

DRAFT 4

Insurance and Payment for Health Care

Harold S. Luft, PhD

Paper prepared for

FRESH Thinking—Focused Research on Efficient, Secure Healthcare

Current policy discussion is turning to serious consideration of covering more of the uninsured. With that topic now on the agenda, it is time to ask what should be covered and how coverage should be implemented in terms of payments borne by plans and patients. Discussions of the first question—what should be covered—often become debates about exactly what services should be included and how new technologies should be assessed. Discussions about the second question—patient vs. plan responsibilities—often become debates about equity issues inherent in proposed new high deductible plans. While these issues are important, they will be addressed in other papers in this series. I will argue here that a more fundamental reconsideration is needed of the appropriate roles of insurance and payment in achieving more efficient and responsive health care delivery. Such reconsideration may also make more fruitful parallel discussions about technology assessment and equity issues.

Most health care coverage is designed to insure against the financial risk due to an accumulating stream of medical expenses. The cost sharing policy options typically used—deductibles and coinsurance—differentially impact price at the margin. It is often argued that coinsurance results in the consumption of more health care than would otherwise be optimal (Pauly, 1968). Deductibles are known to enhance price sensitivity,

but are particularly burdensome to the poor. Complete insurance, however, makes patients indifferent to resource use, leading to growing expenditures. This problem has been recognized for decades and is one reason for stalemated policy change.

I will put forward an alternative way of thinking about coverage, redirecting the primary focus from insuring *expenditures* to insuring *events* likely to be expensive. I will argue, moreover, that such coverage addressing the *occurrence of expensive events* should be mandatory. The residual uncertainty associated with what it costs to care for such events and for the more common lower cost events, may be addressed through voluntary coverage offered to those bearing the risks—patients *and* providers. Equity concerns should be addressed with income-based subsidies of such coverage. Changing how we address uncertainty and risk-bearing adds important tools in redesigning incentives for patients and providers. Mandatory coverage of a targeted *segment of health events* addresses critical adverse selection problems, while allowing more individual preference in other health events that are not as subject to selection.

This paper begins with a review of the rationale for insurance from the perspectives of both patients and providers, briefly discussing other issues associated with coverage, including tax preferences and equity. The next section summarizes the advantages and disadvantages of common cost sharing solutions—deductibles and coinsurance for patients with fee-for-service to providers, or nominal patient copayments combined with capitation for providers. It also discusses the selection problems when those two strategies are offered as options, or when coverage is based on groups in which enrollment is voluntary. The third section imposes a structure—one conceptually and clinically based—on what is usually simply seen as an endless stream of medical

services. Instead, this new structure sees clinical events as components of particular episodes of care. Viewed this way, one can ask: for each episode, who is the most appropriate decision-maker to whom incentives should be directed? The fourth section uses data on a large employed population to illustrate episodes of care and the incentive problems implied by standard approach using deductibles and coinsurance. The fifth section outlines an alternative in which comprehensive, lump-sum payments are made from a mandated pool for certain types of episodes, substantial periodic payments are made for others, and the remainder may or may not be insured. The final section discusses likely incentive effects under the classic models, capitation, low and high deductible insurance, and episode-based insurance.

Why Have Health Insurance?

Health care is rife with uncertainty and some medical care is very costly. The uncertainty, moreover, exists both with respect to the occurrence of a medical problem *and* what it may cost to treat the problem. This contrasts with automobile and property losses, in which the event is uncertain, but once it has occurred an ‘insurance adjuster’ can reasonably estimate what it will cost to make the policy-holder “whole.” In medical care, physicians are often uncertain as to exactly what needs to be done to treat a problem—both before and after the fact. Arrow’s 1963 classic article, “Uncertainty and the Welfare Economics of Medical Care” set the stage for thinking about this problem. Arrow laid out the welfare gains for risk averse people of having fairly priced health insurance (Arrow, 1963). Pauly’s 1968 article on “The Economics of Moral Hazard” highlighted the problem with insurance coverage—by reducing the price at the margin, patients demand more care than if without insurance (Pauly, 1968).

Medical problems, however, bring more than financial costs and lost income. Life itself may be in jeopardy and even medical problems with low mortality often involve significant discomfort, pain, and other aspects that make tradeoffs “in the moment” different than what one might have expected in advance. When choices need to be made by a proxy (e.g., a parent for child or husband for wife), models of rational economic decisions are unlikely to reflect behavior. Nyman argues that an important rationale for health insurance is to transfer extra purchasing power in the event of health problem (Nyman, 2003). Put another way, the perceived “moral hazard problem” in which patients “consume too much because the marginal price is zero” may be just what people want—to be able to avoid making economic choices in extreme situations.

Not only do prospective patients have an interest in insurance, but providers do as well. Health insurance was developed during the Depression, first by the hospitals and then by medical associations, to reduce their bad debt problem (Starr, 1982). Interestingly, coverage for people long pre-dated such insurance schemes. Prepaid plans began in the mid-19th century but were opposed by most physicians, at least in part because it shifted the risk to the physician (Trauner, 1985). Physicians use the term “reimbursement” for what economists call “payment”. Although physician fees include profit as well as input costs, and may be negotiated or arbitrarily set by a payer, such fees still have a flavor of “reimbursement” in that they are intended to pay for whatever was done. Reimbursement implicitly assumes that the services provided were necessary or otherwise agreed upon by the patient as an informed decision-maker. Whether this is a plausible assumption will be discussed below.

In the United States there are additional reasons for the presence of health insurance. Beginning with decisions after World War II that employer contributions for health insurance were exempt from the then-existing wage and price controls, the current tax subsidy of employer-based insurance resulting from this historical happenstance now exceeds \$200 billion (Selden & Gray, 2006). Abstracting from insurance loads, this subsidy makes it roughly 30% cheaper to buy medical care through an insurance plan than out-of-pocket. Some employers structure their compensation packages around relatively high benefits and lower wages. Medical care, moreover, is privileged in our social value hierarchy; public funding is more readily available for a safety net for medical care than for shelter, although the latter may actually impact health more.

Many believe expanding health insurance coverage to reduce the number of uninsured is simply the right thing to do. Unfortunately, the focus of the debate is on deciding whether a voluntary vs. mandatory strategy is to be preferred in reaching that goal, rather than *how* coverage does—or should work.

Current Models of Coverage

In contrast to the (hypothetical) situation in which an insurer adds a specific sum to a person's bank account in the event of a medical problem, current insurance models involve a "third party" in the payment nexus. Organized medicine traditionally railed against the involvement of such third parties in the physician-patient relationship (Starr, 1982). Their presence, however, creates a new policy tool because payments *from* patients need not be in the same form as payments *to* providers. For example, Medicare patients pay a fixed amount on admission to the hospital (a deductible) and then an amount *per day* after 60 days, analogous to the per diem payments used by many health

insurers. These payments actually go to Medicare, which pays hospitals a fixed amount for *each admission* based on the patient's diagnoses, and in some cases, procedures.

Insurance can transfer large or small amounts of resources, but its incentives largely operate at the margin—the incremental costs or revenues associated with each decision. In economic models of medical care use, patients are expected to demand additional medical care as long as its incremental benefits exceed its incremental costs to them. “Normal businesses” offer additional services as long as the incremental revenue exceeds their incremental cost of production. Many clinicians, however, play a dual role, serving both as agents for their patients and as providers of care.ⁱ In the first role they should make decisions purely reflecting their patient's interests. In the second, they are business operators considering the marginal costs and revenues of services they provide.

Effects of Fee-for-Service Payment on Providers and Patients

The classic insurance model relies on patient incentives using deductibles, coinsurance and a maximum out-of-pocket limit. The notion is that the patient can afford to cover small amounts of care, represented by the deductible, paying the full incremental cost during that period. Thereafter, the patient pays only the coinsurance rate, typically 20%, until a maximum out-of-pocket (MOOP) limit is reached when the marginal cost then falls to zero. The combination of coinsurance and MOOP maintain some price sensitivity over a wider range of care while capping financial risk. For a prototypical policy with a \$500 deductible, a 20% coinsurance rate and a MOOP of \$2500, the patient's marginal cost is \$1 per dollar of care for the first \$500, dropping to \$0.20 per dollar for the next \$10,000 in expenses, and then falls to zero.ⁱⁱ At the beginning of a new year the cycle begins again, unless a lifetime benefit has been reached.

The incentive situation for the clinician is markedly more complex due to the dual role as provider and agent. Simply from the perspective of a provider of services, insurance is desirable because it increases effective demand, although the administrative problems in getting paid are not trivial. Most physician practices have substantial fixed costs for rent, staff, malpractice, and other costs, so the marginal cost to the clinician of additional services is usually well below the marginal revenue (Conrad & Christianson, 2004). To set aside the potential conflict of interest as both agent and provider, imagine an alter ego standing behind the physician with access to all the information and expertise the physician has. The alter ego acts purely as the patient's agent signalling yes or no for each choice to be made. The alter ego would still face a near impossible task—patients do not come for treatment with diagnoses painted on their foreheads. Even Medicare's Diagnosis Related Group system of paying hospitals uses the diagnosis principally responsible for the patient's admission—*as determined by the time of discharge*. Furthermore, even with a definitive diagnosis the best course of treatment is not always clear and idiosyncratic patient problems may warrant diverging from “standard practice.” There is substantially more uncertainty in what medical care is needed to deal with a patient's health problem than with a car's collision damageⁱⁱⁱ.

Given this uncertainty, medical providers have preferred to be reimbursed for whatever they do, in some instances getting paid for the complications associated with poor quality of care. Not all care, however, is on this strict fee-for-service (FFS) basis. Surgical fees for a procedure commonly include follow-up visits and prenatal care has traditionally covered extended series of visits for an uncomplicated pregnancy. Most significantly, Medicare pays hospitals a fixed amount for a patient stay depending on the

diagnosis. With important exceptions in the case of very long stays, the hospital is “at risk” for all the services it provides during that admission. Individual cases may be more or less expensive than their associated DRG payment, but on average the large number of cases washes out variability.

Although about a third of all costs are associated with hospital care, hospital entities have little control over the costs incurred by individual patients. Decisions about the specific services required, the length of stay and most importantly, whether a patient should be admitted, are made by physicians--ideally as well-intentioned agents for their patients. In addition to uncertainties related to appropriate diagnosis and treatment, neither clinician nor patient is likely to know the cost to the patient for various interventions. The outpatient clinician may know his or her fees, but typically does not know the patient’s deductible and coinsurance rate, nor where the patient is in the accumulation of expenditures determining whether the marginal cost is \$1.00, \$0.20, or \$0 per dollar. The complications of insurer and health plan-specific negotiated fees, exclusions, and after-the-fact adjudication render prospective price-sensitive decision-making impossible. Both patients and clinicians probably roughly categorize services as insured vs. out-of-pocket, treating the first as “free” to the patient, and the second as “uncovered.” With this simplified notion, agency becomes a plausible, but not poorly-informed, goal and recommendations can simply focus on whether there is a positive medical benefit expected from the services “covered.”

Effects of Capitated Payment of Providers and Patients

The major alternative to conventional insurance for the patient and FFS payment for providers is the HMO model of nominal copayments for the patient combined with

capitation for the providers. HMOs focus incentives on the clinicians and take responsibility for providing all necessary medical care for a population of patients. Per visit copayments reduce use somewhat or shift it to more appropriate settings (e.g., away from emergency room use) (Reed et al., 2005; Selby, Fireman, & Swain, 1996). The patient's marginal cost for a visit with a copayment may be less, roughly the same, or more than under a conventional plan, depending on when the visit occurs in the accumulation of expenses in the latter.

At one level, the annual capitation payment to providers to cover the necessary care of all enrollees simplifies incentives: marginal revenue for additional services is zero. In practice, however, the operative incentives are more complex. While a large group can be paid on a capitation basis, individual providers within the group may be paid in various ways, ranging from strict salary, to some form of profit sharing, to strict fee-for-service. These issues are addressed in more detail by Shortell and Casalino in their paper. Although there is ample evidence that systems such as Kaiser Permanente operate effectively with capitation and provide high quality care, their size and integrated group structure are important features making this possible. Size allows the organization to smooth out random variation in patient need *within* the group; integration allows implicit guidelines and standard operating procedures to substitute for formal practice rules. Efforts by other HMOs to capitate primary care physicians to manage all the risk for their patients have not been successful, largely because they are unable to control the costs of hospitalizations. Capitation for out-patient pharmaceuticals has been resisted because the physicians feel they cannot control the prices of drugs or the approval of new, expensive medications (Lipton, Agnew, Stebbins, Kuo, & Dudley,

2005). Even for a limited set of services, individual physician-level capitation is often resisted because there is too much variability in use across patients; a physician may find his or her whole risk pool used up by one or two unusually sick patients.

Parallel Fee-for-Service and Capitated Models

The markedly differing incentives in the two combinations of coverage and payment yield differing expectations with respect to “practice styles” or the characteristic ways in which clinicians approach and treat a problem. The (largely) reimbursed FFS model of coverage and payment facilitates wide variation in practice because additional practitioners can be supported through higher utilization on a per capita basis (Fisher et al., 2003a; Fisher et al., 2003b). Reimbursed FFS is, on average, likely to generate resource use above the optimal (but unknown) level. In contrast, capitation has incentives to constrain resource use, although if the group has long term responsibility for the enrolled population, they may give more attention to preventing downstream disease and exacerbations. Until better research identifies optimal treatment patterns, it is socially beneficial to have the two systems operating side by side to provide comparative measures of resource use and outcomes.

The parallel operation of the two approaches, however, can be quite problematic in certain situations. The capitated health plan usually seeks to actively manage the costs of care offered its enrollees with nominal cost sharing provisions. The reimbursement plan usually is more passive and relies on deductibles and coinsurance to dampen patient demand for care, then pays for services provided, sometimes at negotiated fees. When both systems are available, people often have the choice of enrolling in one or the other. In theory, either type of system could handle high or low risk enrollees, but there is

substantial evidence that HMOs tend to have a lower risk enrollee mix^{iv}. Without adequate and precise risk adjustment, premiums will not be fairly set. Risk adjustment techniques have improved over time, but none are really adequate to prospectively adjust for differences in clinical need (Dudley et al., 2003). Some have therefore argued that choice should be eliminated, but this is politically unacceptable and would probably reduce quality of care. If anything, patients with difficult problems should be encouraged to join those plans and see those clinicians best able to address their medical needs; the challenge for policy analysts is to design a system to pay fairly for their care.

A Clinical Focus on Services

The current payment system imposes little “structure” on how medical services are paid. Insurers may approve or deny payment for specific services after considering the nature of the patient’s problem, but such reviews are the exception rather than the norm. Deductibles and coinsurance have a “structure” based on arbitrary accumulation of claims over a period of time, usually January-December. While this has been the standard way of paying for care for decades, it has not been examined with respect to its differential incentive impacts. To begin this assessment, we use a large data set of about 800,000 people covered by employer-based insurance during 2003-04^v. All these people are under the age of 65, largely eliminating services provided by Medicare. With the exception of home health care, all covered services (including prescription drugs) are included in the claims. When assessing cost sharing incentives, however, we exclude the prescription drug data because those plans often have separate cost sharing requirements.

Table 1 presents the data for the 488,135 people with paid claims for 2003. Actual cost sharing rates are not known, nor are the specifics of claims adjudication possibly affecting payment. To illustrate the point of this paper, a plan is assumed to have a deductible of \$500 per person, followed by a 20% coinsurance rate until the maximum out of pocket expense of \$2500 is reached. These figures are in the mid-to-high range reported in the 2004 Medical Expenditure Panel Survey (Sommers, 2006).^{vi} By definition, all the claimants experienced claims subject to the hypothetical deductible, with an average cost of \$371. (Note that this exceeds the average amount across all at-risk members, many of whom had no claims reported in the data set.) Somewhat over half of those with some claims experienced costs at the 20% coinsurance rate, and only 4% went above the MOOP limit of \$2500. Costs accrued above the MOOP limit, however, accounted for over a quarter of *all* costs incurred by this population.^{vii} It is well-known that a small fraction of the population with high expenditures account for most medical care costs (Conwell & Cohen, 2005). What is less well appreciated is that much of these costs occur a plausible maximum out-of-pocket limit.

The current payment system also ignores the fact that *who* makes critical decisions varies with the nature of the problem. Patients typically decide to initially see a clinician for a problem and they often choose to not fill, or to not take as directed, prescriptions (DiMatteo, 2004). Patients may request or refuse hospitalization and may choose among major alternative treatment options, but most of resource use decisions during an admission are controlled by the physician, and to a lesser extent, the hospital.

This consideration leads to a different way of viewing payment for health care services: categorizing them by episode of care. The episode notion links all the services

associated with a given problem, regardless of when or by whom the services are provided. For example, a soccer injury might lead a patient to seek care, resulting in an initial recommendation of conservative treatment and a prescription for anti-inflammatory medication. If the pain doesn't subside after a period of time, a second office visit may lead to imaging studies, perhaps a referral to an orthopedist, and a series of physical therapy visits. Various software programs, or "groupers," can collect these disparate claims together into an episode of acute care. "Windows" of time without a service distinguish different episodes of acute problems, e.g., repeated ankle sprains. During our hypothetical patient's initial visit for her injury, she may also have had her HgA1c checked as part of the regular monitoring of her diabetes. Several days later her physician changes her prescription and requests a follow-up visit in two months. As diabetes is a chronic condition, once identified it has no logical end date, so all services for the diabetes are lumped into an episode that may be arbitrarily stopped at 12 months for coverage purposes and then restarted. The groupers can extract individual services, such as the HgA1c test, from a visit and apply each to the appropriate episode.

People often think of conditions as acute vs. chronic, recognizing that some services are classified as 'well' or 'preventive'. Although making sense for some purposes, when considering incentives, a fourth category (Major) should be added to account for major acute events, significant exacerbations of chronic conditions, and major procedural interventions. In essence, these are the hospital-based components of care, broadened to include major intervention involving multiple clinicians undertaken in a facility, even if the patient stayed less than 24 hours.^{viii} Acute episodes involving a hospitalization or similar facility use are reclassified as Major.^{ix} If a chronic episode

involved such use, the services associated with that period of time are reclassified as a Major episode and the other services remain in the chronic episode category^x. In the standard approach, a patient with congestive heart failure (CHF) would have a year-long episode of CHF which might include one or more hospitalizations for the CHF. Here those hospitalizations will be reclassified as Major episodes and the remainder of the CHF services comprise the CHF chronic care management episode.

To provide a sense of the implications of recategorizing services into episodes of care, we use the data set of about 800,000 people for 2003-04; unlike the case before, prescription drugs are included in the claims. Table 2 presents overall indicators of the four types of episodes. Major events account for only 4% of all episodes, but 34% of the cost experienced by these people. In contrast, minor acute episodes were very common, accounting for 61% of all episodes and 34% of total cost.

Episodes of care, of course, do not happen in isolation, but occur within individuals. Knowing the sequence of claims during the year allows the application of the deductible, coinsurance, and MOOP rules as the claims occur; then reorganizing the data by type of episode illustrates how the different types of cost sharing would affect different categories of episodes of care. As seen in Table 3, slightly less than ten percent of the people in our data set experienced a Major episode during 2003. For nearly all of these people, some care was provided during the major episode at the standard coinsurance rate; only a little more than half began their episode while still in the deductible range.^{xi} Somewhat more than a quarter had charges for that episode after exceeding their maximum out-of-pocket. Nearly half of all the dollars in these major episodes, however, were expended when the cost sharing was zero. In contrast, for

chronic care management, when adherence to regimen may be most important, 84% of the patients (187,779/222,941) had some chronic care expenditures in the deductible range. Overall, 19.7% of the expenditures for chronic care would have been paid in full by the enrollee with this hypothetical coverage.

The large costs and low probability of occurrence associated with the major events argues for insurance on the same welfare grounds originally presented by Arrow. Unfortunately, many of these major events are acute exacerbations of underlying chronic conditions. Even if not formally uninsurable, insurers would seek to avoid enrollees with such conditions if they do not receive higher premiums to cover predictably higher costs.

Episodes and Incentives

Episodes of care provide a way of grouping services and incentives around situations in which the decision makers differ. Most decisions are out of the patient's hands during hospitalization. Even if patients could control resource use, only half of all expenditures in that category would be subject to coinsurance most of the rest is "free" at the margin. Because the patient has relatively little control over resource use, cost-sharing merely shifts the cost burden to the sick. Literally, not just figuratively, the patient already has his or her "skin in the game"—the more significant economic problem is that the physician does no incentive to consider the value of the resources used at the margin. Medicare's DRG payments provide incentives for hospitals to attempt to contain costs, but these incentives do not apply to the physicians making the decisions. As an alternative, we could bundle together the payments for physicians and hospitals into an Expanded DRG paid to what might be called the Care Delivery Team, or CDT. Such

teams could use a wide range of internal mechanisms to manage resources and enhance quality

Paying for ongoing chronic care is more complex because the patient has more opportunities to make decisions, especially with respect to adherence to regimens. As bearing the full cost of drugs, tests, or visits during a deductible reduces use, this may have deleterious impacts on overall health and may result in more acute exacerbations that are ultimately paid with more heavily insured dollars (Parente, Feldman, & Christianson, 2004). Thus, deductibles may be inappropriate for such episodes. Contrast this with the situation with respect to minor acute care. People often have several such unrelated episodes in a year. If cost sharing should provide incentives for the patient to consider whether professional care is necessary rather than home treatment, a better alternative would be a small per episode deductible, say \$100, followed by coinsurance at the usual rate. In this way, patients having sprains occurring in December would have the same incentives as those in January.

In a separate proposal for health reform, I argue for a coverage and payment scheme that builds on this conceptualization of episode based insurance.^{xii} In brief, everyone would have coverage through a universal pool for major events and chronic illness management, eliminating nearly all the selection problems arising when people choose among plans and providers. This pool would provide care delivery teams a fixed sum for each Expanded DRG (EDRG) and CDT members would allocate funds amongst themselves. Hospitals could house multiple CDTs focusing on different clinical areas, such as orthopedics vs. cardiovascular. People might choose to buy supplemental

policies to cover additional payments for CDTs that felt the EDRG rates were insufficient, perhaps because their practice styles was more resource intensive.

The pool would provide monthly payments for chronic illness management (CIM), based on the condition and risk factors *within* the condition. Except for those in relatively large groups, clinicians are unlikely to want to bear the risk associated with variations in resources used for such patients, even if hospitalizations were handled by the CDTs. The clinician managing the patient's care will have chosen a carrier to handle his or her claims and these monthly CIM payments will cover most of the costs, except for those associated with acute exacerbations, which are covered directly by payments from the pool to the CDTs. In effect, the carrier provides reinsurance to smooth out the variance across individual cases. The physician's carrier will offer patients a menu of plan options with varying cost sharing approaches, such as the size of the deductible and copayments. The net premium to the enrollee takes into account these cost sharing incentives, the CIM payments, the physician's practice patterns and the reinsurance cost. This provider-based insurance allows the patient to see (and bear) the net effects of physician practice style and fee differentials on overall care, with the pool carrying most of the risk associated with chronic illness and major events.

Every payment scheme includes opportunities for gaming. Here, physicians and plans would have incentives to use diagnosis codes that would classify episodes as chronic, rather than acute, and thus get payments from the pool. The ongoing nature of the chronic payment from the pool and universal capture of billing data, however, allows the pool to identify patients with chronic diagnoses, such as diabetes, who are not being appropriately monitored. Fraudulent coding would be quickly identified as poor quality

care subject to more detailed auditing. Moreover, the monthly pool CIM payments would be risk adjusted *within* chronic conditions with only nominal payments made for mild and undocumented diagnoses. Additional clinical data, laboratory values and prescriptions allow monitoring condition severity for appropriate payment as well as processes and outcomes for quality. Such enhanced data are available in electronic medical records and can be input from more conventional systems for selected patients—in the absence of additional data, patients are assumed to be at the minimum risk (and payment) level. The enriched data set will improve practice and rapidly distinguish occasional miscoding of a diagnosis from patterns suggesting outright fraud.

The pool is not involved in minor acute or preventive care, except for those preventive services demonstrated to save the pool money. Tax advantaged HSA-like accounts could be used by individuals to manage their expenses for minor acute episodes, or they could buy coverage through plans like those handling chronic illness (in fact, they could be the same). These plans would have significant flexibility in designing payment approaches, perhaps using episode-based deductibles and fixed copayments instead of conventional deductibles. (Because practice-style and patient-preference sensitive oriented approaches have not been widely used, substantial experimentation is likely to occur.) The plans will compete to be the carriers chosen by doctors—low administrative loads, transparent pricing, and the provision of data useful to improve practice efficiency will be important features in how physicians choose carriers. Premiums across physicians, however, will reflect the physicians' own choices about their fees, referral patterns, and practice styles, not their case mix.^{xiii} Income-based subsidies for premiums and cost sharing would replace categorical programs for the poor.

Table 4 compares how different reform plans could use coverage, cost sharing, and payment approaches. Because the episode based insurance model uses different payment approaches for each type of care, it requires four columns, but most of these cells are blank. The first two columns describe the situations under a standard plan with deductibles, coinsurance, and FFS payments to providers, as well as an HSA-style plan with a high deductible and high maximum out-of-pocket payment. The rows use the categories of episodes discussed above with one important difference; major episodes are split into those that are acute and those that can be scheduled. Heart attacks would be in the former category, hip replacements in the latter. The distinction allows a more nuanced assessment of when price shopping might be possible. Note, however, that some recent work suggests little evidence of effective price shopping even in situations in which most costs are out-of-pocket (Tu & May, 2007).

Conventional insurance typically involves a small disincentive for hospitalization; to the extent that a deductible followed by coinsurance has less impact than a large deductible, HSAs may have a slightly greater disincentive for admissions. Episode Based Insurance (EBI) has no admission disincentive, but strong efficiency incentives for providers within the episode. For schedulable hospitalizations, HSAs may result in some time shifting in demand and there might be some unit price shopping^{xiv}. The cost sharing under conventional insurance may affect adherence to treatment regimens for chronic conditions, and this problem is greater for HSAs. Recurring costs may cause selection problems unless a universal pool offsets the pressures on plans to avoid high cost enrollees. For minor acute problems, cost sharing may be appropriate, but the structure of conventional plans means the marginal cost is usually either \$1/\$1 or \$0.20/\$1.00.

Most people with HSAs will bear the full cost of most minor acute care, potentially a problem if some early interventions can reduce a worsening of problems. The incentives for preventive care depend on whether such services are covered.

Summary and Conclusions

Payment incentives influence patient and provider behavior and in the long run even the direction of technological innovation. Our current payment system is built on a model of insurance against financial risk and would be reasonably appropriate if the occurrence of events were largely random, infrequent and with few choices about resource use *given* a medical problem. Unfortunately, this does not reflect the real world of medical care practice. Providers' natural aversion to accepting economic risk for treating specific cases leads them to prefer FFS payment. Combining conventional insurance and FFS payment generates incentives that fuel expenditure inflation, support widely varying clinical practice, and are probably suboptimal for patient care quality.

Various combinations of insurance and payment create different patterns of risk-bearing and incentives. The fundamental nature of medical problems and medical care is that most expenditures are concentrated in a small number of people, that chronic illness and its exacerbations are highly predictive of such costs and that there is still enormous uncertainty in exactly what is needed to appropriately care for specific patients. Patients can reasonably influence certain medical decisions, but many decisions are beyond the technical understanding of the patient or need to be made under circumstances when preferences are likely to be skewed by the illness itself.

The proposed episode based insurance uses a universal pool to cover the occurrence of major episodes of care and chronic illness, thereby eliminating the

selection problems that plague voluntary insurance approaches. For these episodes, providers (CDTs for major episodes and ambulatory physicians managing chronic illness care) bear a small degree of risk due to random variation associated with patient factors *not* already accounted for in the risk models. Much more important will be the consistent effects of their own fees, referral patterns, and practice styles. Reinsurance can offset the variation due to random factors; CDTs may choose to purchase this and it is likely to be bundled into the residual premiums quoted patients by ambulatory care physicians. The universal pool would not cover minor acute episodes, but even though the mean cost for such episodes is small, most people would not want to bear the risk for such care. The premiums of plans purchased by patients reflect the risk shifted to the carrier.

Reform of the health care system means more than just providing coverage for everyone—how that coverage pays for care and how patients are engaged in the cost implications of their treatment is critical. Capitation is a viable model in large integrated group practices, but has not been well-received by the majority of physicians who are in solo or small group practice. Most proposals promising payment simplicity ignore the complexity of medical care decisions and delivery. Fee-for-service appears simple until one examines the dizzying array of administrative systems needed to make it work. Capitation is simple until one has to address how payments should be adjusted across plans and over time to account for new treatments and health problems. Competition between capitated plans and the standard reimbursed FFS with a financial model of patient cost sharing, however, places enormous pressure on effective risk-adjustment.

An alternative using episode based insurance and provider based premiums has markedly better incentives, yet does not require large integrated medical practices. It

takes as given the existing mix of providers and practice arrangements, as well as the connections between patients and physicians, thereby increasing its political viability. Even so, transition to and implementation of such a system will not be a simple undertaking; a discussion of those issues, however, is beyond the scope of this paper. Other options for payment are also conceivable and they should be encouraged to be put on the table as true reform is considered. Careful assessments of their likely impact on practice, patients, and overall costs should inform our choices among the reform options.

- Arrow, K. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53(5), 941-973.
- Conrad, D. A., & Christianson, J. B. (2004). Penetrating the "black box": financial incentives for enhancing the quality of physician services. *Med Care Res Rev*, 61(3 Suppl), 37S-68S.
- Conwell, L., & Cohen, J. (2005). *Characteristics of Persons With High Medical Expenditures in the U.S. Civilian Noninstitutionalized Population, 2002. Statistical Brief #73*. Rockville, MD: Agency for Healthcare Research and Quality.
- DiMatteo, M. R. (2004). Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med Care*, 42(3), 200-209.
- Dudley, R. A., Medlin, C. A., Hammann, L. B., Cisternas, M. G., Brand, R., Rennie, D. J., et al. (2003). The best of both worlds? Potential of hybrid prospective/concurrent risk adjustment. *Med Care*, 41(1), 56-69.
- Fisher, E. S., Wennberg, D. E., Stukel, T. A., Gottlieb, D. J., Lucas, F. L., & Pinder, E. L. (2003a). The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care. *Ann Intern Med*, 138(4), 273-287.
- Fisher, E. S., Wennberg, D. E., Stukel, T. A., Gottlieb, D. J., Lucas, F. L., & Pinder, E. L. (2003b). The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care. *Ann Intern Med*, 138(4), 288-298.
- Luft, H.S., (2007). Universal Health Care Coverage: A Potential Hybrid Solution. *Journal of the American Medical Association*. 297(10), 1115-1118.
- Lipton, H. L., Agnew, J. D., Stebbins, M. R., Kuo, A., & Dudley, R. A. (2005). Managing the Unmanageable: The Nature and Impact of Drug Risk in Physician Groups. *Journal of Health Politics Policy and Law*, 30(4), 719-750.

- Nyman, J. A. (2003). *The theory of demand for health insurance*. Stanford, Calif.: Stanford University Press.
- Parente, S. T., Feldman, R., & Christianson, J. B. (2004). Evaluation of the effect of a consumer-driven health plan on medical care expenditures and utilization. *Health Serv Res*, 39(4 Pt 2), 1189-1210.
- Pauly, M. (1968). Economics of Moral Hazard: Comment. *American Economic Review*, 58(3), 531-537.
- Reed, M., Fung, V., Brand, R., Fireman, B., Newhouse, J. P., Selby, J. V., et al. (2005). Care-seeking behavior in response to emergency department copayments. *Med Care*, 43(8), 810-816.
- Selby, J. V., Fireman, B. H., & Swain, B. E. (1996). Effect of a Copayment on Use of the Emergency Department in a Health Maintenance Organization. *N Engl J Med*, 334(10), 635-642.
- Selden, T. M., & Gray, B. M. (2006). Tax Subsidies For Employment-Related Health Insurance: Estimates For 2006. *Health Aff*, 25(6), 1568-1579.
- Sommers, J. P. (2006). *Co-pays, Deductibles, and Coinsurance Percentages for Employer-Sponsored Health Insurance in the Non-Federal Workforce, by Industry Classification, 2004. Statistical Brief #145*. Rockville, MD: Agency for Healthcare Research and Quality.
- Starr, P. (1982). *The social transformation of American medicine*. New York: Basic Books.
- Trauner, J. B. (1985). The California health care market--where is it headed next? *Front Health Serv Manage*, 1(4), 4-30.
- Tu, H. T., & May, J. H. (2007). Self-Pay Markets In Health Care: Consumer Nirvana Or Caveat Emptor? *Health Aff*, hlthaff.26.22.w217.

ⁱ This dual role varies with the situation. For example, some clinicians merely apply their expertise in settings in which their interests are unlikely to be in conflict with their patients. Thus, a radiologist interpreting an MRI is not in the same situation as an orthopedist potentially recommending back surgery based on that MRI.

ⁱⁱ While the notion of incentives working at the margin is generally accepted, the empirical impact of deductibles, coinsurance, and maximum out-of-pocket limits is more complex. In theory, consumers assess the marginal cost to them taking into account not only insurance coverage for the specific item, but whether this expenditure will affect the price of future one. A patient may have a “fresh” \$500 deductible in January, but expect that he or she is likely to be in the “coinsurance” range long before the end of the year. If so, even the first dollar has an effective marginal cost close to \$0.20. In practice, while some carefully calculating economists may behave according to this theory, the Rand Health Insurance Experiment found that most people acted as if expenditures during the deductible cost them dollar for dollar.

ⁱⁱⁱ The auto insurance case is even simpler because for many complex repair problems, the insurer can simply choose just to buy the car for its “Blue Book value” and

then either resell it as damaged or for scrap. This is obviously not an option for medical care.

^{iv} HMOs typically offer more comprehensive coverage and benefits, so one might expect them to attract higher risk enrollees. On the other hand, HMOs,—especially the integrated group practice model—tend to rely on a more narrow network of providers, which may be unattractive to those thinking they will need care. Perhaps more importantly, people beginning in a reimbursement system who know they are sick are likely to have established relationships with clinicians that they feel uncomfortable breaking merely to access better financial coverage.

^v These data, and the Symmetry episode grouper used in the analysis, were provided by Ingenix, Inc. Under the terms of Ingenix’s agreements with the carriers for these health plans, we are not able to know the source of the data, except that they come from multiple carriers with enrollees across the country.

^{vi} These data and this plan yield results similar to national data for 2003 from the Medical Expenditure Panel Survey. Overall covered charges in this data set are \$1398 per person, vs. \$1603 for MEPS, excluding dental and prescription drug costs for those under the age of 65 with employer-sponsored coverage. Overall out-of-pocket costs implied by the 500/20%/2500 policy are 27.9% of the total; the figure from MEPS is 31-35% for children and adults, but this includes dental and prescriptions. Given the differences in definitions of services, coverage, and the potential role of supplemental plans, the Ingenix data and the hypothetical plan are reasonable approximations.

^{vii} People in this range, of course, incurred an even larger share of total costs, because some were incurred at the deductible and coinsurance ranges.

^{viii} The intent here is to make no distinction between similar surgical and other procedures that are done on either an inpatient or ambulatory basis, but to exclude the minor surgical procedures that are commonly done in a physician’s office. If payments are to be associated with the nature of the episode, it is important that non-substantive differences in venue or coding not allow “gaming” of the payment system.

^{ix} The Symmetry software does not categorize episodes into acute, chronic and preventive. We utilized the Johns Hopkins University ACG software to do this, coding each diagnosis appearing within the claims bundled into an episode defined by Symmetry into one of the 32 Hopkins Adjusted Diagnosis Groups, such as “stable chronic.” The pattern of diagnoses within the ETG was used to classify it as Acute, Chronic, or Preventive.

^x We plan to extend the definition of the Major (inpatient) episode within a chronic episode of care to include services received for the problem up to 3 days prior to admission and 7 days post discharge. This is to capture the visits, imaging, and other services that may be associated with an acute exacerbation, as well as immediate post

discharge care that may reflect step-down units and home care substituting for inpatient care. The data below do not reflect these proposed changes, but the impact on our findings is likely to be qualitatively small.

^{xi} Unlike the situation for all expenses, when broken down by type of episode, the first dollars within a category may occur later in the year after the deductible has been met. Not surprisingly, many major episodes are preceded by other care that “takes the patient past the deductible range” before admission.

^{xii} The complete description of the proposal is still in progress as a book. A short version of the concept is presented in Luft, 2007.

^{xiii} Legitimate questions can be raised about the influence of patient factors in what may appear as the physician practice style or quality. For example, if patients do not adhere to recommended drug, diet, or smoking cessation regimens, should the physicians be held responsible? The answer is complex, partly because we do not know to what extent the adherence problems can be overcome with better physician communication, the use of outreach workers, or other innovations. The import of the problem also depends on whether adherence issues are randomly distributed across physician ‘panels’ or are concentrated among certain physicians due to either geographic or referral factors. If consistent risk factors for poor adherence can be identified, ranging from limited medical literacy to chaotic family surroundings to substance abuse, it is better to identify them and address them than to use them as arguments against risk-adjusting payments or quality measures. Providers feeling they are adversely affected by having a disproportionate share of patients with adherence problems have strong incentives to develop appropriate codes, have the data collected and have such measures included in the risk models. Once this is done, others will develop better ways to address the identified problems.

^{xiv} If people covered with HSAs expect the cost of their admission to exceed their maximum out-of-pocket expense they will be price insensitive. If the absence of good quality information leads patients to assume that high price means high quality, then the ‘shopping’ behavior encouraged by HSAs may actually have a price increasing effect.

I am indebted to Laura J. Eaton, MD, MPH, UCSF Dept. of Family and Community Medicine for her assistance in conceptualizing and developing the episode measure and to Beth Newell, BA, for her general research and editing assistance.