

Doing Better by Doing Less: Approaches to Tackle Overuse of Services

Timely Analysis of Immediate Health Policy Issues

January 2013

Robert A. Berenson and Elizabeth Docteur

This analysis presents what we know about the provision of medically inappropriate and unnecessary services that drive up health care spending without making a positive impact on patients' health outcomes. It also describes approaches that have already been used to address this issue—with limited success. We suggest that broader payment reforms are needed to minimize incentives to overdiagnose and overtreat and to better support the other approaches.

Introduction

The United States is an outlier in terms of its per capita spending on health care, surpassing other developed countries by a considerable margin. High prices and fees in the U.S. health sector have been identified as important drivers of these spending variations. The complexity and fragmentation of our system also plays a role in generating higher-than-average administrative costs and in creating an environment in which waste, fraud, and abuse can thrive.¹

The volume and mix of health care services provided in the United States also play a role in explaining higher health spending, although the story is complicated. By many measures of service volume, the United States is not an outlier. In fact, we have fewer physician visits,

fewer hospital admissions, and shorter lengths of stay in comparison with many developed countries.² According to many indicators of preventive service use, U.S. performance is only average or below average.³ But there are certain areas, including some relatively high-tech and high-cost services such as imaging and cardiac surgery, where the United States appears among the most prolific users.⁴

Extrapolating from studies focusing on particular conditions or services, some analysts have estimated that as much as a third of U.S. health care spending is unnecessary and wasteful.⁵ This estimate includes the provision of medical services and the prescription of medicines that are medically inappropriate—in other words, health care from which the patient derives no medical benefit or for which the potential harms exceed the potential benefits. More than a quarter of all wasteful spending in health care—an estimated \$210 billion out of \$765 billion in wasteful spending in 2009—is attributed to overuse of services that are medically inappropriate or otherwise unnecessary, which includes services that are provided more frequently than warranted and higher-cost services that are no better than lower-cost alternatives.⁶

There is variation in the level of inappropriate use by type of service, and it is evident that some services are subject to a great deal of overprovision. A recent review of the research literature found rates at which particular therapeutic procedures, tests, or medications were performed or prescribed when clinically inappropriate ranged from a low of 1 percent to a high of 89 percent.⁷ For example, a 2007 study found that 60.8 percent of colon cancer screenings undertaken were medically inappropriate, while a 2005 study found that 27 percent of physical therapy prescriptions for low back pain did not meet threshold standards for appropriate service use.

While any overuse of services drives up spending, inappropriate service use is particularly important because it has the potential to harm patients. One in every four patients admitted to the hospital is prescribed an inappropriate medicine, sometimes leading to adverse drug reactions that are responsible for 20 percent of inpatient deaths.⁸ Other examples of harmful overuse include radiation exposure from imaging scans such as CT scans, elective C-sections performed for convenience, and prescribing aggressive treatment options to those with terminal illnesses without disclosing the likely futility of such interventions.



Robert Wood Johnson Foundation



Reducing the overuse of just seven services known to be subject to high rates of inappropriate use could save between \$33 billion and \$62 billion annually.⁹

The potential of achieving cost reductions while substantially improving health care quality and outcomes can only tantalize policy-makers in this era of rapidly rising health spending and very strong constraints on the financing side. Meanwhile, current support for experimentation with new forms of health service delivery and concurrent changes in payment and regulatory oversight make the present a most opportune moment in which to consider the prospects for ensuring that those changes help, rather than hinder, efforts to reduce inappropriate service use.

This analysis focuses on what we know about the provision of medically inappropriate and other unnecessary services, which drive up health spending without making a positive impact on the health outcomes of patients. We begin by reviewing the literature on the extent of medically inappropriate overuse, continue by describing approaches used to tackle inappropriate service use, and then discuss the implications of current and proposed provider payment methods on the provision of inappropriate services and in spurring excess volume of services. In the review of payment approaches, we discuss opportunities to modify the various payment methods to address the provision of medically inappropriate services and overuse of unnecessary services, as more broadly construed.

What do we know about inappropriate service use?

Although inappropriate service use has been studied by researchers since at least the late 1970s, measuring it is technically challenging and costly. As a result, there is still a lot we do not know about its incidence and the reasons for its persistence. Despite the limited evidence base, the evidence that we do have suggests that there is a great deal of costly overuse of services that has defied efforts to address the problem. The problem is difficult to study for reasons discussed below, and those same reasons also make implementing policies to tackle the problem difficult.

What has the research shown?

After reviewing the research literature on inappropriate services, Korenstein and colleagues uncovered 172 articles measuring overuse that were published between 1978 and 2009 and met minimum quality standards in terms of methodology and other criteria.¹⁰ Of these, 53 articles addressed therapeutic procedures, 38 concerned diagnostic tests, and 81 pertained to medications. Within each area, a relatively small number of procedures, tests, and medicines had been studied. For example, 59 studies (more than one-third of the total) addressed antibiotics for upper respiratory tract infections, 17 studies looked at coronary angiography, and 7 studies investigated use of upper endoscopy. In sum, only 18 unique therapeutic procedures, 24 diagnostic tests, and 13 medications were evaluated in terms of the incidence of inappropriate use.

Based upon their findings, Korenstein and colleagues concluded that inappropriate use is often a problem for the services included in the published articles, although there is wide variation in rates of overuse documented by the research. At the same time, for the vast majority of procedures, tests, and medications in use today, no studies have assessed the extent to which they are overused in practice. Because of limitations in the scope of the research, there are important gaps in our understanding that limit opportunities to reduce health care spending without adversely affecting public health.

The review demonstrated that new and costly procedures were particularly underrepresented in the research literature.¹¹ In part, this dearth is explained by shortfalls in comparative effectiveness research for many services, particularly new ones. Filling these gaps in the evidence base would need to be prioritized if this research is to serve as a resource for targeting potential opportunities for enhancing the cost-efficiency of spending on health care services.

Certain services, including antibiotic use and several coronary procedures, have been subject to repeated study over time. For some of these services, such as carotid endarterectomy, the rate of inappropriate use has declined considerably.¹² In the case of inappropriate antibiotic use, there is evidence of reductions in overuse, but a high level of overuse persists, despite significant outreach and education efforts. This finding indicates that publication of research documenting overuse is not necessarily sufficient to ensure changes in practice patterns.

Why do we know so little about this critical aspect of health care quality?

Korenstein and colleagues pointed to a number of reasons why inappropriate use of services is relatively less studied in comparison with underuse of medically appropriate services.¹³ Measuring overuse via the assessment of the rate of provision of medically inappropriate services is technically challenging. It requires defining circumstances in which a particular service is inappropriate for use with patients with certain characteristics or under certain circumstances, generally through a process that involves reviewing relevant scientific evidence and developing consensus of clinical experts to produce guidelines and performance standards. Furthermore, because administrative data sets lack the type of information and level of specificity required, assessing inappropriateness generally requires resource-intensive review of medical records. In addition to these technical considerations, identification of inappropriate use is controversial and sensitive, in part because there are often stakeholders with strong economic interest in defending a particular procedure or medication. This is the case with vertebroplasty, a surgical procedure on the spine that continues to be widely performed despite findings questioning its medical benefit.¹⁴

For many services, if not most, measuring the extent to which they are subject to inappropriate use has not been attempted because of limited research dollars to support the effort, even though the pay-off might be far greater than the cost. Prominent gaps in scientific knowledge make the development of evidence-based standards difficult for many types of services,

particularly ones that are relatively new or in use for new indications, and mean that practice standards may reflect differences in expert opinion where the science is unresolved. These standards require updating to be consistent with the current state of knowledge.¹⁵ For example, recent studies determined that the accepted practice of raising end-stage renal disease patients' blood counts to levels that are normal in the general population lead to serious, unforeseen events such as heart attacks.¹⁶

Further, for every indication in which use of a service can either be labeled clearly appropriate or inappropriate, there are many more for which the benefits are unknown, unclear, or uncertain, leaving physicians and patients to navigate a vast gray area by relying on some combination of judgment, instinct, experience, and tradition. This gray area stymies both health services researchers in their efforts to assess the prevalence of inappropriateness and health policy-makers in their efforts to target the problem for reduction. We also lack research on whether overuse of services varies by race or ethnicity, and we know little about the impact the overuse of services has on health care disparities.

Although the issue of inappropriate overuse received significant research attention in the 1980s and 1990s, the challenges inherent in this type of research have led to a shift toward more population-based studies, which demonstrate substantial geographic variations in service use without differences in outcomes.¹⁷ While there has been some recent renewal of interest in appropriateness research, its challenges, particularly the need to review detailed clinical information that is unavailable in administrative

claims data, continue to limit its broad application.

Do geographic areas where service use is relatively high have more inappropriate services?

In a recently published review of the literature, Keyhani and colleagues identified only five studies addressing the question of whether geographic areas with high rates of service use had higher rates of inappropriate care. They concluded that the limited available evidence fails to support the hypothesis that inappropriate use of procedures is a major factor explaining geographic variations in intensity or cost of care.¹⁸ Most of the studies reviewed found that geographic areas with high rates of service use and areas with relatively low rates of service use had similar rates of inappropriate use of services.

Work by researchers affiliated with the RAND Corporation has been very prominent in establishing the information base in this area.¹⁹ Chassin and colleagues found that geographic areas where Medicare beneficiaries were 2.3 times more likely to receive a coronary angiography had only modestly higher rates of inappropriateness than was found in areas where use of the procedure was dramatically lower (82 percent versus 71 percent). RAND work has also found that areas of the United States did not have higher rates of inappropriate service use than areas of Canada, despite having higher rates of provision of certain services.²⁰

While further research to explore the relationship between high volume and the rate of inappropriate service use is needed, the available evidence has certain implications for policy-makers. If it is true that

the rate of inappropriate service provision is similar across areas with different overall levels of service use, it is likely that the volume differential can be explained largely as a difference in use of so-called discretionary services—services with relatively low benefit-to-risk ratios (e.g., elective knee and hip replacements for arthritis) or with indications for use that are uncertain, including services for which provider and patient preferences play a decisive role. Thus, policy-makers could reduce incentives for provision of both inappropriate and discretionary services, as well as duplicative services reflecting poor care coordination or case management, to increase efficient use of resources and reduce cost pressure.

Why does overuse of services persist?

Possible reasons why overuse of inappropriate or unnecessary services persists in spite of some initiatives by policy-makers and health care administrators to target the problem include lags in the diffusion of new scientific evidence; incentives that promote overuse of services, including financial incentives (e.g., payment, ownership of equipment or practices), legal incentives (e.g., malpractice), and administrative incentives (e.g., quality standards); limitations in available measures and public reporting of information on appropriate use; patient demand, including demand that is fuelled by direct-to-consumer advertising of pharmaceuticals; professional and cultural biases that favor employing a treatment even if there is doubt as to its potential benefits; and the default standard in health care that an intervention is presumed effective until proven not to be in particular circumstances. Past

efforts to impose standards of practice through managed care plans' utilization review processes or even through public program coverage policies have met strong resistance from patients, medical professionals, and the drug and medical device industry stakeholders.

Approaches used to tackle inappropriateness

Efforts of policy-makers to tackle inappropriate service use have, to date, come up short. Below we discuss general approaches that have been tried, including monitoring, reporting, and efforts to affect demand and supply, and we consider why these efforts failed.

Efforts to monitor the rates of provision of inappropriate services

Although measurement of health care quality at the provider and health plan levels has become an increasingly common part of health care in the United States today, routine measurement of the rates of provision of inappropriate services is not a focus of these efforts. By one estimation, only four of the 39 quality measures in the 2011 Healthcare Effectiveness Data and Information Set (HEDIS) of quality measures that are commonly used by payers and other parties to assess care quality explicitly address overuse.²¹

One reason is that the determination of appropriateness requires review of clinical information, which generally cannot be ascertained from claims data. The result is that the few appropriateness measures that have been adopted do not require clinical information but rather can rely on demographic information that is available on claims. For example, recent guidelines call for cessation of

screening colonoscopies in people over age 74. Because there are specific procedure codes that distinguish between screening and diagnostic colonoscopies, it is possible from claims data to calculate a rate of inappropriate screening colonoscopies for those over 74—anything over zero is inappropriate. Of course, this all-or-nothing measure assumes that colonoscopies are coded correctly.

Most of the time, the clinical information needed to establish an appropriateness measure is not available from claims data. This shortfall is important in that quality measurement is central to most efforts to set baseline standards and to motivate and ensure improvement in health care quality, through approaches such as informing patient choice and establishing administrative or regulatory rewards and sanctions.

Efforts to educate physicians to comply with evidence of appropriateness

Medical specialty societies, in an effort to fulfill their role of educating their members to stay current with the body of knowledge the public expects professionals to have at their command, have long devoted resources to develop evidence-based care guidelines, including identifying specific clinical indications for performing diagnostic and treatment interventions. Some, including the American College of Physicians, the American College of Cardiology, and the American College of Radiology, go further to identify clear contra-indications to performance of interventions—that is, to specify inappropriate use.

Typically, the process of setting appropriateness criteria begins with a thorough literature review. But

since the available literature is usually not definitive in addressing all of the clinical variations that guidelines have to address, the review is followed by an expert consensus process to set the actual guidelines. Other entities, such as health insurance plans and authorities like the U.S. Preventive Services Task Force, also develop evidence-based guidelines, which may provide advice about inappropriate use.

Different users of evidence may come to somewhat different conclusions about the implications of the research findings on appropriateness of specific services, especially when the literature is not definitive. Payers, for example, are more likely to interpret the findings more strictly than providers, requiring a higher threshold of evidence to support use of a diagnostic or treatment intervention.

There are many facets involved in a comprehensive appropriateness determination. For example, it may not be possible to determine how frequently a screening test should be performed, even if the screening test has been proven useful and appropriate. The result of these and other considerations is that there may be variation in criteria and in guidance about what constitutes inappropriate, as opposed to uncertain, use.

Although physicians have a duty to adhere to the professional standards that evidence-based practice guidelines help establish, it is clear from the literature that professionalism and the efforts of professional societies have not succeeded in substantially reducing the provision of inappropriate services. To increase attention to the issue, nine specialty societies recently joined the American Board

of Internal Medicine Foundation and Consumer Reports in the first phase of the Choosing Wisely campaign.²² Through this initiative, each society has developed a list of five tests, treatments, and services that are commonly used in that specialty and for which there was a judgment that the use was often not necessary for patients. Examples include unnecessary CT scans, antibiotics for an acute sinus infection, and stress imaging tests during annual checkups. All of these services were deemed to be provided at times when they offer no benefit to the patient or may actually cause harm. It remains to be seen if mounting a campaign to promote discussions between physicians and patients about the merits of particular interventions, rather than the more traditional approach of promulgating guidelines for physician use, proves more successful in reducing inappropriate service use.

Efforts to affect demand for inappropriate services

Prior to the recent launch of the Choosing Wisely campaign, efforts to activate patients to protect themselves against overuse have been negligible, with the exception of instituting cost-sharing arrangements that give patients incentives to limit their use of services. After studies of the impact of increased across-the-board cost sharing showed that patients tended to reduce use of both appropriate and inappropriate care,²³ interest has grown in moving to value-based benefit designs in insurance policies. Value-based benefit designs increase cost-sharing for services of uncertain benefits and those prone to overuse (e.g., imaging), while reducing cost-sharing for services that are viewed as high value under most patient

circumstances (e.g., immunizations).

Implementing value-based benefit designs that assign higher cost-sharing rates to services of lower value has proved challenging. The value of an MRI, for example, depends on the specific clinical circumstances in which the scan is used. So it is one thing to believe that in aggregate there are many inappropriate MRIs being performed, but hard to define a policy to reduce inappropriate use as there are also many MRIs that are appropriate and even essential. There is no easy way to operationalize a priori a higher co-payment (e.g., for the weekend athlete with knee discomfort), because of the discretion needed for individual cases.

The relative sparseness of patient education and information campaigns focusing on the problem of overuse may reflect both market failures and conflicts of interest, although recent investments in areas such as comparative effectiveness research and informed shared decision-making may offer promise for the medium- and long-term future. However, it may require considerable effort to convince patients that more is not always better when it comes to service use, as recent controversies relating to new and more restrictive guidelines for appropriate use of breast and prostate cancer screenings have demonstrated.

Efforts to use administrative levers to reduce inappropriate services

The same methodological problems that make it difficult to study appropriateness also make it difficult to target policy remedies to this problem, particularly those that depend on defining standards and overseeing compliance with

evidence-based, clinical practice guidelines or criteria derived from such guidelines.

Prior authorization, a process through which providers seek advance approval from a payer that the service will be paid for when provided to the patient, is the administrative approach most widely used by payers to try to reduce the provision of inappropriate services. The approach tries to address individual patients' specific, clinical circumstances, usually not through full review of medical records but with review of the clinical information most relevant to the appropriateness of the clinical intervention requested.²⁴ Based upon some early reports of success when used to assess the appropriateness of a patient's proposed hospitalization,²⁵ prior authorization was expanded to outpatient surgical procedures, and then again to routine ambulatory care referrals from one physician to another for fairly routine evaluation and management services.²⁶ However, its pervasive, often intrusive application made it ripe for criticism by physicians and patients, contributing to the managed care backlash. Health plans have retrenched and now apply prior authorization more selectively. Common applications include elective surgery, referral for advanced imaging (such as MRI and CT scans), and as part of the management of pharmaceutical formularies.

When applied selectively to high-cost, discretionary services where objective information can be reviewed by qualified third parties, prior authorization can play a role in reducing inappropriate services, although with less success for services for which actual benefits

are uncertain or are believed to be minimal. Following reports of private health plans' use of radiology benefit managers to review and approve requests for advanced imaging, the Medicare Payment Advisory Commission recently recommended the targeted application of prior authorization in Medicare for office-based referrals for advanced imaging.²⁷

Given the administrative complexity involved, the changing clinical evidence of what works and when, and the large amount of uncertainty in the gray area, prior authorization's role in reducing inappropriate overuse is somewhat limited. Yet this form of utilization management can certainly complement other approaches addressing overuse.

In summary, the various approaches described above either have not been used or have been used without much impact on reducing overutilization. Few quality measures exist to measure overuse, while patient education and information campaigns focusing on the problem of overuse have been half-hearted at best. Meanwhile, professional societies' efforts to educate their members to reduce overuse are still fairly nascent and have yet to demonstrate a significant impact—to some extent because of the power of fee-for-service incentives as described below. Prior authorization has been successful to limit overuse for particular clinical services but becomes intrusive and counter-productive when used outside of its sweet spot.

This reality suggests the need to rely more on payment incentives—combined with investment in production and dissemination of evidence on what works under what

circumstances—to influence clinical decisions and support efficient delivery that avoids overuse and reduces the use of services offering low benefits, relative to risks, rather than on approaches that rely on measurement of inappropriate services or enforcing adherence to standards.

How payment methods affect overuse

Although research on the effects of payment methods on volume of services has demonstrated that fee-for-service tends to incentivize more use of services than does capitation,²⁸ the recent research literature is surprisingly sparse on the question of how payment methods affect the provision of inappropriate services.

One recent study found evidence to suggest that physicians who were paid on the basis of capitation were less likely to indicate that they would provide discretionary services, as compared with physicians who were paid fee-for-service.²⁹ This study's findings are consistent with the broad perceptions that fee-for-service rewards not only more service use but also excessive volume of services, including both unnecessary and inappropriate services, and are supported by observed anecdotal examples of real medical practice phenomena. For example, the abusive overdosing of the anti-anemia drug erythropoietin to raise blood counts took place under generous fee-for-service payments. The practice ceased immediately when the drugs were bundled into the dialysis payment, because payments were fixed regardless of whether and how much of the erythropoietin drugs were used.³⁰

Studies that have attempted to establish a relationship between payment method and service use, whether inappropriate or not, are challenging because many accompanying characteristics of a payment method impact clinical decision-making.³¹ Of course the generosity of the payment could influence provider behavior. For example, under capitation, a low payment level considered inadequate by physicians might lead to stinting on services and altered referral patterns, whereas higher payments might not. But the context in which the payment method applies should also strongly affect clinician decision-making; the particular culture of the provider organization in which the clinician practices can cause any particular payment method to behave differently.

As the following discussion makes clear, provider responses to payment incentives can be quite nuanced. Understanding those nuances can help design payment approaches that would address the problem of unneeded services and possibly the provision of inappropriate services—especially the provision of care in the gray area that lacks evidence-based guidelines and prior authorization rules. There, payment incentives can play an important role in influencing decisions, for better or worse. We next review some basic issues in how commonly used and proposed payment approaches can affect the volume of services in general and inappropriate overuse in particular.

Fee-for-service

There is a growing consensus that fee-for-service represents payment for volume, regardless of appropriateness, and needs to be

replaced. However, how physicians respond to fee-for-service incentives depends on whether or not a particular service is included for payment in the fee schedule and on how profitable particular services are. A growing body of evidence has documented excesses associated with current fee-for-service practice, including abusing physician self-referral of imaging and other noninvasive office testing, selecting the highest cost (and most profitable) intervention for treatment of prostate cancer, and performing inappropriate major spine surgery on patients with chronic low back pain.³² Again, while the literature does not actually document that inappropriate services occur more commonly in fee-for-service than in capitated or salaried practice, it is clear that at least some physician self-referral behavior would be financially self-defeating without fee-for-service.

The precise fee levels for specific services can also play a role in incentivizing overuse of particular services. Longstanding research has shown that physicians respond to reductions in fees by increasing the services affected by the payment reductions. In the context of the Medicare Fee Schedule, CMS actuaries have calculated a behavioral offset to partly account for expected volume increases that physicians generate in response to fee reductions or freezes.³³ However, other research, some quite recent, suggests that physicians in fact respond to fees more like other economic actors do. That is, if a service becomes less profitable, the incentive to produce it declines.³⁴

In short, the physician response to fee-for-service payment levels is actually not so simple, as well summarized in the title of a recent

issue brief on the topic, “Expect the Unexpected? Physician Responses to Payment Changes.” From their reading of the accumulating evidence, the authors argue that cutting fees leads to higher utilization when the targeted services account for a large share of physician income.³⁵ However, if the services involved do not account for substantial income, physicians reduce their output of the services whose fees were cut, plausibly reducing both appropriate and inappropriate services.¹

In the Medicare Fee Schedule, payments for services such as test and imaging interpretations and minor procedures often far exceed the resource costs, making them highly profitable and, therefore, provided to excess. Other services and patient care activities that are uncompensated or relatively poorly paid, such as engaging in care

¹ In the Deficit Reduction Act of 2005, Congress decided to pay physician practices for imaging services no more than what is paid to hospital outpatient departments. Fees were reduced significantly, as much as 40 percent for a common MRI of the brain. As the Government Accountability Office documented in the first year with the reduced fees, the approach generated first-year savings of 13 percent for the affected imaging services, largely from the direct reduction in prices paid. (See: *Trends in Fees, Utilization, and Expenditures for Imaging Services before and after Implementation of the Deficit Reduction Act of 2005*. Washington: Government Accountability Office, 2008, www.gao.gov/products/GAO-08-1102R.) The volume of the affected services continued to rise but at a lower rate than in the most recent prior years. There was no behavioral response to make up for pay reductions with volume. Now, years later, the volume of imaging services paid for under the Medicare fee schedule has flattened out, albeit still at a high level with many, presumably unneeded, inappropriate services. The point is that even fee-for-service can be modified to reduce provision of volume, some of which likely represents inappropriate overuse.

planning activities and phone consultations, are not prone to overuse. These distortions are not inevitable, although there are inherent practical and political challenges in getting administered prices correct.

Episodes and bundled episodes

The episode-based payment approach involves a single payment covering all services provided over a defined period of time. In this discussion, we distinguish episodes from bundled episodes, since bundling implies merging payment streams that had previously been paid separately to different providers. In the case of Medicare's diagnosis-related group (DRG) payments, which are a form of episode-based payment, the period of time for a hospitalization is from one to three days prior to admission through the hospital discharge. In Maryland, which runs an all-payer rate-setting program for hospitals, the period of time covered by a single payment to the hospital has been extended to 30 days post-discharge for many diagnoses.³⁶ Another common example of an episode payment is the global payment made to surgeons for major surgery. The single fee-for-service payment covers the cost of the actual operation and routine post-operative care for up to 90 days after surgery.

The theoretical merit of paying for an episode of care is that the payment amount is fixed, regardless of the quantity and mix of services actually provided inside the episode period. The costs of care are internalized to the providers, who essentially are at financial risk for spending inside the episode, though exceptions are made for outlier cases. This means that the basic incentive for providers is to provide

a large number of efficiently provided episodes.

With episode-based payment, the behavioral response differs for discretionary and nondiscretionary services. As a general rule, providers cannot induce demand for maternity care or treatment of a fractured femur—they are nondiscretionary. Yet, providers can and do induce demand for discretionary services, such as those found in the appropriateness literature review cited earlier, including joint replacements, heart and large vessel procedures, spine surgery, and deliveries by Cesarean section.³⁷

The concern is that with episode-based payment, discretionary services will increase in volume—sometimes representing inappropriate overuse—offsetting the savings achieved by internalizing the costs of producing the services to the provider. In short, some inappropriate services will decrease and others—the interventions that create the episode—might increase.

Bundled episodes. With bundled episode-based payments, where the various payment streams to different providers are merged into one, the bundling can be actual (i.e., a single payment is made to a recipient provider who then is responsible for distributing it to the constituent providers who provide services inside the bundled episode) or it can be virtual (i.e., the separate payments continue but with an accounting of how much the actual payment distributions varied from the target payment amount for the bundled episode, as the basis for determining surpluses or deficits, which become the responsibility of the bundled entity). Only if costs are less than the bundled episode

payment level does one have to worry about distribution of surpluses to the various parties providing care in the bundled episodes.

The reason for the virtual approach is that maintenance of traditional fee-for-service claims and payments is viewed by many as operationally and politically easier for providers to manage and accept. However, this approach may have negative consequences on provider behavior to generate inappropriate services.

The objective of bundling is to break down provider silos to promote more cooperation and coordination among the various providers that deliver care. A major challenge in this approach is that bundling across providers is operationally difficult.³⁸ Indeed, many consider it much more challenging administratively than capitation.

Our interest here is in how bundled episodes address the problem of inappropriate overuse of services, particularly discretionary procedures. It is notable that some analysts have not even raised the concern that bundled episodes have fee-for-service-like incentives for generating unneeded services.³⁹ There is an apparent assumption that incentives for overprovision of interventions when paid with bundled episodes (i.e., up-coding to claim payment for a more complex episode) are no worse than the incentives to prescribe unnecessary services within fee-for-service. For example, a recent projection of substantial Medicare savings under a regime of bundled episode payments did not consider the possibility that the volume of bundles could be different from that under the baseline, fee-for-service use patterns.⁴⁰

Although there is little evidence one way or the other, given the rare application of bundled episode payments to date, a plausible argument can be advanced that bringing providers together to collaborate on care for specific acute care episodes could actually increase volume to levels that could counteract any spending reductions from the new internalized incentives to reduce the cost of the bundled episode. Hospitals and specialists can brand and market service lines with the hope of attracting patients from other providers and creating new demand for the service line product. Such a strategy produces non-price competition that raises costs by increasing demand for technically oriented interventions, thereby contributing to a “medical arms race.”⁴¹ In the last decade, hospitals have done just this—developing and marketing profitable service lines by closely affiliating with specialist physicians essential to hospitals’ service line products.⁴²

Ultimately, whether bundled episode payments raise the volume of inappropriate services even more than under straight fee-for-service will be determined empirically. The Innovation Center at CMS and some private health insurers have begun experiments with bundled episodes, mostly involving hospitalization for an acute condition involving a procedure. Those tests will help tell us whether the incentives of bundled episodes reduce discretionary services within episodes and whether they increase the number of episodes.ⁱⁱ Whether

ⁱⁱ In 2009, CMS began implementing the acute care episode demonstration of bundled payments for physician and hospital services furnished to patients as part of hip and knee replacements and cardiac procedures. In its early results, Hillcrest Medical Center in Tulsa reported cost savings, especially from increased bargaining power for equipment and

those changes involve a decrease or increase in inappropriate services would need to be separately assessed through medical record review. Further, it is possible that some of the approaches cited earlier, such as requiring prior authorization, adherence to evidence-based appropriateness criteria, or shared decision-making with patient decision aids⁴³ could be used to constrain further overuse of the bundled episodes for which payment is claimed.

Pay-for-performance

As is true with other types of quality problems (e.g., underuse) the potential for using pay-for-performance methods to discourage the provision of inappropriate and unnecessary services is limited by accountability issues. As noted above, there are very few measures of appropriateness in current use. This is not for lack of interest. Rather, there are inherent limitations in being able to develop valid measures relying on administrative data sources in a fee-for-service payment environment. For the most part, individual patients are not assigned or formally associated with particular practices or delivery systems, so there is no relevant population denominator on which to establish norms for rates of services provided.

There may be some ability to relate rates of one service to another. For example, the Maryland Health Services Cost Review Commission

supplies as physicians and the hospital consolidated their negotiations with vendors. At the same time, Hillcrest reported about 30 percent increases in volume for the orthopedic and cardiac procedures. Whether that large volume increase represents redistribution of cases from other facilities or increased total interventions for beneficiaries in the area remains to be determined. (See endnote 27.)

is measuring the rate of stent placement to coronary catheterizations to try to identify inappropriate overuse of coronary artery stent placement. In this case, the denominator in calculating comparative rates is not the population served (information that is not available from claims data), but rather the population of patients who have undergone a cardiac catheterization (data that is derived from claims). Yet, this creative ratio-based attempt to identify overuse is the exception that proves the rule. Namely, in the absence of a population denominator, performance rates as the foundation for measuring appropriate provision of services generally are not available. Further, even if there were a population denominator, it would still be necessary to do case-mix adjustment for the population’s health status, another challenging requirement.

Also, without a population denominator, one could still try to measure the rate of performance of particular services in relation to clinical indications for that patient. But as noted earlier, claims data used to assess performance generally lack the clinical information needed to permit an accurate assessment of appropriateness.

The challenge of accounting for variations in provision of discretionary services was well captured in an article reviewing the difficulties of basing payment policy on measurement of spending and health outcomes. The article noted: “Hospital performance measurement will be most biased if the decision to admit varies systematically across hospitals and regions. Such variation may reflect greater illness levels in the population. However, higher

utilization also may be due to a lower threshold for admission, driven for example by higher per capita supply of physicians and hospital beds. In such hospitals, sufficient slack capacity allows physicians to work ‘down an **appropriateness curve**’ in their admission decisions” [emphasis added].⁴⁴ The stated implication is that unless one can account for different appropriateness thresholds for admission, measures of hospital spending and quality outcomes are likely to be inaccurate.⁴⁵

Shared savings and global payment

Global payment or global capitation is a population-based payment approach in which a fixed per-person, per-month prospective payment is made to an organization responsible for providing services to individuals who elect or are assigned to receive care from the providers in that organization. With global capitation payment (in contrast to primary care and professional capitation), most of the services that comprise the payer’s benefit package are included in the global payment. Sometimes prescription drugs, mental health benefits, and other specialized benefits may be carved out of the global payment.

Global capitation is the payment model that most fundamentally changes fee-for-service incentives that reward provision of inappropriate services. Simply put, unneeded procedures and hospital stays constitute what financial officers call “profit centers” under fee-for-service and its variations, whereas they are “cost centers” under global payment. In contrast to episode-based payment, under global payment there is no distinction between services that create an episode (and are therefore

financially desirable) and those that take place within an episode (and are therefore financially undesirable). In short, global capitation penalizes volume of services, whether appropriate or not. Indeed, a major policy concern is that organizations and their constituent members will underserve patients because the incentives are too strong in the opposite direction from fee-for-service. This concern has led to proposals to mix fee-for-service and capitation to try to balance incentives (i.e., partial capitation)⁴⁶ or to adopt other approaches that mitigate the extent of financial risk, softening the incentives to withhold needed services without eliminating the incentive to economize that is inherent in global payment.

Adherence to professional standards is supposed to serve as an important brake on some amount of inappropriate underuse just as it surely restrains some overuse under fee-for-service. Further, the clinical domains with the greatest number of valid clinical quality measures are in the area of primary and secondary prevention, permitting a complementary use of global payment and performance measures to detect some underuse. Unfortunately, there are not good measures to detect underuse at the other end—for patients with serious health problems. For example, some patients might be best served with referral to clinicians and/or facilities that are not part of the provider network receiving the capitated payment. However, we cannot easily measure the failure to make an appropriate referral.

Shared savings. Shared savings within an accountable care organization (ACO) is being promoted as a possible transition approach for organizations seeking

to move from fee-for-service but are not willing or able to make the move to global or even partial capitation. Under the Medicare Shared Savings Program that was created by the Affordable Care Act, initially fee-for-service to the constituent provider members of the ACO will continue according to existing Medicare payment rules, i.e., fees to physicians and diagnosis-related group payments for inpatient hospitalization. At the end of an accounting period, the actual spending attributed to the ACO is compared to a target spending amount, with any surpluses split between Medicare and the ACO. The ACO is responsible for determining how surpluses (savings to Medicare) will be distributed among the members of the ACO.

In this one-sided risk approach, and in contrast to global payment, there is no financial penalty if actual spending turns out to be more than the target amount. In the Shared Savings Program, there is also an option for ACOs to assume two-sided risk, an approach that would be required by year four of the program. With two-sided risk, both surpluses and deficits would be shared with Medicare. This two-sided approach starts on the road to global payment in that the marginal incentive for the organization is to reduce volume.

However, in considering shared savings to ACOs under the CMS program, one needs to recognize that the incentives for the ACO and those for the ACO’s constituent providers might be quite different, as in the case of managed care organizations and their provider network constituents. Cash flow to the constituent providers continues based on fee-for-service. So for a specialist achieving high incomes

by doing discretionary procedures or a hospital trying to keep beds occupied by generating additional admissions, the financial incentive to generate volume far exceeds the incentive to share in possible savings achieved by the collective ACO. It is not surprising then that in the Physician Group Practice Demonstration, which was the model for the ACA's Shared Savings Program, participating organizations focused on reducing spending associated with management of particular chronic diseases, especially congestive heart failure, but not on reducing the volume of diagnostic and therapeutic procedures that are the core income and revenue generators for important specialist physicians and hospitals, including joint replacements, spine surgery, cardiac procedures, and other procedures found in the literature to be provided inappropriately.⁴⁷

The incentives will actually vary based on the structure of the ACO. In true multispecialty group practices, fee-for-service incentives to do more should be moderated substantially, even in a shared savings payment arrangement based on fee-for-service. It depends on the method the ACO chooses to distribute its global payments. But where the providers in the ACO essentially receive pass-through fee-for-service payments, with only surpluses against target spending amounts subject to the group's distribution formula, it is hard to see how the incentives on the individual clinician vary much from those of fee-for-service. The ultimate effect would likely be similar to that with bundled episode payment, in which cash flow to individual providers continues and only the savings are considered to be the group's to divide up. In both cases, the practical approach of not interfering

with standard fee-for-service cash flow may undermine the objective of the new payment approach to economize on provision of services.

Conclusion

Failing to do whatever we can to reduce the use of medically inappropriate and unnecessary services is indefensible at a time of mounting pressures on health financing. But experience shows that aligning incentives to squeeze out waste and improve the efficiency of health services provision will not be easy. A sophisticated and multifaceted solution is required.

To get at inappropriate service use, we need to invest in researching what works under what circumstances and how alternative treatment approaches compare to one another. That information is sorely lacking, due to an underinvestment in research that is only beginning to be remedied. While research on effectiveness will always lag behind technology, the gap between what providers do and what researchers know can be bridged, and strategic prioritization can identify the most promising areas for investment in further research.

But research alone will not suffice. There is a need for actors at all levels to increase efforts to make use of the available evidence. The best approaches need to be determined, whether they include gentle tools, such as establishing campaigns to inform providers and patients, or stronger yet sometimes cruder approaches like incentives to use clinical guidelines through benefit design and coverage decisions, or finer scalpels, such as prior authorization rules for certain procedures with high rates of inappropriate use. New investments

in electronic health records and tools for shared decision-making can make information and evidence more accessible to providers and patients. Where possible, given the methodological constraints of measuring inappropriate service use, we need to amplify use of quality measurement and monitoring, and we need to increase the amount of data given to providers that shows their provision of overused services relative to their peers.

For many services, particularly commonly used and relatively low-cost services for which prior authorization is infeasible, as well as for all services of unknown or uncertain benefit, administrative and policy tools will not suffice and could be counter-productive. In these cases, the incentives established through payment are likely to be of critical importance. Further, it is likely that the various educational efforts and administrative approaches will be more successful if coupled with altered payment incentives.⁴⁸

Aligning payment incentives with desired outcomes can only help, and may in some cases be essential to reducing the enormous problem of resources squandered through wasteful practice decisions. On the cusp of important changes in payment methodology, policy-makers need to design payment methods in ways that reduce incentives for provision of excess service volume and maximize rewards for provision of beneficial, appropriate care.

Glossary

Bundled episodes—A payment approach in which a single payment is made to cover the cost of services delivered by multiple providers over a defined period of time to treat a given episode of care (e.g., a knee replacement surgery or a year’s worth of diabetes care).

Episode-based payment—A payment approach in which a single payment is made to cover the cost of services delivered by a single provider over a defined period of time to treat a given episode of care.

Fee-for-service—A payment approach in which health care providers receive a separate fee for each service they deliver.

Fee schedule—A comprehensive list of fees used by either a private or public health insurance plan or payer to reimburse health care providers on a fee-for-services basis.

Financial risk—When an entity assumes liability for the financial loss that could occur if actual costs exceed expected revenues.

Global capitation—A single payment made to a provider organization to cover the cost of a predefined set of services delivered to a patient (e.g., an amount paid per member per month to cover the cost of all of a patient’s health care needs). In many cases, the provider organization is responsible for reimbursing other providers for care they deliver to the patient.

Partial capitation—When a payer pays for some types of services on a capitated basis (e.g., by contracting with a group of providers to deliver all of their enrollees’ outpatient care) and pays for other services on a fee-for-service basis (e.g., reimbursing any hospital in their network for inpatient care delivered to their enrollees).

Shared savings—A payment approach whereby a provider or provider organization shares in the savings that accrue to a payer when actual spending for a defined population is less than a target amount. (Typically, performance targets on quality measures must be met to qualify for shared savings.) If actual costs are higher than projections, there are no financial repercussions for providers, unless shared losses are also part of the payment agreement.

Adapted from: Delaware Healthcare Association. “Delaware Healthcare Association Glossary of Health Care Terms and Acronyms.” Dover, Del.: Delaware Healthcare Association, www.deha.org/Glossary/GlossaryA.htm (accessed December 2012).

The views expressed are those of the authors and should not be attributed to the Robert Wood Johnson Foundation, or the Urban Institute, its trustees, or its funders.

About the Authors and Acknowledgments

Robert A. Berenson, MD, is an institute fellow at the Urban Institute, and Elizabeth Docteur is an independent health policy consultant. The authors thank Elizabeth McGlynn, director of the Kaiser Permanente Center for Effectiveness and Safety Research, for her helpful comments on this paper. This research was funded by the Robert Wood Johnson Foundation.

About the Urban Institute

The Urban Institute is a nonprofit, nonpartisan policy research and educational organization that examines the social, economic, and governance problems facing the nation.

About the Robert Wood Johnson Foundation

The Robert Wood Johnson Foundation focuses on the pressing health and health care issues facing our country. As the nation's largest philanthropy devoted exclusively to health and health care, the Foundation works with a diverse group of organizations and individuals to identify solutions and achieve comprehensive, measurable, and timely change. For 40 years the Foundation has brought experience, commitment, and a rigorous, balanced approach to the problems that affect the health and health care of those it serves. When it comes to helping Americans lead healthier lives and get the care they need, the Foundation expects to make a difference in your lifetime. For more information, visit www.rwjf.org. Follow the Foundation on Twitter [www.rwjf.org/twitter](https://twitter.com/rwjf) or Facebook [www.rwjf.org/facebook](https://www.facebook.com/rwjf).

Notes

- ¹ Angrisano C, Farrell D, Kocher B, et al. *Accounting for the Cost of Health Care in the United States*. McKinsey Global Institute, 2007, www.mckinsey.com/insights/mgi/research/americas/accounting_for_the_cost_of_health_care_in_the_united_states (accessed December 2012).
- ² *Health at a Glance 2011: OECD Indicators*. Paris: OECD, 2011, www.oecd.org/health/healthpoliciesanddata/49105858.pdf (accessed December 2012).
- ³ *ibid.*
- ⁴ Squires D. *The U.S. Health System in Perspective: A Comparison of Twelve Industrialized Nations*. New York: The Commonwealth Fund, 2011, www.commonwealthfund.org/~media/Files/Publications/Issue%20Brief/2011/Jul/1532_Squires_US_hlt_sys_comparison_12_nations_intl_brief_v2.pdf (accessed December 2012); *Health at a Glance 2011*.
- ⁵ *The Healthcare Imperative: Lowering Costs and Improving Outcomes: Workshop Series Summary*. Washington: Institute of Medicine, 2010.
- ⁶ *ibid.*
- ⁷ Korenstein D, Falk R, Howell E, et al. "Overuse of Health Care Services in the United States: An Understudied Problem." *Archives of Internal Medicine*, 172(2): 171-178, 2012.
- ⁸ Scott I and Jayathissa S. "Quality of Drug Prescribing in Older Patients: Is There a Problem and Can We Improve It?" *Internal Medicine Journal*, 40(1): 7-18, 2010.
- ⁹ Bentley TG, Effros RM, Palar K, et al. "Waste in the U.S. Health Care System: A Conceptual Framework." *Milbank Quarterly*, 86(4): 629-659, 2008.
- ¹⁰ Korenstein, et al.
- ¹¹ *ibid.*
- ¹² *ibid.*
- ¹³ *ibid.*
- ¹⁴ Swensen S, Kaplan G, Meyer G, et al. "Controlling Healthcare Costs by Removing Waste: What American Doctors Can Do Now." *BMJ Quality & Safety*, doi:10.1136/bmjqs.2010.049213, 2011.
- ¹⁵ Lawson EH, Gibbons MM, Ingraham AM, et al. "Appropriateness Criteria to Assess Variations in Surgical Procedure Use in the United States." *Archives of Surgery*, 146(12): 1433-40, 2011.
- ¹⁶ Pfeffer MA, Burdmann EA, Chen CY, et al. "A Trial of Darbepoetin Alfa in Type 2 Diabetes and Chronic Kidney Disease." *New England Journal of Medicine*, 361(21): 2019-2032, 2009; Singh AK, Szczech L, Tang KL, et al. "Correction of Anemia with Epoetin Alfa in Chronic Kidney Disease." *New England Journal of Medicine*, 355 (20): 2085-2098, 2006.
- ¹⁷ Keyhani S and Siu A. "The Underuse of Overuse Research." *Health Services Research*, 43(6): 1923-1930, 2008.
- ¹⁸ Keyhani S, Falk R, Bishop T, et al. "The Relationship Between Geographic Variations and Overuse of Health Care Services: A Systematic Review." *Medical Care*, 50(3): 257-61, 2012.
- ¹⁹ See for instance: Chassin MR, Koseoff J, Park RE, et al. "Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services? A Study of Three Procedures." *Journal of the American Medical Association*, 258(18): 2533-2537, 1987; Leape LL, Park RE, Solomon DH, et al. "Does Inappropriate Use Explain Small-Area Variations in the Use of Health Care Services?" *Journal of the American Medical Association*, 263(5): 669-672, 1990.
- ²⁰ McGlynn EA, Naylor CD, Anderson GM, et al. "Comparison of the Appropriateness of Coronary Angiography and Coronary Artery Bypass Graft Surgery Between Canada and New York State." *Journal of the American Medical Association*, 272: 934-940, 1994.
- ²¹ Korda H. *Medical Overuse: More Care Is Not Necessarily Better Care*. Washington: Altarum Institute, 2012, altarum.org/forum/post/medical-overuse-more-care-not-necessarily-better-care (accessed December 2012).
- ²² Cassel C and Guest J. "Choosing Wisely: Helping Physicians and Patients Make Smart Decisions About Their Care." *Journal of the American Medical Association*, 307(17): 1801-1802, 2012.
- ²³ Chernew ME and Newhouse JP. "What Does the RAND Health Insurance Experiment Tell Us About the Impact of Patient Cost Sharing on Health Outcomes?" *American Journal of Managed Care*, 14(7): 412-414, 2008.
- ²⁴ Gray B and Field M (eds). *Controlling Costs and Changing Patient Care? The Role of Utilization Management*. Washington: Institute of Medicine, 1989.
- ²⁵ Wickizer T. "The Effect of Utilization Review on Hospital Use and Expenditures: A Review of the Literature and an Update on Recent Findings." *Medical Care Research and Review*, 47(3): 327, 1990.
- ²⁶ Berenson R and Harris M. "Using Managed Care Tools in Traditional Medicare: Should We? Could We?" *Law and Contemporary Problems*, 65(4): 139-167, 2002.
- ²⁷ *Report to the Congress: Medicare and the Health Care Delivery System*. Washington: Medicare Payment Advisory Commission, 2011, medpac.gov/documents/Jun11_EntireReport.pdf (accessed December 2012).
- ²⁸ Berenson R, Sunshine J, Deb A, et al. *The Effect of Provider Payment Systems on Quality, Cost and Efficiency, and Access: A Systematic Literature Review*. Warsaw: InterQuality Research Project, European Union, 2012; Robinson J. "Theory and Practice in the Design of Physician Payment Incentives." *Milbank Quarterly*, 79(2): 149-177.
- ²⁹ Shen J, Anderson R, Brook R, et al. "The Effects of Payment Method on Clinical Decision-Making: Physician Responses to Clinical Scenarios." *Medical Care*, 42(3): 297-302, 2004.

- ³⁰ Sack K, “Unintended Consequence for Dialysis Patients as Drug Rule Changes,” *New York Times*, May 11, 2012, www.nytimes.com/2012/05/11/health/policy/dialysis-rule-changes-followed-by-transfusion-increases.html (accessed December 2012).
- ³¹ *Report to Congress on Incentive Arrangements Offered by Health Maintenance Organizations and Competitive Medical Plans to Physicians*. Washington: U.S. Department of Health and Human Services, 1990.
- ³² “Chapter 8: Addressing the Growth of Ancillary Services in Physicians’ Offices.” *Report to the Congress: Aligning Incentives in Medicare*. Washington: MedPAC, 2010, www.medpac.gov/documents/jun10_entirereport.pdf (accessed December 2012); Jacobs BL, Zhang Y, Skolarus TA, et al. “Growth Of High-Cost Intensity-Modulated Radiotherapy For Prostate Cancer Raises Concerns About Overuse.” *Health Affairs*, 31(4): 750-759, 2012; Waldman P, “Prostrate Patients Suffer as Money Overwhelms Best Therapy,” *Bloomberg*, November 6, 2012, www.bloomberg.com/news/2012-11-06/prostate-patients-suffer-as-money-overwhelms-optimal-therapy.html (accessed December 2012); Brox JJ, Nygaard OP, Holm I, et al. “Four-year Follow-up of Surgical Versus Non-surgical Therapy for Chronic Low Back Pain.” *Ann Rheum Dis.*, 69: 1643-1648, 2010.
- ³³ Codespote S, London W and Shatto J. *Physician Volume and Intensity Response*. Baltimore: CMS, 1998, www.cms.gov/Research-Statistics-Data-and-Systems/Research/ActuarialStudies/downloads/PhysicianResponse.pdf (accessed December 2012).
- ³⁴ Hadley J, Reschovsky J, Corey C, Zuckerman S. “Medicare Fees and the Volume of Physicians’ Services.” *Inquiry*, 46(4): 372-390, 2009.
- ³⁵ Jacobson M and Newhouse J. “Expect the Unexpected? Physicians’ Responses to Payment Changes.” Washington, DC: National Institute for Health Care Management Foundation, November 2010, <http://nihcm.org/pdf/EV-JacobsonNewhouseFINAL.pdf> (accessed December 2012).
- ³⁶ Sommers A, White C and Ginsburg P. “Addressing Hospital Pricing Leverage Through Regulation: State Rate Setting.” Washington: National Institute for Health Care Reform, 2012, www.nihcr.org/11t92 (accessed December 2012).
- ³⁷ Gruber J and Owings M. “Physician Financial Incentives and Cesarean Section Delivery.” *RAND Journal of Economics*, 27(1): 99–123, 1996.
- ³⁸ Hussey P, Ridgely MS and Rosenthal M. “The PROMETHEUS Bundled Payment Experiment: Slow Start Shows Problems in Implementing New Payment Models.” *Health Affairs*, 30(11): 2116-2124, 2011.
- ³⁹ Bailit M, Burns M and Painter M. *Bundled Payment Across the US Today: Status of Implementations and Operational Findings*. Newtown, CT: Health Care Incentives Improvement Institute, 2012, www.hci3.org/sites/default/files/files/HCI-IssueBrief-4-2012.pdf (accessed December 2012); Cutler D and Ghosh K. “The Potential for Cost Savings Through Bundled Episode Payments.” *New England Journal of Medicine*, 366: 1075-1077, 2012.
- ⁴⁰ Cutler and Ghosh.
- ⁴¹ Robinson JC and Luft HS. “Competition and the Cost of Hospital Care, 1972–1982.” *Journal of the American Medical Association*, 257(23): 3241-3245, 1987.
- ⁴² Berenson R, Bodenheimer T and Pham H. “Specialty-Service Lines: Salvos in the New Medical Arms Race.” *Health Affairs*, 25(5): w337-w343, 2006; Casalino L, Devers K and Brewster L. “Focused Factories? Physician-Owned Specialty Facilities.” *Health Affairs*, 22(6): 56-67, 2003.
- ⁴³ O’Connor AM, Llewellyn-Thomas HA and Flood AB. “Modifying Unwarranted Variations in Health Care: Shared Decision Making Using Patient Decision Aids.” *Health Affairs*, doi: 10.1377/hlthaff.var.63: VAR63-VAR72, 2004.
- ⁴⁴ Huesch M. “Payment Policy Based on Measurement of Health Care Spending and Outcomes.” *Journal of the American Medical Association*, 303(23): 2406-6, 2010.
- ⁴⁵ Siu AL, Sonnenberg FA, Manning WG, et al. “Inappropriate Use of Hospitals in a Randomized Trial of Health Insurance Plans.” *New England Journal of Medicine*, 315(20): 1259-66, 1986.
- ⁴⁶ Newhouse J. “Patients at Risk: Health Reform and Risk Adjustment.” *Health Affairs*, 13(1): 132-146, 1994.
- ⁴⁷ *Report to Congress: Physician Group Practice Demonstration Evaluation Report*. Washington: U.S. Department of Health and Human Services, 2009, www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/PGP_RTC_Sept.pdf (accessed December 2012); *Report to Congress: Physician Group Practice Demonstration First Evaluation Report*. Washington: U.S. Department of Health and Human Services, 2006, www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/downloads/Leavitt1.pdf (accessed December 2012); Trisolini M, Aggarwal J, Leung M, et al. *The Medicare Physician Group Practice Demonstration: Lessons Learned on Improving Quality and Efficiency in Health Care*. New York: The Commonwealth Fund, 2008, www.commonwealthfund.org/~media/Files/Publications/Fund%20Report/2008/Feb/The%20Medicare%20Physician%20Group%20Practice%20Demonstration%20%20Lessons%20Learned%20on%20Improving%20Quality%20and%20Effici/1094_Trisolini_Medicare_phys_group_practice_demo_lessons_learned%20pdf.pdf (accessed December 2012).
- ⁴⁸ Ghomrawi HMK, Schackman BR, and Mushlin AI. “Appropriateness Criteria and Elective Procedures – Total Joint Arthroplasty.” *New England Journal of Medicine*, 367(26): 2467-2469, 2012.