaper for insurers and healthier consumers more expensive for them.

   Abstracting away from implementation difficulties, risk-adjustment transfers should be implemented aggressively in almost all health insurance markets if done properly and on an ex-ante basis (i.e. the risk-adjustment occurs prior to the realization of health states). Cutler and Reber (1998), Handel et al. (2015), Mahoney and Weyl (2017), Handel et al. (2019) are examples of papers that study the stylized implications of risk-adjustment transfers for selection markets. Many other empirical papers in insurance markets avoid modeling risk selection by assuming that risk-adjustment transfers are fully effective, though this assumption is unlikely to be entirely realistic.

In practice, risk-adjustment transfers are typically only partially effective. They are quite difficult to implement for both computational and strategic reasons. Ellis and Layton (2014) and Geruso and Layton (2017) provide nice bridges between conceptual discussions of risk-adjustment transfers and empirical implementation of these transfers. One key implementation issue is that it is much easier to implement ex post risk-adjustment (based on actual realized claims) from a statistical standpoint. But, ex post risk-adjustment creates significant incentive problems by giving insurers the incentive to increase patient costs. Ex ante risk-adjustment is clearly preferable, but requires a very detailed and effective statistical model mapping prior usage to expected future usage. While these types of algorithms have improved over time, they are still susceptible to gaming by insurers, as demonstrated by Geruso et al. (2019b), which provides evidence that insurers try to attract consumers whose expected cost is below expected premiums net of risk-adjustment transfers. Moreover, Geruso and Layton (2020) show that once an enrollee joins a plan, insurers have the power to affect the size of risk-adjusted payments from the regulator by inflating the severity of the enrollee's reported illness.

While subsidies and risk-adjustment transfers are similar, since they both provide subsidies (one directly to consumers the other to plans), Einav et al. (2019) show that there are two substantive differences between these policies from a conceptual standpoint. First, subsidies provide a greater ability to target consumers based on their willingness-to-pay for plans, rather than just their costs (which are a component of willingness-to-pay). Second, as a result of this first insight, under imperfect competition subsidies can be used to induce lower markups and greater plan enrollment, relative to risk-adjustment. Using data from the California ACA state exchange, the authors find that, holding government spending constant, shifting spending to subsidies away from risk-adjustment transfers could increase exchange enrollment by six percentage points while keeping all consumers weakly better off.

3. **Price discrimination and risk-rating:** Another key regulatory decision made by market designers in health insurance is the extent to which they should allow price discrimination. This issue is addressed directly in Handel et al. (2015), who discuss the key tradeoff between adverse selection and reclassification risk that is induced by different regulations that impact pricing granularity.

   The tradeoff between adverse selection and reclassification risk is one between short-run and long-run risk protection for consumers. In Handel et al. (2015) the authors assume that the market is a static, one year at a time market, similar to what is present in almost all empirical settings. The authors study welfare as insurers are allowed to price based on finer and finer observables. At one end of the spectrum is pure community rating, where all consumers must be charged the same price. In that setting, there is potential for significant adverse selection (especially without effective risk-adjustment transfers) because all consumers are placed in one risk pool together and the sick drive up the cost of more generous plans making those plans much less desirable for healthier consumers. However, there is no risk over time: consumers know they will face the same community-rated premium every year, regardless of their health status.

On the other end of the spectrum, imagine a case where insurers are allowed to price discriminate based on all observable information. In that case, adverse selection is limited, since risk can be priced quite effective. However, long-run year-to-year risk is substantial: if a consumer becomes sick in a persistent manner their premiums increase substantially and for a long period of time.

The paper studies a range of interim cases (partial price discrimination) and quantifies the tradeoff between adverse selection and reclassification risk. The paper finds that, generally, community-rating is preferable to different forms of price discrimination, but that this does depend on lifetime income paths for consumers. For example, if consumers are very healthy earlier in life but have lower income then as well, some risk-rating may be preferable. The paper discusses pricing based on age, which many exchanges have implemented in practice, and finds that this can be a valuable tool for mitigating this correlation between health and income over the life cycle while still achieving the welfare benefits of reduced reclassification risk from community rating.

In related work, Ericson and Starc (2015) study the interaction between insurer price regulation and imperfect competition on the Massachusetts Connector Exchange. They study age rating regulation that restricts insurers to charging prices that are within 3 to 1 ratio relative to one another, across the range of ages. They find that younger consumers are twice as price sensitive as older consumers, implying that younger consumers face lower markups over cost. The paper explores the linkages and shows that there are two key benefits from pooling older consumers with younger consumers when age-pricing regulation is binding: (i) older consumers benefit from the higher price sensitivity of younger consumers, via lower markups and (ii) older consumers benefit from being pooled with younger consumers who are lower costs, assuming risk-adjustment transfers are imperfect. Thus, stricter age-rating regulations transfer resources from younger, low-cost consumers to older, high-cost consumers, reduce firm profits, and increase consumers surplus overall.

4. **Contract generosity:** Another key design decision in exchange markets is what levels of financial coverage to allow insurers to offer. The key benefits of allowing standardized coverage levels are (i) insurers have a harder time cream skimming profitable consumers with small changes to plan designs and (ii) making it easier for consumers to evaluate plans (which can increase price sensitivity and the overall competitiveness of the market).

   Determining the levels of contract generosity a regulator should allow is quite complicated. As discussed in more depth later, Marone and Sabety (2020) show in a large employer context that self-selection into plans limits the value of vertical plan differentiation on financial characteristics. Consumers self-select based on expected transfers while a social planner would like to target plans based on the surplus generated from risk protection as compared to the surplus lost from moral hazard. How close or far away plans are from one another in the attribute space has meaningful but subtle implications for adverse selection, as shown in Marone and Sabety (2020), Ho and Lee (2020), and Handel et al. (2015).

Moreover, in practice, regulation of the minimum coverage plan has had especially important implications for welfare. In practice, the minimum coverage plan attracts many healthy consumers, either due to liquidity constraints or adverse selection into more generous plans. Thus, the minimum coverage level in many exchanges is almost a direct regulation of the risk protection and moral hazard surplus generated by healthy consumers, as discussed in, e.g. Geruso et al. (2019a). In addition to showing that advere selection within-market (intensive margin selection) can lead to many consumers winding up in the minimum regulated coverage, Geruso et al. (2019a) also show that minimum coverage is especially important for determining who selects into the market on the extensive margin. Especially for the Affordable Care Act exchanges, where market participation is a pressing concern, this second margin of selection is important. As the minimum coverage level is lowered, liquidity constraints and lower premiums may push consumers already in the market towards that lower coverage. But, as the minimum coverage is lowered, consumers outside the market may be more willing to pay the now smaller cost of opting in.

This second margin of selection into the market also relates to another oft-discussed policy: insurance mandates requiring all consumers to purchase health insurance. Many of the studies above (apart from Geruso et al. (2019a) presume that a fully enforced mandate is in effect (close to what is idealized in the Affordable Care Act). The case for a mandate is clear: it reduces dynamic free riding and allows for community rated premiums to include both sick and healthy consumers. In practice, mandates have typically been implemented as taxes, where a consumer pays a tax for not having health insurance. This operates similar to a negative lump-sum subsidy, as discussed earlier in this section, with one key difference being that some may feel a moral obligation to purchase insurance with the framing of the tax as a mandate (see, e.g., Cox et al. (2015)).

There are a number of other regulations that we don't discuss in depth here. Later on in this chapter, we discuss regulation of contract length (i.e. can insurance contracts be five years long?). Another important regulation studies the types of services and doctors that plans can and should cover. Additionally, as discussed soon in the section on market regulation when consumers have choice frictions, there are a number of policies (e.g. restricting the number of plans and default plan design) that may be valuable in the presence of consumer choice difficulties.

4.4 Choice frictions and health insurance market design

A range of demand-focused papers have shown that choice frictions and/or behavioral choice factors can have important implications in health insurance markets. While much of this literature (as described in section 3.2) focuses on identifying issues with choices from a given menu of plans, some of it incorporates supply-side models that show important interactions between demand frictions and key supply outcomes.

4.4.1 Consumer mistakes and adverse selection

One key industrial organization insight from the literature on consumers' choice frictions is that choice frictions have different implications for selection markets than for typical product markets. As is typical, in health insurance markets consumers' mistakes are bad for them given a specific market structure. However, conceptual and empirical research has shown that in insurance markets where adverse selection is a prime concern, improving choices may ultimately make consumers worse off. This presents a challenge for policymakers considering avenues to improve consumers' decisions.

Adverse selection is an important potential inefficiency that arises in health insurance markets where the costs to the insurer depend on who is insured. When sicker consumers choose more comprehensive insurance coverage, the premiums for those plans increase to reflect greater costs to the insurer. As a result, healthier consumers, who could prefer plans with greater network coverage or risk protection, may be priced out of the market.

Handel (2013) studies the interaction between inertia and adverse selection using a counterfactual analysis where, as consumer inertia is reduced, consumers pick different insurance plans and the prices of those plans adjust as a result. When inertia is reduced by 75% of its baseline estimate, the premiums for comprehensive coverage increase sharply as healthy people who had been choosing that coverage and losing value shift to less generous coverage. This leads to a death spiral, where the comprehensive plan essentially disappears from the market with an extremely high premium, and consumers who want higher coverage are forced into lower-coverage options. Quantitatively, reduced inertia leads to a 7.7% unintended welfare *reduction* in this environment: helping consumers make better choices is bad for the sample overall.

Polyakova (2016) studies a similar question in Medicare Part D, but emphasizes that whether reduced inertia will be good or bad for consumers depends on how initial prices in the market are set. This paper shows that in Part D, where initial prices for comprehensive coverage are relatively far away from those for less generous coverage, reduced inertia actually helps the prices of more comprehensive coverage adjust downward over time (under the assumption that insurers use lagged average cost pricing). This is because initial prices were set further apart than steady-state equilibrium prices, so reduced inertia, which helps prices move more quickly towards the steady-state equilibrium, reduces the price gap. Thus, in the Part D environment, reduced inertia may both help consumers conditional on the market environment and, by lowering the price of comprehensive coverage in the market, reduce adverse selection.

Handel et al. (2019) provide a general framework for studying when improved choices exacerbate adverse selection. They use a simple model to analyze a population of consumers that are heterogeneous across many dimensions: costs, willingness-to-pay, true value, and choice frictions driving a wedge between willingness-to-pay and true value. The authors derive several theoretical results for competitive insurance markets where consumers make active choices. They show that, as both the mean and variance of consumer surplus rise relative to the mean and variance of costs, improving consumer choices is more likely to be beneficial. This is because the feedback loop between costs and premiums generating adverse selection becomes dominated by improved matching of consumers to the plans they value the most. For example, if heterogeneous consumer values for insurance as a tool for risk protection are relatively large and varied, then improved decision-making facilitates large improvements to welfare through better matching. If these values are not strongly correlated with costs, then there will be limited incremental selection but substantial gains from better choices. The converse is also true: as the mean and variance of costs become more important contributors to insurance value relative to surplus from risk protection, helping consumers make better decisions is worse for the market: these improved decisions cause additional adverse selection which dominates the benefit from better matching. The authors illustrate the interactions between these key objects with simulations as well as an empirical application based on Handel and Kolstad (2015b).

4.4.2 Choice frictions and firm pricing

In addition to studying the pricing impacts of improved choices in competitive markets with adverse selection, several papers study how firms price in markets with inertial consumers. Ericson (2014) documents the invest-then-harvest pricing patterns in Medicare Part D, finding that firms initially set prices low in order to attract consumers and then raise prices to take advantage of consumers' inertia. Ho et al. (2017) study dynamic firm pricing to inertial consumers in Medicare Part D, with a model of imperfect competition. This paper begins by providing substantial multi-dimensional evidence that inertia is a large force shaping consumer plan demand. With estimates of rational inattention, consumer preferences, and consumer costs in hand, Ho et al. (2017) then go on to study how inertia shapes premiums and costs in the market. They estimate the reduction in steady state plan premiums if all consumers were attentive, i.e. inertia were fully removed. They find that an average consumer could save \$1,050 over three years and that government savings over the same time horizon could amount to \$1.3 billion or 1% of the cost of subsidizing the relevant enrollees. This paper shows that understanding dynamic and behavioral aspects of competition are crucial for painting a full picture of competition in Medicare Part D (and likely in insurance exchanges with a large number of options more broadly).

Overall, there has been quite limited work studying how firms price to behavioral consumers in health-insurance and health-care markets, despite the importance of this topic. See Heidhues and Koszegi (2018) for a broader discussion of behavioral industrial organization and some of the approaches that could be applied to studying related questions in health-care markets.

4.4.3 Choice set restrictions and targeted defaults

One additional policy question that interacts with choice frictions, health insurance, and industrial organization is whether and how choice sets should be curated to help deal with those choice frictions. In most product markets, limiting the number and types of choices available would likely be welfare reducing. However, in health insurance markets where choice is complex and difficult, research has shown that such interventions can improve the efficiency gained from managed competition markets.

Abaluck and Gruber (2016) study whether including more options in the consumers' choice sets is beneficial. The authors study a tradeoff between improved consumer-plan matches from greater choice and increased consumer-choice errors from having more options. They leverage a unique data set from Oregon school district employees where each district had the opportunity to offer any combination of 13 approved plans to consumers. Thus, the overall set of plans each district could offer was fixed, but each district could curate its own set of options. Both cross-sectionally and over time, the authors observe similar consumers with a number of choices ranging from 1 to 13, drawn from the same overall set of plans. They argue that market regulators may want to actively consider and regulate the number and quality of plans allowed to enter a market. If more options confuse consumers, and allow firms to prey on them, regulators could curate these markets closely and serve as intermediaries between plans and consumers.[32] Whether or not restricting choice sets is "good" policy depends on the specifics of a given empirical context.[33] In addition, Abaluck and Adams-Prassl (2021) study targeted defaults in Medicare Part D under a set of assumptions about (i) inattention underlying consideration set formation and (ii) the level of tangible switching costs. They find small benefits from targeted defaults if switching costs are high and everyone with a benefit from switching is defaulted into that ex post better option. But, they find that if people are given a new default option above some value threshold, e.g. at least a $300 expected gain, then targeted defaults increase consumer surplus substantially.

4.5 Imperfect competition and welfare

Several papers not yet mentioned focus primarily on the welfare implications of imperfect competition in insurance markets. Town and Liu (2003) study the welfare effects of the Medicare+Choice HMO market, the predecessor to Medicare Advantage. They find meaningful welfare gains from this program for both firms and consumers from 1993-2000, with bigger gains for firms than for consumers. They find that the welfare gains for consumers are distributed unevenly across geographic

---

[32] Certain ACA state exchanges, e.g., Massachusetts and California, use this kind of model to offer more curated options.

[33] A related topic that researchers should investigate is the extent to which targeted (or "smart") default options can help or hurt consumers. Handel and Kolstad (2015a) propose a targeted default policy and analyze the tradeoffs involved in some simulations. See Chandra et al. (2019) and Handel and Kolstad (2015a) for a more in depth discussion.

regions, with greater consumer welfare impacts occurring in markets with more plans. They find that the main impact of having more plans is increased price competition, transferring surplus from firms to consumers. In related work, Curto et al. (2019) use claims data from the Medicare Advantage program to study whether Medicare Advantage plans generate cost savings relative to traditional Medicare. They find the Medicare Advantage plans spend 9 to 30 percent less on health care, adjusting for enrollee case mix, and that these savings are generated through quantity reductions (for both high and low value care) rather than via lower prices or increased consumer price shopping. They find that plans generate revenue that is 30% higher than health care costs, suggesting some, though not exorbitant, profitability once factoring in non-health expenses.

While the above papers estimate Medicare Advantage plan profits and costs using static frameworks, Miller (2019) argues that these papers overestimate plan markups and underestimate plan costs because they do not account for consumer inertia and the dynamic behaviors of consumers and firms. Miller presents reduced-form evidence of dynamic consumer and firm behavior, similar to that in Ericson (2014) and Ho et al. (2017) for Medicare Part D, and estimate a dynamic model of demand and supply. He finds that costs in Medicare Advantage are higher than traditional Medicare for sick consumers, but not for healthy consumers, finding higher costs on average in this market than prior work that does not account for dynamics.

Importantly, imperfect competition can also matter for product design in addition to mattering for pricing. Starc and Town (2019) study the cross-market effects of insurance benefit design. This is an especially important topic area for health care in the United States, where private insurers often operate across multiple markets simultaneously, with some potential for consumers to substitute from one market to another.

The authors compare insurer behavior in stand-alone Medicare Part D drug plans (which only cover drugs) to their behavior in Medicare Advantage plans, which cover both drugs and medical care. One key insight is that plans covering medical expenses as well as drugs are more likely to cover more drugs and to cover drugs more generously. They show that this additional coverage induces greater drug use and, moreover, that these drugs reduce medical expenditures. Thus, having stand-alone drug or medical plans imposes a negative fiscal externality on traditional Medicare, because strategic private insurers don't consider the entire breadth of coverage when making decisions for plans that only cover one slice of health spending.[34]

Starc and Town (2019) set up an equilibrium model that endogenizes what plans choose to cover and subsequent plan pricing, accounting for adverse selection and asymmetric information. They use the model to show that strategic insurer incen-

[34] In other recent work that studies cross-market effects, Holmes (2020) studies risk selection in private insurance as a result of the Affordable Care Act Medicaid expansion, finding that the expansion of the public Medicaid program has important implications for the cost of private insurance. Holmes (2020) finds that Medicaid expansion reduced private insurance premiums by approximately 10%, due to the selection of sicker consumers into Medicaid post-expansion.

tives are as important, from a regulatory design standpoint, as adverse selection in determining benefit design in the markets they study. Lavetti and Simon (2018) show similar results related to plan formulary design and plan incentives for Medicare Part D and Medicare Advantage.

Of course, the papers mentioned earlier in this section that relate to specific policy regulations (e.g. subsidy design) also often investigate interesting interactions between subsidy design and imperfect competition. Thus, though there are not that many papers in the literature whose main policy question emphasizes market power and addressing market power, there are many papers that provide insight into the nature of imperfect competition in insurance markets and its implications. We now turn to a discussion of large group markets, which operate in a different manner than regulated exchanges but (i) cover more than half of consumers in United States and (ii) surface a range of rich industrial organization questions that researchers have worked on.

4.6 Group markets

Especially in the United States, large group markets are a core method for delivering health insurance and health care benefits to consumers. Roughly 60% of consumers in the U.S. are covered by an employer, many of whom are covered by large group purchasers who have significant input into the design of health insurance menus. With roughly 700 billion dollars at stake annually in this segment of the market, it is crucial to understand how these markets operate and how distinct policies could improve their functioning.

Work in this market generally focuses on large self-insured employers who are at-risk for covered lives and have a lot of input into plan design. We will focus first on this setting but then describe recent research into competition and efficiency in smaller group purchaser markets, where insurers sell plans to employers and insurers are liable for the resulting risk.

Several recent empirical papers study the important question of how large employers should optimally design their menu of health plans. These papers leverage the framework built in Cardon and Hendel (2001) (who also study employer markets) and follow on papers, particularly Einav et al. (2013), but add important additional dimensions to the model and leverage more granular and comprehensive data. The questions these papers ask are more ambitious and contain myriad important positive and normative analyses that provide insights into how employers should design health insurance menus for employees.

Marone and Sabety (2020) study optimal menu design for large employers with a specific focus on what micro-foundations imply about that menu design. Specifically, they ask when offering vertical choice of different levels of financial coverage is fruitful vs. when it adds (or potentially subtracts) value from offering just one plan. The key insight in this paper is that consumers self-select into plans when there are multiple options and that self-selection typically will not line up with the plan selection that is socially optimal. They perform this analysis conditional on a range of assumptions about how employers subsidize plans and how plan premiums are set, though the insights also apply to regulation of competitive markets (e.g. those described in Handel et al. (2015)).

The authors model consumer willingness-to-pay for a plan x, relative to a less generous financial baseline plan x_0, as:

$$WTP(x,\theta) = E_l[c_0(l,\omega,x_0) - c_0(l,\omega,x)] + E_l[v(l,\omega,x)] + \Psi(x,\theta) \tag{23}$$

Here, $E_l[c_0(l,\omega,x_0) - c_0(l,\omega,x)]$ is the expected out-of-pocket cost benefit for a given consumer from enrolling in the more generous plan x relative to x_0. $E_l[v(l,\omega,x)]$ is the expected value of moral hazard spending in plan x relative to x_0, i.e. the net benefit of additional spending, factoring in the health benefits and financial costs. $\Psi(x,\theta)$ is the incremental value of risk protection in plan x relative to plan x_0. With this private benefit of additional coverage in mind, the authors also define the social surplus from additional coverage:

$$SS(x,\theta) = \Psi(x,\theta) + E_l[k*(l,\omega,x)] + k_0(l,\omega,x) - v(l,\omega,x)] \tag{24}$$

Here, the social planner has the same value of additional risk protection as the consumer but a different cost of moral hazard, equal to the excess spending in more generous coverage net of the value of that spending.

The authors show that because consumers value changes to their expected spending (and the planner does not) while the planner values the total cost of moral hazard spending (while the consumer does not) there is a wedge between the private and social optima when plan choice is possible. Vertical choice should only be offered if consumers with higher willingness-to-pay have a higher efficient level of coverage, i.e. if the consumers who value insurance the most individually also are those with the highest social surplus from additional coverage.

The authors implement their model empirically using data on public school employees in Oregon between 2008 and 2013 (roughly 45,000 households). The context is useful for identifying moral hazard separately from risk aversion because each of 187 school districts independently selects plans to offer to its specific employees, from a broader menu of potential options negotiated at the state level. The authors find a negative correlation between willingness-to-pay and financial risk, in part because plan out-of-pocket maximums in all the options studied limit risk for the higher spending consumers, who gain substantial expected value from transfers under relatively generous coverage. In addition, lower spending individuals are slightly more prone to moral hazard. Consequently, in their setting, they find that households with higher willingness-to-pay do not have higher efficient coverage levels among the plans studied, implying that allowing for choice is not optimal.

In related work, Ho and Lee (2020) study optimal menu design using a new data set on employee health plan choice and utilization at Harvard. In 2015, Harvard changed their menu of plan options, requiring a subset of (non-unionized) employees to enroll in plans with higher out-of-pocket spending. Facing push back from its employees, Harvard then instituted a more expensive plan option with essentially no out-of-pocket spending that employees could self-select into. This presents a nice setting for identifying heterogeneous moral hazard for consumers in care utilization, as well as risk aversion and plan preferences.

Ho and Lee (2020) begin by studying the implications of allowing for choice of insurance plans that differ only on financial dimensions, similar to the exercise in Marone and Sabety (2020). They find, similarly, that the benefits from offering choice only just outweigh the costs that arise due to self-selection, implying only a small gain in consumer surplus from offering the best possible version of two plans rather than the best possible single plan.

In addition, Ho and Lee (2020) also model preferences for non-financial plan characteristics, e.g. preferences for a specific plan carrier / brand and preferences for a different model of insurer care provision (HMO vs. POS plan). The Harvard context has nice variation on these two dimensions and the authors estimate plan preferences for these non-financial characteristics as a function of underlying health status. This is an important added dimension when considering menu design for large employers, since non-financial plan features are typically quite important when assessing the value of health insurance in U.S. The authors find that allowing for this additional dimension of heterogeneity in menu design makes offering choice much more attractive than when only financial plan characteristics are considered. This insight suggests that, if features like carrier and insurance model differentially appeal to different types of households, then offering a choice of plans that differ on these dimensions may be very useful in helping capture more of the potential gains from plan choice.

Tilipman (2021) (also outlined briefly in Section 2 above) studies optimal menu design for large employers but with a different focus than Marone and Sabety (2020) and Ho and Lee (2020). He holds the financial dimension of insurance fixed and focuses on the non-financial dimension of plan choice (as discussed above in Ho and Lee (2020)). While Ho and Lee (2020) show that this dimension can be important for menu design in counteracting selection on financial dimensions, Tilipman (2021) focuses in on consumer demand for providers and plan networks and on network design. Then he investigates why narrow network plans, that limit access to doctors but are able to extract better prices from those doctors, are less prevalent than one would expect based on consumer preferences. He finds that plan switching costs are a key potential reason why the large employer he studies continues to offer broad network options, despite the meaningful additional cost of offering these options relative to consumer preferences for them net of switching costs.

To perform this study, Tilipman (2021) investigates the plan menus offered over time by the Massachusetts Group Insurance Commission (GIC), a large group purchaser that covers roughly 300,000 lives, many of whom are employed by state agencies. The context is ideal for identification of provider switching costs, plan switching costs, and preferences for broad vs. narrow networks. During the time period studied, the GIC offered some narrow network plans and quite a few broad network plans. In 2012, the GIC offered a three month premium holiday where they (i) forced all state employees to actively re-enroll in a health plan, with no default option and (ii) provided three months of free coverage if a consumer switched from a broad network insurance plan to a narrow network plan.

With this variation, the paper estimates a model of consumer demand for hospitals, physician practices, and insurance plans (conditional on the provider preferences). There are a number of key findings. First, a single-member household values a broad network over a narrow network by between \$15 to \$50 a month (holding insurance carrier fixed). 85% of this incremental value is due to preferences for physician access, rather than preferences for hospital access. Of the value placed on access to physicians in broad networks vs. narrow networks, roughly 50% comes from consumers who are loyal to a physician they have already seen and who they would lose access to under the narrow network option. The author shows that consumer price sensitivity for plans is biased downwards significantly if provider preferences and inertia are not taken into account. In terms of plan demand, the paper finds very high plan switching costs, in line with other papers in the literature (see, e.g., Handel (2013) or Abaluck and Gruber (2016)). Consumers value staying in the same plan at roughly \$250 per month, a large proportion of overall premiums. While this switching cost may reflect a range of micro-foundations, its implications for stickiness in plan choice are clear.

With these ingredients in hand, the paper studies optimal insurance menu design with respect to (i) the number of plans offered and (ii) the types of plans offered (broad or narrow network). The paper sets up an employer objective function for the GIC that trades off consumer surplus from a plan menu, GIC spending on premiums (net of consumer contributions), and the fixed cost of offering additional plans. The estimation of fixed costs uses a moment inequalities framework similar to that used in Ho (2009) to analyze the components of hospital profits; it is a nice addition to the literature considering how many (and what kind of) choices employers should offer.

More precisely: the supply model makes the revealed preference assumption that GIC surplus from offering the observed set of plans is greater than that from potential alternative configurations, at each point in time. The decision at time t depends on the following objective function:

$$E[W_t(\delta_{Jt}, \theta)|\Lambda] = E[S_t(\delta_{jt}, \theta)|\Lambda] - \Sigma_j FC_j \tag{25}$$

Here, S is the marginal surplus from offering plan menu δ_J and FC is the fixed cost of introducing a new plan j into the set of plans offered. Using the revealed preference assumption, this translates into the following estimation equation:

$$\frac{1}{T}\Sigma_t[(W_t(\delta_{Jt}, \theta) - W_t(\delta^a_{Jt}, \theta)) \times g(z)] \geq 0 \tag{26}$$

This framework identifies a set of (i) weights on consumer surplus vs. premiums and (ii) fixed costs that are consistent with marginal surplus maximization, net of fixed costs. The estimates suggest that the GIC values consumer surplus at four times the weight it values premiums and that fixed costs are small relative to premiums, equivalent to roughly \$8 million per new plan introduced.

The author uses all of these ingredients as inputs into several counterfactuals. First, he studies what happens in the market in the absence of consumer loyalty to their prior physicians. He finds that this would ultimately lead to similar plan configurations, and not lead to more narrow network plans generally, because consumers place high value on seeing their prior providers and these providers are typically not the most expensive in the network. Thus removing consumer surplus in this way is not worth the resulting premium gains from narrow network bargaining. Next, he studies the impact of removing plan switching costs, and finds that this has a large impact on menu design, leading to both a greater number and greater variety of plans being offered (in terms of network configurations). The number of broad network plans goes down, but the variety and number of narrow network plans increases substantially. The intuition is that, with more elastic consumers, the fixed costs of plan design are now worth paying, but many narrow networks are important to ensure consumers can access their prior physicians within these lower premium options. Finally, the author shows that a subsidy design that makes consumer pay more on the margin for broad networks is ultimately good given the employer's objective function, because it reduces premiums by so much that is compensates for the loss in surplus for consumers who genuinely want broader networks.

Taken together, the Ho and Lee (2020), Marone and Sabety (2020), and Tilipman (2021) papers provide a useful set of insights about optimal menu design. They show that, on purely financial dimensions, it is not obvious that more choices will lead to higher surplus, because of the implications of self-selection into plans. In addition, we learn that consumer switching costs, both for providers and for plans, are big potential impediments for offering a menu of narrow network plans that are able to lower costs via better bargained rates with providers. Plan switching costs are instrumental: if these are removed (holding provider switching costs constant) a lot of the surplus from network-based menu design can be realized. Additionally, Ho and Lee (2020) shows that, for design on the financial dimension, heterogeneity in design on the non-financial dimension that Tilipman (2021) studies can be quite beneficial in relaxing the constraints imposed by self-selection. Finally, in the context of the earlier discussion of consumer choice difficulties in the chapter, it is not obvious that having more choices is always better, even if the criteria set in the papers here that relate to selection and moral hazard are satisfied. It could be, e.g., the conditions in Marone and Sabety (2020) are met for offering multiple plans but that doing so is still sub-optimal because of choice frictions. See Ericson and Sydnor (2017) for an extended discussion of the questionable value of having more choices in health insurance markets.

There are several other papers that relate to employer menu design in large group markets. Bundorf et al. (2012) study premium setting by employers and find that requiring uniform / community rated premiums can lead to meaningful short-run welfare losses due to adverse selection, though the authors note that this kind of pricing could be beneficial due to reductions in reclassification risk (as studied in depth in Handel et al. (2015)). Einav et al. (2010b) studies adverse selection with data from a large employer, Alcoa, and find limited welfare consequences from adverse selection.

Though this is a valuable empirical contribution, perhaps the most important contribution is the framework for studying competitive provision of insurance, as discussed earlier in this chapter and in the chapter in this volume focusing on selection markets specifically (Einav et al. (2021)). Einav et al. (2013) study selection on moral hazard in the same large-employer context, showing that heterogeneity in moral hazard can interact with adverse selection in important ways when considering the welfare implications of adverse selection and menu design more broadly. This paper is also discussed in depth in Einav et al. (2021). Also of note are Handel (2013), which studies inertia and adverse selection in a large employer setting, and Handel and Kolstad (2015b), which studies menu construction in a large employer setting. We discuss these papers in more depth in other sections of this chapter, since the primary contributions relate to the underlying concepts discussed for health insurance markets broadly, rather than the specific contribution to the large employer context.

While this section has focused on papers studying large group markets where employers self-insure, there is a smaller but important literature studying the sizeable part of the group markets where employers contract with insurers to provide coverage for their employees and insurers bear risk for the spending that results. As noted in Section 2, Dafny (2010) studies this market with a specific focus on whether it is highly competitive. Using a proprietary dataset of firms nationwide, Dafny (2010) tests for competition by seeing whether premiums rise more quickly, ceteris paribus, for firm that have increasing profits over time. In a competitive market for insurance, employer profits should have no impact on insurance premiums. However, if insurers have some market power, higher firm profits should lead to higher premiums. This paper finds that when firm profits rise, insurance premiums rise as well, signaling insurer market power. Moreover, the paper studies firms operating in multiple states, and finds that, for those firms, premiums rise even more steeply in states where insurance markets are more concentrated. While a natural reaction is to presume that this kind of market power should be reduced, in health insurance markets it is unclear whether this is true or not since health insurers with market power can also extract lower prices from medical providers through improved bargaining position.

Fleitas et al. (2020) study the small group health insurance market in the U.S., where employers contract with insurers to provide coverage for their employees. The authors find that, though contract premiums are linked closely to the risk of the employees at the firm up front, this link breaks down dynamically, as only 16% of cost changes over time are passed through into premiums. Thus, though not explicit, employer-insurer relationships operate like guaranteed renewable contracts that provide insurance against reclassification risk for the firm's employees. The authors note that the benefits of these implicit properties are limited by the duration of employer-insurer contracts, which may also partially explain why insurers are fine with providing this kind of implicit coverage in the first place. We discuss dynamic aspects of insurance contracts directly in the next section.

4.7 Long-term guaranteed renewable contracts

Thus far, we have focused primarily on the managed competition paradigm with near community rating when discussing the industrial organization of insurance markets. This is the dominant setup for major regulated markets in the United States, and is also present in several other countries around the world (e.g. the Netherlands, as discussed in Handel et al. (2021b) or Switzerland). The health and industrial organization literature has studied several other market designs as well. Here, we focus on the alternative design provided by long-term guaranteed renewable insurance contracts.

Ghili et al. (2020) study long-term guaranteed renewable contracts in depth, first developing the micro-foundations for characterizing and computing these contracts and then estimating their welfare effects empirically.[35] As discussed above, most managed competition markets allow for insurance contracts that are one year in duration. With these contracts, as discussed in Handel et al. (2015), there is an inherent tradeoff between adverse selection and reclassification risk when considering regulation on permissible price discrimination.

One way to get around this tradeoff conceptually is to allow for long-term guaranteed renewable contracts with one-sided commitment. These contracts are similar to those used frequently in life insurance markets (see, e.g., Hendel and Lizzeri (2003)), where the insurer commits to a lifetime premium path and the consumer can leave the contract at any point in time. This one-sided commitment is consistent with legal restrictions in practice, where firms can commit to these kinds of arrangement but typically it is difficult to commit consumers to paying large sums of money years in advance. Importantly, insurers can risk-rate consumers when they sign up for the contract, an aspect the authors discuss in depth.

The paper sets up a dynamic optimization problem with symmetric information and symmetric learning for insurers and consumers. Consumers have a health status transition matrix and lifetime income paths. The paper presents a series of proofs to characterize the optimal contracts, which have several key features. First, contracts are front-loaded, in the sense that consumers pay more than their expected costs early in the contract and, as a result, cover the expected losses for the insurer down the road created by consumers being allowed to leave the contract at any point in time (i.e. if they have had a positive health shock and can find a better contract elsewhere). The authors show that the potential value of these contracts is higher when income paths are flatter, since front-loading is worse for consumers with credit constraints who have low income initially.[36,37]

[35] For earlier conceptual work in this space, see, e.g., Cochrane (1995) and Pauly et al. (1995).

[36] It is also important to note that the long-term contracts studied in Ghili et al. (2020) can also be interpreted as guaranteed renewable contracts coupled with potential consumer-led renegotiation. Evidence from Fleitas et al. (2020) supports this interpretation, that consumers renegotiate or search for better options when their health improves, something that is not allowed in other papers studying guaranteed renewability (see, e.g., Pauly et al. (1995)).

[37] In extension of the base model, the authors also allow for consumer myopia, consumer switching costs, limited access to credit, and several other extensions.

The paper estimates the lifetime welfare implications of these contracts and compares them to the managed competition year-to-year contracts typical of the Affordable Care Act exchanges. The authors develop an algorithm to compute the optimal dynamic contracts and use all-payer claims data from the state of Utah to characterize a second-order Markov process for health state transitions for a large population of policy interest. The authors show a range of contract comparative statics with respect to underlying health and underlying income and find, overall, that optimal dynamic contracts can reduce 94% of the welfare loss from reclassification risk for healthy age 25 consumers (at the presumed start of the market), relative to year-to-year contracts that allow for full risk-rating. Moreover, these contracts fully remove the issue of adverse selection on observable information. However, for consumers that are initially sick, these contracts cannot add much value, because they are already starting off in an expensive state.

In the paper, the authors show that, with ex ante health state insurance at the start of the market (age 25) there are some scenarios where dynamic long-term contracts add significant value relative to the typical managed competition setup with one year contracts. The scenarios where guaranteed renewable contracts are most appealing include:

1. Cases where consumer income paths are flatter over the life cycle
2. Cases where consumers have greater access to credit
3. Cases where consumers have higher contract switching costs, relaxing the lapsation constraint that incorporates the cost of consumers' inability to commit to a contract
4. Cases where consumers have low myopia and better incorporate the long-term value of the contract relative to higher up-front payments

The authors close with a discussion of why these contracts are only seen infrequently in practice. One potential reason is that long-run medical cost inflation risk is hard to diversify, especially for specific sub-populations of consumers (e.g. those with cancer or diabetes). Another possible impediment is that consumers whose health state degrades may have the quality of the benefits strategically degraded by insurers, as discussed in, e.g. Atal (2019). Finally, another key issue is contextual: in the United States, the employer-based market enrolls the majority of consumers, in part due to the tax exemption given to premiums for these plans. As a result, the state-based exchanges typically "fill in the gaps" of employer provided coverage, rather than supplant it. In that kind of context, with substantial churn, long-term contracts will be less appealing than short-run contracts, unless they are put on equal footing as employer plans from a tax standpoint.

In practice, these contracts are seen in the private market in Chile (as studied in Atal (2019)) and the private market in Germany (as studied initially by Hofmann and Browne (2013)). In Germany, Atal et al. (2020) leverage the theoretical framework developed in Ghili et al. (2020) to study the welfare impacts of the long-terms contracts found in practice. They find that, despite the contracts having a simpler dynamic structure than the optimal contracts derived in Ghili et al. (2020), these long-term contracts deliver much of the welfare gains that the optimal contracts would in their context, relative to short-term contracts that are one-year in duration. Thus, though empirical case studies of long-term contracts are limited, these contracts hold some promise in future market designs, especially as the technology to design and implement them improves.

5 Going forward

We close by noting some opportunities for future research. While some parts of the literature are remarkably complete, others are sparse and represent opportunities for researchers. There are also opportunities to cross-pollinate the different literatures discussed in this chapter and bring insights from models of selection markets into the study of provider competition, pricing, and antitrust. We now discuss several of these directions for future research.

Value creation in insurance markets: A key premise of the managed competition paradigm that underlies many regulated markets is that private insurers will compete on prices and on product characteristics. While there is a substantial literature on the impacts of competition on prices, there is much more limited work on whether insurer competition leads to product innovation on other dimensions.

Assessing whether and how insurers innovate is an important topic for future work. There are some working papers that make progress on this topic. Abaluck et al. (2020) study differences in mortality effects for Medicare Advantage plans. They use a design that leverages plan exits from the market to show meaningful heterogeneity in plan mortality effects across plans. They show that consumers place little weight on these differential mortality effects in plan choices, potentially because these effects are not generally known to consumers. Moving beneficiaries away from the bottom 5% of plans by mortality effects could save tens of thousands of elderly lives every year.

This work shows that (i) plans have differential causal effects on a key health outcome measure and (ii) consumers have low willingness to pay for this dimension, likely because of a lack of information. Future work that unpacks the different potential domains for insurer innovation and investigates the micro-foundations for why consumers don't respond to these differences will be quite valuable.

For example, in ongoing work, Handel et al. (2021) use the Utah all-payer claims database along with a series of natural experiments to assess different domains for insurer innovation. They observe employer insurance menus and identify a large number of cases where employers required employees to switch insurance carriers if they wanted to take advantage of their benefits. As a result, many consumers switch from one carrier to another due to these broad policy changes. The authors decompose insurer strategies into (i) network configuration (ii) consumer steering within-network (iii) consumer price shopping, and (iv) other constraints (e.g. prior authorization) and find that across many private insurers, there is quite limited differentiation on these key dimensions.[38] In other recent work, Dunn et al. (2021) studies prior authorization requirements that require insurer permission for certain types of medical procedures, imposing hassle costs on consumers and providers with the goal of reduced spending. They document that providers lose 16% of medical billing due to prior authorization restrictions in Medicaid and 4% in private markets, though they don't unpack the heterogeneous value (or lack thereof) generated by these restrictions across private insurers. For prior authorization and other dimensions of insurers productivity, unpacking this heterogeneity and the resulting value creation or destruction is an important avenue for future work.

Modeling provider quality: We discuss in Section 2 the literature that considers provider quality investments, in markets where prices are administered and those with negotiated prices. As noted there, a potential next step in this literature would be to specify and estimate a model of provider investment in quality, given expectations of the impact on outcomes; negotiated prices; utilization and profits. Garthwaite et al. (2020) is a very recent working paper that is relevant here. The authors discuss a theoretical model of strategic quality choice by hospitals making investments to increase their private revenue. They argue, in the context of this model, that the welfare loss from high provider markups could be offset by the implied incentives for hospital investment. In some situations, price regulation might be welfare-reducing. Their empirical analyses confirm some of the model's predictions across several quality measures.

A full model-based empirical analysis of quality investment would be complex, in part because the problem is inherently dynamic: an investment in quality or technology in year t may affect outcomes and patient demand for many future years. However, it could also be important for understanding the impact of market design and regulation on equilibrium quality.

Public option: Many countries around the world have a two-tiered insurance system where there is (i) a robust public insurance option and (ii) the public option can either be replaced by or supplemented by private insurance. In the United States, this structure exists in Medicare Advantage relative to traditional Medicare, but not in the Affordable Care Act exchanges, where it has been a topic of much debate. While the papers referenced in this chapter on Medicare Advantage make some progress in modeling how having a public option impacts market outcomes and welfare, there is still limited work formally studying these kinds of two-tiered systems, especially with an eye towards implementation in the under-65 U.S. private market.

Cross-insurance market interactions: One particular feature of privatized health insurance markets is that insurers often operate across multiple product markets simultaneously. For example, in a typical U.S. context, an under-65 individual could

[38] Earlier related work in this space includes, e.g., Gaynor et al. (2004) and Cutler et al. (2000).

enroll in an ACA-exchange plan, an employer plan, or Medicaid (including privatized Medicaid) if eligible. These markets are typically studied in isolation from one another, though there are potentially meaningful interactions. For example, Holmes (2020) studies the impact of ACA Medicaid expansion on private market premiums and finds that the expansion reduced those premiums, facilitating movement into those markets. In ongoing work, Dickstein et al. (2021) use a model of insurance demand and insurer pricing, and detailed data from Oregon, to study the effects of recent policies to (partially or completely) combine the small-group market and the individual market.

There are quite a few interesting topics for future work to explore in this space including (i) optimal multi-market regulation (ii) insurer strategy that accounts for substitution across markets they operate in, and (iii) how tax policy (e.g. for employer-sponsored plans) impacts other health care markets, such as the ACA exchanges.

Firm pricing to consumers with choice frictions: Heidhues and Koszegi (2018) discuss theoretical and empirical work in IO generally that relates to firm pricing with behavioral consumers, and/or consumers with choice frictions. Given the preponderance of evidence that consumers have choice issues in health insurance markets, there is quite limited work that rigorously studies how insurers respond to those choice issues in pricing (and in forming product characteristics more broadly). Ho et al. (2017) and Ericson (2014) are two notable examples that investigate firm pricing in the presence of inertia in Medicare prescription drug markets. In addition, it will be interesting to study product design with behavioral consumers, similar to, e.g., Gabaix and Laibson (2006). While there is minimal work to date on this area in health care markets, the literature on choice frictions and limited information about product characteristics referenced earlier in this chapter suggests that product design may respond to these effects. These existing papers, and the broader industrial organization literature, show that modeling pricing and product design with both behavioral / information-constrained consumers and adverse selection is quite challenging. However, given the importance of this topic and the general advances in IO models when consumers have choice frictions, we view this as a promising area for future work.

Firm learning and firm dynamics: In the IO literature broadly, it has been challenging to study firm learning and firm dynamics, due to the data and modeling requirements. In the health care space, which is often in flux, firm learning and dynamics have the potential to be important for both insurers and providers. This comment applies both conceptually when assessing policy impacts and also empirically when considering whether a given analysis that assumes a "steady state" is valid. While we recognize that this is a difficult area for future work, it is also an important one.

Provider price negotiations and network formation: While the literature on insurer-provider price negotiations is quite well-developed, there are still some unanswered questions in this space. One example is the topic of insurer-provider negotiations over tiered provider networks. Prager (2020) shows that consumers respond to tiered hospital networks, tending to favor hospitals on preferred tiers. This consumer behavior is likely to generate incentives for insurers to establish tiered networks, and to affect the negotiations that determine prices and equilibrium networks in interesting ways that may well be important for policy. Starc and Swanson (2021) provide some evidence on this point, demonstrating that Medicare Part D plans with more restrictive preferred pharmacy networks pay lower retail drug prices. Research to develop an equilibrium model of tiered network negotiation would be a useful and interesting avenue for future work. It would be useful for such work to also model and estimate consumer frictions in provider selection and network valuation, a complex topic that has been understudied but which is important for understanding provider price negotiations and network formation.

Payment reform, organizational structure, and gaming: As discussed earlier in Section 2.6, there are now many important documented instances where providers respond forcefully to pricing incentives for delivering care on the margin. While the papers discussed earlier make excellent headway in this area, payment reform and organizational structure is a broad and important topic with significant potential to impact efficiency in the health care system. Different potential reforms, like bundled payments, capitated payments, Accountable Care Organizations (ACOs) and more aggressive price regulation are all quite present in the current policy debate around reducing health care spending. Still, there is limited IO work investigating such reforms, including the papers on LTCH contract design discussed earlier and a recent working paper by Einav et al. (2020) that studies selection into voluntary payment regulation by hospitals. Given the policy importance of this area and the implications of such reforms for market structure, pricing and care delivery, we believe this is a very fruitful area for future work.

The role of sales intermediaries in health care markets: In some markets, relationships between sales agents, brokers, and employer benefits managers may be important determinants of consumers' allocations to health plans. Sales intermediaries are also likely to influence the extent of price competition between insurers and hence the level of insurance premiums in these cases. This seems particularly likely in the small group market where employers are not commonly self-insured and are likely to rely on brokers to help choose and access plans. Dickstein et al. (2021) use their detailed data from Oregon to show that average plan markups in the small group market are substantially higher than in the individual market, where brokers are less prevalent and competition may be more effective. However, they do not observe brokers and cannot investigate their impact further. Additional research on this issue would be very useful if data become available.

Market power and monopsony: As discussed in Section 2, there are many existing academic papers and merger retrospectives on the impact of health care sector mergers on prices and quantities of the traded good. Excluding fully structural work like Ho and Lee (2017), there has been much less work on market power and monopsony, where worker surplus is reduced. This is true generally (see, e.g., Berry et al.

(2019)) and also specifically for health care. A notable exception is recent work by Prager and Schmitt (2021), which uses hospital mergers as events to study monopsony. They find evidence for reduced wage growth in instances where the increase in concentration induced by the merger is large and workers' skills are specifically tailored to the health care sector. They also find that unions attenuate these wage reductions. Future work in this area would be valuable in a number of key directions including (i) additional evidence on hospital market power and monopsony (ii) new evidence on monopsony related to other large non-hospital health care organizations (e.g. insurers, provider organizations), and (iii) the role of licensing on preventing job substitution in the presence of market power. The labor economics literature also has quite a few recent papers on monopsony across a range of other sectors (e.g. Benmelech et al. (2020), Arnold (2020)) and future work on this topic has the potential to contribute jointly to the labor and IO literatures.

Vertical contracting, integration, and welfare: Section 2 outlines the literature that considers the impact of vertical integration in health care markets. As noted there, most of these papers consider integration between types of providers: e.g., primary care practitioners and hospitals; or primary care practitioners and specialists. There is also a very small literature analyzing the effects of integration between insurers and providers. There is clearly room for more research in this area, particularly given the degree of consolidation observed in the health care industry in recent years. Papers that account for the complementarities between medical providers in primary care and in various specialties, and for geographic dispersion of physicians as well as hospitals, would be particularly welcome. There is also considerable scope for papers considering potential agency problems and other sources of inefficiencies in vertical contracting, as in Frandsen et al. (2019).

Nuanced objective functions: Almost all papers that study welfare in the health and IO literature use objective functions that assume away potentially important factors like heterogeneity in marginal utility of income and liquidity constraints. These factors are potentially quite important to consider in many health care contexts, where consumer financial risk is high and health care is often viewed as a "right."

One key modeling choice that assumes away these income-based issues is the use of constant absolute risk aversion (CARA) preferences that assume that the concavity of the utility function in income is invariant to the starting wealth point. One primary reason for this assumption is that it removes the need for detailed income or wealth data. Of course, in certain contexts, such as subsidies for the uninsured, Medicaid policies, and Medicare Part D subsidies, the distributional consequences of policies for low-income groups are especially important to regulators.

Recent work by Handel et al. (2021b) does explicitly consider policy impacts across the income distribution using tools from public finance. However, there is room for much more work in this area, especially work that considers the interactions between these distributionally-oriented micro-foundations and competition policy.

Overall, the health care and industrial organization literature has made substantial progress, especially over the last decade as new data have unlocked the ability to study myriad questions of interest. This list of potential topics for future research is a small slice of the many promising conceptual and empirical advances that we expect researchers to make going forward as data, methods, and institutions evolve.

References

Abaluck, Jason, Adams-Prassl, Abi, 2021. What do consumers consider before they choose? Identification from asymmetric demand responses. The Quarterly Journal of Economics 136 (3), 1611–1663.

Abaluck, Jason, Gruber, Jonathan, 2011. Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program. The American Economic Review 101 (4), 1180–1210.

Abaluck, Jason, Gruber, Jonathan, 2016. Evolving choice inconsistencies in choice of prescription drug insurance. The American Economic Review 106 (8), 2145–2184.

Abaluck, Jason, Gruber, Jonathan, Swanson, Ashley, 2018. Prescription drug use under Medicare Part D: a linear model of nonlinear budget sets. Journal of Public Economics 164 (C), 106–138.

Abaluck, Jason, Caceres Bravo, Mauricio, Hull, Peter, Starc, Amanda, 2020. Mortality Effects and Choice Across Private Health Insurance Plans. NBER Working Paper 27578.

Agarwal, Nikhil, Budish, Eric, 2021. Market design. In: Ho, K., Hortaçsu, A., Lizzeri, A. (Eds.), Handbook of Industrial Organization, vol. 5. Elsevier. Chapter 10 (this volume).

Agarwal, Rajender, Liao, Joshua M., Gupta, Ashutosh, Navathe, Amol S., 2020. The impact of bundled payment on health care spending, utilization, and quality: a systematic review. Health Affairs 39 (1), 50–57. PMID: 31905061.

Akerlof, George, 1970. The market for 'lemons': quality uncertainty and the market mechanism. The Quarterly Journal of Economics 84 (3), 488–500.

Alcott, Hunt, Glied, Sherry, Handel, Benjamin R., Kim, Grace, 2021. Price Transparency and Medical Provider Pricing: Evidence from a Randomized Evaluation in New York. UC Berkeley Working Paper.

Allende, Claudia, 2020. Competition Under Social Interactions and the Design of Education Policies. Working Paper.

Allende, Claudia, Gallego, Francisco, Neilson, Christopher, 2020. Approximating the Equilibrium Effects of Informed School Choice. Princeton University Working Paper.

Arnold, David, 2020. Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes. Princeton University Working Paper.

Aron-Dine, Aviva, Einav, Liran, Finkelstein, Amy, Cullen, Mark, 2012. Moral Hazard in Health Insurance: How Important is Forward Looking Behavior? NBER Working Paper no. 17802.

Aron-Dine, Aviva, Einav, Liran, Finkelstein, Amy, Cullen, Mark, 2015. Moral hazard in health insurance: do dynamic incentives matter? Review of Economics and Statistics 97 (4), 725–741.

Arrow, Kenneth, 1963. Uncertainty and the welfare economics of medical care. The American Economic Review 53 (5), 941–973.

Atal, Juan Pablo, 2019. Lock-in in Dynamic Health Insurance Contracts: Evidence from Chile. University of Pennsylvania Working Paper.

Atal, Juan Pablo, Fang, Hanming, Karlsson, Martin, Ziebarth, Nicolas, 2020. Long-Term Health Insurance: Theory Meets Evidence. University of Pennsylvania Working Paper.

Atwood, Alicia, LoSasso, Anthony, 2016. The effect of narrow provider networks on health care use. Journal of Health Economics 50, 86–98.

Autor, David, Dorn, David, Katz, Lawrence F., Patterson, Christina, Van Reenen, John, 2020. The fall of the labor share and the rise of superstar firms. The Quarterly Journal of Economics 135 (2), 645–709.

Baicker, Katherine, Mullainathan, Sendhil, Schwartzstein, Joshua, 2015. Behavioral hazard in health insurance. The Quarterly Journal of Economics 130 (4), 1623–1667.

Baker, Laurence C., Kate Bundorf, M., Kessler, Daniel P., 2016. The effect of hospital/physician integration on hospital choice. Journal of Health Economics 50 (1), 1–8.

Barseghyan, Levon, Molinari, Francesca, O'Donoghue, Ted, Teitelbaum, Joshua, 2013. The nature of risk preferences: evidence from insurance choices. The American Economic Review 103 (6), 2499–2529.

Benmelech, Efraim, Bergman, Nittai, Kim, Hyunseob, 2020. Strong employers and weak employees: how does employer concentration affect wages? The Journal of Human Resources. 0119–10007R1.

Berry, Steven, Gaynor, Martin, Scott-Morton, Fiona, 2019. Do increasing markups matter? Lessons from empirical industrial organization. The Journal of Economic Perspectives 33 (3), 44–68.

Bhargava, Saurabh, Loewenstein, George, Sydnor, Justin, 2017. Choose to lose: health plan choices from a menu with dominated option. The Quarterly Journal of Economics 132 (3), 1319–1372.

Binmore, Ken, Shaked, Avner, Sutton, John, 1989. An outside option experiment. The Quarterly Journal of Economics 4 (104), 753–770.

Brand, Keith, Rosenbaum, Ted, 2019. A review of the economic literature on cross-market healthcare mergers. Antitrust Law Journal 82 (2), 533–549.

Bresnahan, Timothy, 1989. Empirical studies of industries with market power. In: Schmalensee, Richard, Willig, Robert (Eds.), Handbook of Industrial Organization, vol. 2(1), pp. 1011–1057.

Brot-Goldberg, Zarek, de Vaan, Mathijs, 2020. Intermediation and Vertical Integration in the Market for Surgeons. Working Paper.

Brot-Goldberg, Zarek C., Chandra, Amitabh, Handel, Benjamin R., Kolstad, Jonathan T., 2017. What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. The Quarterly Journal of Economics 132 (3), 1261–1318.

Brot-Goldberg, Zarek, Layton, Timothy, Vabson, Boris, Wang, Adelina, 2021. The Behavioral Foundations of Default Effects: Theory and Evidence from Medicare Part D. University of Chicago Working Paper.

Brown, Jason, Duggan, Mark, Kuziemko, Ilyana, Woolston, William, 2014. How does risk selection respond to risk adjustment? New evidence from the Medicare advantage program. The American Economic Review 104 (10), 3335–3364.

Brown, Zach, 2019. Equilibrium effects of health care price information. Review of Economics and Statistics 101 (4), 699–712.

Brown, Zach, 2020. An Empirical Model of Price Transparency and Markups in Health Care. University of Michigan Working Paper.

Bryan, Kevin, Williams, Heidi, 2021. Markets for innovation: market failures and public policies. In: Ho, K., Hortaçsu, A., Lizzeri, A. (Eds.), Handbook of Industrial Organization, vol. 5. Elsevier. Chapter 13 (this volume).

Bundorf, Kate, Levin, Jonathan, Mahoney, Neale, 2012. Pricing and welfare in health plan choice. The American Economic Review 102 (7), 3214–3248.

Capps, Cory, 2005. The Quality Effects of Hospital Mergers. Economic Analysis Group Discussion Paper.

Capps, Cory, Dranove, David, Ody, Christopher, 2018. The effect of hospital acquisitions of physician practices on prices and spending. Journal of Health Economics 59 (1), 139–152.

Capps, Cory, Dranove, David, Satterthwaite, Mark, 2003. Competition and market power in option demand markets. The Rand Journal of Economics 34 (4), 737–763.

Cardon, James, Hendel, Igal, 2001. Asymmetric information in health insurance: evidence from the national medical expenditure survey. The Rand Journal of Economics 32 (3), 408–427.

Carlin, Caroline, Town, Robert, 2009. Adverse Selection, Welfare, and Optimal Pricing of Employer-Sponsored Health Plans. University of Minnesota Working Paper.

Chandra, Amitabh, Handel, Benjamin R., Schwartzstein, Josh, 2019. Behavioral economics and health-care markets. In: Handbook of Behavioral Economics - Applications and Foundations, vol. 2, pp. 459–502.

Chiappori, Pierre-Andre, Salanie, Bernard, 2000. Testing for asymmetric information in insurance markets. Journal of Political Economy 108 (1), 56–78.

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

Clemens, Jeffrey, Gottlieb, Joshua D., 2017. In the shadow of a giant: Medicare's influence on private physician payments. Journal of Political Economy 125 (1), 1–39.

Clemens, Jeffrey, Gottlieb, Joshua D., Molnar, Timea Laura, 2017. Do health insurers innovate? Evidence from the anatomy of physician payments. Journal of Health Economics, 153–167.

Cochrane, John, 1995. Time-consistent health insurance. Journal of Political Economy 103 (3), 445–473.

Cohen, Alma, Einav, Liran, 2007. Estimating risk preferences from deductible choice. The American Economic Review 97 (3), 745–788.

Collard-Wexler, Allan, Gowrisankaran, Gautam, Lee, Robin S., 2019. "Nash-in-Nash" bargaining: a microfoundation for applied work. Journal of Political Economy 127 (1), 163–195.

Cooper, Zack, Scott-Morton, Fiona, 2016. Out-of-network emergency-physician bills - an unwelcome surprise. The New England Journal of Medicine 375 (20), 1915–1918.

Cooper, Zack, Morton, Fiona Scott, Shekita, Nathan, 2020. Surprise! Out-of-network billing for emergency care in the United States. Journal of Political Economy 128 (9), 3626–3677.

Cooper, Zack, Gibbons, Stephen, Jones, Simon, McGuire, Alistair, 2011. Does hospital competition save lives? Evidence from the English NHS patient choice reforms. The Economic Journal 121 (554), 228–260.

Cooper, Zack, Craig, Stuart V., Gaynor, Martin, Van Reenen, John, 2019. The price ain't right? Hospital prices and health spending on the privately insured. The Quarterly Journal of Economics 134 (1), 51–107.

Coughlin, Maura, 2021. Insurance Choice with Non-Monetary Plan Attributes: Limited Consideration in Medicare Part D. Rice University Working Paper.

Cox, Natalie, Handel, Benjamin R., Kolstad, Jonathan, Mahoney, Neale, 2015. Messaging and the mandate: the impact of consumer experience on health insurance enrollment through exchanges. The American Economic Review 105 (5), 105–109.

Craig, Stuart, Ericson, Keith, Starc, Amanda, 2021. How important is price variation between health insurers? Journal of Health Economics 77 (4).

Crawford, Gregory, Lee, Robin S., Whinston, Michael, Yurukoglu, Ali, 2018. The welfare effects of vertical integration in multichannel television markets. Econometrica 86 (3), 891–954.

Crawford, Gregory S., Yurukoglu, Ali, 2012. The welfare effects of bundling in multichannel television markets. The American Economic Review 102 (2), 643–685.

Cuesta, Jose Ignacio, Noton, Carlos, Vatter, Benjamin, 2020. Vertical Integration Between Hospitals and Insurers. Working Paper.

Curto, Vilsa, Einav, Liran, Finkelstein, Amy, Levin, Jonathan, Bhattacharya, Jay, 2019. Health care spending and utilization in public and private Medicare. American Economic Journal: Applied Economics 11 (2), 302–332.

Cutler, David, Reber, Sarah, 1998. Paying for health insurance: the trade-off between competition and adverse selection. The Quarterly Journal of Economics 113 (2), 433–466.

Cutler, David M., Dafny, Leemore S., 2011. Designing transparency systems for medical care prices. The New England Journal of Medicine 364 (10), 894–895.

Cutler, David, McClellan, Mark, Newhouse, Joseph, 2000. How does managed care do it? The Rand Journal of Economics 31 (3), 526–548.

Cutler, David M., Huckman, Robert, Kolstad, Jonathan, 2010. Input constraints and the efficiency of entry: lessons from cardiac surgery. American Economic Journal: Economic Policy 2 (1), 51–76.

Cutler, David, Huckman, Robert, Landrum, Mary Beth, 2004. The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery. The American Economic Review 94 (2), 342–346.

Dafny, Leemore S., 2009. Estimation and identification of merger effects: an application to hospital mergers. Journal of Law, Economics, & Organization 52 (3), 523–550.

Dafny, Leemore S., 2015. Evaluating the impact of health insurance industry concentration: learning from experience. Commonwealth Fund Issue Brief 33.

Dafny, Leemore S., Dranove, David, 2008. Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. The Rand Journal of Economics 39 (3), 790–821.

Dafny, Leemore S., Hendel, Igal, Wilson, Nathan, 2015. Narrow networks on the health insurance exchanges: what do they look like and how do they affect pricing? A case study of Texas. The American Economic Review: Papers and Proceedings 105 (5), 110–114.

Dafny, Leemore S., 2005. How do hospitals respond to price changes? The American Economic Review 95 (5), 1525–1547.

Dafny, Leemore S., 2010. Are health insurance markets competitive? The American Economic Review 100 (4), 1399–1431.

Dafny, Leemore S., Ho, Kate, Lee, Robin S., 2019. The price effects of cross-market mergers: theory and evidence from the hospital industry. The Rand Journal of Economics 50 (2), 286–325.

Dafny, Leemore S., Duggan, Mark, Ramanarayanan, Subramaniam, 2012. Paying a premium on your premium? Consolidation in the U.S. health insurance industry. The American Economic Review 102 (2), 1161–1185.

Dalton, Christina M., Gowrisankaran, Gautam, Town, Robert J., 2019. Salience, myopia, and complex dynamic incentives: evidence from Medicare Part D. The Review of Economic Studies 87 (2), 822–869.

Decarolis, Francesco, 2015. Medicare Part D: are insurers gaming the low income subsidy design? The American Economic Review 105 (4), 1547–1580.

Decarolis, Francesco, Polyakova, Maria, Ryan, Stephen P., 2020. Subsidy design in privately provided social insurance: lessons from Medicare Part D. Journal of Political Economy 128 (5), 1712–1752.

De Loecker, Jan, Syverson, Chad, 2021. An industrial organization perspective on productivity. In: Ho, K., Hortaçsu, A., Lizzeri, A. (Eds.), Handbook of Industrial Organization, vol. 4. Elsevier. Chapter 3.

Demsetz, Harold, 1973. Industry structure, market rivalry, and public policy. Journal of Law, Economics, & Organization 16 (1), 1–9.

Desai, Sunita, Hatfield, Laura, Hicks, Andrew, Chernew, Michael, Mehrotra, Ateev, 2016. Association between availability of a price transparency tool and outpatient spending. Journal of the American Medical Association 315 (17), 1842–1843.

Dickstein, Michael, Ho, Kate, Mark, Nathaniel, 2021. Market Segmentation and Competition in Health Insurance. Working Paper.

Dranove, David, Sfekas, Andrew, 2008. Start spreading the news: a structural estimate of the effects of New York hospital report cards. Journal of Health Economics 27 (5), 1201–1207.

Dranove, David, Ody, Christopher, 2019. Employed for higher pay? How Medicare payment rules affect hospital employment of physicians. American Economic Journal: Economic Policy 11 (4), 249–271.

Dranove, David, White, W.D., 1994. Recent theory and evidence on competition in hospital markets. Journal of Economics & Management Strategy 3 (1), 169–209.

Dranove, David, Kessler, Daniel, McClellan, Mark, Satterthwaite, Mark, 2003. Is more information better? The effects of "report cards" on health care providers. Journal of Political Economy 111 (3), 555–588.

Dunn, Abe, Gottlieb, Joshua D., Shapiro, Adam, Sonnenstuhl, Daniel J., Tebaldi, Pietro, 2021. A Denial a Day Keeps the Doctor Away. National Bureau of Economic Research Working Paper 29010.

Einav, Liran, Finkelstein, Amy, Levin, Jonathan, 2010a. Beyond testing: empirical models of insurance markets. Annual Review of Economics 2 (1), 311–336.

Einav, Liran, Finkelstein, Amy, Cullen, Mark, 2010b. Estimating welfare in insurance markets using variation in prices. The Quarterly Journal of Economics 125 (3), 877–921.

Einav, Liran, Finkelstein, Amy, Mahoney, Neale, 2018. Provider incentives and healthcare costs: evidence from long-term care hospitals. Econometrica 86 (6), 2161–2219.

Einav, Liran, Finkelstein, Amy, Mahoney, Neale, 2021. The IO of selection markets. In: Ho, K., Hortaçsu, A., Lizzeri, A. (Eds.), Handbook of Industrial Organization, vol. 5. Elsevier. Chapter 14 (this volume).

Einav, Liran, Finkelstein, Amy, Schrimpf, Paul, 2010. Optimal mandates and the welfare cost of asymmetric information: evidence from the U.K. annuity market. Econometrica 78 (3), 1031–1092.

Einav, Liran, Finkelstein, Amy, Tebaldi, Pietro, 2019. Market Design in Regulated Health Insurance Markets: Risk Adjustment vs. Subsidies. Working Paper.

Einav, Liran, Finkelstein, Amy, Ryan, Stephen, Schrimpf, Paul, Cullen, Mark, 2013. Selection on moral hazard in health insurance. The American Economic Review 103 (1), 178–219.

Einav, Liran, Finkelstein, Amy, Ji, Yunan, Mahoney, Neale, 2020. Voluntary Regulation: Evidence from Medicare Payment Reform. NBER Working Paper no. 27223.

Eliason, Paul J., Heebsh, Benjamin, McDevitt, Ryan C., Roberts, James W., 2020. How acquisitions affect firm behavior and performance: evidence from the dialysis industry. The Quarterly Journal of Economics 135 (1), 221–267.

Eliason, Paul J., Grieco, Paul L.E., McDevitt, Ryan C., Roberts, James W., 2018. Strategic patient discharge: the case of long-term care hospitals. The American Economic Review 108 (11), 3232–3265.
Ellis, Randall P., Layton, Timothy J., 2014. Risk selection and risk adjustment. In: Encyclopedia of Health Economics. Elsevier Press, Boston, pp. 289–297.
Enthoven, Alain, Garber, Alan, Singer, Sara, 2001. Near-universal coverage through health plan competition: an insurance exchange approach. In: Meyer, J., Wicks, E. (Eds.), Covering America: Real Remedies for the Uninsured, pp. 155–172. Washington DC.
Ericson, Keith, 2014. Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange. American Economic Journal: Economic Policy 6 (1), 38–64.
Ericson, Keith Marzilli, Sydnor, Justin, 2017. The questionable value of having a choice of levels of health insurance coverage. The Journal of Economic Perspectives 31 (4), 51–72.
Ericson, Keith M. Marzilli, Starc, Amanda, 2015. Pricing regulation and imperfect competition on the Massachusetts health insurance exchange. Review of Economics and Statistics 97 (3), 667–682.
Farrell, Joseph, Balan, David, Brand, Keith, Wendling, Brett, 2011. Economics at the FTC: hospital mergers, authorized generic drugs, and consumer credit markets. Review of Industrial Organization 39 (1), 271–296.
Fleitas, Sebastian, Gowrisankaran, Gautam, LoSasso, Anthony, 2020. Reclassification Risk in the Small Group Health Insurance Market. Working Paper.
Frandsen, Brigham, Powell, Michael, Rebitzer, James B., 2019. Sticking points: common-agency problems and contracting in the US healthcare system. The Rand Journal of Economics 50 (2), 251–285.
Frech, Ted, Whaley, Christopher, Bowers, Liora, Simon, Carol, Handel, Benjamin R., Scheffler, Richard, 2015. Market power, transactions costs and the entry of accountable care organizations in health care. Review of Industrial Organization 47 (2), 167–193.
Gabaix, Xavier, Laibson, David, 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. The Quarterly Journal of Economics 121 (2), 505–540.
Galbraith, John K., 1952. American Capitalism: The Concept of Countervailing Power. Houghton Mifflin, Boston.
Garmon, Christopher, 2017. The accuracy of hospital merger screening methods. The Rand Journal of Economics 48 (4), 1068–1102.
Garthwaite, Craig, Ody, Christopher, Starc, Amanda, 2020. Endogenous Quality Investments in the US Hospital Market. NBER Working Paper 27440.
Gaynor, Martin, Town, Robert J., 2011. Competition in health care markets. In: Pauly, Mark V., McGuire, Thomas G., Barros, Pedro (Eds.), Handbook of Health Economics, vol. 2, pp. 499–637.
Gaynor, Martin, Vogt, W.B., 2003. Competition among hospitals. The Rand Journal of Economics 34 (4), 764–785.
Gaynor, Martin, Propper, Carol, Seiler, Stephan, 2016. Free to choose? Reform, choice, and consideration sets in the English national health service. The American Economic Review 106 (11), 3521–3557.
Gaynor, Martin, Rebitzer, James B., Taylor, Lowell J., 2004. Physician incentives in health maintenance organizations. Journal of Political Economy 112 (4), 915–931.
Gaynor, Martin, Ho, Kate, Town, Robert J., 2015. The industrial organization of health-care markets. Journal of Economic Literature 53 (2), 235–284.
Gaynor, Martin, Moreno-Serra, Rodrigo, Propper, Carol, 2013. Death by market power: reform, competition, and patient outcomes in the national health service. American Economic Journal: Economic Policy 121, 228–260.

Geruso, Michael, Layton, Timothy J., 2017. Selection in health insurance markets and its policy remedies. The Journal of Economic Perspectives 31 (4), 23–50.
Geruso, Michael, Layton, Timothy, 2020. Upcoding: evidence from Medicare on squishy risk adjustment. Journal of Political Economy 128 (3), 984–1026.
Geruso, Michael, Layton, Timothy J., McCormack, Grace, Shepard, Mark, 2019a. The Two Margin Problem in Insurance Markets. National Bureau of Economic Research Working Paper 26288.
Geruso, Michael, Layton, Timothy, Prinz, Daniel, 2019b. Screening in contract design: evidence from the ACA health insurance exchanges. American Economic Journal: Economic Policy 11 (2), 64–107.
Ghili, Soheil, 2020. Network Formation and Bargaining in Vertical Markets: the Case of Narrow Networks in Health Insurance. Yale University Working Paper.
Ghili, Soheil, Handel, Benjamin R., Hendel, Igal, Whinston, Michael, 2020. Optimal Long-Term Health Insurance Contracts: Characterization, Computation, and Welfare Effects. UC Berkeley Working Paper.
Goeree, Michelle Sovinsky, 2008. Limited information and advertising in the US personal computer industry. Econometrica 76 (5), 1017–1074.
Gowrisankaran, Gautam, Nevo, Aviv, Town, Robert, 2015. Mergers when prices are negotiated: evidence from the hospital industry. The American Economic Review 105 (1), 172–203.
Grennan, Matthew, 2013. Price discrimination and bargaining: empirical evidence from medical devices. The American Economic Review 103 (1), 145–177.
Grennan, Matthew, 2014. Bargaining ability and competitive advantage: empirical evidence from medical devices. Management Science 60 (12), 3011–3025.
Grennan, Matthew, Swanson, Ashley, 2020. Transparency and negotiated prices: the value of information in hospital-supplier bargaining. Journal of Political Economy 128 (4), 1234–1268.
Grieco, Paul L.E., McDevitt, Ryan C., 2016. Productivity and quality in health care: evidence from the dialysis industry. The Review of Economic Studies 84 (3), 1071–1105.
Grubb, Michael, Osborne, Matthew, 2015. Cellular service demand: biased beliefs, learning, and bill shock. The American Economic Review 105 (1), 234–271.
Gruber, Jonathan, McKnight, Robin, 2016. Controlling health care costs through limited network insurance plans: evidence from Massachusetts state employees. American Economic Journal: Economic Policy 8 (2), 219–250.
Gruber, Jonathan, Handel, Benjamin R., Kina, Sam, Kolstad, Jonathan, 2020. Managing Intelligence: Skilled Experts and AI in Markets for Complex Products. UC Berkeley Working Paper.
Haas-Wilson, Deborah, Garmon, Christopher, 2011. Hospital mergers and competitive effects: two retrospective analyses. International Journal of the Economics of Business 18 (1), 17–32.
Hackmann, Martin, Kolstad, Jonathan, Kowalski, Amanda, 2015. Adverse selection and the individual mandate: when theory meets practice. The American Economic Review 105 (3), 1030–1066.
Handel, Benjamin R., Hendel, Igal, Whinston, Michael D., 2015. Equilibria in health exchanges: adverse selection versus reclassification risk. Econometrica 83 (4), 1261–1313.
Handel, Benjamin R., 2013. Adverse selection and inertia in health insurance markets: when nudging hurts. The American Economic Review 103 (7), 2643–2682.
Handel, Benjamin R., Kolstad, Jonathan, 2015a. Getting the Most from Marketplaces: Smart Policies on Health Insurance Choice. Brookings Hamilton Project Discussion Paper 2015-08.

Handel, Benjamin R., Kolstad, Jonathan, 2015b. Health insurance for humans: information frictions, plan choice, and consumer welfare. The American Economic Review 105 (8), 2449–2500.

Handel, Benjamin R., Schwartzstein, Joshua, 2018. Frictions or mental gaps: what's behind the information we (don't) use and when do we care? The Journal of Economic Perspectives 32 (1), 155–178.

Handel, Benjamin R., Holmes, Jonathan, Kolstad, Jonathan, Lavetti, Kurt, 2021. What do Insurers do Differently than One Another? Managed Competition and Value Added. UC Berkeley Working Paper.

Handel, Benjamin R., Kolstad, Jonathan, Minten, Thomas, Spinnewijn, Johannes, 2021b. The Social Determinants of Choice Quality: Evidence from Health Insurance in the Netherlands. UC Berkeley Working Paper.

Handel, Benjamin R., Kolstad, Jonathan T., Spinnewijn, Johannes, 2019. Information frictions and adverse selection: policy interventions in health insurance markets. Review of Economics and Statistics 101 (2), 326–340.

Heidhues, Paul, Koszegi, Botond, 2018. Behavioral industrial organization. In: Bernheim, Douglas B., DellaVigna, Stefano, Laibson, David (Eds.), Handbook of Behavioral Economics, vol. 1, pp. 517–612.

Heiss, Florian, McFadden, Daniel, Winter, Joachim, 2010. Mind the gap! Consumer perceptions and choices of Medicare Part D description drug plans. In: Research Findings in the Economics of Aging, pp. 413–484.

Heiss, Florian, McFadden, Daniel, Winter, Joachim, Wupperman, Amelie, Zhou, Bo, 2016. Inattention and Switching Costs as Sources of Inertia in Medicare Part D. NBER Working Paper No. 22765.

Hendel, Igal, Lizzeri, Alessandro, 2003. The role of commitment in dynamic contracts: evidence from life insurance. The Quarterly Journal of Economics 118 (1), 299–328.

Hofmann, Annette, Browne, Mark, 2013. One-sided commitment in dynamic insurance contracts: evidence from private health insurance in Germany. Journal of Risk and Uncertainty 46, 81–112.

Ho, Kate, 2006. The welfare effects of restricted hospital choice in the US medical care market. Journal of Applied Econometrics 21 (7), 1039–1079.

Ho, Kate, 2009. Insurer-provider networks in the medical care market. The American Economic Review 99 (1), 393–430.

Ho, Kate, Pakes, Ariel, 2014. Hospital choices, hospital prices, and financial incentives to physicians. The American Economic Review 104 (12), 3841–3884.

Ho, Kate, Lee, Robin S., 2020. Health Insurance Menu Design for Large Employers. Princeton University Working Paper.

Ho, Kate, Lee, Robin S., 2017. Insurer competition in health care markets. Econometrica 85 (2), 379–417.

Ho, Kate, Lee, Robin S., 2019. Equilibrium provider networks: bargaining and exclusion in health care markets. The American Economic Review 109 (2), 473–522.

Ho, Kate, Hogan, Joseph, Morton, Fiona Scott, 2017. The impact of consumer inattention on insurer pricing in the Medicare Part D program. The Rand Journal of Economics 48 (4), 877–905.

Holmes, Jonathan, 2020. Does Medicaid Make Private Health Insurance Cheaper? UC Berkeley Working Paper.

Horn, Henrick, Wolinsky, Asher, 1988. Bilateral monopolies and incentives for merger. The Rand Journal of Economics 19 (3), 408–419.

Ho, Vivian, Hamilton, Barton, 2000. Hospital mergers and acquisitions: does market consolidation harm patients? Journal of Health Economics 19 (5), 767–791.

Jackson, Matthew O., Wolinsky, Asher, 1996. A strategic model of social and economic networks. Journal of Economic Theory 71 (1), 44–74.

Jaffe, Sonia, Shepard, Mark, 2020. Price-linked subsidies and imperfect competition in health insurance. American Economic Journal: Economic Policy 12 (3), 279–311.

Kaiser Family Foundation, 2010. Employer health benefits: 2010 annual survey. http://ehbs.kff.org/pdf/2010/8085.pdf.

Kaiser Family Foundation, 2011. Focus on health reform: summary of the affordable care act. http://kaiserfamilyfoundation.files.worldpress.com/2011/04/8061-021.pdf.

Kaiser Family Foundation, 2020. A dozen facts about Medicare advantage in 2020. https://www.kff.org/medicare/issue-brief/a-dozen-facts-about-medicare-advantage-in-2020/.

Ketcham, Jonathan, Lucarelli, Claudio, Miravete, Eugenio, Roebuck, Christopher, 2012. Sinking, swimming, or learning to swim in Medicare Part D. The American Economic Review 102 (6), 2639–2673.

Kolstad, Jonathan, 2013. Information and quality when motivation is intrinsic: evidence from surgeon report cards. The American Economic Review 103 (7), 2875–2910.

Kowalski, Amanda, 2016. Censored quantile instrumental variable estimates of the price elasticity of expenditure on medical care. Journal of Business & Economic Statistics 34 (1), 107–117.

Lavetti, Kurt, Simon, Kosali, 2018. Strategic formulary design in Medicare Part D plans. American Economic Journal: Economic Policy 10 (3), 154–192.

Lee, Robin S., Fong, Kyna, 2013. Markov-Perfect Network Formation: an Applied Framework for Bilateral Oligopoly and Bargaining in Buyer-Seller Networks. Unpublished.

Lewis, Matthew S., Pflum, Kevin E., 2017. Hospital systems and bargaining power: evidence from out-of-market acquisitions. The Rand Journal of Economics 48 (3), 579–610.

Lieber, Ethan, 2017. Does it pay to know prices in health care? American Economic Journal: Economic Policy 9 (1), 154–179.

Liebman, Eli, 2020. Bargaining in Markets with Exclusion: an Analysis of Health Insurance Networks. University of Georgia Working Paper.

Mahoney, Neale, Weyl, E. Glen, 2017. Imperfect competition in selection markets. Review of Economics and Statistics 99 (4), 637–651.

Manea, Mihai, 2018. Intermediation and resale in networks. Journal of Political Economy 126 (3), 1250–1301.

Marone, Victoria, Sabety, Adrienne, 2020. Should There be Vertical Choice in Health Insurance Markets? UT Austin Working Paper.

Marton, James, Yelowitz, Aaron, Talbert, Jeffrey, 2015. Medicaid Program Choice and Participant Inertia. University of Kentucky Working Paper.

Mehta, Nitin, Rajiv, Surendra, Srinivasan, Kannan, 2003. Price uncertainty and consumer search: a structural model of consideration set formation. Marketing Science 22 (1), 58–84.

Miller, Keaton, 2019. Estimating Costs When Consumers Have Inertia: Are Medicare Private Insurers More Efficient? University of Oregon Working Paper.

Miller, Keaton, Petrin, Amil, Town, Robert J., Chernew, Michael, 2019. Optimal Managed Competition Subsidies. UT Austin Working Paper.

Moscelli, Giuseppe, Gravelle, Hugh, Siciliani, Luigi, 2021. Hospital Competition and Quality for Non-emergency Patients in the English NHS.

Neilson, Christopher, 2020. Targeted Vouchers, Competition among Schools and the Academic Achievement of Poor Students. Princeton University Working Paper.

Newhouse, Joseph P., 1993. Free for All? Lessons from the RAND Health Insurance Experiment. Harvard University Press.

Pauly, Mark, Kunreuther, Howard, Hirth, Richard, 1995. Guaranteed renewability in insurance. Journal of Risk and Uncertainty 10 (2), 143–156.

Pauly, Mark V., 1968. The economics of moral hazard: comment. The American Economic Review 58 (3), 531–537.

Polyakova, Maria, 2016. Regulation of insurance with adverse selection and switching costs: evidence from Medicare Part D. American Economic Journal: Applied Economics 8 (3), 165–195.

Prager, Elena, 2020. Healthcare demand under simple prices: evidence from tiered hospital networks. American Economic Journal: Applied Economics 12 (4), 196–223.

Prager, Elena, Schmitt, Matt, 2021. Employer consolidation and wages: evidence from hospitals. The American Economic Review 111 (2), 397–427.

Prager, Elena, Tilipman, Nicholas, 2020. Regulating Out-of-Network Hospital Payments: Disagreement Payoffs, Negotiated Prices, and Access. Working Paper.

Rabin, Matthew, 1998. Psychology and economics. Journal of Economic Literature 36 (1), 11–46.

Raval, Devesh, Rosenbaum, Ted, Wilson, Nathan, 2020. Using Disaster Induced Closures to Evaluate Discrete Choice Models of Hospital Demand. Working Paper.

Romano, P., Balan, D., 2011. A retrospective analysis of the clinical quality effects of the acquisition of highland park hospital by Evanston northwestern healthcare. International Journal of the Economics of Business 18 (1), 45–64.

Rothschild, Michael, Stiglitz, Joseph, 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. The Quarterly Journal of Economics 90 (4), 630–649.

Schmalensee, Richard, 1989. Inter-industry studies of structure and performance. In: Schmalensee, Richard, Willig, Robert (Eds.), Handbook of Industrial Organization, vol. 2(1), pp. 951–1009.

Schmitt, Matt, 2018. Multimarket contact in the hospital industry. American Economic Journal: Economic Policy 10 (3), 361–387.

Scott Morton, Fiona, Kyle, Margaret, 2011. Markets for pharmaceutical products. In: Pauly, Mark V., Mcguire, Thomas G., Barros, Pedro P. (Eds.), Handbook of Health Economics, vol. 2. Elsevier, pp. 763–823. Chapter 12.

Shepard, Mark, 2020. Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange. Unpublished.

Song, Zirui, Safran, Dana, Landon, B., He, Yulei, Mechanic, Robert, Day, Matthew, Chernew, Michael, Ellis, Randall, 2011. Health care spending and quality in year 1 of the alternative quality contract. The New England Journal of Medicine 365 (10), 909–918.

Song, Zirui, Wallace, Jacob, Neprash, Hannah, McKellar, Michael, Chernew, Michael, McWilliams, J. Michael, 2015. Medicare fee cuts and cardiologist-hospital integration. JAMA Internal Medicine 175 (7), 1229–1231.

Starc, Amanda, Swanson, Ashley, 2021. Preferred pharmacy networks and drug costs. American Economic Journal: Economic Policy 13 (3), 406–446.

Starc, Amanda, Town, Robert J., 2019. Externalities and benefit design in health insurance. The Review of Economic Studies 87 (6), 2827–2858.

Strombom, Bruce, Buchmueller, Thomas, Feldstein, Paul, 2002. Switching costs, price sensitivity, and health plan choice. Journal of Health Economics 21 (1), 89–116.

Tebaldi, Pietro, 2020. Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design Under the ACA. Working Paper.

Tilipman, Nicholas, 2021. Employer Incentives and Distortions in Health Insurance Design: Implications for Welfare and Costs. Working Paper.

Town, Robert, Liu, Su, 2003. The welfare impact of Medicare HMOs. The Rand Journal of Economics 34 (4), 719–736.

Town, Robert J., Vistnes, Gregory, 2001. Hospital competition in HMO networks. Journal of Health Economics 20 (5), 733–753.

Trish, Erin E., Herring, Bradley J., 2015. How do health insurer market concentration and bargaining power with hospitals affect health insurance premiums? Journal of Health Economics 42, 104–114.

Vistnes, Gregory, Sarafidis, Yianis, 2013. Cross-market hospital mergers: a holistic approach. Antitrust Law Journal 79, 253–291.

Wallace, Jacob, 2019. What Does a Provider Network Do? Evidence from Random Assignment in Medicaid Managed Care. Yale University Working Paper.

Weyl, Glen, Veiga, Andre, 2017. Pricing institutions and the welfare cost of adverse selection. American Economic Journal: Microeconomics 9 (2), 139–148.

Whaley, Christopher M., 2015. Searching for Health: the Effects of Online Price Transparency. Working Paper.

Zeckhauser, Richard, 1970. Medical insurance: a case study of the tradeoff between risk spreading and appropriate incentives. Journal of Economic Theory 2 (1), 10–26.