Pay for Performance Contracts in Healthcare Markets
ECON 590/890

Luca Maini*

March 6, 2022

See most recent version here

CONTENTS

*These notes are a work in progress. Please check back frequently for an updated version, and email lmaini@email.unc.edu for corrections.

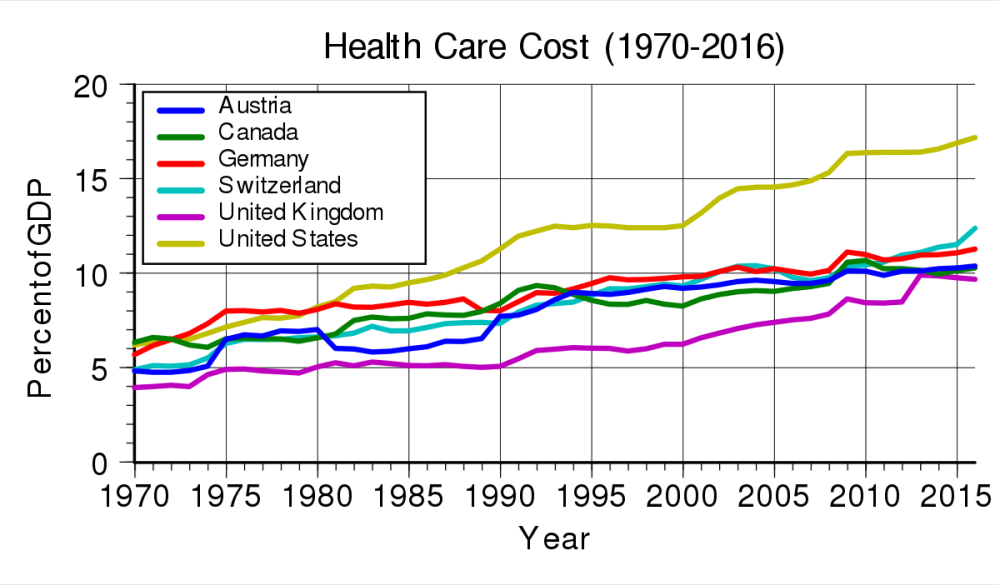

Figure 1: HEALTHCARE SPENDING AS A FRACTION OF GDP

1 OVERVIEW OF ISSUES

One of the themes that emerge from the study of physician agency and learning is that physicians—like all other economic agents—respond to incentives, financial and otherwise.

Historically, the US healthcare system has used a fee-for-service (FFS) mechanism to compensate medical care providers: doctors are paid for each patient they see, surgery performed, prescription doled out. While this mechanism makes intuitive sense (reward physicians who do more/work harder), it has crucial pitfalls. If doctors receive higher compensation for doing more, they will see more patients, perform more surgeries, prescribe more drugs. That is not always desirable because more care leads to more spending, frequently diminishing returns.

An excellent historical example of this phenomenon is days spent in the hospital. Medicare used to reimburse hospitals for each day a patient spent in the hospital. However, it turns out that staying in the hospital can be harmful. The risk of infection is very high, and as a general rule, patients should be discharged as soon as it is medically advisable. At some point, Medicare switched to a reimbursement model that did not depend on the length of a hospital stay. After that reform was implemented, inpatient stays became shorter, and health outcomes did not get worse (Lave, 1989). So the concern here is that if we incentivize spending, we will spend too much, and what we purchase with the additional spending may not be valuable.

The US has consistently ranked as the country with the highest health care spending in the world (Figure 1). Healthcare spending is approximately 20% of GDP and rising. Overutilization of care is a big reason why.

2 Physicians' Response to Financial Incentives

I will apply, for the benefit of the sick, all measures [that] are required, avoiding those twin traps of overtreatment and therapeutic nihilism.

Modern Hippocratic Oath

If you talk to healthcare providers, many of them might tell you that they do not change their behavior based on financial incentives: their treatment choices are based on the best course of action for the patient. That may be true, but a large body of research has determined that financial incentives affect the amount and type of services that patients receive. We tackle that here, and we start by providing a quick overview of physician payment schemes.

2.1 Overview of payment schemes

Physicians can be paid using three possible forms of compensation: fee-for-service, capitation, and salary.

Fee-for-service (FFS) is how physicians receive most of their compensation in the US. Under FFS, physicians receive a fee for every service they deliver: an office visit, surgery, X-rays, etc. FFS implies that physicians are paid proportionally to what they do: the more a physician does, the more money she gets paid.

Capitation is a collective word that describes a vast array of payment schemes that compensate physicians using lump-sum payments that cover bundles of services. For example, Medicare compensates hospital stays using fixed payments tied to a patient's diagnosis. Hospitals that care for patients with more severe conditions receive more money. However, the amount is fixed regardless of what the hospital does. Similar payment schemes can be tied to a period of time. For example, in North Carolina, Blue Cross Blue Shields pays many primary care physicians a lump sum payment for each patient they see. The payment covers the entire year, regardless of how many times the patient comes into the office, what tests they get, etc.[1] The incentives of capitation are the opposite of FFS: providers actually receive more money if they do less (more on this later).

Finally, physicians that are paid a **salary** receive the same amount of money regardless of what they do. This is the most neutral of all payment schemes because physicians do not have incentives to over provide or under provide care. However, it is pretty hard to pull off. Many physicians in the US are paid a salary. However, this salary is often tied to productivity, either directly (through performance bonuses) or indirectly (because the hospital pays the salary, and hospitals earn more money if they perform more services). The clearest way to detach compensation from productivity and make salaries completely neutral is to have a neutral entity pay for the cost of all services. Generally, this entity is the government, so salary payment schemes are the most common way physicians are compensated in countries with a single-payer system.

[1] This program is called Blue Premier: you can read about it here and here.

2.2 Do Physicians Respond to Financial Incentives?

Before we worry about the effect of payment systems, we should stop and consider whether physicians respond to financial incentives at all. We start by discussing a paper that uses a natural experiment to answer this question. Clemens and Gottlieb (2014) exploits a 1997 reform of the Medicare payment structure to study (i) whether medical treatment responds to financial incentives, and (ii) how patient health is affected by the change.

Using policy changes as natural experiments is a tried and true tradition for economists. In health care, where virtually every aspect of the market is subject to some form of regulation, such reforms are ubiquitous. In this case, the paper's case for causal identification is based on two elements. First, prices for services received by Medicare patients are set unilaterally by the government. As a result, these prices are plausibly exogenous. Second, these prices change pretty frequently as the government updates its formulas. In this particular case, the source of price variation comes from a reduction in the number of Medicare "regions". Medicare prices are region-specific because CMS (the Center for Medicare and Medicaid Services) wants to make sure that its prices reflect geographic variation in the cost of inputs such as rent, provider salaries, etc. In 1997, Medicare cut the number of regions from over 200 to less than 100. As a result, physicians in certain areas saw an increase in their compensation, and physicians in other areas saw a decrease in compensation.

Their paper is split into two parts. The first part is a simple event-study analysis of what happened before and after the reform. They set up the event study as a parametric diff-in-diff and regress a measure of total services provided at the area level on change in the reimbursement rate and various controls. They find pretty good evidence of a positive response in services, which appears to increase over time before flattening.

To understand the timing and size of these supply responses, they develop a model of physicians' joint supply and investment decisions. Doctors have heterogeneous productivity and can invest in a productivity-enhancing technology. They value their own income as well as patient health.

Physicians face a binary investment choice. They can practice medicine using a standard practice style (S) that has a variable cost of therapy $\bar{c}$ per unit of care, or an intense practice style (I) that reduces unit costs to $\underline{c}$ but costs $k > 0$ to adopt. Notice that this is not a quality-enhancing technology, but rather a cost-reducing technology.

Physicians differ in terms of their productivity γ_i, which is distributed on $(0, \infty)$ according to some known distribution $F(\cdot)$. The interpretation of γ_i is essentially "speed". Doctor i takes $\frac{1}{\gamma_i}$ units of time to produce one unit of care. Utility in the standard and intense practice regimes is

Clemens and Gottlieb (2014)—Paper at a glance

Positioning Statement: FFS medicine could lead to higher spending, but it can also expand access to valuable services. Understanding the tradeoff between these two forces is important.
Research Question: Do physicians' financial incentives affect medical treatment and patient health?
Answer: Physicians respond to financial incentives by increasing provision of services, and investing in new technology. The expansion in services does not lead to a significant change in patient outcomes.
Identification Strategy: 1997 regional consolidation of Medicare payment structure which harmonized payments from over 200 regions down to less than 100.

specified as

$$U_S(q;\gamma_i) = (r-\bar{c})\,q - e\left(\frac{q}{\gamma_i}\right) + \alpha b(Q)\,q \tag{1}$$

$$U_I(q;\gamma_i) = (r-\underline{c})\,q - k - e\left(\frac{q}{\gamma_i}\right) + \alpha b(Q)\,q \tag{2}$$

where r is the (exogenous) Medicare reimbursement rate, $e(\cdot)$ is an increasing and convex function of physician time that captures decreasing returns to leisure, α is a weight for patient welfare, and $b(Q)$ captures the health benefit of the marginal unit of care. They assume $b'(Q) < 0$—i.e. marginal benefits are decreasing in Q.

Under this setup, they show that physicians will invest above a certain productivity threshold $\gamma^\star$ which is decreasing in the reimbursement rate r, and in the weight placed on patient benefits α.

The point of the model is that the supply response of physicians needs to be broken down in two components: the marginal response of physicians who adjust their behavior, and the discrete response of physicians who invest in productivity-enhancing technology. When r grows to r', the threshold for technology adoption falls from $\gamma^\star(r)$ to $\gamma^\star(r')$. Physicians whose productivity lies on this interval will switch from the "standard" to the "intensive" mode of care, and will exhibit a very large response. Notice that the other way around is not quite symmetric: lowering r may not result in dis-investment of existing technology.

One implications of this model (and any model where patient welfare enters in the choice utility function) is that elective procedures should respond more strongly to financial incentives. They test this hypothesis and find some pretty good evidence in support of the theory (Figure 2).

2.3 How Do Physicians React to Incentives?

Further readings:

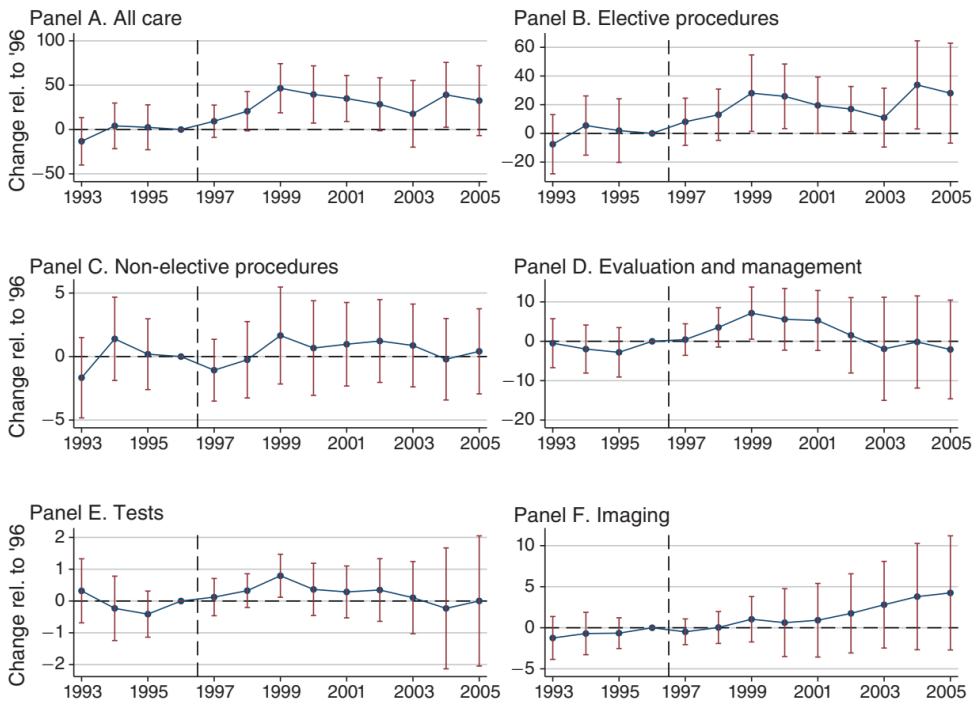

FIGURE 5. SUPPLY RESPONSE BY SERVICE CATEGORY

Figure 2: CHANGE IN MEDICAL CARE PROVISION (CLEMENS AND GOTTLIEB, FIGURE 5)

- Finkelstein (2007), on impact of Medicare expansion (dynamic incentives for technology adoption);
- Dafny (2005) on how *hospitals* respond to price changes by upcoding diagnoses;
- Jacobson et al. (2010) on how physicians may react to lower reimbursement by prescribing more drugs (income targeting).
- Melichar (2009) on how physicians treat capitated patients and FFS patients differently.
- Work on Medicare payment reform: Dranove and Ody (2019), Gross et al. (2021).
- A more direct literature estimates models of physician altruism where physicians care about financial incentives and patient health, following a model first developed by Ellis and McGuire (1986). Chandra et al. (2011) provides a good overview of this literature. More recent work in the area includes Dickstein (2017) and Xiang (2020). Also see Section 4 of the chapter on competition and consolidation in healthcare markets.

3 THE DESIGN OF PAY FOR PERFORMANCE SCHEMES

We now consider the optimal design of financial incentives for providers. The issues with FFS should be apparent by now: as Clemens and Gottlieb (2014) (and several other papers) shows

pretty convincingly, higher reimbursement leads to the provision of more care, with very little change in outcomes to show for it.

The last bit of the paper results appears particularly tantalizing: if increasing reimbursement to providers does not improve health outcomes, could the same be true if we decreased reimbursement? Two points about this question. First, it is not rhetorical. It is entirely possible that increasing compensation does not lead to better outcomes because we are already very close to the frontier of what is possible in medicine. Decreasing spending, however, could lead to worse health outcomes because the system no longer has the necessary resources to provide the right amount of care to everyone. Second, this is a very emotional and loaded question to many people and therefore must be handled with the proper care. When discussing spending reduction, we are not discussing cutting physician pay indiscriminately. Instead, we think about how to reduce overall healthcare spending in the context of the graph in Figure 1. Realistically, part of it will probably require addressing the structure of physician payments.

3.1 Issues in the Design of Payment Schemes

Most attempts at payment reform in the US have been forms of capitation. The appeal of capitation is relatively simple: if providers are overutilizing care, we can curb that behavior by making them responsible for the cost of additional care. The downside should be equally obvious: if we incentivize providers to do less, how can we ensure they will not *under provide* care? To understand why this question does not have a simple answer, consider a hypothetical capitation scheme that provides providers with a fixed payment for each patient for a given amount of time (say, one month). The payer has to make several decisions: (i) how is the payment determined? (ii) how are patients attributed to a specific provider? (iii) how can we ensure that patient health will not suffer? We tackle each issue separately.

How is the payment determined?

How does a system decide how much money to allocate to each patient? Ideally, the payment should reflect the cost of providing the appropriate level of care to the patient. In practice, coming up with a number that accurately reflects this goal is impossible for two reasons. First, our data will never be accurate enough to reflect all possible patient characteristics. Second, even if we did have that much data, it still reflects how physicians and patients are behaving in the current system and not in the "ideal" system that provides the appropriate amount of care.

Payers usually take a two-step approach to solving this problem. The first is to design a risk-adjustment model predicting spending based on observable patient characteristics, such as demographics, pre-existing conditions, and past illnesses. The second is to determine a benchmark—based on the model—against which to judge the hospital's performance. The simplest benchmark could be the predicted spending according to the model. In practice, payers often pick benchmarks that can be higher (say, to account for drug spending inflation) or lower (because the goal

is generally to spend less than in the past). A variety of additional considerations factor into these decisions:

- Should benchmarks be provider-specific, or should all providers share the same benchmark? Provider-specific benchmarks can account for unobserved diversity in patients (for example, teaching hospitals like Duke or UNC usually get patients with more expensive conditions), but also reduce the incentives to improve if providers with poor performance are held to a lower standard.
- If hospitals lower spending in a given year, should their benchmark be lower in the following year?
- Providers can be exposed to significant risk if, say, they receive many patients that are unobservably expensive. Should we include financial instruments that limit the exposure to this risk? What kind of instruments are most appropriate? (A few examples: cap payments at a certain threshold, exclude patients that exceed a certain spending threshold, use one-sided risk—e.g. hospitals can only get bonus payments, and are not on the hook for extra spending.)

How are patients attributed to providers?

In some cases, it is easy to see which provider is responsible for a patient's care. For example, in the case of Medicare's DRG payments for hospital stays, the payment goes to the hospital that admits the patient. In other cases, the attribution is less clear. Recently, Medicare launched a new program targeted at treating episodes of oncology care (Oncology Care Model, or OCM). Patients who start chemotherapy at selected hospitals around the country are included in this model, which provides bonus payments to providers that save money against a pre-established benchmark. However, some patients receive chemotherapy from more than one provider. How should those cases be handled?[2]

The issue of attribution is very complicated and tends to be resolved using relatively simple methods, like attributing the patient to the first provider they visit or to the provider that delivers the majority (or plurality) of care. The questions one should keep in mind are:

- If a patient sees two providers, which one is responsible for its care? How is the other provider paid? And should the main provider be responsible for the cost of care that another provider prescribes?
- How do we ensure that providers will not turn away unobservably expensive patients for fear of losing money?[3]

[2] OCM chose to attribute the patient to the provider who delivers the plurality of cancer-related office visits. Do you think this is a good idea? What are some pros and cons?

[3] There is contrasting evidence on this issue: McWilliams et al. (2019) does not find significant evidence of selection from the Medicare Shared Savings Program, but Ouayogodé et al. (2021) finds selection when examining Accountable Care Organizations.

How can we ensure that patient health will not suffer?

This is the last and perhaps most important question. Assuming we have created reliable benchmarks and a reasonable attribution strategy, how can we ensure that our payment model will not lead to a drop in patient health?

To ensure this does not happen, most payment models include quality controls that providers have to maintain to be eligible for bonus payments. This sounds straightforward in theory, but it is tough to do in practice.

The problem is that quality in healthcare settings is exceedingly hard to measure. To a certain degree, this is the same issue we encountered with risk adjustment: we do not have enough data about a patient, and therefore we cannot tell if they receive appropriate care or not.

To put these problems in a concrete setting, we will use the case of Accountable Care Organizations (ACOs). ACOs—created through incentives contained in the Affordable Care Act—are groups of doctors, hospitals, and other health care providers who come together to give coordinated care to Medicare patients. ACOs receive bonus payments from Medicare if they achieve spending savings while maintaining a high standard of care.

Figure 3 list all quality measures used by Medicare for their ACO model. To rate the quality of care provided by ACOs, Medicare uses a blend of ten patient experience measures (ACO-1 through ACO-46), measurable health outcomes (ACO-8 through ACO-43 and ACO-40 through ACO-28), and adherence to the standard of practice (ACO-43 through ACO-42).

Each set of measures captures something that the other measures do not. Patient experience is necessary because it is the only way to measure intangible aspects of care that are nonetheless important, such as the spread of information to patients, the ability of patients to receive timely care, etc. Measurable health outcomes are important because they capture the most crucial aspect of quality of care: a patient's health status. Finally, adherence to established standards of care can help ensure providers provide the proper care in cases where measuring outcomes directly is not possible.

It is also easy to see why each of these measures is imperfect. Patient experience is subject to a patient's assessment, and patients are not always the best judges of medical care. 1 Measures of health outcomes are probably ideal but also incredibly limited. You will see that the list for ACOs covers only a fraction of all possible relevant outcomes. Moreover, these outcomes are not always under the control of the physician. A provider can prescribe the correct medication or perform surgery to perfection, but what if the patient does not take the drugs or follow post-operative care guidelines? Adherence to medical standard measures are a better reflection of physician input but are also imperfect: standards of care are rarely universal, and physicians can sometimes improve patient outcomes by deviating from these standards.

Domain	Measure	Description	Pay-for-Performance Phase In			30th perc.	40th perc.	50th perc.	60th perc.	70th perc.	80th perc.	90th perc.
			PY1	PY2	PY3							
Patient/Caregiver Experience	ACO-1	CAHPS: Getting Timely Care, Appointments, and Information	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Patient/Caregiver Experience	ACO-2	CAHPS: How Well Your Providers Communicate	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Patient/Caregiver Experience	ACO-3	CAHPS: Patients' Rating of Provider	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Patient/Caregiver Experience	ACO-4	CAHPS: Access to Specialists	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Patient/Caregiver Experience	ACO-5	CAHPS: Health Promotion and Education	R	P	P	56.26	57.57	58.86	60.08	61.39	62.83	64.90
Patient/Caregiver Experience	ACO-6	CAHPS: Shared Decision Making	R	P	P	56.82	57.98	59.17	60.20	61.46	62.77	64.90
Patient/Caregiver Experience	ACO-7	CAHPS: Health Status/Functional Status[1]	R	R	R	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Patient/Caregiver Experience	ACO-34	CAHPS: Stewardship of Patient Resources	R	P	P	24.23	25.47	26.68	27.90	29.19	30.88	32.90
Patient/Caregiver Experience	ACO-45	CAHPS: Courteous and Helpful Office Staff[2]	R	R	R	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Patient/Caregiver Experience	ACO-46	CAHPS: Care Coordination[2]	R	R	R	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Care Coordination/ Patient Safety	ACO-8	Risk-Standardized, All Condition Readmission[3]	R	R	P	15.75	15.62	15.50	15.38	15.23	14.97	14.56
Care Coordination/ Patient Safety	ACO-38	Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions[3]	R	R	P	66.46	62.37	58.85	55.49	52.15	48.57	43.74
Care Coordination/ Patient Safety	ACO-43	Ambulatory Sensitive Condition Acute Composite (AHRQ Prevention Quality Indicator (PQI) #91)[3,4]	R	R	R	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Care Coordination/ Patient Safety	ACO-13	Falls: Screening for Future Fall Risk	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Preventive Health	ACO-14	Preventive Care and Screening: Influenza Immunization	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Preventive Health	ACO-17	Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention[5]	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Preventive Health	ACO-18	Preventive Care and Screening: Screening for Depression and Follow-up Plan[6]	R	R	R	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Preventive Health	ACO-19	Colorectal Cancer Screening	R	R	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Preventive Health	ACO-20	Breast Cancer Screening	R	R	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00
Preventive Health	ACO-42	Statin Therapy for the Prevention and Treatment of Cardiovascular Disease[1]	R	R	R	N/A	N/A	N/A	N/A	N/A	N/A	N/A
At-Risk Population Depression	ACO-40	Depression Remission at Twelve Months[1]	R	R	R	N/A	N/A	N/A	N/A	N/A	N/A	N/A
At-Risk Population Diabetes	ACO-27	Diabetes Mellitus: Hemoglobin A1c Poor Control[3]	R	P	P	70.00	60.00	50.00	40.00	30.00	20.00	10.00
At-Risk Population Hypertension	ACO-28	Hypertension (HTN): Controlling High Blood Pressure	R	P	P	30.00	40.00	50.00	60.00	70.00	80.00	90.00

Figure 3: ACO QUALITY MEASURES BENCHMARKS FOR 2020-21

3.2 Some examples of pay-for-performance and capitation schemes in the US

The following is a (very incomplete) list of alternative payment models and pay-for-performance contracts implemented at various levels of the US healthcare system. For a broader overview looking beyond the US, see Mjåset et al. (2020).

- **Alternative Quality Contract (Blue Cross Blue Shield of Massachusetts):** Song et al. (2012, 2014, 2019).
- **Oncology Care Model and Oncology Care First (Medicare):** Kline et al. (2015, 2017); Brooks et al. (2019).
- **Accountable Care Organizations (Medicare):** Fisher and Shortell (2010); Frech et al. (2015); McWilliams et al. (2016).
- **Medicare ESRD Bundled Payment:** Eliason et al. (2020)

3.3 Exercise: an alternative payment model for maternity care

In this exercise we apply what we have learned so far to a practical problem: the design and analysis of an alternative payment model based around maternity care.

Design

We consider an insurance company who is designing an alternative payment model for a large hospital system (e.g. UNC). This hospital system employs several gynecologists and PCPs that operate in an outpatient setting, but also receives referrals from physicians outside the system. The hospital system is made up of a flagship hospital, which is equipped to care for all types of patients and a series of satellite hospitals, which may lack the facilities to take care of the most complicated patients. Currently, the system is paid on a FFS basis, and services received in the flagship hospital cost more than identical services received in the satellite hospitals because the flagship hospital has higher maintenance costs (due to the extra equipment and staffing needs). The hospital system also has competitors in the area, including at least one other large hospital system (e.g. Duke). The two systems compete for patients, but can also refer patients to each other if appropriate.

Discussion:

- What care does a pregnancy require?
- What is the single determinant of the cost of a pregnancy?
- Why would a pregnant woman choose to deliver in the flagship vs. satellite hospitals?
  - Distance, comorbidities, copays

Ho and Pakes (2014)—Paper at a glance

Positioning Statement: Capitated payment models like ACOs aim to reduce spending by altering physicians' financial incentives, but little is known about how physicians trade off quality and monetary incentives.
Research Question: Do physicians under capitated payment models send patients to cheaper, but lower-quality hospitals?
Answer: Physicians under capitated payment models do send patients to cheaper hospitals, that are farther away but of similar quality.
Identification Strategy: Structural model that allows for flexible differences in quality across hospitals, but imposes some functional form assumptions on physician-patient choice utility.

We consider the following:

1. Who is included in the design model? (points for largest number of patients included)
2. What costs does the model include? (points for large number of costs/time span)
3. What is the attribution method? (points for accuracy, limiting selection concerns)
4. How is payment determined? (points for minimum gameability)
5. How is quality ensured? (points for minimum gameability)

4 ESTIMATING THE IMPACT OF PAY-FOR-PERFORMANCE SCHEMES

Suppose we have now implemented an alternative payment model. How do we evaluate its success? A good starting point is to look at changes in cost/utilization patterns relative to expectations. If you have a reasonable control group, this can be a relatively straightforward exercise. Checking the impact of the alternative payment model on quality is a bit harder. If you find lower quality according to the metrics that the model targets explicitly you can be pretty certain that quality dropped. More often than not, however, you will find that quality metrics remained the same, or even improved. This is not surprising: because those metrics are targeted by the model, they become focal to the operations of the hospital and therefore will rarely suffer. However, untargeted (and potentially unobserved) measures of quality might suffer. Another issue that you might worry about is patient selection. If the attribution algorithm allows it, providers may be inclined to turn costly patients away. The selection concern is particularly troublesome because it could potentially affect the interpretation of results even if one set up a perfect randomized controlled trial. These considerations raise two issues: (i) how do we account for patient selection, and (ii) how do we check for changes in quality?

The identification problem at hand is not a simple one. We want to compare the choices of physicians paid through capitated models and physicians paid through traditional FFS. You

should realize immediately that we cannot simply compare the two. Since physicians can choose where to work, capitated physicians may be different than physicians paid through FFS. In this case, however, the identification issues run deeper than selection into treatment and control. Suppose that we had a quasi-experimental, or even a totally random assignment of physicians to financial models. If we compare spending and utilization we can say that any difference is due to the financial model—that's what random assignment guarantees. But what do these differences reflect? A key problem here is what we are able to measure. Our estimates will reflect a (risk-adjusted) difference in what happens to patients treated by each group of physicians. But are we observing two comparable samples of patients. Physicians can respond to financial incentives along several margins: they can change their utilization patterns, yes, but they can also upcode patients, or turn them away. This is going to alter the characteristics of their group of patients in ways that are not easily observable to the physician. On top of that, patients choose their doctors, and their decision may depend on the financial model. If a capitated physician is reluctant to prescribe a test or medication that they deem unnecessary, a patient might simply go to a different physician to get it.

All of this points to the need for additional structure for identification to be meaningful. This is very hard, and, to the best of my my knowledge, no paper has managed to do all of this. Here, I discuss a paper—Ho and Pakes (2014)—that makes significant progress in estimating quality in a flexible way (though it essentially punts of the issue of patient selection).

The paper uses a structural model to compare the choices of capitated physicians and physicians paid through FFS. They do not have a source of plausibly exogenous variation that assigns physicians to a given set of financial incentives, so they simply assume away the issue of physician heterogeneity. They also—largely—assume away the issue of patient heterogeneity.[4] What they do is think carefully about quality.

To make the research question feasible, they narrow their focus to the choice of hospital for child delivery. Their starting point is a random utility choice model. The choice utility of patient i, whose health plan is from insurer π, and choosing hospital h is

$$W_{i,\pi,h} = \theta_{p,\pi} p\left(c_i, h, \pi\right) + g_\pi\left(q_h\left(s\right), s_i\right) + f_\pi\left(d\left(l_i, l_h\right)\right) + \varepsilon_{i\pi h}. \tag{3}$$

This functional form allows for a large degree of flexibility:

- Both price sensitivity and the weight given to distance between hospital and patient $d\left(l_i, l_h\right)$ are insurer-specific (to allow for differences in the characteristics of patients across insurers)
- Price can depend on hospital and insurer, but also on the medical condition (diagnosis) c_i of patient i
- Hospital quality $q_h\left(\cdot\right)$ is severity-specific (severity is indicated by a measure of sickness

[4] They runs a few balance tests to compare patients who are treated by the two groups physicians and finds no significant differences between the two. But since the main concern is about unobservable differences, I think addressing this issue remains beyond the scope of the paper.

level s_i), to indicate that some hospitals may be better equipped than others to deal with certain types of patients. Notice that this means that hospitals need not be purely vertically differentiated.

- The weight given to quality is flexibly expressed by the function $g_\pi(\cdot)$, and can vary by severity. This allows for quality to become a secondary concern if say, a patient is healthy and presents no complications.

Keep in mind that the goal is to estimate $\theta_{p,\pi}$ for physicians who are under capitation models (the variation is at the level of the insurer π), and then to compare the choices these two types of physicians make.

You might notice that the kind of data required to estimate this model is unlikely to exist. Actual prices paid by insurers are hard to come by. Hospital quality is usually unobserved. This can sometimes be addressed by using a hospital fixed-effect, but in this model that is not enough, because quality is multi-dimensional. Finally, notice that all functions are expressed flexibly, which would require non-parametric estimation.

As a starting point, you might want to try making some functional assumptions and trying to estimate a simple logit model. For example, assuming you have access to a dataset with claims, you could set actual price to be an insurer-specific fraction of observed allowed amounts:

$$p(c_i, h, \pi) = \delta_h l p^o(c_i, h) \tag{4}$$

(notice that this means you will only be able to jointly estimate $\theta_{p,\pi} \times \delta_h$, unless you somehow observe an estimate of δ_h). Next, you can approximate hospital quality as a combination of hospital fixed effects hospital characteristics interacted with patient sicknes level:

$$g_\pi(q_h(s), s_i) = q_h + \beta z_h x(s_i). \tag{5}$$

Finally, $f_\pi(\cdot)$ could be approximated as a polynomial:

$$f_\pi(d(l_i, l_h)) = \theta_{d1} d(l_i, l_h) + \theta_{d2} d(l_i, l_h)^2. \tag{6}$$

These assumptions impose two restrictions that are worth discussing. First, you assume that there is no error in the price measure. Even simple measurement error is going to attenuate your estimates. This is a larger issue in a context where measurement error definitely exists. Second, you severly restrict the functional form of hospital quality. To be fair, the function specified in equation 5 can be fairly flexible in theory. In practice, however, one can only include a limited number of characteristics in the estimation before running into computational issues or the incidental parameter problem, whereby estimated coefficients become increasingly unreliable. Since the functional form is restricted, we expect some form of measurement error in this term as well. Measurement error in quality is concerning because quality is going to be correlated with both prices and choices and generate an upward bias in the estimated price coefficient.

To avoid these assumptions and still recover a consistent estimate for $\theta_{p,\pi}$ they use an inequality estimator based on revealed preferences. It is useful to take a step back and think about why you might be inclined towards this type of estimator. The core issue at stake is that we do not observe quality, but keep in mind that we also don't need to estimate any of the parameters associated with quality (we are interested mainly in $\theta_{p,\pi}$). A common strategy to take care of unobservables that generate bias but are not the focus of the analysis is to difference them out (think about first-difference models in time series). A natural way to build differences in random utility choice models is to compare the observed choice to its possible alternatives. Since agents maximize utility, we know, by revealed preferences, that the chosen action yields higher utility than all other options. In our case, we know that for patient i choosing hospital h

$$\theta_{p,\pi} p\left(c_i, h, \pi\right) + g_\pi\left(q_h\left(s\right), s_i\right) + f_\pi\left(d\left(l_i, l_h\right)\right) + \varepsilon_{i\pi h} > \\ \theta_{p,\pi} p\left(c_i, h', \pi\right) + g_\pi\left(q_{h'}\left(s\right), s_i\right) + f_\pi\left(d\left(l_i, l_{h'}\right)\right) + \varepsilon_{i\pi h'} \quad (7)$$

for all other hospitals h' in the choice set. This inequality in and of itself does not help us too much because we cannot cancel out $g_\pi\left(q_h\left(s\right), s_i\right)$ and $g_\pi\left(q_{h'}\left(s\right), s_i\right)$. So we need to find a way to add the first term to the RHS of the inequality and the second term to the LHS of the inequality. To see how we can do that, imagine that we can find another patient, patient i', who is also covered by insurer π and has severity level $s_{i'} = s_i$, but who chooses to go to hospital h' (while having hospital h in its choice set). For patient i',

$$\theta_{p,\pi} p\left(c_{i'}, h', \pi\right) + g_\pi\left(q_{h'}\left(s\right), s_{i'}\right) + f_\pi\left(d\left(l_{i'}, l_{h'}\right)\right) + \varepsilon_{i'\pi h'} > \\ \theta_{p,\pi} p\left(c_{i'}, h', \pi\right) + g_\pi\left(q_{h'}\left(s\right), s_{i'}\right) + f_\pi\left(d\left(l_{i'}, l_{h'}\right)\right) + \varepsilon_{i'\pi h'} \quad (8)$$

If we sum inequalities 7 and 8 we get

$$\theta_{p,\pi}\left(p\left(c_i, h, \pi\right) + p\left(c_{i'}, h', \pi\right)\right) + \boxed{g_\pi\left(q_h\left(s\right), s_i\right) + g_\pi\left(q_{h'}\left(s\right), s_{i'}\right)} \\ + f_\pi\left(d\left(l_i, l_h\right)\right) + f_\pi\left(d\left(l_{i'}, l_{h'}\right)\right) + \varepsilon_{i\pi h} + \varepsilon_{i'\pi h'} > \\ \theta_{p,\pi}\left(p\left(c_i, h', \pi\right) + p\left(c_i, h', \pi\right)\right) + \boxed{g_\pi\left(q_{h'}\left(s\right), s_i\right) + g_\pi\left(q_{h'}\left(s\right), s_i\right)} \\ + f_\pi\left(d\left(l_i, l_{h'}\right)\right) + f_\pi\left(d\left(l_i, l_{h'}\right)\right) + \varepsilon_{i\pi h'} + \varepsilon_{i\pi h'}$$

The two boxed parts of the new inequality cancel out because hospital quality is assumed to be severity-specific, so $g_\pi\left(q_h\left(s\right), s_i\right) = g_\pi\left(q_{h'}\left(s\right), s_i\right)$. Notice that prices do not necessarily cancel out because patients with the same severity can have different conditions c_i, which result in different hospital prices. Hence, we are left with

$$\theta_{p,\pi}\left(p\left(c_i, h, \pi\right) + p\left(c_{i'}, h', \pi\right)\right) + f_\pi\left(d\left(l_i, l_h\right)\right) + f_\pi\left(d\left(l_{i'}, l_{h'}\right)\right) + \varepsilon_{i\pi h} + \varepsilon_{i'\pi h'} > \\ \theta_{p,\pi}\left(p\left(c_i, h', \pi\right) + p\left(c_i, h', \pi\right)\right) + f_\pi\left(d\left(l_i, l_{h'}\right)\right) + f_\pi\left(d\left(l_i, l_{h'}\right)\right) + \varepsilon_{i\pi h'} + \varepsilon_{i\pi h'}$$

At this point we almost have something we can estimate: we can measure everything in this equation except for the logit error terms. To average out the logit error terms all we need is a group of patients like patient i and a group of patients like patient i' (this also takes care of any eventual measurement error in price) our final condition is

$$\sum_{i} \theta_{p,\pi} \left(p\left(c_i, h, \pi\right) + p\left(c_{i'}, h', \pi\right) \right) + f_{\pi}\left(d\left(l_i, l_h\right)\right) + f_{\pi}\left(d\left(l_{i'}, l_{h'}\right)\right) > \sum_{i'} \theta_{p,\pi} \left(p\left(c_i, h', \pi\right) + p\left(c_i, h', \pi\right) \right) + f_{\pi}\left(d\left(l_i, l_{h'}\right)\right) + f_{\pi}\left(d\left(l_i, l_{h'}\right)\right)$$

We can build an inequality restriction for all pairs of patients i and i' we can find. Put together, the inequalities identify a set of values of $\theta_{p,\pi}$ that satisfy all the inequalities (if no value of $\theta_{p,\pi}$ satisfies all the inequalities, then the single value that minimizes violations is chosen as the estimate).

5 BIBLIOGRAPHY

Brooks, Gabriel A., Shalini Jhatakia, Amanda Tripp, Mary Beth Landrum, Thomas J. Christian, Gabriella Newes-Adeyi, Susannah Cafardi, Andrea Hassol, Carol Simon, and Nancy L. Keating, "Early Findings From the Oncology Care Model Evaluation," *Journal of Oncology Practice*, 2019, *15* (10), e888–e896.

Chandra, Amitabh, David Cutler, and Zirui Song, "Who Ordered That? The Economics of Treatment Choices in Medical Care," in "Handbook of Health Economics," Vol. 2, Elsevier B.V., 2011, pp. 397–432.

Clemens, Jeffrey and Joshua D. Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

Dafny, Leemore S., "How Do Hospitals Respond to Price Changes?," *American Economic Review*, 2005, *95* (5), 1525–1547.

Dickstein, Michael J., "Physician vs. Patient Incentives in Prescription Drug Choice," 2017.

Dranove, David and Christopher Ody, "Employed for Higher Pay? How Medicare Payment Rules Affect Hospital Employment of Physicians," *American Economic Journal: Economic Policy*, 2019, *11* (4), 249–271.

Eliason, Paul J, Benjamin Heebsh, Riley J League, Ryan C Mcdevitt, and James W Roberts, "The Effect of Bundled Payments on Provider Behavior and Patient Outcomes," 2020.

Ellis, Randall P. and Thomas G. McGuire, "Provider behavior under prospective reimbursement. Cost sharing and supply," *Journal of Health Economics*, 1986, *5* (2), 129–151.

Finkelstein, Amy, "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare," *The Quarterly Journal of Economics*, feb 2007, *122* (1), 1–37.

Fisher, Elliott S. and Stephen M. Shortell, "Accountable Care Organizations: Accountable for What, to Whom, and How Elliott," *JAMA*, 2010, *304* (15), 1715–1716.

Frech, H. E., Christopher Whaley, Benjamin R. Handel, Liora Bowers, Carol J. Simon, and Richard M. Scheffler, "Market Power, Transactions Costs, and the Entry of Accountable Care Organizations in Health Care," *Review of Industrial Organization*, 2015, *47* (2), 167–193.

Gross, Tal, Adam Sacarny, Maggie Shi, and David Silver, "Regulated Revenues and Hospital Behavior: Evidence from a Medicare Overhaul," 2021.

Ho, Kate and Ariel Pakes, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, dec 2014, *104* (12), 3841–3884.

Jacobson, Mireille, Craig C. Earle, Mary Price, and Joseph P. Newhouse, "How medicare's payment cuts for cancer chemotherapy drugs changed patterns of treatment," *Health Affairs*, 2010, *29* (7), 1391–1399.

Kline, Ron, Kerin Adelson, Jeffrey J. Kirshner, Larissa M. Strawbridge, Marsha Devita, Naralys Sinanis, Patrick H. Conway, and Ethan Basch, "The Oncology Care Model: Perspectives From the Centers for Medicare & Medicaid Services and Participating Oncology Practices in Academia and the Community," *American Society of Clinical Oncology Educational Book*, 2017, *37*, 460–466.

Kline, Ronald M., Carol Bazell, Erin Smith, Heidi Schumacher, Rahul Rajkumar, and Patrick H. Conway, "Centers for Medicare and Medicaid Services: Using an episode-based payment model to improve oncology care," *Journal of Oncology Practice*, 2015, *11* (2), 114–116.

Lave, J. R., "The effect of the medicare prospective payment system," *Annual Review of Public Health*, 1989, *10*, 141–161.

McWilliams, J. Michael, Laura A. Hatfield, Bruce E. Landon, and Michael E. Chernew, "Spending Reductions in the Medicare Shared Savings Program: Selection or Savings?," 2019.

_ , _ , Michael E. Chernew, Bruce E. Landon, and Aaron L. Schwartz, "Early Performance of Accountable Care Organizations in Medicare," *New England Journal of Medicine*, 2016, *374* (24), 2357–2366.

Melichar, Lori, "The effect of reimbursement on medical decision making: Do physicians alter treatment in response to a managed care incentive?," *Journal of Health Economics*, 2009, *28* (4), 902–907.

Mjåset, Christer, Navraj S Nagra, and Thomas W Feeley, "Value-Based Health Care in Four Different Health Care Systems," *NEJM Catalyst*, 2020.

Ouayogodé, Mariétou H., Ellen Meara, Kate Ho, Christopher M. Snyder, and Carrie H. Colla, "Estimates of ACO savings in the presence of provider and beneficiary selection," *Healthcare*, 2021, *9* (1).

Song, Zirui, Dana Gelb Safran, Bruce E. Landon, Mary Beth Landrum, Yulei He, Robert E. Mechanic, Matthew P. Day, and Michael E. Chernew, "The 'alternative quality contract,' based on a global budget, lowered medical spending and improved quality," *Health Affairs*, 2012, *31* (8), 1885–1894.

_ , Sherri Rose, Dana G. Safran, Bruce E. Landon, Matthew P. Day, and Michael E. Chernew, "Changes in Health Care Spending and Quality 4 Years into Global Payment," *New England Journal of Medicine*, 2014, *371* (18), 1704–1714.

_ , **Yunan Ji, Dana G. Safran, and Michael E. Chernew**, "Health Care Spending, Utilization, and Quality 8 Years into Global Payment," *New England Journal of Medicine*, 2019, *381* (3), 252–263.

Xiang, Jia, "Physicians as Persuaders: Evidence from Hospitals in China," 2020.