FOR MORE THAN THREE AND A HALF DECADES, health care expenditures in the United States have grown at a much higher rate than those in other wealthy nations. Evidence clearly shows that high prices are the primary driver of this growth; the U.S. pays more than comparable nations for the same health care procedures, services, drugs and devices. In order to control expenditures, we must get a handle on prices. Price control strategies to date have relied on administering prices through the public insurance system, or using market forces in the private insurance system to drive down the price and cost of care through consumer action—with each effort encountering a series of obstacles to achieving its desired result. In addition, payment in the United States is fractured across more than one million health plans sponsored by private employers, thousands of plans sponsored by public employers, thousands of Marketplace plans, 51 state Medicaid programs offering scores of managed care plans, dozens of separate CHIP plans, and Medicare. The result is that no one has the power or the incentive to control overall expenditures. Rate setting, by contrast, consolidates buying power. Given that our four decades-old experiments with market competition has failed to control expenditures, it seems time to consider the use of state rate setting.

Key Features of Rate Setting

- Rates must be updated periodically with an eye toward only those factors over which providers exercise no control, such as the costs of inputs.
- The system must include all payers.
- The necessary Medicare waiver must help fend off the demands of providers and local interests.
- The system must also be simple and transparent.
- Stakeholders, including insurers and providers, must buy into the system.

Rate setting in the U.S. has traditionally been applied to public utilities, which sell standardized products for mass consumption, like water and electricity, and possess natural monopolies. By contrast, health care rate setting necessarily involves extremely heterogeneous goods and services that raise complex, important, and difficult quality issues. As a result, rate setting must pertain to some unit of payment that averages across types of services and patients, such as Medicare’s inpatient DRG-based system, which bundles an array of services into a single consolidated unit of payment, the DRG. However, payment units like DRGs allow providers to shift services to more lucrative settings, not subject to the set rates, and to increase either or both the volume and intensity of care. To prevent such behavior, global units of payment are preferable.

State rate setting, when properly designed and implemented, has a strong track record. Properly designed, rate setting can control health care prices and the volume and intensity of services. It also promises large administrative savings. Key features and governance structures are described below.

Designing and Governing Rate Setting

- The rate regulatory board should be independent of other executive agencies and should have a dedicated funding source, separate from general revenues.
- The board’s members should be appointed to relatively long terms, be experts in the field and not be employed by or have affiliations with the entities they regulate.
- A strong conflict of interest policy also is necessary.
- A rate-setting board must also be supported by highly trained professional staff insulated from day-to-day lobbying.

Framing the Issue

Properly designed rate setting has been successful in controlling expenditures in the United States and elsewhere. What we have been doing is not working. Rate setting should be high on the policy and political agenda.
Introduction

EVEN AS IT has flattened in recent years, the growth in U.S. health care spending remains a major concern. National expenditures remain much higher than those of comparable nations. If growth picks up again—as now seems to be occurring and many predict will continue—this gap may widen, with health care spending as a percentage of GDP reaching 20 percent within a decade. This level of spending heavily burdens the economy, particularly workers, and it will strain government budgets while threatening the viability of public insurance programs on which well over 100 million Americans depend.

Spending growth is driven by a number of factors, including unmanaged, and potentially avoidable, chronic illnesses such as diabetes that trigger a higher need for health care; advances in medical technology that drive up the volume and intensity of care; and the price of health care goods and services. It is widely recognized, however, that price is the principal reason why U.S. health care expenditures are so much higher than those in other wealthy nations. To forestall continued growth in spending, we must gain control over price.

Like other wealthy nations (with limited exceptions), the United States relies on the private sector to deliver health care. Unlike other nations, however, the United States has taken a unique approach to financing care that relies on competition among payers, insurers and providers to control health care spending. As a result, payment in the United States is fragmented while in other wealthy countries payment is consolidated. Additionally, over time and particularly recently, U.S. health care providers have become increasingly consolidated. This combination of fragmentation on the payment side and consolidation on the provider side makes it more likely that providers will be able to command higher prices and even less likely that our decades’ long strategy of relying on competition—essentially a grand experiment—will control spending.

Many people now put their faith in payment-centered strategies such as changing the method of payment from one based on volume to one based on “value.” However, any new method of payment will be implemented in this underlying structure that will remain unchanged and spending will continue to rise. Health care expenditures will reach a breaking point—if they haven’t already—such that spending for health care will crowd out spending needed to meet other vital needs, for families, communities, and governments alike. The pressure on workers is particularly significant, with a combination of inflation and workers’ contribution to premiums eating up the lion’s share of wage increases (e.g., Claxton et al. 2015, Exhibit B). It may be time to rethink our approach and to turn to rate setting, a method that has had success both abroad and in the United States.

Following an overview of the nation’s health care spending problem, we examine the two principal ways the United States has tried to control expenditures—direct regulation of prices paid to providers in public insurance programs such as Medicare, and market-based strategies in the case of private insurance. We describe the ways that each approach is falling short and then turn to state rate setting to explore its potential.

Before proceeding, it is important to define our terms precisely. We use the term “payers” to describe the purchasers of health care, most importantly plan sponsors, whether private—e.g., employers—or public—e.g., Medicare. In our terminology, “insurers” are not payers but instead are intermediaries between payers and providers. Insurers may bear risk or when they service self-insured plans, their intermediary function is limited to furnishing administrative services. The term “all-payers rate setting,” often used below, is somewhat of a misnomer, because in rate setting the entities regulated are neither purchasers—“payers”—nor their insurer/administrator intermediaries but instead are providers—sellers—which must sell their goods or services at set rates to “all payers.” Therefore, it would make sense to use the term, “all-seller rate setting” instead but we use the traditional term, “all-payer rate setting.”
Background

Health care expenditures in the United States are high and growing higher. Data from 2013 (Mossialos et al. 2016) show that, whether measured as a percentage of gross domestic product (GDP) (17.1 percent) or as dollars spent per capita ($9,086), expenditures are nearly 50 percent higher than the countries with the next highest level of expenditures, France (11.6 percent GDP) or Switzerland ($6,325 per capita). Data from a year earlier starkly illustrate the historical trend (Figures 1 and 2).

Growth in expenditures has flattened somewhat in recent years, averaging 3.7 percent—2.4 percent per capita—over the 2009-2013 five-year time period. However, recently national health spending has accelerated, growing by 5.3 percent between 2013 and 2014, 4.5 percent per capita (Martin et al. 2016). The Office of the Actuary for the Centers for Medicare & Medicaid Services (CMS) projects that roughly this rate of growth will continue, averaging 5.8 percent—4.9 percent per capita—from 2014-2024, and consuming 19.6 percent of GDP by 2024 (CMS 2016b), although views differ on the amount of continued escalation (e.g., Getzen 2015; Roehrig 2015). Continued growth at such levels will put great pressure on federal and state budgets and greatly suppress the growth of wages (e.g., Holahan et al. 2011).²

Recent research has placed blame for high spending squarely on the prices we pay to providers (e.g., Anderson et al 2003; Angrisano et al. 2007; Laugesen and Glied 2011), as well as our high administrative expenses, while ruling out other drivers of expenditures (e.g., Casalino et al. 2009; Morra et al. 2011; Woolhandler, Campbell, and Himmelstein 2003). The population in the United States is no older than that of other advanced, industrialized nations; in fact, in many cases, the converse is true. In the United States we don’t utilize higher amounts of health care goods and services, with the possible exception of very costly technology and some specialty care; and we don’t substitute more intensive for less intensive care any more than in comparable nations. Nor are the higher expenditures in the United States explained by higher quality; again, the converse is often the case. The simple truth is that in the United States we pay providers more than others do (e.g., Reinhardt 2012); and medical prices have been the largest component of growth of expenditures (Figure 3).

The domination of price is highest in the private insurance sector’s expenditure growth, as illustrated by figures from 2014, which show price rising even in the face of decreased utilization (Figure 4).
Figure 3. Factors Accounting For Growth In Per Capita National Health Expenditures, 2004-2014

Source: Centers for Medicare & Medicaid Services, Office of the Actuary, National Health Statistics Group.

Figure 4. Changes in Utilization and Prices of Medical Subservice Categories: 2014

Source: HCCI, 2015
Notes: All data weighted to reflect the national, younger than 65 ESI population.
Data from 2013 and 2014 adjusted using actuarial completion.
Medicare’s Administered Pricing Approach for Controlling Hospital and Physician Spending

STARTING IN THE 1970S, with federal support approximately 30 states turned toward some form of rate regulation of hospital prices. Despite evidence that some forms of rate regulation had the capacity to limit expenditures, by 1997 state rate regulation had ended in all but two states, Maryland and West Virginia, in part swamped by the antiregulatory fervor that began to sweep across the country in the late 1970s and particularly in the 1980s after the election of President Ronald Reagan, but also, if not much more, the result of a number of factors, some of which are discussed below, that caused the demise of collective support from the hospital industry, business and labor interests, insurers and state officials (e.g., Murray and Berenson 2015).

It was rather ironic, then, that Medicare’s systems of administered pricing began in this era, with the implementation of the Inpatient Prospective Payment System (IPPS) in 1983. At that time the system was revolutionary because hospitals were no longer to be reimbursed for their properly reported costs, but were instead to be paid prospectively—put on a diet with the number of calories set in advance. Hospitals would then “compete” in the sense that those whose costs were less than the prospective budget would keep the surplus, while those whose costs exceeded the budget would absorb the loss. Over time this competition would drive the prospective budget downward to the costs of “efficient” production, and expenditures thereby would be controlled. Medicare has now extended prospective payment to most portions of its programs.

For five principal reasons Medicare’s administered pricing systems have met with limited success in controlling overall expenditures. These reasons constitute cautionary tales for state rate setting, as we explain below.²

1. Fragmented modes of payment. Medicare payment is structured around silos pertaining to different sites of care or types of providers; it is fragmented in this fashion. Thus when pressure is placed on the expenditure balloon in one place, it pops out in another. The examples are legion: hospitals have shifted care from inpatient settings, paid under Part A, to outpatient ones, paid under Part B, because payment for the latter has been more lucrative; hospitals buy up physician practices so that physician services can be billed, not simply as professional services, but as hospital outpatient care (and through contractual arrangements physicians can share in the gains); hospitals conducted pre-surgical testing prior to admission to get paid for it twice, once under Part B for outpatient testing and then under Part A for what should have been testing during the inpatient stay.

More recent efforts by the federal government to move Medicare more decisively to alternative payment methods that emphasize bundled payments for entire episodes of care (such as hip and knee replacements) represent attempts to counteract this fracturing, but the tortured path to such reforms is evidenced by the fact that even in the case of a single type of care—joint replacement—the government has had to move slowly, with a limited number of demonstrations dotting the landscape rather than a single unified effort. Moreover, just this single effort pertaining to one type of procedure, albeit an important one, required CMS to issue long and complex rules (CMS 2015). These efforts are largely grafted onto the preexisting fee-for-service structure, imposed on just one entity, typically hospitals, which are then supposed to push the expenditure constraint down the line to other types of providers, e.g., rehabilitation facilities, and thereby creating conflicts of interest among providers. Most likely the different silos of payment will continue to predominate, as will the concomitant opportunities to shift costs and sites of care. The fact that so many gaps exist in payment systems led one seasoned observer to put it, “games, games, games” (McDonough 1997a, 148).

2. Partial provider capture of payment systems. Providers often have some degree of control over the payment system. Medicare’s physician payment system, under which physicians are paid according to a Resource-Based Relative Value Scale (RBRVS), provides perhaps the starkest example. The scale itself was initially designed by researchers at Harvard, working with CMS’s predecessor, the Health Care Financing Administration. Like any such scale it must periodically be updated to take into account changes such as new technology and new modes of practice. Although CMS makes the ultimate decisions, it relies heavily on the American Medical Association’s Relative Value Scale Update Committee (RUC) (albeit with modifications in recent years in how CMS reviews the Committee’s recommendations). Substantial evidence shows that over time the Committee’s dominance by specialists has skewed the scale toward procedures and imaging, furnished most often by specialists, and thereby has actually increased the differential between payment for specialty and primary care, an effect converse to one of the purposes in the adoption of the payment system (e.g., MedPAC 2006, ch. 3; 2011a, 12-17). Berenson and Goodson (2016) show convincingly that the scale is thirty years behind, freezing in place technology and modes of practice for specialty care that no longer exist, thereby greatly inflating payment for specialist services, while grossly underpaying the “cognitive work—the critical thinking involved in data gathering and analysis, planning, management,
decision making, and exercising judgment in ambiguous or uncertain situations” (2), particularly in caring for patients who present with “several chronic conditions, take multiple medications, and see numerous clinicians” (3).

3. Providers gaming the payment system. Providers can game the payment system. “Upcoding,” for example, is common behavior because fuzziness exists among coding categories and providers can code to receive higher payment. Medicare’s IPPS, based on diagnosis-related groups (DRGs), has been the subject of “DRG creep” over its history (Frankford 2017). Medicare Advantage plans are paid in part based on their enrollees’ health status, and substantial evidence exists that Part C plans are upcoding by increasing the number of diagnosis codes in enrollees’ medical records (Geruso and Layton 2015; Kronick and Welch 2014). MedPAC has recently expressed concern that inpatient rehabilitation facilities are engaging in similar behavior (MedPAC 2016). Similarly, services, like imaging, can be unbundled and billed separately, with a concomitant increase in volume, as has happened with physician-owned, in-office imaging facilities (e.g., Baker 2010). The examples are legion.

4. Failure to control volume and intensity. Systems of administered pricing that pay unit-prices alone lack control over volume and intensity of care. The most salient example is provided by Medicare’s physician payment system, which is fee-for-service. In theory, the relative-value scale objectively limits the amount any physician can obtain for a particular service. However, physicians, more than any other types of providers, can increase the volume of services or their intensity, or simply shift to more lucrative services, to make up for income diminished because of the cap on the amount paid for any service.

The RBRVS-based system was designed so that this “volume effect” would be countered by a formula—last named the Sustainable Growth Rate (SGR) formula—to take back any increase in aggregate expenditures resulting from increased volume or intensity of care. From 2002 to 2015, actual expenditures exceeded the target, but starting in 2003, Congress intervened each year to reduce or eliminate the take-back for that year. The formula then compounded the shortfall in successive years—just like an interest rate compounds each year by adding to the base over which it is taken—and updates in any given year were to be based on what spending would have been had the full reductions occurred. The effects would have been very substantial. For 2015, the take-back would have cut fees by a whopping 21.2 percent (CMS 2014, 67742). Because that level of shock to providers is unreasonable, in 2015 Congress effectively gave up and repealed the SGR formula (Medicare Access and CHIP Reauthorization Act of 2015).

5. “Medicare-only” payment. Medicare’s systems of administered pricing follow a “Medicare-only” policy—design and payment are focused on Medicare alone without regard to other payers (e.g., MedPAC 2011b, 51-57). However, the fragmentation in payment outside Medicare’s systems of administered pricing pulls up expenditures within the systems. While Medicare has been able to reduce its expenditures relative to other payers (e.g., Gaynor 2016; Russell 1989), expenditures continue to escalate in the health care system as a whole. The gap between what is paid to hospitals by Medicare and private payers continues to increase (e.g., Selden et al. 2015), as illustrated by the fact that Medicare margins are now negative 5 to negative 6 percent and expected to sink to negative 9 percent in 2016, while their overall profits have reached a 30-year high (MedPAC 2016). Experts have debated whether Medicare has shifted costs to other payers and thereby pushed up their prices (e.g., Dobson, DaVanzo, and Sen 2006), or whether the problem is that private payers have failed to control their expenditures and pulled up Medicare’s (e.g., Frakt 2011; White and Wu 2014). While the weight of evidence and opinion now pretty much refutes the cost-shifting claim, nonetheless this “push-pull” debate obscures the fact that the continuing escalation of expenditures is a joint product of all of Medicare and private payers—they’re all in it together because the problem is fragmented payment (e.g., Reinhardt 2006, 2012). We return more generally to this point below.

Using Market Forces

SOMETHING IRONICALLY, at the same time that Medicare was shifted to a system of administered pricing for hospital and physician care, the private sector was being moved to a system of competition to control spending, making the United States “the odd man out” internationally (Abel-Smith 1985). Probably most important in this transformation was a belief that the nation’s vast sea of payers would gain control over spending if health care providers could only be forced to negotiate with them as individuals rather than as collectives. This approach in turn depended on more rigorous application of the antitrust laws against organized medicine, aimed at eliminating its control over health care payment. The net effect of this change was to separate payment for health care from its delivery, outside of a few fully integrated systems like Kaiser, Group Health of Puget Sound, and Geisinger.

Most important to the unleashing of market forces has been the creation or facilitation of markets. Part of this task has been the attempts over the years to create providers linked in some way to bear risk and furnish care over the continuum of care.
There are many iterations of this theme, with cynics pointing to recycled ideas characterized only by new branding (e.g., Marmor and Oberlander 2012); Enthoven's managed care; independent delivery systems; clinics without walls; accountable care organizations (ACOs). Another part of the effort to create or facilitate markets has been to enable comparison shopping, of which the ACA's Marketplaces are the most prominent example. The effort to make price and quality transparent is of course necessary to operation of these shopping centers, because a consumer cannot make choices without adequate price and quality information. Finally, for the markets to work, shoppers must have incentives to shop, and the leading mechanism of doing this is to give “consumers” “skin in the game” through means like higher deductibles, copays and coinsurance, and the recent experiments with reference-pricing.

Problems Faced by Market-Based Solutions

This effort to use markets to control spending faces numerous obstacles because it is widely recognized that health care markets deviate substantially from more ordinary markets, even relatively complex ones that produce custom products. If health care markets fail, then their use cannot control expenditures. Six features of health care markets, discussed below—and many of which overlap—contribute to this failure.

1. **Heterogeneity everywhere.** Compare the purchase of eyeglasses to health care.\(^3\) There are many distinctive features to different eyeglasses: size, shape, color, material, texture, weight and brand names (like fancy designers). Ordinary people, looking for eyeglasses, can easily discern these features and have much or all of the necessary knowledge to make educated choices among them; and shoppers can try eyeglasses prior to purchase simply by putting them on. Eyeglasses are relatively “homogeneous” and quintessentially “shoppable.”

By contrast, health care goods and services are most often extremely “heterogeneous.” A multitude of possible treatments exist for any given “condition”—e.g., for back pain, one might buy different furniture, use aspirin, use an NSAID like ibuprofen, try physical therapy, or be given multiple surgical options. Moreover, a “condition” can be treated in a multitude of ways in different persons—e.g., compare stroke treatment in an otherwise healthy and active 45-year-old man (with a family history) to treatment for a 75-year-old woman who is institutionalized quite often but otherwise lives alone on the third floor of a building without an elevator, and who also has hypertension and diabetes and moderate congestive heart failure. Finally, there is often vast variation in the price and quality offered by different providers. Purchasing in the face of such a huge number of contingencies and options is extremely difficult at best, especially for laypersons who know only that they need care.\(^4\)

2. **Dependence and agency.** To make even relatively rudimentary choices in treating complex, high-cost conditions, patients need expertise, but they are not experts. They might know something is wrong—being tired, feeling a lump, experiencing pain and so on—but they don’t know what the problem is (if they even know that a problem exists). Furthermore, one’s condition—purposely put in quotes above—cannot be known without diagnosis. To even know which of the multitude of different goods and services are needed, one needs to know the diagnosis before sorting through the options. To diagnose and treat most illness, therefore, patients need experts, aka “agents,” but they are dependent on those agents because they don’t know what the experts know. This problem is referred to as “information asymmetry.”

Necessarily, the problem arises, “How am I to choose an agent if I am not an expert?” Below we discuss the possibility of using insurers or employers, sophisticated buyers, as agents, but here we focus on consumers’ directly making choices among providers.

One way for consumers to overcome their ignorance is to try a product before choosing, as in our eyeglass example above. Most often, however, consumers don’t know what they need and therefore what they should try. Moreover, one cannot try a service in advance of the production of a service. One can’t try a surgeon without having surgery. The simultaneous production and consumption of health care makes such a strategy quite perilous.

Some health economists argue that providing consumers with information directly will lead to less reliance on agents. However, the dependence on agents stems from much more than lack of expertise. Patients are vulnerable not just because of their lack of knowledge but because they are necessarily emotionally involved. For this reason, some type of agent becomes essential, given the impossible task of choosing when one is sick and vulnerable. The necessary decisions do not resemble the discretionary purchase of a high-end television; it is selecting among costly approaches with unknown effects, for conditions whose true nature might not even be understood and the consequences of which can be disability or death.

The extreme heterogeneity coupled with the lack of consumer wherewithal to navigate among options (with perhaps limited exceptions for relatively discretionary health purchases that entail information that can be used by lay people) means that very few services are “shoppable.” A recent definition of “shoppable” is the following: “For a health care service to be ‘shoppable,’ it must be a common health care service that can be researched (‘shopped’) in advance; multiple providers of that service must be available in a market (i.e., competition); and sufficient data about the prices and quality of services must be available” (Frost and Newman 2016). One recent study has estimated that *at most* roughly one-third of total expenditures on health care are for shoppable services (White and Eguchi 2014).
Even for those services, consumers will shop only if the amount of their “skin in the game” makes the effort worthwhile. Insurance with first-dollar coverage obviously makes nothing worth shopping, while complete lack of insurance makes much, much more shopable. A recent study from the Health Care Cost Institute examined employer-provided insurance from 2011 and reached the following conclusions again as maximums: (1) of the $524.2 billion in expenditures, 43 percent, $225.4 billion, was spent on shoppable services; (2) 15 percent of total expenditures were out of pocket; and (3) only $37.7 billion, 7 percent of total out-of-pocket expenditures, was spent on shoppable services (Frost and Newman 2016). The take-away is that unless out-of-pocket costs are very significantly increased—with the limit being complete de-insurance—relatively little is to be gained from consumer shopping. Moreover, recent evidence shows that even when consumers are given the incentive to shop by the use of high deductible plans, their choices are perverse and their choices don’t improve with experience. Consistent with the findings from the RAND Health Insurance Experiment of many years ago, relatively high-income, sophisticated consumers did not respond to the incentives of high deductibles by shopping but instead simply reduced the quantity of their care across the board, failing to differentiate among services that have value from those that did not (Brot-Goldberg et al. 2015). Consequently, consumers chose to forego even high-value care.

3. **Shortage of sophisticated purchasers/agents.** The situation might be totally different, however, if consumers could rely on sophisticated purchasers as their agents in choosing among providers and in choosing among services. A leading choice for this position is, of course, insurers. However, that just pushes the agency problem back by a degree, because now consumers need agents to pick their insurers as the agents that in turn pick their covered providers and treatments. In other words, consumers lack the knowledge or information to pick their insurer agents (who pick their providers as agents) because consumers lack the needed knowledge—much less the information—to choose this agent.

The evidence that this problem exists—regardless of how consumer-friendly mechanisms like the ACA Marketplaces are made—is more than substantial. Numerous studies show that even relatively sophisticated persons choosing among insurance products get overloaded by too many choices and too much information, and they lack even rudimentary understanding of basic features of insurance products such as coinsurance, deductibles and networks (e.g., Bhargava, Loewenstein, and Sydnor 2015; Loewenstein et al. 2013).

For the most part in the United States employers have been the dominant agents shopping for insurance for their employees. However, this solution has proved inadequate for a number of reasons. First, the employer system leaves out millions, who in turn depend on public programs such as Medicare and Medicaid to act as their agents. Furthermore, employer-provided insurance and the segment of the population it covers has been declining. At its zenith in the late 1970s and early 1980s, employer-sponsored insurance covered approximately 70 percent of the non-elderly population (e.g., Gabel 1999, 65). Since then there has been a relatively steady march downward, with coverage of the non-elderly population falling as low as 58.4 percent in 2011 during the Great Recession but rebounding to, and possibly stabilizing at, 60 percent in 2013-2014 (Fronstin 2015b, 7).

Moreover, even when employers sponsor insurance, huge variation exists in terms of their incentives or capacity to serve as adequate agents for their employees. To function as agents, employers must be willing to invest in their employees, which means that in order to benefit from that investment they must have expectations of a fairly stable workforce (e.g., Fronstin and Helman 2003). These incentives vary across sectors of the economy and employers within one sector may have different incentives depending on such factors as firm size, the size, mixture and age of a firm’s workforce, and local labor market conditions (e.g., Buchmueller, Carey, and Levy 2013; Christianson and Trude 2003; Fronstin 2007; Morran, Chernew, and Hirth 2001). Thus, even among many employers offering insurance, incentives are relatively short term (e.g., Adams and Salisbury 2014); what may be good for the employer may not be good for employees over the long run.

Large employers and employer coalitions engage in serious efforts to understand their employees’ preferences and incorporate those preferences into plan designs (Peel et al. 2000). By focusing on dimensions of care like quality and satisfaction, instead of just on their own bottom line, these plan sponsors have acted as “responsible purchasers” (Lo Sasso et al. 1999) when they arrange for insurance or buy care directly from providers (e.g., Eggbeer, Morris, and Sukenik 2016). However, most employers lack this incentive or capability, particularly as scale decreases (e.g., Cebul et al. 2011; Rosenthal et al. 2007), and most workers are not employed by the large firms that have the capacity—technical skill and market power—to control expenditures. As a result, most employers have little choice but to turn to tools developed by the insurance industry itself to hold down premiums—high deductible plans and narrow networks—simply to cut their expenditures.

4. **Fragmentation among payers.** This then brings us to the most important point, the extreme fragmentation of the financing system. With more than one million health plans sponsored by private employers, thousands of plans sponsored by public employers, thousands of Marketplace plans, 51 state Medicaid programs, nearly as many CHIP programs, and Medicare, payment is fractured and lacks more than minimal cooperation despite efforts by some states and CMS to begin to develop multi-payer initiatives.
Even if these initiatives are successful to some degree in controlling expenditures—and the evidence so far is not very promising (e.g., Dale et al. 2016)—they remain few in number and, moreover, they are voluntary, relying on governments’ enticement of private parties rather than a direct exercise of state power by such means as rate setting.

Thus, almost universally every payer fends for itself and has incentives with regard only to the insurance pools for which its plans are responsible. As discussed above, even a payer as large as Medicare attends only to its budget, obligated from its inception to pay only its “fair share” of costs. No payer has the incentive or capability to control expenditures overall. Instead, payers—and risk-bearing providers too—most often take a path of least resistance, which is to reduce expenditures by pushing risk and cost to someone else, rather face down powerful interests, overcome problems of collective action or challenge an organization’s fundamental way of doing things (e.g., Evans 1990; Marmor 2010). Externalities abound.

This (non)system of payment stands in stark contrast with those of other wealthy democracies, which accomplish risk pooling using government sponsorship or government-regulated social organizations (or both), thereby creating systems grounded in social solidarity and social security. In Western Europe, with its strong tradition of solidarity and mutual aid, the risk of illness is pooled in the sickness funds, which cooperate voluntarily and with varying degrees of state coercion in collecting revenues, spreading risk, and paying providers. In national health system countries, like the United Kingdom and the Scandinavian countries, national governments unite all citizens into a single risk pool. In Canada, Canadian Medicare brings the population together into the risk-pooling systems of the provinces and territories, with some degree of federal subsidization. All advanced, industrialized countries socialize the risk-pooling function, either as large as Medicare attends only to its budget, obligated from its inception to pay only its “fair share” of costs. No payer has the incentive or capability to control expenditures overall. Instead, payers—and risk-bearing providers too—most often take a path of least resistance, which is to reduce expenditures by pushing risk and cost to someone else, rather face down powerful interests, overcome problems of collective action or challenge an organization’s fundamental way of doing things (e.g., Evans 1990; Marmor 2010).

Externalities abound.

The insurance side of the market has likewise become much more consolidated. (e.g., Dafny 2010; Dafny, Duggan, and Ramanarayanan 2012). The most visible evidence of this trend is the possibility that the Aetna-Humana and Anthem-Cigna mergers, if allowed by federal and state antitrust departments, other state officials and the courts, will reduce the big five insurers down to three.

Even without these mergers, the figures are very troubling. In the commercial market “[i]n 2004, the largest insurers controlled more than half the market in 16 states and at least one-third of the market in 38 states. Between 1998 and 2006, the fraction of health care markets that were concentrated to levels high enough to raise antitrust concerns, according to the U.S. Department of Justice’s Horizontal Merger Guidelines, increased from 68 percent to 99 percent” (NASI 2015, 11). In more detail, “AMA data show that 64 percent of commercial health insurance markets are already highly concentrated. Twenty percent of these markets [greatly exceed the standard criteria for high concentration]. Fifty-three percent of those markets have two insurers that account for 65 percent or more of the combined market for HMO, PPO, and POS insurance services. Other studies indicate that in 74 percent of states, the three largest insurers hold 80 percent or more of the market share in each of the individual, small group, and large group market segments. Nationally, the share of the largest four insurers increased from 74 to 83 percent from 2006 to 2014” (Greaney and Moss 2016, 3-4). In the Medicare Advantage market, Biles, Casillas, and Guterman (2015, 1) found that “97 percent of markets in U.S. counties are
highly concentrated and therefore lacking in significant MA plan competition. Competition is considerably lower in rural counties than in urban ones. Even among the 100 counties with the greatest numbers of Medicare beneficiaries, 81 percent do not have competitive MA markets. Market power is concentrated among three nationwide insurance organizations in nearly two-thirds of those 100 counties."

Many reasons have been given for increased consolidation. On the provider side there are claims that greater scale is necessary to bargain effectively against consolidating insurers, to obtain efficiencies, to bear risk, to take advantage of information technologies, to prepare for and implement systems of payment supposedly based on value, to take care of populations, and to reduce duplication (e.g., Burns, Goldsmith, and Sen 2013; Gaynor 2016). On the insurer side, there is likewise a claim that enhanced bargaining power is needed to counter concentrated providers—that size will enable insurers to push down provider prices—but there are also claims of reduced administrative costs, in particular improved risk bearing, higher quality and that investment in IT can be spread across a larger base (e.g., Dafny 2015). However, aside from highly centralized hospital systems—particularly when the system's hospitals are few in number, close to each other geographically and tightly integrated—there is little if any empirical evidence to support these claims either on the provider side (e.g., Burns, Goldsmith, and Sen 2013; Burns et al. 2015; Gaynor and Town 2012; Goldsmith et al. 2015) or the insurer side (e.g., Dafny 2015).

There is substantial empirical evidence, by contrast, that provider consolidation confers greater market power and the possibility of increasing price. Substantial evidence shows that prices increase after hospitals merge (e.g., Gaynor and Town 2012). One recent study found increased prices even when hospitals in different, within-state local markets merge, a finding that is particularly troubling because antitrust law and officials are usually concerned only with concentration in local markets (Dafny, Ho, and Lee 2016). More generally, concentrated hospital markets are strongly correlated with higher prices (e.g., Cooper et al. 2015). Concentrated physician markets have been studied less often than hospital markets but evidence correlates higher fees with greater concentration (e.g., Austin and Baker 2015; Baker, Bundorf, and Royalty 2014; Sun and Baker 2015). More recent evidence shows that vertical integration among hospitals and physicians likewise increases prices (e.g., Baker, Bundorf, and Kessler 2014; Conti, Landrum, and Jacobson 2016; Goldsmith, Kaufman, and Burns 2016; Neprash et al. 2015; Robinson and Miller 2014). Finally, evidence shows that insurance premiums are higher in concentrated insurance markets (e.g., Dafny 2010; Dafny, Duggan, and Ramanarayanan 2012).

The evidence of the effects of consolidation on payers is not good—for payers at least. In local markets where fragmented providers face an insurer with market power, providers’ prices either fall or stabilize. However, some evidence shows that these prices are not reflected in lower premiums for plan sponsors (Dafny, Duggan, and Ramanarayanan 2012). When the situation is reversed—when consolidated providers face fragmented insurers—providers’ prices rise. Insurers pass these increases onto payers in the form of higher premiums (e.g., Town et al. 2006; Trish and Herring 2015). Finally, when both sides of the provider-insurer market are consolidated, one can infer from available evidence that the concentrated insurers do not pass along any profits they might wrest from consolidated hospitals (Dafny, Duggan, and Ramanarayanan 2012; Scheffler et al. 2016). Indeed, some evidence exists that the two sides just shake hands, sharing together the increased premiums imposed on plan sponsors (Greaney 2016; Rosenbaum et al. 2012, 1339-41). It seems that consolidation at either or both levels results in higher premiums.

If plan sponsors lacking market power resist the premium increases, they then have to absorb the extra cost, which will be passed back to their workers in some form such as lower wages or benefits (e.g., Baicker and Chandra 2006), or accept plans with higher out-of-pocket costs for plan members, more shallow coverage, narrower networks, or some combination of the above—all forms of less comprehensive insurance. The effect of this “de-insurance” of plan members is that the consequences of power possessed by providers, insurers or both effectively get absorbed either by plan sponsors, plan members or some combination of the two.

In short, growing market consolidation seems to make it even less likely that growth in expenditures will fall, as increasingly concentrated providers and insurers face fragmented payers. At some point, giant insurers and mega-health systems may begin to integrate into the fully integrated models represented by Kaiser or Geisinger, but this does not mean that prices will fall, especially if few competitors in any given market are left standing. It is difficult not to conclude that our four decade-old experiment with using markets to control spending has failed, that what we have to show for it is the steady erosion in coverage for tens of millions of insured Americans, and a health care market that is harder than ever for normal humans to understand. We can move the coverage and payment levers forever, but this will not change the basic underlying fact of provider and insurer consolidation and greater spending for us, the ultimate payers. As Vladeck and Rice have put it, “The current U.S. health care system can be described, in at least one respect, as a massive engine for the redistribution of resources from employers, taxpayers, and households to the organizations that provide health care goods and services, and the people (including us) who work for such organizations” (Vladeck and Rice 2009, 1314)."
State Regulation of Prices

GIVEN THE FAILURE of a market-based payment system to control spending and prices, and given the extant and growing consolidation among providers and insurers, state regulation of provider prices may offer a promising alternative strategy. As we indicated above, the experience with Medicare’s and Medicaid’s administered pricing systems (as well as many systems across the world not delineated here) provides lessons for the design of a state system of rate regulating provider prices. To summarize, those problems are: (1) fragmented modes of payment because different sites of care or types of providers are paid under different payment systems, such as Medicare Part A and Part B; (2) provider capture of a payment system, such as RUC’s substantial control over updating the RBRVS for physician payment; (3) providers’ gaming—manipulating—the coding within a payment system to obtain payment under more lucrative codes; (4) a system’s inability to control volume—e.g., the number of hospital admissions—or intensity of care—e.g., substituting a PET scan for a CT scan; and (5) a system’s failure to control overall expenditures because the system has little or insufficient control over all payments. The design features we discuss below eliminate or ameliorate these problems.

1. Does it work? When done correctly, yes, rate setting controls prices and expenditures. Rate setting has a strong track record in the United States, particularly in Maryland (e.g., Murray and Berenson 2015), and moreover internationally. At different times it controlled prices to various degrees in a number of states, and one can reasonably conclude that its failure at other times—for example, in New Jersey—was due to capture or features of design, some of which are discussed below, that with very careful design can be avoided. Indeed, a report in 2011 from the Urban Institute canvassed various policy options for bringing the growth in national health expenditures closer to growth in GDP, and the potential savings from rate setting greatly eclipsed any other option (Holahan et al. 2011). It is the most promising policy option available.6

2. Why does it work? Rate setting works, quite simply, because it consolidates the demand-side of health care payment into one unit, thereby asserting collective bargaining power against the provider side. It is the aggregation of purchasing power such that wealth is transferred to consumers/taxpayers. Properly designed, it can control health care prices and the volume and intensity of services. It also promises large administrative savings in that rates are transparent and they are standardized (e.g., White 2009), in contrast to the bedlam that now characterizes payment in the United States in which everyone has to keep track of a myriad of conflicting, complex rules negotiated among a horde of payers, insurers and providers (e.g., Reinhardt 2006).

3. How does it work? Rate setting typically conjures up the image of a regulated public utility like water, natural gas or electricity companies, which sell standardized products for mass consumption. These regulated public utilities have natural monopolies because only one set of large fixed costs—say, the water pipes connecting the water treatment plant with all homes in a city—is sufficient to service all conceivable demand; and in fact, competition would clearly result in duplication of resources—it doesn’t make sense to have two sets of water pipes. Given that public utilities necessarily possess monopoly power and that everyone needs access—everyone needs water—public utility regulation rests on an implicit bargain that the utilities are guaranteed a stable, relatively low profit in return for an obligation to service all customers with a price set by government.

To some extent, rate setting in health care is similar. Health care providers do not have power as natural monopolies, but, as discussed above, their power stems from market failure and concentration in certain geographic markets, with the result that, like public utility regulation, purchasers must exercise collective power to control providers’ prices. The deal struck in rate-setting health care providers, therefore, resembles that in public utility regulation: providers are supposed to be guaranteed some level of financial stability, a point to which we return below, and the quid pro quo is that they are deprived of their power in that rates are set and their profits regulated.

Rate setting for public utilities is relatively simple because quality is relatively easily assured and, as examples, “a kilowatt hour of electric energy is a kilowatt hour, and a therm of natural gas having prescribed characteristics is just that” (Priest 1970, 842). Thus, public utility regulation can set a single price for a homogeneous product that raises few issues about quality.7

Rate setting in health care, by contrast, necessarily involves extremely heterogeneous goods and services in which quality is a complex, important and difficult issue—this challenge faced by a market-based payment system doesn’t magically vanish with rate setting—although any payment system must be based on some unit of payment that averages, to some degree, across services and institutions—again a problem not faced by public utility regulation because there is no need to derive an average among “types” of, say, electricity, because no different types exist. The question in rate regulation of health care, then, is how broad or narrow to make the regulated unit of payment. For example, hospitals could be paid based on global budgets or they could be paid on some activity-based unit, like a per diem or DRGs. Physicians could be salaried, capitated or paid, as they now most often are, based on fee for service.
Rate setting in health care can use a uniform rate for all rate-regulated providers. For example, administered pricing schemes, like Medicare’s IPPS and its RBRVS-based physician payment systems, impose a single rate on all providers, subject to adjustments for variations across groups of providers for such factors as geographic location. However, rate setting can allow different providers to charge different prices. For example, until recently each year Maryland’s rate-setting system established, for each hospital, an overall case-mix adjusted, per-case revenue constraint—a mandated target—and a set of approved service-specific unit rates (like an operating room charge per minute or various ancillary charges). Then, if the hospital successfully reduced its costs so that its average charge per case fell below the predetermined case-mix adjusted charge-per-case target, the hospital could increase its unit charges so that its average revenue then equaled the charge-per-case target. This feature more accurately reflected actual resource use for individual patients than a DRG-based system like Medicare’s IPPS, and it allowed unit charges to vary among hospitals based on their competition around the benchmarks they were given. In a somewhat similar, analogous fashion, Medicare pays Part C Medicare Advantage plans based on risk-adjusted benchmark prices for geographic regions, but plans bid around those benchmarks and can structure those bids around varying coverage packages and other plan features. By means of such methods, the rate-regulated entities are given the flexibility to compete around the chosen “yardstick”—hence the name “yardstick competition.”

The mechanism used for rate setting can be fairly simple. The initial rates are typically based solely or to some degree on historical prices. These prices can then be adjusted in various ways to achieve inter-institutional comparability and chosen policy goals. A hospital’s prior budget, for example, can be chosen as that hospital’s payment for the next year, but this payment then can be adjusted up or down depending on how that budget compares with an average for a chosen peer group—urban, rural, teaching, non-teaching, large, small and the like—or some selected measure of efficiency. Medicare’s Shared Savings Program has paid participating ACOs in part around a nationally derived benchmark, while CMS (2016a) just announced it will shift to regional benchmarks instead. The initial basis of Medicare’s RBRVS-based physician payment system, by contrast, was somewhat unique in that the units of payments—the relative value units—were measures of the value of different services that were derived not from historical charges but instead from rigorous comparison of the work involved, among other factors, in performing different services. Any system can then use bonuses or penalties added separately, such as the amounts tied to quality in Medicare’s payment systems, to achieve a variety of policy goals.

A rate-setting system typically separates factors over which providers have control from those outside of their control. For example, hospitals are thought not to be in control of the local wage markets in which they hire, e.g., nurses. Therefore, the IPPS adjusts the uniform DRG-based rates, applicable to all hospitals, for differences in local labor prices. Likewise, physicians are not in control of various practice expenses, which vary by location. Thus the uniform RBRVS-based prices Medicare pays physicians for various services are adjusted to take those local variations into account. After these types of factors have been removed from uniform prices and adjusted for local conditions the adjusted prices reflect just the costs over which providers do have control. Therefore, even though the prices are set, providers still have incentives to reduce the costs they can control and through this resultant process of yardstick competition the systems achieve some measure of efficiency.

Set rates then have to be updated periodically to take into account such factors as changes in input prices. Typically, as in Maryland or Medicare’s IPPS, some type of formula is used to make this adjustment. For example, both Maryland’s system and Medicare’s IPPS use changes in a “market-basket index,” which simply measures changes in the prices of hospitals’ inputs. As another example, the means by which physician services are performed can change due to new technology. To account for these changes, CMS periodically revalues the relevant relative-value units.

4. What are the desirable, if not essential, features of design?

   a. Prospective. First and foremost, to control expenditures a rate-setting system must be fully prospective. Providers must be given a budget in advance and they must be expected to stay within that budget. Moreover, when the rates are updated the governmental unit in charge of the system—to be discussed more fully in a moment—must too abide by those budgets. If it doesn’t and simply allows providers’ “excess” costs to be built into future rates, then the system is not prospective and expenditures remain uncontrolled. Necessarily, there are unforeseen changes that occur after rates are set—e.g., a new technology that improves quality but is more expensive or a sudden demand for more intensive and costly care in the face of a public health emergency—and therefore rate-setters must decide what changes warrant altered rates. However, these are relatively minor adjustments made to projected expenditures that must largely be made to stick.

   b. All-payer and, ideally, no price discrimination. To control expenditures the rate-setting system has to set rates for all payers. The example of Medicare’s IPPS, discussed above, provides the best illustration. Part of the reason for the increase in its expenditures—and moreover, all payers’ expenditures—is the fact that its rates apply to only its expenditures, not to those of private payers. The latter continue to rise out of control; the rate-setting systems of other nations work because they set rates for everyone.
Not only is this necessary for there to be the assertion of collective economic power but also to prevent providers or insurers with market power from cutting special deals only for themselves. A substantial factor in the destabilization of rate setting in New Jersey, New York and Massachusetts was the ability of some payers to obtain discounts, while, by contrast, tight regulation of discounts made the Maryland system much more stable (McDonough 1997a, 1997b). Additionally, by disallowing price discrimination all actors can shift focus from using or protecting against price discrimination to paying more attention to (yardstick) competition around quality and efficiency. In the ideal, competition occurs only around those two dimensions.

In allowing discounts, West Virginia’s system is therefore something of a compromise in that it establishes what are effectively rate corridors, with, to simplify, the maximums set as the most that hospitals can charge private payers and the minimums set as floors based on hospitals’ average costs. This type of system places a limit on what a dominant hospital can charge, thereby protecting payers without market power, and it prevents a dominant insurer from driving a hospital’s prices below cost, thereby protecting hospitals without power. While this sort of compromise might enhance a system’s political acceptability, an issue we address more generally below, it also creates greater administrative costs and quite arguably locks in continued market power at the price of diminished innovation and efficiency—on both the provider and insurer sides—problems often created when state law protects the power of dominant firms.

To create all-payer systems a state needs to obtain a waiver from CMS to include Medicare payments within the state’s system. Indeed, an important, albeit not sole, reason for the demise of some state rate-setting systems in the 1980s and 1990s was that the states could no longer satisfy the waivers’ requirements that their systems be budget-neutral (McDonough 1997b). Part of the stability of Maryland’s system rested on the fact that the initial grant of its waiver was written into federal law. Moreover, as Murray and Berenson (2015) point out in their comprehensive report on rate setting, in the waiver process CMS can impose fiscal discipline on a state’s system as political cover to stave off provider demands and other local pressure, and to prevent provider capture, a problem discussed below.

**c. Global units of payment to include a broad range of services in multiple settings.** To the extent technically feasible, units of payment for the set rates should be global in order to prevent providers from shifting the site of care to, or raising rates for, activities not included in the system. Much of the bite of Medicare’s Part A IPPS has been diminished because hospitals have moved care to outpatient settings, thereby drawing on Medicare’s more lucrative payments under Part B. More global units of payment, such as global budgets, prevent such games by bringing all activities into the rate-setting system.

Because they are divorced from activity to varying degrees, more global units of payment also prevent providers from increasing their revenue by churning volume or substituting more intensive for less intensive care. While Maryland’s and West Virginia’s systems have to some degree controlled expenditures per inpatient admission, both also have had extremely high rates of admissions and readmissions, as well as outpatient utilization, all volume effects (e.g., Murray and Berenson 2015). Control over volume, as well as affording hospitals flexibility within established rates, are reasons that under a revised federal waiver and an innovation project Maryland has just implemented global budgets, previously applied only to rural hospitals, but now extended to all of its hospitals (e.g., Murray and Berenson 2015). Vermont will do the same in its All-Payer ACO Payment Model, approved by its regulatory board and CMS on October 26, 2016 (CMS 2017; Green Mountain Care Board 2017; NASHP 2016; Zemel and Riley 2016). In its own way, Massachusetts may be “slouching” (Morone and Dunham 1985) toward an all-payer system, as its legislature has, through a series of enactments, incrementally expanded governmental control over expenditures, including the setting of annual state-wide expenditure targets, tied to the state’s economic growth, and the creation of an infrastructure for the enactment of stronger controls, with the establishment of an independent commission to monitor whether the expenditure targets are being met, to set expenditure targets for individual providers and insurers and to require outliers to create plans for corrective action, with potential reporting to the state attorney general for abuse of market power (Mechanic, Altman, and McDonough 2012). There are signs that Massachusetts’ framework is having a positive effect. In 2013 state-wide spending increased by 2.3 percent, 1.3 points below the target, while in 2014 the target was exceeded by 1.2 percent, but largely due to the full implementation of the ACA (Zemel and Riley 2016). Likewise, Maryland has reported some early success with the use of global budgets, with expenditures for the first year 2.11 percent lower than the growth rate agreed upon with the federal government (Patel et al. 2015), as has Vermont through its process of budget review (e.g., NASHP 2016; Zemel and Riley 2016).

Global units of payment, like global budgets, are a two-edged sword, precisely because they are divorced from activity. Therefore, to control volume and activity, many nations use hybrid systems, such as combining global budgets, to set overall expenditure constraints and to police against volume effects, with episode-based measures like DRGs to monitor activity and prevent providers from “shirking”—reducing the level of activity built into the budgets (e.g., Frankford 2017). Some of the state rate-setting systems obtained similar results by adjusting rates up or down by taking into account only some portion—say 50 percent—of changes in revenue resulting from changes in volume on the rationale that 50 percent of...
hospital costs are fixed, i.e., do not vary with changes in volume. If volume increased, allowing a hospital to retain only half of the additional revenue diminished its incentives to increase volume because effectively the hospital was only paid for the variable costs of the additional volume—allowing it only to break even—thereby eliminating its ability to churn volume to increase profit. If volume decreased, allowing only half of the savings in subsequent years’ rates would just allow the hospital to recover its fixed costs, again cutting the knot between volume and profit (Murray and Berenson 2015, 19–20).

Global units of payments can also aggregate providers across the continuum of care, thereby preventing the cost-shifting among them that mars our present payment system, as illustrated by the problems created by Medicare’s payment silos discussed above. Indeed, rate-setting systems are quite amenable to the use of methods like ACOs and bundled payments to integrate payment among providers. This integration is contemplated in the Maryland project (e.g., Patel et al. 2015) and Vermont’s all-payer system will use ACOs as a unit of payment (e.g., CMS 2017; Green Mountain Care Board 2017; NASHP 2016; Zemel and Riley 2016). Full integration has enabled systems like Kaiser, Intermountain and Geisinger to control their expenditures more than other parts of our health system. A rate-setting system that pays on such an integrated basis gives providers much greater incentives actually to integrate their operations in order to stay within the parameters that a rate-setting system maintains on the payment side.9

It is particularly important to resolve a basic conflict that exists, when, as in Maryland’s innovation project, hospitals are paid under global budgets while physicians continue to have incentives to enhance their payments—not controlled by the rate-setting system—through increased volume and the substitution of more lucrative services for less highly paid ones. This contradiction could be solved by bundling together into one unit hospital and physician payment, analogous to Japan’s combined payment in its all-payer rate-setting system (Ikegami and Anderson 2012). This course would face challenges, given the heterogeneity of physician practices and potential opposition (Murray 2012). A less radical solution would be for a state to use a hospital rate-setting system alongside something like Medicare’s physician fee schedule, with a mechanism to control for volume that avoids the fate of Medicare’s SGR formula. Unlike the SGR mechanism, the adjustment of physicians’ rates must be relatively rapid and automatic in response to increases in volume. In Germany, this goal is accomplished through use of a point system, in which each service in the relative value scale is assigned a number of points, which are then added up quarterly to get the total points billed by physicians. If the total exceeds the number of points projected in the budget for that year, then prices are adjusted to bring the total back within budget (e.g., White 2013). Such a system is rapid and automatic.

d. Simple, transparent and flexible. Maryland’s system, in contrast to almost all others, was initiated through legislation that established broad principles but not particular methods of rate setting with any degree of particularity. As a result, the system has been nimble, able to develop incrementally modified methods to adapt to changing needs and contexts, while remaining transparent and relatively simple. By contrast, almost all other states’ legislative grants contained complex detail that hemmed in the rate-setting process such that change could occur only by legislating radical, abrupt changes, thereby creating instability, or by adding layers of complexity onto layers of complexity such that, eventually, the systems collapsed under this weight. The systems became incomprehensible to almost everyone—as in Massachusetts, New Jersey and New York—and therefore were robbed of essential political support.

e. Buy-in of stakeholders while avoiding provider capture. The successful all-payer state systems were established through the efforts of broad coalitions of state actors, hospitals, insurers, business and labor.10 Today getting the buy-in of providers with market power will be perhaps the greatest challenge. When Massachusetts, Maryland, New Jersey and New York established their all-payer systems in the 1970s and 1980s, in some of these states hospitals went along for such reasons as the threat of even more stringent cost-control measures, the problems created by substantial levels of uncompensated care, and a general tradeoff of giving up revenues that might be higher without rate setting, for the stability that rate setting provides (McDonough 1997b). It is not clear that today providers with market power would strike such a balance.11

Insurers with market power likewise might balk because depending on its design, rate setting might prevent them from asserting their market power to drive discounts from providers, thereby depriving them of earning greater profits but also preventing them from using discounts as barriers to entry against insurers unable to obtain similar bargains (Pauly 1988). When the state all-payer systems were established, the Blue Cross organizations were far more dominant than they are now (McDonough 1997b), although they were getting hurt because they were forced to take all comers and were thereby saddled with poorer risk pools than the competing commercial insurers. Rate-setting systems like that of New York obtained their buy-in by imposing surcharges on plans not likewise burdened, thereby leveling the playing field. Nothing of the sort exists today. Leaving aside the remaining regional nonprofit Blues (like Horizon, Independence, Highmark), the dominant players—UnitedHealth, Aetna, Anthem, Cigna and Humana—are all for-profit firms, with no reason to give up the benefit of their market power. Some of the remaining nonprofit Blues, e.g., Highmark, possess similar, regional market power. As we have seen, even when dominant insurers face consolidated providers, they are able to pass those higher expenditures along to payers, too fragmented to resist. Today, perhaps a dominant insurer can be bought off by giving it a monopoly over claims.
administration in an all-payer system (see Fox and Blanchet 2015, 480), but it is not clear that an insurer with market power would trade fragmented payers as its dancing partners for bargaining with the state.

Given that the current context differs greatly from that during which state rate setting was established and flourished, one might speculate that the movement toward state all-payer rate setting is likely to be driven by state governments fiscally pressured with (ever-)expanding Medicaid budgets and, moreover, business interests and coalitions, which see no end to continued growth in expenditures, and might now overcome their skepticism toward—or at least ambivalence about—governmental involvement (Smyrl 2014), either because they no longer perceive any other option to control expenditures or because rate setting alone leaves them free to design their benefit packages with the inclusion of vehicles like network selection, cost-sharing and health savings accounts (Blanchet and Fox 2013). Therefore, perhaps the system in Maryland and the planned one in Vermont can be generalized (Murray 2012), but one must take seriously the claim that Maryland is anomalous because of its particular politics and market setting, dominated by Johns Hopkins and the University of Maryland, and the unusual skill and political aptitude of its regulators (Pauly and Town 2012). These factors may not exist in many states. Add to this structural and cultural lacuna an anti-government ideology heightened by the continuing debate over the ACA, the political row to hoe in many states is simply going to be tough, something not to be minimized.12

In states able to establish rate setting, its governance must be carefully structured to be responsive to stakeholders while avoiding provider capture (Murray and Berenson 2015). The rate regulatory board should be independent of other executive agencies and have its own funding source, separate from general revenues, such as user fees or an assessment imposed on providers, as in both Maryland and West Virginia, respectively. The board’s members should be appointed to relatively long terms, be experts in the field and not be employed by or have any affiliations with the entities they regulate; and a strong conflict of interest policy is necessary. Perhaps even more important, the board members must be supported by an extensive highly trained and technically sophisticated professionalized, paid staff. Formal channels of communication with stakeholders should be established—open meetings and some process for comments—but steps must be taken to insulate staff from day-to-day lobbying. With this sort of structure, the regulatory board is not dependent on stakeholders for resources, information or expertise, while its processes are open and provide some degree of voice to affected parties. These features can endow the board with legitimacy among all stakeholders, something that is crucial because the perception that the system played favorites among some had in part led to the demise of earlier state rate-setting efforts (McDonough 1997a).13

Conclusion

EXPERIENCE IN THE UNITED STATES, particularly Maryland, and abroad clearly demonstrates that properly designed, rate setting can reduce the growth of expenditures. These are facts. We also have solid facts that our experiment with market-based payment has failed to control expenditures and that the systems used abroad have done a much, much better job. There are numerous reasons that our experiment has not worked. A number of observers have argued that the experiences in Maryland and abroad cannot be transferred to other states or to the United States more generally. This is supposition. In our view, this juxtaposition of fact with supposition leads to the conclusion that state rate setting should be high on the policy and political agenda.

About the Authors

David M. Frankford is Professor of Law at Rutgers Law School; Professor at the Rutgers Institute for Health, Health Care Policy and Aging Research in New Brunswick; and Faculty Director at Camden of the Rutgers Center for State Health Policy.

Sara Rosenbaum is the Harold and Jane Hirsh Professor of Health Law and Policy and Founding Chair of the Department of Health Policy, George Washington University Milken Institute School of Public Health. She also holds a Professorship by Courtesy in the GW Law School and is a member of the faculty of the School of Medicine and Health Sciences.
Acknowledgements

We are grateful for the support provided by the Robert Wood Johnson Foundation and for comments from David Adler, Robert Berenson and Joel Cantor.

References


1. There are differences in data and methods used by The Commonwealth Fund, the OECD (2015) and the Office of the Actuary but the points in text hold nonetheless.

2. Much more detail about administered pricing is provided in Rosenbaum et al. (2012, ch. 12) and Frankford (2017).

3. We choose eyeglasses purposely as our example because it was one of the first products the purchase of which was separated from professional control in a proceeding by the Federal Trade Commission forcing eye-care practitioners to release eyeglass prescriptions to their patients (FTC 1997).

4. To be sure, for purpose of payment systems and quality metrics “conditions” are collected into groupings, such as the DRGs used for inpatient hospital stays. However, those classifications are statistical norms in which there often is variation from the norm and, moreover, in generating the statistical norm many distinguishing features of patients are stripped away—e.g., the fact that the 75-year-old stroke victim is in and out of institutionalized settings.

5. Scheffler and Glied (2016) have recently proposed that states could act as selective purchasers, as California does now, and allow plans to sell on the ACA Marketplaces only if their premiums do not exceed stipulated levels. We can therefore have the best of competition and regulation. Regardless of whether this proposal would work on Marketplaces other than California’s, particularly since in many states the Marketplaces are concentrated, for the following reasons it is difficult to see how it could apply to plans sold outside of the Marketplaces, particularly in the commercial insurance market: (1) there are no mechanisms comparable to the Marketplaces; (2) it would be administratively expensive given the huge number of plans in existence; and (3) there is a chance that ERISA would preclude the states from taking such action against ERISA plans, i.e., the entire employer-sponsored market. Moreover, it is clear that ERISA would preempt such regulation of self-insured plans, which now constitute almost 60 percent of employer-sponsored plans—a number that is growing (Fronstin 2015a). Large self-insured plans would therefore be allowed to exercise power against providers, to the extent they could, and other payers would then likely pay higher prices, unless they all were to flock to the regulated dominant insurers. Payment would remain fragmented, with the effects we describe. Perhaps Scheffler and Glied are correct that states need to regulate insurers’ premiums, particularly given the concentration among them. However, their proposal is that states effectively put a cap on premiums while trying to ensure competition among plans. Given the fragmentation that will remain on the payer side, payers will not be able to impose price discipline on providers, particularly those with market power. The result is like sealing the top of a boiling pot and adding more fuel to the fire. Premiums are the sealed top and provider prices are the boiling water inside. The whole thing has to explode. Regulation of providers is a prior condition to successful regulation of premiums, likewise particularly needed because of the extant and growing concentration among providers.

6. Since our focus is on slowing the growth of expenditures we do not discuss the other potential benefits of rate setting, unless they have a bearing on controlling expenditures: the unfairness of price discrimination, the financial stability of providers, spreading the costs of uncompensated care, improving access and improving quality. Additionally, we don’t address the fact that controlling expenditures alone leaves intact many pathologies of our current insurance and delivery systems, and the difficult choices we face (cf. Pollack 2015).

7. We simplify here because rate regulators rationally can, and often do, charge different users different prices. Take, for example, pricing for crossing a bridge used by commuters at rush hour and by “casual” users at other times of the day. The costs of the bridge are predominantly as fixed as fixed can be, the costs to construct the bridge. Nonetheless, there are variable costs imposed by the amount of use, the traffic, such as wear and tear on the road surface. Casual users might just stay home if they are charged very much, whereas commuters must go to work. Rate regulators might decide, therefore, to charge more to the commuters, the “ captive” users, and thereby recover the fixed costs from them, while charging the casual users less, down to the point of recovering from them only the variable costs they impose in crossing the bridge. Necessarily, there are social choices to be made here (e.g., Kahn 1970).

8. Clearly there exist many fallback positions from much of what we state here (e.g., Emanuel et al. 2016; Sommers, White, and Ginsburg 2012). Space only allows us to set forth the desirable or essential most important features of design. Additionally, to be politically and technically feasible, undoubtedly any system, particularly one approaching our ideal, would have to be phased in incrementally, varying from state to state, something we do not have space to discuss.

9. We do not address whether the integration of operations within ACOs is occurring.

10. In the following discussion we just hit some major points because obviously we can neither fully analyze the politics of enacting rate setting nor engage in the necessary, accompanying empirical work.

11. Uncompensated care presents a challenge for any rate-setting system because it raises both difficult distribution issues among hospitals and payers, and the possibility that hospitals game a system in their categorization of bad debt and charity care and in their diligence in collecting debt. If these issues become disputed, hospitals and their associations become divided, as do payers, and the political legitimacy of the system erodes or disappears, likewise dissipated the collective support that is necessary or at least helpful for the system to survive. Indeed, this problem was a significant factor in the New Jersey and New York systems’ demise. By contrast, Maryland has handled uncompensated care in a way such that hospitals and payers have largely perceived a lack of favoritism, bolstering their support (McDonough 1997a; Murray and Berenson 2015). Of course, the difficulties increase as the level of uncompensated care increases, and so among other things, states’ accepting the Medicaid expansion would help greatly.

12. Because of their strong preference for market-based payment, Pauly and Town (2012) also argue that rate regulation is unnecessary in states with workable competition. Obviously we disagree.

13. State rate setting is unlikely to overcome the huge problems created by the Supreme Court’s decision in Gobeille v Liberty Mutual Insurance Company, decided in 2016, in which the Court concluded that ERISA’s federal reporting requirements—which address compliance with the law’s
fiduciary standards rather than the price and quality of health care—nonetheless preempted Vermont’s all-payer claims reporting system, an essential first step to getting a handle on variable pricing. Insurers’ claims contain information about prices paid to all providers of all types. By contrast, in rate setting a state gets price information only from the entities so regulated, e.g., hospitals. Therefore, to duplicate the price information contained in an all-payer claims database—to get the information that is in all of the insurance claims—a state would have to regulate every provider of every type. Rate regulation of aggregations of providers, like ACOs, can have this effect but only if all providers are within the ACOs. That’s effectively what insurers do in creating their networks—aggregate every provider of every type—which is why getting all of their claims is essential.

The power of states to regulate providers’ prices to all payers, including ERISA plans, is very likely not threatened by Gobeille or the Supreme Court’s action in a subsequent case, Self-Insurance Institute of America v. Snyder, in which the Court ordered the court of appeals to reconsider its decision in light of Gobeille, decided after the circuit court had already ruled. Both Gobeille and potentially Snyder involve state laws that, at least according to the Supreme Court, impose record-keeping and reporting requirements that are governed by ERISA and therefore state law is preempted. By contrast, state rate setting imposes requirements only on providers, the regulated entities, and not on plans. To be sure, rate regulation affects the prices ERISA plans pay providers, but to hold that such an indirect economic effect causes preemption would require the Supreme Court to overrule its decision in New York State Conference of Blue Cross & Blue Shield Plans v. Travelers Insurance Company. In that case, the Court held that there was no preemption from such an indirect effect on ERISA plans because the state regulation there, which actually involved New York’s regulation of hospital prices under an all-payer system, falls within the core of states’ traditional and strongly protected power to regulate health care. An analogy would be that state licensure of doctors is preempted because it precludes ERISA plans’ purchase of the services of unlicensed practitioners, an absurd result. Nothing indicates that the Court’s decision in Gobeille and its order in Snyder signal an intention to go so far, which is really far, in displacing state power. (In fact, as this writing was going to press, the Sixth Circuit ruled that the Supreme Court’s decision in Gobeille is limited to states’ direct regulation of ERISA plans’ reporting and recordkeeping obligations and does not extend to “incidental reporting” required to comply with statutes like states’ tax laws.)