INTRODUCTION

The standard theoretical framework of health economics, as summarized in public finance undergraduate textbooks (e.g., Rosen, 1995, chapter 11), draws attention to four problems. First, illness and accidents are unpredictable but costly events, and therefore health insurance is desirable. Unfortunately, the existence of health insurance means that people are insulated from the true costs of medical care, resulting in too much service use and excessively high medical spending. This moral hazard problem is exacerbated by the tax–subsidy for the purchase of insurance. Second, adverse selection causes low–risk individuals to leave the insurance market, so that a single–pooling equilibrium with one premium for both low– and high–risk is frequently unsustainable and low–risk people are often underinsured. Third, physicians are better informed than patients and the resulting information asymmetries cause market failures. Finally, concerns about equity may lead to redistribution related to health care.

This standard theoretical framework provides the underpinnings of most empirical health economics and is a valuable guide for policy analysis. However, it has significant limitations that recent contributions have sought to address. In this paper, we examine these developments and their relation to the important problems confronting health care policy. The next section of the paper first describes the contributions and limitations of the standard theoretical framework, highlighting the significant empirical phenomena that it does not explain. The following four sections examine in turn each of the major areas of health economics: moral hazard, patient–provider informational asymmetries, selection and equity. Each section explains the new developments and discusses their implications. The final section concludes.

CONTRIBUTIONS AND LIMITATIONS OF THE STANDARD THEORY

Health economics theory has provided a fertile ground for empirical analysis. Health economists are data rich, profiting from the availability of large administrative datasets,
quasi–experimental policy changes, demonstration projects, and randomized experiments. These resources have allowed empirical analysts to assess the price elasticity of demand (a key component of moral hazard), the extent of adverse selection, and the effects of asymmetric information in health care markets.

Research on the price elasticity of demand for medical care began with cross-sectional studies that examined health service use as a function of the price of medical care, measured as co-payment levels in health insurance (Rosett and Huang, 1973). The results of these studies were confounded by the correlation between health status and co-payment levels induced by adverse selection. Later studies tried to avoid this by turning to quasi–experimental studies, evaluating, for example, a change in the health insurance coverage of a defined population (Scheffler, 1984). Finally, in 1975, the Department of Health and Human Services funded the RAND health insurance experiment, a randomized, controlled study of the responsiveness of health care use to out–of–pocket prices that included over 5,000 families, who were followed for 3–5 years. The results of the RAND study suggest that the price elasticity of demand for medical care is about –0.2. Estimates of moral hazard–related welfare losses that are based on this elasticity suggest that the welfare losses associated with voluntarily–purchased private health insurance are very large—comprising between 8 percent and 28 percent of national health expenditures.¹

Studies of adverse selection in health insurance have focused almost exclusively on situations in which people face a choice among health plans. A pioneering study (Price and Mays, 1985) examined the Federal Employees Health Benefit Plan, whose beneficiaries could choose among a range of coverage options. That study found that employees who selected the plan with the lowest co-payment levels were much less healthy than those who selected less generous plans. Twenty one percent of the price of the generous plan could be attributed to selection. The plans examined by Price and Mays were fee–for–service and HMO options. More recent studies examine selection between fee–for–service and HMO options (for a review, see Glied, 2000). One such study examined how Medicare beneficiaries choose between these coverage modalities (Brown et. al., 1993). It found that Medicare beneficiaries who switched to HMOs had 21 percent lower utilization in the year before switching than did those who remained in the fee–for–service plan.

The literature on adverse selection provides less evidence of the extent of adverse selection in the choice between coverage and no coverage. Although theory suggests that policies that limit the spread between prices charged to high– and low–risk people may lead to the undermining of the insurance market, one recent study, which examines the effect of a move to full community rating in New York State’s individual and small group insurance markets, concludes that this change did not generate the expected adverse selection spiral (Buchmueller and DiNardo, 1999).

The theory of informational asymmetries in medical care has engendered a substantial literature on the effects of payment incentives on the provision of medical services. If purchasers were fully informed about the costs, quality, and consequences of physician actions, competitive forces should imply that changes in payment mechanisms would have little effect on provider behavior. In fact, how-

¹ This figure is the authors’ calculation based on Feldman and Dowd’s (1991) estimate that the size of the welfare loss in 1984 was between $33.4 and $109.3 billion. 1984 National health expenditures were $390 billion.
ever, numerous studies show that providers do respond appreciably to payment incentives. One strand of the empirical literature examines physician responses to reductions in fee–for–service payment rates. Reductions in these rates appear to lead to proportionate increases in the volume of services provided (Christensen, 1992). A second strand of the literature examines how providers behave when payment changes from fee–for–service to either an annual flat capitation payment (in the physician market) or a case rate, such as a DRG payment (in the hospital market). This literature finds that changes in payment mechanisms generate corresponding changes in physician effort (Glied and Graff Zivin, 2002; Hillman, Pauly, and Kerstein, 1989).

Standard theory is also helpful in guiding analysis of policy. For example, the standard treatment of moral hazard has led economists to emphasize the value of withdrawing the favorable tax treatment of employer-sponsored health insurance. This tax treatment costs a great deal and is both distortionary and inequitable (Steuerle and Hoffman, 1979; Selden and Moeller, 2000). Since no other constituency champions withdrawing this tax exemption, economists have contributed by making its problems clear. Nonetheless, empirical estimates of the effect of withdrawing the favorable tax treatment suggest that withdrawing the exemption would reduce spending by only about 5–15 percent, with at least half of this reduction occurring through a drop in the number of people with any coverage, not a reduction in the generosity of coverage among the insured (Chernick, Holmer, and Weinberg, 1987). These estimates are strikingly modest in comparison with the estimated welfare losses associated with moral hazard described above. This is puzzling, since private health insurance is voluntarily purchased and the tax subsidy is the principal incentive for inefficient purchasing decisions.

The standard theory has been important, but recent developments in health economics have tended to suggest modifications of it. The principal reason for interest in modifying theory is that it fails to capture many of the key stylized facts about American health care and therefore provides insufficient guidance for health care policy.

Consider four of the most significant issues for policy makers. First, some 40 million Americans do not have health insurance coverage. Why do so many people lack coverage? Standard theory points to adverse selection as the market failure that might lead to this outcome. Many of the uninsured (about half) are, indeed, young, or are older but in excellent health. But nearly half are neither young nor very healthy. Many uninsured people lack coverage because they have health conditions that insurers refuse to cover or will cover only at very high prices. According to public opinion polls and legislative actions, this group represents the strongest motivation for intervention in the insurance market. Concern over these groups explains the existence of the largest public health insurance programs. Medicare covers the elderly and permanently disabled, and Medicaid, while often thought of as a program covering welfare-eligible women and children, paid 71 percent of its budget for care for the disabled and elderly in 1998 (Health Care Financing Administration, 2002). The standard theory of adverse selection provides no explanation for why there might be a market failure demanding government intervention on behalf of these groups.

Second, the health insurance market has shifted away from cost sharing and toward a set of alternative strategies for controlling moral hazard, known as managed care. Between 1970 and 2000, the share of health care expenditures paid out-of-pocket fell from 34 percent to 15 percent (http://www.cms.hhs.gov/statistics/nhe/). By contrast, between 1985 and
1993, the fraction of privately insured Americans with managed care style insurance contracts increased from 20 percent to 70 percent (Glied, 2000). Managed care plans use three tools—“micromanagement” of service use, such as utilization management and protocols, where the insurer directly monitors treatment decisions; financial incentives for providers, such as capitation payment and bonuses; and restricted panels of providers—to control health care spending. Although these alternative methods of reducing medical care expenditures are consistent with efforts to reduce moral hazard, standard theory does not explain why they appear to be preferred. The theory of moral hazard suggests that people would be better off if they faced greater financial risk, but the market seems intent on giving us less.

Third, there is a widely documented and substantial prevalence of “flat of the curve” medicine, care for which the expected benefit is zero (Enthoven, 1980, pp. 45–51). In many cases, flat–of–the–curve medicine occurs because neither patients nor providers have the information needed to make optimal health care decisions. However, in other cases, the information needed to eliminate this care exists. Despite the substantial gains that might result from getting rid of flat–of–the–curve medicine, it appears to be very difficult to do so.

Finally, concerns about equity in health care never seem to be resolved. Unlike other apparent necessities, such as food, where the level of redistribution can be fixed at one point in time—the food stamps allocation was fixed in 1965—the necessary level of redistribution is constantly at issue (Glied, 1997). For example, while the Medicare program has existed since the mid–1960s, policy makers are again revisiting the parameters of coverage in the current debate over inclusion of prescription drugs.

New theoretical developments, extensions of the existing framework, aim to provide economic explanations for these public policy problems and to focus policy proposals in new directions. Understanding these developments is important not only for health economists, but for all public sector economists. Medicare and Medicaid alone represent 18 percent of Federal government spending (Centers for Medicare and Medicaid Studies, 2002a). Moreover, health care is a topic that is perennially on the minds of politicians and policy–makers.

MEDICAL COMPLEXITY AND MORAL HAZARD

Unlike other insurance contracts, which transfer money in case of an adverse event, health insurance operates by reducing the price paid for medical care in the event that it is needed. Given this structure, the term moral hazard in health care refers to “overconsumption” of medical services “because insurance pays for some or all of the cost” (Rosen, 1995, p. 219).

This reduction in price increases the quantity of health care consumed (Figure 1). $Q_{FC}$ is the quantity of medical care chosen when consumers must pay the full cost. $Q_{INS}$ is the larger quantity of medical care chosen under insurance where consumers pay only a fraction of the cost of care at the point of service. The shaded triangle represents the overconsumption of medical care that occurs because the price has fallen. The price elasticity of demand for health care and the generosity of insurance coverage determine the extent of overconsumption.

In computing the welfare losses due to health insurance, this overconsumption is separately weighed against the reduction in financial risk associated with insurance coverage. Thus, the value of this reduction in financial risk depends on the extent of risk aversion. Absent risk aversion, according to the standard theory, social welfare would be maximized if people paid the full cost of medical care out of
With risk aversion, “the best that can be done . . . is find a happy compromise with some risk–spreading and some incentive” (Zeckhauser, 1970).

The policy prescriptions generated by the standard framework emphasize the importance of reducing moral hazard through increases in cost–sharing, such as the use of medical savings accounts coupled with catastrophic insurance. Yet very few people, even in the individual market (where coverage is not distorted by the tax treatment of insurance), choose policies of this type. New theoretical approaches to the demand for health insurance help to explain why people seem to prefer managed care over cost–sharing for addressing this moral hazard problem. They also suggest that the welfare losses associated with overconsumption in health insurance may be smaller than previously estimated. These theories emphasize the difference in the value of medical care across health states and the informational problems of differentiating among states.

The standard framework implicitly assumes that the marginal utility of medical spending is independent of the state of health. Since it does not allow the demand for medical care to be state–dependent, this formulation omits the value of insurance in enabling people to buy care that, without insurance, they could not afford given their lifetime wealth (Nyman, 1999a, forthcoming).

An alternative way to represent the demand for health insurance is through a complete contingent contract framework, a general equilibrium treatment that allows for both the value and the harm of insurance in a single welfare analysis. As Marshall (1976) noted, this treatment avoids the “difficulty of estimating state–dependent compensated price elasticities.” Note that while the complete contingent claims contract approach is convenient, the same results can be obtained from more complete formulations of the standard treatment of moral hazard that incorporate health–state–dependent compensated demand curves and large income transfers across states (Nyman, 1999b, 2001, forthcoming; Blomqvist, 2001a, 2001b; Manning and Marquis, 1993, 2001).

Consider a model to illustrate the complete contingent claims approach. There are three periods. In period one, everyone is identical. In period two, uncertainty is

---

2 See Remler (1994a) for an accessible treatment of the complete contingent contract for health insurance and medical care markets.
realized and everyone discovers his state of health. The effectiveness of medical treatment depends on the particular state of health. In period three, medical treatment is provided.

Medical care insurance is desirable in this world both because some forms of medical care are very expensive and because the demand for care varies by state. Medical insurance can be desirable even if individuals are not risk averse at all. It can even be optimal to redistribute large sums—possibly more than one’s lifetime income—into rarely occurring states, where the value of medical care is very high. For example, consider a state in which the only way to save your life is through a $1,000,000 kidney transplant, but the kidney transplant is almost sure to be successful. If this state occurs only very rarely, it may be worthwhile to purchase actuarially fair coverage for the transplant even if one’s lifetime income is less than $1,000,000 (Nyman, 1999a, forthcoming). This coverage is desirable only because utility depends on general health level and the effectiveness of medical care in improving health is state–dependent. Some recent empirical research suggests, consistent with this theory, that the demand for medical care is, indeed, health state dependent (Remler and Atherley, forthcoming).

In these circumstances, the first–best solution is a complete contingent contract that specifies the treatment provided depending on the state of nature realized. In this contract, the marginal utility of general consumption and medical spending—which depends on health state—are equated across all states of nature. The marginal utility of medical spending depends on the value of medical care in each state.

Under this contract, there can be no overconsumption of medical care. The quantity of medical care specified for a given state in the contract as written ex ante—and incorporated in the ex ante premium—is, by definition, optimal.

The complete contingent contract could be implemented through indemnity insurance that transferred income into each realized disease state. This contract would not lower medical care prices at the point of service. With the appropriate income transfer, each person could purchase the quantity of medical care he deemed ex ante to be optimal in that state.

It is not practically possible to write and enforce this optimal complete contingent contract. A tremendous amount of clinical detail is required to specify what medical care merits coverage in each state. Consider the example of cardiac catheterization, an imaging technique to see blood vessel blockages. Cardiac catheterization is often valuable and most of us would willingly pay higher premiums to include it under our contract. But would we pay an extra $1,000/year to ensure cardiac catheterization coverage for any kind of chest pain in any situation? Would we pay anything to ensure coverage in states where we are terminally ill with an unrelated disease? Spelling out all potential cases is impossible. Furthermore, the parameters of this contract would change constantly as technology changes, new techniques are invented, and medicine gets better at the old techniques.

The sensitivity of the value of medical care to clinical detail makes the size of the income transfers needed for the optimal contingent claims contract excessively sensitive to clinical detail. Reducing the price of medical care in the event of illness is a second–best solution to this contracting problem.

Cost sharing, in its standard form, provides a very weak facsimile of the optimal complete contingent claims contract. One difficulty in using cost sharing to approximate the optimal contract is that most health care spending is concentrated among a few particularly expensive individuals, many of whom are likely to be in states with a high value of medical care.
As the standard framework points out, the financial protection offered by insurance is eroded if such individuals face substantial cost sharing. In addition, if the value of care is state-dependent, high cost-sharing means that some sick people will not purchase care that they would have identified as highly valuable \textit{ex ante}. However, if those with the greatest expenditures do not face substantial cost-sharing, their expenditures will not be controlled.

As the cost of care rises relative to income, the magnitude of the difference between the optimal contingent claims contract and the optimal cost-sharing contract grows. Marshall, writing in 1976, suggested that “better incentives can also be produced by using contracts which more nearly specify states of nature.” One way to improve the contract would be to make cost-sharing levels state-dependent (Zeckhauser, 1970). Many insurance contracts now incorporate variable cost sharing through, for example, tiered cost sharing for pharmaceutical drugs. The informational limits that prohibit implementation of the complete contingent contract restrict the extent of variation in copayments that can be incorporated into contracts.

Some forms of managed care also move in the direction of the complete contingent contract by making medical care treatment decisions sensitive to particular patient characteristics (Remler, 1994a, 1994b). For example, clinical guidelines and utilization review of specific treatment decisions can allow contracts to match care to states of nature.

The complete contingent claims contract framework suggests a redefinition of welfare losses due to moral hazard in health care. The standard framework splits the decision to consume medical care from the analysis of the benefits of reductions in financial risk, and so it does not yield a clear measure of what level of medical care is optimal and not overconsumption. In the complete contingent contract framework, all care that would be chosen \textit{ex ante} is efficient. Only care delivered because of the inability of the insurer to prevent patients from consuming medical care that they would not have chosen \textit{ex ante} represents moral hazard induced inefficiency. If the demand for health care is sensitive to health state, estimates of the welfare losses due to health insurance that do not take state-dependent utility into account will be too high.

\section*{Patient–Provider Informational Asymmetries}

Even without insurance, medical care suffers from a standard principal–agent (i.e., physician–patient) informational asymmetry. Because doctors know more than patients do, they can, in principle, manipulate patterns of care to take best advantage of payment arrangements. They can do this by generating a demand for too much care (induced demand) or shirking (underservice). They can also manipulate the provision of care, for example, by treating a condition with more short visits rather than few long visits to maximize payment.

Patients’ difficulty in assessing the quality and value of the care they receive may contribute to the prevalence of “flat-of-the-curve” medicine, which delivers very little or no benefit at the margin (Cutler, 2000). It may also be a factor in generating the substantial variations in patterns of delivery of medical care in different regions of the country (Wennberg and Gittelsohn, 1973; Phelps, 2000). Information asymmetries, flat-of-the-curve medicine, and variations in care are not necessarily policy problems. Physicians will always know more than patients. Information problems in medicine go well beyond asymmetries. Physicians, too, lack information. Much of the variation in care across regions can be attributed to general uncertainty about treatment options (Phelps, 2000). Often physicians cannot observe patient compliance with their
directives (Leonard and Graff Zivin, 2002). Thus, the relevant issue for policy is whether some contracting arrangements generate more opportunities for manipulation of quantity, quality, and price than do others.

Until recently, most of the theoretical literature on provider payment focused on problems of overservice. An extensive literature proposed theoretical explanations of the phenomenon of physician-induced demand (see McGuire, 2000 for a summary). A similar literature, examining the hospital market, focused on the so-called “medical care arms race” wherein hospitals competed for patients by investing in ever more costly technologies (see Dranove and Satterthwaite, 2000 for a summary). In both of these markets, theory tied increases in utilization to insurance, which insulated patients from the cost of their care and led them to sacrifice lower costs in favor of higher quality, combined with fee-for-service payment, which gave providers incentives to encourage overutilization. Empirical research, however, suggests that altering consumer incentives through cost sharing has little effect on the value of medical care services purchased. While cost-sharing affects overall levels of service use, it cuts out demonstrably good care just as much as it cuts out care of ambiguous value (Newhouse and the Insurance Experiment Group, 1993).

With the empirical rejection of the hypothesis that flat-of-the-curve medicine is principally a response to insufficient cost sharing in health insurance, recent theoretical research has focused, instead, on the provider payment side of the equation. The informational advantage of physicians over patients makes the agency problem of ensuring good effort by physicians a central concern, whether or not patients use insurance to pay their bills (Gaynor, 1994; Leonard and Graff Zivin, 2002; McGuire, 2000; Pauly, 1980).

New models of physician payment incorporate unobservable (or noncontractible) physician effort into traditional models of the demand for medical services in the presence of health insurance. Using these models, theorists have examined the implications of a variety of alternatives to fee-for-service payment (Ma and McGuire, 1997; Ellis and McGuire, 1986, 1993; Selden, 1990; Blomqvist, 1991). These alternative payment methods are often collectively referred to as instances of “supply-side cost-sharing” and they share the characteristic that they put physicians at some financial risk for the costs incurred by their patients.

The results of these models suggest that under both supply-side cost-sharing payment methods and traditional fee-for-service, the inability to contract on physician effort, especially when combined with the existence of health insurance that generates overconsumption, results in a third-best equilibrium (Ma and McGuire, 1997). While providers paid on a fee-for-service basis have incentives to provide too much care, under supply-side cost sharing, providers have incentives to shirk on the services they provide and to select healthy patients rather than sicker ones.

The use of supply-side cost sharing has proliferated in the health care market over the past two decades. Commonly observed examples of supply-side cost sharing include prospective payment of hospitals and capitation payment of physicians. Since 1984, the Federal government has paid for Medicare hospital stays prospectively, using a system of Diagnostic Related Groups (DRGs). Under DRG payment hospitals receive a single payment that depends on the diagnosis of the patient, rather than additional payment for each day in the hospital and each service provided. The payment is independent of the cost of treating the specific patient, leaving the hospital at financial risk for actual treatment costs.

Under capitation, physicians are paid, a “per member per month” payment irrespective of the amount of care, if any, they
provide to patients (Remler et. al., 1997). In some cases, the scope of services included under the capitation payment includes referrals and hospitalizations. In such cases, physicians are at financial risk both for their own services and for costs incurred by other providers to whom they refer their patients.

Theory suggests that increasing the share of expenses for which providers are responsible leads to greater incentives for cost containment but also increases incentives for selection and underservice. The optimal form of insurance is therefore a blend of demand-side and supply-side cost sharing (Selden, 1990; Newhouse, 1996). Consistent with these findings, payment systems often incorporate limits on provider risk-bearing. For example, hospitals receive “outlier payments” for particularly costly Medicare hospital patients.

Empirical evidence shows that changing financial incentives for providers alters patterns of care. However, there is little evidence to show that changing financial incentives affects the prevalence of flat-of-the-curve medical care. Instead, regulation of the appropriateness and quality of medical care appears to rely heavily on non-financial methods, including utilization management and the dissemination and enforcement of guidelines for physician and hospital practice.

**SELECTION**

New theoretical research on adverse selection has centered on the question of why getting health insurance coverage to less healthy people is a public policy priority. The first step in understanding this phenomenon is developing a theory that explains why people in poor health might find it difficult to obtain health insurance coverage, even at high prices. According to the most recent Current Population Survey (2001), 15.2 million of the 38.7 million uninsured report that their health is not “excellent” or “very good”. Standard economic theory suggests that pooling equilibria cannot be sustained in health insurance markets (Rothschild and Stiglitz, 1976). A competitive insurer can always attract the healthiest share of a heterogeneous pool by offering a slightly narrower coverage package at a lower cost. Similarly, a competitive insurer can offer coverage to the least healthy customers at very high prices. In fact, however, populations covered by health insurance contracts (even in the individual market), are quite heterogeneous. Furthermore, rather than charging high prices for high-risk customers, insurers often deny them coverage altogether. One way to explain this pattern of facts is to incorporate transactions costs into the standard Rothschild-Stiglitz model (Newhouse, 1996).

Assume that there are transactions costs associated with writing insurance contracts. If these costs are large enough, competing insurers will not find it profitable to write a new contract unless it will attract a sufficiently large share of the population. There will be comparatively few contracts and choice will be limited. If costs of writing contracts that correspond to the health risks of sick customers are high enough, no insurer will find it profitable to develop a contract that includes these customers and will screen them out directly, rather than through contract design. In this way, transactions costs can generate a market failure in the health insurance market where high-risk customers obtain inefficiently little coverage.

The modified selection model provides a theoretical explanation for why coverage denials may be evidence of a market failure. However, public policymakers also seem to have an additional concern that people at high risk can obtain coverage only at very high prices. Nor is this exclusively a concern about making sure...
that poor people have access to coverage. For example, in 46 states, regulators now impose rate bands that limit the spread between the most and least costly health insurance contracts in the small firm insurance market. Rate bands have also been imposed in the individual market in 16 states (Chollet, Kirk, and Simon, 2000). These bands are typically combined with guaranteed issue laws that compel insurers to provide coverage to anyone willing to pay the premium. Empirical research on these market reforms suggests that they have not affected the total number of people covered by insurance, although some studies find that they alter the composition of the insured population toward less healthy people (Sloan and Conover, 1998; Zuckerman and Rajan, 1999; Marquis and Long, 2002). These market regulations limit the prices paid by unhealthy high–income people as well as unhealthy low–income people.

Why are policy makers concerned about the premiums paid by people in poor health? One explanation for the introduction of rate bands is that they protect even currently healthy people from health insurance premium increases they would otherwise face in the event that their health later deteriorates. Risk–averse people would prefer to purchase protection against both the risk that they will experience a negative health event in the coming year and also against the risk that they will develop a chronic health condition that will permanently raise their health care costs (Diamond, 1992). Existing health insurance contracts, however, are annual. An event that permanently raises health care costs will permanently raise health insurance premiums. Rate bands prevent such increases—albeit in a rather crude fashion. A better solution would be a long–term health insurance contract.

The lack of long–term contracts in health insurance provides an efficiency argument for some form of intervention in markets. Several recent papers have posited how long–term health insurance contracts might look. One way that such contracts could be structured is by requiring guaranteed renewability in health insurance (Pauly, Kunreuther, and Hirth, 1995). This approach has two defects. Guaranteed renewability can generate adverse selection spirals if only the least healthy customers continually renew their contracts. To avoid this result, contracts may need to have very high initial prices. Even if this can be accomplished, renewability merely ensures that insured people can keep the contracts they now have, but it does not permit them to select new contracts if they are dissatisfied with their current option.

An alternative approach would offer protection against changes in actuarially fair premiums due to changes in health status, by incorporating an option–value payment into a health insurance contract (Cochrane, 1995). Each year health insurers would make a payment to subscribers equal to the discounted value of the change over that year in the predicted cost of their future health care. Subscribers who became sicker would receive lump sum payments; those who had become healthier would be required to make payments to the insurer. To avoid having healthy consumers renege on their contracts, for example via bankruptcy, this model envisions accounts into which consumers would make payments at the beginning of each year. This approach would obviate the need for government intervention in the health insurance market.

Since such contracts would solve the concerns of policy–makers to sustain the insurance coverage of those who become ill, a natural question is why they do not exist today. One reason may relate to the transactions problem described by Newhouse. A system of lump–sum payments requires an agreed–upon method of determining the future costs associated with any given health condition. Over the
past decade, the development of such sys-
tems of lump sum payments has been a central aspect of health services research-
ers, albeit based on annual, not lifetime, costs, and for use in a somewhat different context.

Efforts to generate competition between health plans within employment contexts or competition between managed care options and the traditional fee–for–service option in Medicare have, as described above, often floundered because of adverse selection. Health services researchers, drawing on medical knowledge and statistical methods, have attempted to develop systems of “risk adjustment” that could be used to adjust annual premiums for the composition of the risk pool and eliminate selection (see, for example, Shen and Ellis, 2002). Medicare now incorporates a form of risk adjustment in its payment to health plans. There is, however, considerable skepticism about the success of these systems. The problems of developing annual risk–adjusters suggest that it might be very difficult to develop lifetime risk adjusters that would be acceptable to most health insurers. If there is no agreement on the risk–assessment methodology, lump sum payments provided by one plan may not correspond to those demanded by other plans.

The problem of developing adequate risk adjusters is compounded by the pace of technological change in medical care. In the context of complete contingent claims contracts, a technological advance that improves the quality of care for a given condition would lead people to wish to transfer more resources to that condition state. Lump–sum payments made today may not cover the cost of optimally chosen care if technological change leads to such quality improvements in the future. This problem can be addressed through annual increases in all premiums if the costs of technological change are fully diversifiable. To the extent that future technological change affects an entire cohort, its costs may not be intragenerationally diversifiable.

While employer–sponsored health insurance and pay–as–you–go financing for Medicare present other problems, their persistence can be explained, in part, by this combination of difficulty in risk adjustment and non–diversifiable technological change. Employer–sponsored health insurance creates groups that are not based on health status, reducing potential problems of adverse selection at each point in time. In large group insurance, employer sponsorship also generates a form of long–term health insurance contract (Moran, 1997). People with employer–sponsored coverage whose health insurance costs rise (because the health status of someone in their family deteriorates) do not themselves incur the increased cost (Newhouse, 1996). Some evidence of this protection against increased costs, which is limited to people who remain in their current job, is provided by the literature on “job lock” (Madrian, 1994). Unlike hypothetical long–term contracts, however, employer–sponsored contracts are renegotiated annually. Thus, the costs associated with technological change do not need to be assessed in advance. While the tax treatment of employer–sponsored coverage certainly encourages this organizational form, recent empirical research suggests that variations in marginal tax rates generate only small changes in the propensity to hold employer–sponsored coverage among those employed in large firms (Gruber and Lettau, 2000; Stabile, 2002; Finkelstein, 2002). Instead, employer–sponsored coverage may be an efficient institutional response to short–term and long–term problems in the health insurance market.

The possibility of cost–altering technological change also creates a problem for efforts to fully pre–fund the Medicare program. Pre–funding would require that today’s working age population purchase insurance against the possibility of such
technological change. To the extent that the technological change affects the entire cohort (as has been true in the past), it is not clear how the risk could be diversified within that cohort. Under a partially pay–as–you–go system, younger generations “insure” older generations against the risk of technological progress.

Even absent technical problems in risk adjustment and questions about the diversifiability of technological risk, accidents of birth—health conditions that originate prenatally and persist permanently—might still yield a case for government intervention. Better understanding of the human genome will vastly increase the number of such conditions. If genetic conditions can be assessed at (or before) birth, the scope of conditions that can be covered in competitive health insurance markets may become very narrow.

EQUITY CONCERNS IN HEALTH CARE AND THE GROWTH OF MEDICAL TECHNOLOGY

Any careful examination of the government role in the US health care system would be forced to conclude that market failures are insufficient to explain the significant and enduring roles of local, state, and federal governments in funding medical care services. In 2000, the public sector (excluding tax expenditures) was responsible for 45 percent of US health care costs (Health Care Financing Administration, 2002). Most of this funding addresses redistributive, not efficiency, concerns.

One unusual aspect of redistributive policy in health care is that, in contrast to food stamps or welfare programs, public health insurance programs do not redistribute based on a pre-established and unchanging need standard (Glied, 1997). Rather, public health insurance programs continually expand the array of medical treatments and technologies available to their beneficiaries as the menu of choices available to the privately insured expands. Public policy appears to seek to maintain a measure of relative equity rather than providing a fixed absolute level of health to all. Earlier analysis provided formal descriptions of the implications of such a relativistic redistributive policy (Lindsay, 1969). Lindsay concludes that optimal policy in the face of a relative standard combines subsidies to the poor with constraints on the rich. The literature on equity in health care lay dormant for some 30 years, but a new economics literature on the value of recent improvements in medical technology is likely to rekindle interest in concerns over redistribution and relative equity.

At each point in time, allocations of medical care can be made by equating marginal cost and marginal value. A particular procedure or drug may provide enormous health benefits to a small subset of the population with a defined set of health conditions, but as it diffuses into a broader population, whose problems are more severe, or less severe, or simply a little different, its effectiveness typically diminishes. In a well–functioning market, only people who place a very high value on small improvements in health would purchase these treatments. Public policy might then ensure that everyone had access to the limited set of services that generated significant health improvements. This type of rationing–by–need system may be very difficult to implement in practice, but it is conceptually clear.

Problems arise, however, when technological improvements shift the health care production possibilities frontier outward. A series of recent papers, stemming from the work of the Boskin Commission on revisions to the consumer price index, has examined the value of recent innovations in medical technology. The consensus across these papers is that new medical technologies have generated improvements in average health that outweigh the
new expenditures they generate. This conclusion can be stated in several ways. Some papers show that the quality–adjusted price of medical treatment for certain widespread conditions has been declining (Cutler, McClellan, Newhouse, and Remler, 1998; Frank, Berndt, and Busch, 1999). Others argue that the cost of specific new medical technologies is less than their value measured in terms of quality–adjusted lives saved (Cutler and Richardson, 1999; Cutler and McClellan, 2001). A third group finds that the aggregate value of improvements in medical care exceeds their cost (Murphy and Topel, 2002; Nordhaus, 2002).

Two aspects of this line of research draw attention to the problem of maintaining equity in medical care. First, these results indicate that the medical care price index significantly overstates the cost of maintaining a stable medical care standard of living. Thus, if per capita spending in public health insurance programs increases at the rate measured by the medical care component of the consumer price index, as has been the pattern historically, it would generate ever–increasing standards of living for the beneficiary population. By contrast, if per capita spending in public health insurance programs increases only at the rate implied by a standard–of–living index, beneficiary populations would lag behind those purchasing coverage on their own. How should society make this choice?

Second, these estimates of the value of technological improvements in medical care are all based on an estimate of the value of life calculated from average willingness–to–pay. The value of life is sometimes adjusted to reflect differences in the quality of life caused by illness and disability. The results imply that the average person has been made better off by exchanging money for new medical care innovations at prevailing prices. As with all willingness–to–pay measures, however, estimates of value are likely to vary with income and other characteristics. Some people will have lower valuations; others would be willing to pay a great deal more. Again, policymakers must decide whether to include a technology whose benefits exceed average willingness to pay.

Maintaining equity over time creates difficulties with respect to existing public insurance programs. As the size of the population covered by public insurance swells with the aging of the baby boom, this problem is likely to draw even more attention. To the extent that technological developments exacerbate problems of adverse selection and lead to a broadening of the scope of public insurance, or public insurance regulation, equity concerns are likely to become yet more important.

CONCLUSIONS

Recent developments in health economics theory shed new light on important policy problems, help to explain several persistent anomalies, and provide a rationale for novel policy approaches. Theories of the demand for health insurance that incorporate state–dependent utility help explain why health insurance contracts increasingly rely on cost–containment methods other than pure demand–side cost sharing. Cost–sharing alone is not an efficient way to ration the quantity of care provided to people with very costly–to–treat health conditions. A single cost–sharing parameter is insufficient to ensure both that wasteful care is not consumed and simultaneously provide people with as much valuable care as they desire. Instead, contracts that combine multiple rationing methods, including cost sharing, supplier incentives, and utilization management may come closer to the optimal complete contingent contract.

Theories of provider payment show how different types of payment mechanisms can generate different quantities,
qualities, and prices of care. The possibility of influencing patterns of care through altering payment methods has contributed to the proliferation of payment methods now in use for physician and hospital payment. Payment methods that place health care providers at significant financial risk for patient service utilization can generate strong incentives for selection of healthier patients. They can also induce providers to skim on the provision of treatment. This latter possibility has led, for example, to regulations in the Medicare program that regulate the mechanisms participating health insurance plans may use to pay providers.

Adding transactions costs to a standard Rothschild–Stiglitz framework generates the result that, consistent with empirical fact and policy concern, high-risk people, not low-risk people, are underinsured in the health insurance market. The absence of long-term health insurance contracts means that not only the underinsurance of high-risk people, but also the high prices that these people pay when they can obtain coverage, are not efficient. Some market-based approaches to long-term contracting have been suggested, but the difficulty of making appropriate side-payments, especially in the presence of uncertainty about future technological progress, makes these contracts difficult to write. These problems help explain the persistence of employer-based private health insurance and pay-as-you-go financed public insurance.

Finally, society’s concern about relative equity in health care means that technological advances in health care generate new public policy problems. As more new and expensive treatments are developed that many people would willingly pay for, how are we to determine what we are willing to guarantee to everyone in society? We must decide whether to limit the amount of medical care available to all (a component of the health care systems of many other nations), tolerate growing inequality in health care, or accept the burden of ever greater redistributional expenditures to cover ever improving medical care. These questions are normative, not positive, but economic theory can be useful in describing the costs and consequences of each option.

One important feature of health economics is that all the problems described above, overconsumption, informational asymmetries, selection, and redistributional concerns, occur at the same time. This confluence of problems has led to an interest in developing policies that address multiple objectives at once. Prominent among these, Enthoven (1980) proposed a system of managed competition. In this model, plans would compete ex ante for consumers on the basis of both price and quality. A system of risk-adjustment would control selection. Enthoven’s model, substantially modified to incorporate redistributional concerns, became the theoretical basis of the ill-fated Clinton health reform plan. The Clinton plan floundered for many reasons, but one substantial difficulty in formulating it was—as the discussion above suggests—the problem of grafting a redistributional mechanism onto an efficiently configured market in the context of rapid, cost-increasing, technological change (Glied, 1997).

Unfortunately, achieving the optimal distribution of medical care services appears to be an extremely complex problem that no one yet has solved. Payment arrangements between insurers and providers are myriad and endlessly changing; the appeal of cost containment strategies, such as utilization management, go through regular cycles of enchantment and disappointment; medical technology continually improves; and health care remains an enduringly potent political issue. Despite our recent advances, there is much we still do not understand about the economics of the health care market. Public finance econo-
mists will find this a fruitful field for further theoretical and empirical research.

Acknowledgments

We would like to thank Joshua Graff Zivin, Therese McGuire, and John Nyman for very helpful comments on earlier drafts.

REFERENCES

Blomqvist, Ake.

Blomqvist, Ake.

Blomqvist, Ake.


Buchmueller, Thomas, and DiNardo, John.

Chernick, Howard, Martin Holmer, and Daniel Weinberg.

Chollet, Deborah J., Adele M. Kirk, Kosali I. Simon.

Christensen, Sandra.

Cochrane, John H.

Cutler, David.

Cutler, David, and Mark McClellan.
“Is Technological Change in Medical Care Worth It?” *Health Affairs* 20 No. 5 (September/October, 2001): 11–29.

Cutler, David, Mark McClellan, Joseph P. Newhouse, and Dahlia Remler.

Cutler, David, and Elizabeth Richardson.

De Meza, David.

Diamond, Peter.

Dranove, David, and Mark A. Satterthwaite.

Ellis, Randall P., and Thomas G. McGuire.
Ellis, Randall P., and Thomas G. McGuire.

Enthoven, Alain.

Feldman, Roger, and Bryan Dowd.

Finkelstein, Amy.

Frank, Richard G., Ernst R. Berndt, and Susan H. Busch.

Gaynor, Martin.

Glied, Sherry.

Glied, Sherry A.

Glied, Sherry, and Joshua Graff Zivin.
“How do Physicians Behave When Some (but Not All) of their Patients are in Managed Care?” *Journal of Health Economics* 21 No. 2 (March, 2002): 337–53.

Gruber, Jonathan, and Michael Lettau.

Health Care Financing Administration.

Hillman Alan L., Mark V. Pauly, and Joseph J. Kerstein.

Keeler, Emmet B.

Leonard, Kenneth, and Joshua Graff Zivin.

Lindsay, Cotton M.

Ma, Ching–to Albert, and Thomas G. McGuire.

Madrian, Brigitte C.


Marshall, John M.

McGuire, Therese G.
Moran, John R.

Murphy, Kevin M., and Robert H. Topel.

Newhouse, Joseph P.

Newhouse, Joseph P.
“Why is There a Quality Chasm?” Health Affairs 21 No. 4 (July/August, 2002): 13–25.

Newhouse, Joseph P. and the Insurance Experiment Group.

Nordhaus, William D.

Nyman, John A.

Nyman, John A.

Nyman, John A.

Nyman, John A.

Pauly, Mark V.

Pauly, Mark V., Howard Kunreuther, and Richard Hirth.

Phelps, Charles E.

Price, James, and James Mays.

Remler, Dahlia K.

Remler, Dahlia K.

Remler, Dahlia, Karen Donelan, Robert Blendon, George D. Lundberg, Lucian L. Leape, David R. Calkins, Katherine Binns, and Joseph P. Newhouse.
“What Do Managed Care Plans do to Affect Care? Results from a Survey of Physicians.” Inquiry 34 No. 3 (Fall, 1997): 196–204.

Remler, Dahlia K., and Adam J. Atherly.

Rosen, Harvey.

Rosett, Richard, and Lien–Fu Huang.


