Economic modeling of methods to stimulate quality improvement

KAREN EGGLESTON
Department of Economics, Tufts University, Medford, MA, USA

Abstract

Objective. This paper uses an economic model to compare three methods for stimulating quality improvement: payment incentives, competition for patients, and emphasis on professional ethics.

Design. Use an economic model to simulate the impact on quality distortions (risk selection) of differences in payment incentives, competition for patients, and emphasis on professional ethics.

Setting. Health care policymakers in many countries seek to use incentives and competition to spur quality improvement. However, strong incentives often promote risk selection: insurers and providers financially benefit from distorting quality to attract profitable patients.

Results. The analysis suggests that intense competition for patients and strong financial rewards for cost control can exacerbate quality distortions and compromise social solidarity.

Conclusions. Carefully regulated competition and mixed forms of provider payment (risk sharing) appear to be the best options. Moreover, designing competition, regulation, payment, and other forms of health policy to promote suppliers’ professional ethics can help society to reap the quality and efficiency benefits of competition and incentives without sacrificing social solidarity.

Keywords: altruism, competition, payment incentives, provider ethics, quality distortions, quality improvement, risk selection

Policymakers in many countries seek to improve health care quality and efficiency with market forces. Countries introducing competitive reforms include, for example, Belgium, China, Colombia, the Czech Republic, Germany, Ireland, Israel, The Netherlands, Poland, Russia, and Switzerland [1–3]. Policy debate in the United States also focuses on a stronger role for competition, particularly in the Medicare program, and better alignment of incentives with quality goals, including pay-for-performance [4]. Cutler [5] argues that promoting incentives and competition constitutes a current ‘third wave’ of international health policy reform.

Most nations also recognize, however, the many limitations of free-market competition and high-powered financial incentives in the health sector. These limitations arise predominately because many important dimensions of quality health care are difficult to observe, monitor and motivate. Provider competition to attract consumers can spur providers to raise quality along dimensions that consumers value. But consumers frequently lack sufficient information and market power to be discriminating purchasers of high quality, low cost services. Competition and payment incentives featuring payment risk or ‘supply-side cost sharing’ [6] may give even well-intentioned health plans and providers incentive for socially undesirable behaviors. One such behavior is manipulation of offerings to deter the sick and attract the healthy, variously known as ‘risk selection’, ‘plan manipulation’, ‘cream skimming’, or ‘cherry picking’ [7,8]. For example, in European countries that allow consumers to choose among sickness funds, designing supplementary insurance to attract preferred risks appears to be a widespread tool for risk selection [3].

When insurers compete to avoid the sick rather than provide quality care, the most vulnerable patients may experience access problems, and consumers may be unable to buy insurance against becoming a bad risk in the future or having a child who is a bad risk [7]. Insurers lack incentive to be responsive to the chronically ill or other high-cost patients.

This article uses simulation of an economic model of insurer and provider behavior to compare three methods for stimulating quality improvement: payment incentives, competition for patients and provider altruism or emphasis on professional ethics. The decision maker in the model is an insurer, such as a European sickness fund or US health plan. The conclusions apply equally to insurers that are also integrated with providers ('managed care plans') and, with some modifications, to provider organizations such as hospitals that are paid at least partially prospectively (i.e., are not reimbursed dollar-for-dollar for all services).

Address reprint requests to Karen Eggleston, Tufts University Department of Economics, Braker Hall Medford, MA 02155, USA.
E-mail: karen.eggleston@tufts.edu
The simulations illustrate how forms of payment with strong incentives to control cost, such as capitation, penalize investment in quality for unprofitable patients (and reward quality investments that attract profitable patients). Similarly, competition for patients often increases responsiveness but also can exacerbate quality distortions, especially when insurers or providers are paid by capitation. Carefully regulated competition and mixed forms of provider payment emerge as better policy options. Furthermore, an insurer or provider’s innate concern for patient welfare, arising through professional norms or otherwise, can reduce selection-motivated quality distortions, albeit sometimes at the expense of cost control. Designing competition, regulation, payment, and other forms of health policy to promote suppliers’ professional ethics can thus help society to reap the quality benefits of competition and incentives without sacrificing social solidarity.

The first section below describes the conceptual model. The following sections discuss how payment incentives, competition, and altruism affect insurer and provider behavior, and thus the quality, equity, and efficiency of healthcare delivery.

The model

The economic model of behavior has the following structure. A healthcare purchaser, such as a government program or employer, contracts with multiple insurers. Consumers—the government program’s beneficiaries, or the employer’s employees—choose from among the competing insurers. Each insurer must cover a broad package of healthcare services but also has some discretion over the quality of care. The model focuses on how financial incentives and competition for consumers influence the insurer’s decisions about quality improvements for various services.

Compared with a standard economic model of a multiproduct firm, the economic model developed here incorporates several features specific to healthcare. First, demand may consistently exceed supply. The model focuses on how the insurer rations services to patients. Few countries allow life-saving medical services to be rationed exclusively by ability to pay. Because health is so important, or because insured consumers tend to want ‘everything someone else will pay for’, consumers’ demand, given regulated prices, often exceeds current supply. Hence, patients must wait for available appointments with specialists, or for elective surgery. A natural example is an industrialized country that provides near-universal coverage with moderate user fees. The logic of the argument continues to apply, and indeed with greater force, when some patients are uninsured. For example, an extension of the model describes how providers in China have incentive to overprovide services attractive to the insured minority [9].

Second, the supply side may involve multiple actors: insurers, provider organizations, individual clinicians. Insurers do not directly choose what to provide to each patient; clinicians do. The insurer influences clinical decisions indirectly, by for example deciding how large a budget to make available for each service.

Third, the model is tailored to healthcare through its assumptions about how providers prioritize patient access to limited capacity—such as appointments for magnetic resonance imaging (MRI) scans or surgery. I draw upon a model developed by Frank, Glazer, and McGuire [10] (hereafter FGM) that assumes clinicians prioritize the available resources so that they first serve patients with greatest clinical ‘needs’—in economics jargon, the largest marginal benefits from treatment. The available resources determine a threshold for access to care. High thresholds represent stringent rationing, meaning that only the most severely ill patients receive immediate care.

I will refer to quality investments as budget increases that allow lower thresholds for access to treatment. That is, improved quality means more patients, including those with less severe symptoms, receive timely care.

Thresholds, and thus quality investments, usually differ among health services. For example, insurers invest in disease management programs and advertising campaigns for certain medical conditions and not others. Hospital administrators devote more resources to upgrading quality and staff for some services compared with others. Often, only insiders know the exact process governing resource allocation: rationing is implicit. By contrast, some public insurance systems have instituted explicit processes for rationing, such as by requiring that a patient reach or exceed a clinical scoring threshold to be included on a waitlist for elective surgery, as in New Zealand or the Western Canada Waiting List Project [11]. The model assumes that insurers ration care as if setting such explicit thresholds, through generosity of budgets or otherwise. If the insurer allocates relatively large budgets to certain services, the thresholds for access of those services can be set quite low. By contrast, with tighter budgets, thresholds increase. For example, Fitzgerald [12, p. 339] describes the case of angioplasty in one hospital group in southern New Zealand. Because of budget cuts, the threshold for access to angioplasty (i.e., the score necessary for a patient to be placed on the angioplasty waitlist) more than tripled between 1996 and 1998. This is one example of an increase in threshold used to ration access to a specific service [12,13].

Finally, the model assumes that consumers choose among insurers at least partly based on the quality of services they expect to need. Thus, a chronically ill person seeks an insurer with large budgets allocated to chronic disease care. If payments to the insurer are not accurately risk adjusted—as is the case for most current systems [3,7]—insurers do not wish to invest in services that attract costly and unprofitable patients. The model assumes that the insurer allocates budgets strategically to maximize net revenues. For example, the insurer may find it financially rewarding to limit access to mental health services but encourage access to, and quality investment in, preventive care. As some public systems have discovered, consumers can often discern the direct relationship between low budgets and stringent rationing. For example, ‘there have been examples in New Zealand where District Health Boards have been able to point to the differences in thresholds for access to elective surgery in order to argue for a greater allocation of resources’ [11, p. 9]. The model assumes that
thresholds influence consumer choice of insurer, even if only slightly.

Table 1 summarizes the simple version of the model used for simulations. Insurers provide two medical services: chronic care (service 1) and acute care (service 2). Half of consumers are chronically ill (high risk, H types); the other half does not currently have a chronic condition (low risk, L types). The only difference among them is their expected need for chronic care, which for the chronically ill is five times that for other consumers. The simulation is based on parameters and functional forms drawn from Keeler, Carter, and Newhouse [12], without any stochastic elements. (See the Appendix for details.) The reported results arise from solving the model equations for different values of key parameters. For example, to examine payment incentives, I solve the model varying only one parameter, supply-side cost sharing. Simulations for competition and provider altruism likewise vary only those respective parameters.

Note that the socially optimal threshold for each service is 1: then patients benefit $1 from every $1 spent on care. The ideal situation would be for every insurer to invest in quality so that thresholds for all services are 1, and each insurer serves equal numbers of chronically ill and nonchronically ill patients. However, reality rarely achieves this ideal. To focus on selection incentives remaining under imperfect risk adjustment, as is currently the norm, the model assumes that prepayments to the insurer are not adjusted to account for the higher risk of the chronically ill. Thus, when payment involves prepayment, the insurer reap positive net revenues from every enrollee that is not currently chronically ill, but may lose money on each chronically ill person.

Because the chronically ill disproportionately value low thresholds for access to chronic care, an insurer can discourage their enrollment by setting a high threshold for access to chronic care. The insurer does not want to become known as a high-quality provider of chronic services, because that will attract unprofitable enrollees. Selection through service-specific quality distortions takes the form of stinting on chronic care (high \( q_1 \)), whereas offering high quality acute care (low \( q_2 \)) to attract—"cream"—preferred enrollees. Thus the ratio of thresholds, \( q_1 / q_2 \), is one natural measure of quality-distortion selection. An alternative measure is the selection index defined in the Appendix. This measure rescales all possible threshold disparity to between 0 (no disparity) and 1 (maximum disparity).

The simulations extend the FGM model to incorporate payment incentives, competition, and provider altruism. I discuss each in turn.

---

### Table 1: The model: a simple example of multiple services and multiple consumers

<table>
<thead>
<tr>
<th>Services</th>
<th>Threshold (stringency of rationing)</th>
<th>Consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic care</td>
<td>( q_1 )</td>
<td>Chronically ill (high risk, ( H )) to Not chronically ill (low risk, ( L ))</td>
</tr>
<tr>
<td>Acute care</td>
<td>( q_2 )</td>
<td>5* (spending 1) to spending 1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>spending 2</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>spending 2</td>
</tr>
</tbody>
</table>

---

### Payment incentives

To illustrate the importance of financial incentives for quality improvement decisions, I use a simple payment formula encompassing most common payment methods. The insurer may receive a fixed prepayment for each enrollee, and/or reimbursement for some or all of the costs for each service provided. Pure cost reimbursement is one extreme: no prepayment, just 100% cost reimbursement. The opposite extreme is purely prospective payment, such as capitation: 100% insurer responsibility for spending above the prepaid amount. Mixed payment falls in between: the insurer and the payer share some risk, with part of costs covered by prepayment and part covered by reimbursements.

This model captures different payment methods by varying supply-side cost sharing, cost sharing between the purchaser and insurer. Another commonly used term is risk sharing according to van de Ven et al. [3, p. 98], risk sharing is when ‘sickness funds are retrospectively reimbursed by the solidarity fund for some of the acceptable costs of some of their enrollees.’ Having the insurer bear part of the cost of care gives insurers incentive to control overall spending, but to do so differentially by service. Stinting disproportionately on the services attractive to expensive consumers discourages their enrollment and hence achieves risk selection. Similarly, overspending on services valuable to profitable consumers lures them to enroll. Throughout the discussion I will call such selection-motivated disparities ‘quality-distortion selection’.

The model simulations illustrate how higher supply-side cost sharing induces more quality-distortion selection, for any given intensity of competition for patients (Table 2). Under cost reimbursement \((i = 0)\), the insurer equally welcomes both high-cost and low-cost patients and spends generously—indeed, excessively—for both services. The insurer does not feel compelled to ration care at all \((q_1 = q_2 = 0)\), because all costs are reimbursed. Although in this case quality-distortion selection does not arise—the selection index is 0—aggregate spending is excessive. The aggregate distortion index, which quantifies how far quality (as measured by thresholds) diverges from the ideal, reaches a maximum.

One way to give incentive for cost control is to make the insurer bear part of the cost of care. The simulation reveals that even a small amount of such cost sharing induces the insurer to start rationing care: thresholds exceed 0 for both services when the insurer is responsible for one-tenth or more of costs.

However, such financial incentives also encourage quality distortions to attract preferred enrollees. Table 2 summarizes...
Table 2 Simulation of how strong payment incentives lead a profit-maximizing insurer to control cost but distort quality to attract profitable \((L)\) consumers

<table>
<thead>
<tr>
<th>Parameter varied in the simulation: supply-side cost sharing, (s) (fraction of costs not reimbursed)</th>
<th>0 (Insurer reimbursed 100% of costs of care)</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5 (Insurer reimbursed only 50 cents for every dollar spent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resulting values of variables determined within the model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insurer’s chosen threshold for chronic care, (q_1) (socially efficient threshold is 1)</td>
<td>(0)</td>
<td>0.96</td>
<td>0.99</td>
<td>1.08</td>
<td>1.14</td>
<td>1.15</td>
</tr>
<tr>
<td>Insurer’s chosen threshold for acute care, (q_2) (socially efficient threshold is 1)</td>
<td>(0)</td>
<td>0.47</td>
<td>0.50</td>
<td>0.41</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>Quality distortion as measured by ratio of thresholds, (q_1/q_2)</td>
<td>(-)</td>
<td>2.06</td>
<td>1.99</td>
<td>2.66</td>
<td>3.80</td>
<td>3.90</td>
</tr>
<tr>
<td>Percentage of chronically ill served</td>
<td>100%</td>
<td>99%</td>
<td>96%</td>
<td>52%</td>
<td>8%</td>
<td>5%</td>
</tr>
<tr>
<td>Percentage of nonchronically ill served</td>
<td>100%</td>
<td>97%</td>
<td>95%</td>
<td>94%</td>
<td>92%</td>
<td>91%</td>
</tr>
<tr>
<td>Quality distortion index (how far quality varies from the ideal level for both services; 0, no distortion; 1, maximum distortion)</td>
<td>0.76</td>
<td>0.14</td>
<td>0.12</td>
<td>0.18</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Selection index (0, equal quality for chronic and acute care; 1, maximum difference in quality between chronic and acute care)</td>
<td>0</td>
<td>0.12</td>
<td>0.12</td>
<td>0.22</td>
<td>0.34</td>
<td>0.35</td>
</tr>
<tr>
<td>Spending in chronic care for the chronically ill ((H))</td>
<td>12 000</td>
<td>8172</td>
<td>8025</td>
<td>7688</td>
<td>7438</td>
<td>7390</td>
</tr>
<tr>
<td>Spending on chronic care for other patients ((L))</td>
<td>2400</td>
<td>1634</td>
<td>1605</td>
<td>1538</td>
<td>1488</td>
<td>1478</td>
</tr>
<tr>
<td>Per-patient spending on acute care (for both (H) and (L))</td>
<td>2400</td>
<td>2028</td>
<td>2001</td>
<td>2074</td>
<td>2160</td>
<td>2164</td>
</tr>
</tbody>
</table>

Source: Author’s simulations of an economic model of insurer–provider behavior as described in the text and the appendix.
how the threshold for unprofitable chronic care \(q_1\) increases continuously with supply-side cost sharing, but the threshold for profitable acute care \(q_2\) does not. Thus, the more at risk for the costs of care, the greater the incentive for insurer–provider to distort quality by over-investing in profitability of chronic care \(q_1 = 1.15\) and low threshold for acute care \(q_2 = 0.3\), translating into a relatively meager budget for chronic care compared with acute care. As the intended result, the insurer attracts healthy consumers and few chronically ill. (The model assumes that a backstop insurer, unable to restrict services in this way, absorbs the remaining chronically ill.)

Moreover, insurers who want to ‘cherry pick’ have methods beyond service-specific quality distortions at their disposal. Although enforcing open enrollment periods can be an inexpensive and effective way to reduce explicit dumping of high risks [14], open enrollment requirements cannot prevent many other possible selection behaviors. These include choice of location, selective advertising, size and composition of affiliated providers and other staff and so on.

Van de Ven and colleagues note that in several European countries, particularly Germany and Switzerland, anecdotal evidence suggests that insurers engage in numerous kinds of ‘explicit dumping’ selection activities: bundling sickness fund insurance with life insurance after making health inquiries; encouraging unprofitable enrollees to switch to a different sickness fund; ignoring telephone calls from undesired applicants and so on. To examine incentives for these explicit selection activities as well as quality-distortion selection within the same framework, I integrate into the multiservice model a simple model of explicit dumping, drawing on [15] (also see discussion in [16], chapter 5).

In the simulation, for high enough incentive to risk select, insurers may refuse certain patients (i.e., ‘explicit dumping’) alongside quality-distortion selection (Figure 1). The combined effect (‘total selection’) increases substantially as supply-side cost sharing increases. This reminds us of the potentially large selection problems associated with high supply-side cost sharing, such as capitation payment.

**Figure 1** Quality distortions and other selection problems increase with supply-side cost sharing. Notes: Selection is the incentive for a provider to treat profitable patients and turn away unprofitable patients. Explicit dumping means refusing to treat a specific patient. Quality distortions means that the provider accepts all patients, but provides better quality for profitable services and lower quality for services that attract unprofitable patients. A high value on the vertical axis represents large differences in service quality and/or refusing to treat many unprofitable patients. The horizontal axis plots supply-side cost sharing, the fraction of cost (beyond some fixed prepayment) that is not reimbursed. 0 means all incremental costs are reimbursed. When supply-side cost sharing is 0.5, the provider receives 50 cents for each extra dollar’s worth of services provided; and when supply-side cost sharing is 1, the provider does not receive any reimbursement for providing an extra dollar’s worth of services. The figure shows that under cost reimbursement, providers do not discriminate between patients, but under capitation, providers have strong incentives to distort quality as well as ‘dump’ unprofitable patients.
Although policymakers could prevent both kinds of selection by requiring everyone to join a monopoly insurer, this way of achieving social solidarity sacrifices consumer choice and the efficiency-promoting benefits of competition. In the model, a monopoly profit-maximizing insurer would want to set all thresholds to their maximum. This would remove service-specific quality distortions, at the expense of uniformly low quality. (Only the most severely ill receive prompt treatment.) Once consumers have choice, plans may configure themselves to appeal to desirable consumers. Thus, strong payment incentives reward not only cost control but also selection-motivated disparities in access to medical care. Policymakers who do not wish to promote such disparities should consider health-based risk adjustment and mixed forms of payment. Moreover, they should be aware that payment incentives and competition interact to reinforce both desirable cost control and undesirable selection activities.

### Competition

Many economists and policymakers advocate competition to promote efficiency, allowing consumer sovereignty in choosing insurers, whereas simultaneously exerting pressure on insurers to deliver quality care at reasonable cost. Competitive reforms can flounder or thrive for many reasons. Quasi-market reforms introduced in New Zealand in the mid-1990s, for example, proved short-lived [17–19], whereas quasi-market reforms in the United Kingdom have continued along some dimensions, and several European countries have introduced consumer choice among competing sickness funds [3]. The United States perhaps most clearly embraces competition as a method to spur quality improvement, as evident in the title of the recent Department of Justice and Federal Trade Commission report: *Improving Health Care: A Dose of Competition* [20].

Consistent with this varied international experience, evidence on the quality impact of competition is also mixed. Kessler and McClellan [21] study the welfare effects of hospital competition using US Medicare data on beneficiaries’ treatment and outcomes for heart attacks between 1985 and 1994. After taking account of differences in case mix and other possible confounding factors, they find that by 1990s, competition unambiguously improved welfare, because competition led to both lower treatment costs and improved patient outcomes. By contrast, Propper and colleagues [22] find that more intense competition is associated with higher hospital mortality rates in the United Kingdom, although the magnitude of this negative quality impact is small.

More than one theory can explain the ambiguous impact of competition on quality. This article highlights one specific pathway: competition’s impact on disparities. If competitive pressure causes disparities in quality, and damage to quality for unprofitable patients outweighs quality increases for favored patients—by virtue of larger numbers of patients or a larger change in quality—then competition can decrease average quality, as found in the UK study. Even if competition increases quality for all patients, as found in most US studies, competitive pressures might spur greater quality improvements for the most profitable patients, contributing to widening disparities. If such disparities are a policy concern, then competitive reforms require careful regulation to avoid or minimize such unintended side effects.

The model helps to illustrate these pros and cons of relying on competitive pressures to spur quality improvements. If patients were all the same and needed only one service, the quality-promoting benefits of competition would be reasonably clear. As long as patients are aware of differences in quality to some extent, consumers can switch to insurers offering better quality for similar cost. With consumers flocking to the highest quality insurer, all insurers have incentive to invest in quality improvement. When competitive pressures decrease (captured in the model as a higher cost of switching insurers), consumers have fewer viable alternatives, so an insurer can ignore quality without losing too many enrollees to make it profitable. Thus, in the simplest context, strong competition forces insurers to improve quality as a way to attract and retain customers.

In the more realistic context of multiple services and diverse patient needs, however, competition also spurs insurers to engage in selection activities. Competitive pressure will generally increase incentives for quality improvement disproportionately for services that attract profitable enrollees and may give incentive to lower quality for services valuable to high-cost enrollees.

Figure 2 shows how competition can exacerbate quality distortions, especially when an insurer has strong financial incentive to control costs, such as under capitation ($\gamma = 1$). As competition increases (i.e., travel or switching costs decrease), the index of quality distortion increases.

Moreover, supply-side cost sharing and competitive pressures interact, compounding the effect of either policy in isolation. As shown in Figure 2, for any given level of competition, higher supply-side cost sharing induces higher selection. This kind of interaction seems particularly evident in Germany and Switzerland, where selection is a serious problem [3]. Unlike other countries with competing sickness funds—such as Belgium, the Netherlands and Israel—these two countries do not have risk sharing. Strong payment incentives combined with imperfect risk adjustment give large incentives for selection activities. Using mixed payment (risk sharing) would give less incentive for quality distortions, while still promoting responsiveness and some attention to efficient use of resources.

Another way to mitigate selection distortions among competing insurers is through mandatory high-risk pooling (MHRP), outlier payments or other blended payment schemes [16,23]. Figure 3 shows how MHRP can reduce the incentive to improve quality exclusively (or disproportionately) for the most profitable services. Under such a scheme, the insurer can claim reimbursement of costs for patients that the insurer anticipates will be costly. This pooling of high risks clearly reduces the insurer’s financial risk from enrolling such consumers, and therefore will decrease incentive to underinvest in quality services for them. The simulation assumes cost reimbursement for all expenses of pooled high-risk enrollees, and capitation payment for all nonpooled enrollees. As expected, the model predicts less selection when the insurer bears less of the financial risk of costly patients.
Comparing methods to stimulate quality improvement

...such reduction of selection may come at the cost of considerably higher spending, particularly if the pooled patients’ expenses are fully reimbursed. This suggests the importance of some risk sharing—i.e., mixed payment—to help control excessive spending. Once again, mixed forms of payment emerge as preferable for balancing quality improvement across services and with cost control [also see 24].

Altruism

As distinguished health economist Victor Fuchs noted in his 1996 presidential address to the American Economics Association, ‘one of the greatest errors of health policy-makers today is their assumption that market competition or government regulation are the only instruments available to control health care. There is room for, and indeed need for, a revitalization of professional norms as a third instrument of control’ [25, p. 17]. A simple extension of the model shows how benevolence or altruism [26]—an insurer’s innate concern for patient welfare, arising through professional norms or otherwise—can promote quality for all, mitigating the effects of financial rewards for quality distortions. The larger the weight put on patient benefits relative to net revenues when setting treatment thresholds, the more ‘altruistic’ the insurer.

However, such reduction of selection may come at the cost of considerably higher spending, particularly if the pooled patients’ expenses are fully reimbursed. This suggests the importance of some risk sharing—i.e., mixed payment—to help control excessive spending. Once again, mixed forms of payment emerge as preferable for balancing quality improvement across services and with cost control [also see 24].

As intuition would suggest, the model confirms that greater altruism causes insurers not only to choose higher quality for a given service, but also to allocate quality investments more evenly across patients, so that the ‘neediest’ patients benefit from quality improvements despite being unprofitable (Figure 4).

However, such altruism is not an unmitigated benefit for society. An insurer or provider who places great value on patient healthcare benefits may devote excessive resources, from the social point of view, to patient care. Society might be better off with smaller budgets for healthcare than an altruistic provider would prefer, because some of that money could be more fruitfully spent on public health, education, environmental quality and so on. The dilemma resembles that noted above regarding cost reimbursement: a very ‘altruistic’ insurer, or one reimbursed for all costs, does not wish to distort quality, but also lacks incentive to control aggregate costs. Moreover, an insurer’s degree of altruism appears far less amenable to policy intervention than the insurer’s competitive context or payment incentives.

Conclusion

Healthcare policymakers in many countries seek to use incentives and competition to spur quality while controlling cost.
Figure 3  Mandatory high-risk pooling reduces selection. Notes: The Selection Index, the value plotted on the vertical axis, measures overall quality distortions designed to attract profitable patients. A zero value represents no distortions; a high value represents large differences in service quality. The horizontal axis plots three different scenarios for mandatory high-risk pooling (MHRP): the insurer must predesignate some enrollees (high risks) that will be reimbursed more generously than other patients. The simulation assumes cost reimbursement for all expenses of pooled high-risk enrollees and capitation payment for all nonpooled enrollees. The figure shows that the greater the risk pooling for high risks (i.e., the higher the percentage of enrollees with cost reimbursement), the less severe the selection problem becomes.

Figure 4  Altruism improves outcomes: selection-motivated quality distortions decrease as insurer–provider concern for patients increases. Notes: The Selection Index, the value plotted on the vertical axis, measures overall quality distortions designed to attract profitable patients. A zero value represents no distortions; a high value represents large differences in service quality. The horizontal axis plots different values of parameters used in the economic simulation to represent provider concern for patient welfare (relative to profits), denoted by alpha in the mathematical model. Larger values on the horizontal axis (alpha) represent less profit-seeking. The figure thus shows that quality distortions are worst when the provider exclusively seeks profits (alpha = 0), and quality distortions are predicted to be less of a problem when the provider places less weight on profits relative to patient welfare.
However, such reforms can penalize providers that improve quality of care for the least profitable patients. Therefore, one important challenge is to mitigate the sacrifice of social solidarity that reforms entail.

This article presents simulations revealing how competition, altruism, and payment incentives affect quality, efficiency, and equity of treatment. The focus is how to minimize disparities in quality of care (along racial or ethnic lines as highlighted in [27] or otherwise).

The primary result is that intense competition for patients and strong financial rewards for cost control can exacerbate selection-motivated quality distortions and compromise social solidarity. Reforms that combine careful regulation with competition, and use mixed forms of provider payment, are most likely to overcome these problems.

References


Accepted for publication 5 August 2005

Appendix

The simulation model

This appendix describes the model and simulations from which the conclusions in the main text are drawn. For a more detailed discussion of the underlying economic model, please see [10].

A health plan provides various health care services—prenatal care, treatment of heart attack patients, mental health services—indexed by $j$. Let $m^j_i$ represent the spending on health service $j$ provided to individual $i$. Assume that payment includes two components. First, for each enrollee, the plan receives a fixed prepayment (capitation) $r$. The insurer also receives reimbursement $(1-s_j)m^j_i$ for each service $j$, with $s_j \leq 1$. Supply-side cost sharing arises when $s_j > 0$. Pure cost reimbursement corresponds to $r = 0$ and $s = 0$. A mixed payment...
system features \(0 < \epsilon < 1\). Capitation or fully prospective payment arise when \(\epsilon = 1\).

Insurers ration care by setting a threshold, \(q\), for access to care. High thresholds represent lower quality. Patients value quality, so demand increases in \(w\) and decreases in \(q\).

Denoting patient demand by \(n'(q)\), the insurer’s expected net revenues \(\pi(q)\) are

\[
\pi(q) = \sum_i n'(q) \pi'(q), \text{ where } \pi'(q) = \left(1 - \sum_j \epsilon_j m_j \right)
\]

Consider the simple case of two services, \(j = 1, 2\), and two consumer types, high-risk \(H\) and low-risk \(L\), where expected health spending of \(H\) exceeds that of \(L: m_H > m_L\). For simplicity assume the two types are distributed identically in the population, so that demand for a health plan consists of \(n_H(m_H)\) and \(n_L(m_L)\). To focus on selection incentives remaining under imperfect risk adjustment, assume that pre-payment \(\pi\), if any, is not risk adjusted. To induce a health plan to provide services, enrollment of some individuals—here, \(L\)—must be profitable: \(\pi L(q) > 0\). Assume also that for any positive amount of supply-side cost sharing \((s > 0)\) without risk adjustment, \(H\) are always less profitable than \(L\), and depending on the payment system, the plan may incur a net loss for each \(H\) enrolled.

The simulations assume that consumer utility takes the form

\[
v' = \sum_{j=1}^{2} a_j m_j \left(\frac{1}{2} \theta_j (m_j)^2\right)
\]

with \(\theta_H^2 = 5\); \(\theta_L^2 = \theta_H^2 = 1\). With providers allocating quality investment and access among patients and services according to service-specific thresholds, quality allocation for each service becomes

\[
m_j = \theta_j \left(\frac{a - q_j}{b}\right)
\]

Thus, \(H\) enrollees use five times as much service 1 (chronic care) as \(L\) enrollees, whereas use of service 2 (acute services) does not differ across enrollee type. Simulations assume \(a = 3\) and \(b = 0.0025/2\). Demand is simulated with the cumulative distribution of consumers defined as the cumulative distribution function for a logistic random variable, depending on the cost of traveling to a competing provider, \(c\).

The health plan maximizes net revenues taking the competing plan’s thresholds as given. For all simulations except those for Figure 2, the competitor’s thresholds are held fixed at \(q_1 = q_2 = 1\). For Figure 2, the competitor’s thresholds are fixed at the profit-maximizing response of a plan paid capitation \((\epsilon = 1)\) competing with a plan that sets \(q_1 = q_2 = 1\), i.e., \(q_1 = 1.167\) and \(q_2 = 0.311\). Travel costs \(\epsilon\) are set equal to 100 unless noted otherwise (e.g., in Figure 2). Prepayment \(\tau\) to the plan covers the costs that are to be shared at point of service such that if the plan set efficient thresholds \(q_1 = q_2 = 1\), it would make a profit of 1000 per enrollee.

To aid in assessing the profit-maximizing thresholds, I define two indices designed to normalize the range of threshold distortions between 0 and 1. The first, distortion index, measures the extent to which thresholds deviate from social optimality, which is defined by \(q_1 = q_2 = 1\):

\[
\text{Distortion} = \frac{2}{1 + \exp\left[-{(q_1 - 1)^2 - (q_2 - 1)^2}\right]} - 1
\]

Note that when \(q = 1\) for every service, this distortion index is equal to zero. As thresholds diverge from one in either direction (creaming or stining or any combination thereof), the distortion index becomes a larger positive number, approaching 1 in the limit. This index therefore normalizes to between 0 and 1 all distortions from \(q_1 = q_2 = 1\).

The selection index, by contrast, captures the degree to which thresholds are set differently by service to attract the profitable patients \((L)\) and exclude unprofitable patients \((H)\):

\[
\text{Selection} = \frac{2}{1 + \exp\left[-(q_1 - q_2)^2\right]} - 1
\]

This selection index appears on the vertical axis of Figures 1, 3 and 4. In all cases, selection distortions arise because \(q_1 > q_2\), i.e., to cream \(L\) enrollees and dump \(H\) enrollees, as predicted.

I also report the ratio of \(q_1\) to \(q_2\), which when it exceeds 1 also represents quality distortions designed to attract \(L\) and eschew \(H\) enrollees. The range of this ratio reported in the simulations—between 2 and 10—is comparable to that found empirically in FGM [10].

To model MHRP of \(\epsilon\)\% of enrollees (Figure 3), I assume cost reimbursement \((\epsilon = 0)\) for the pooled enrollees and calculate profit-maximizing thresholds for serving the remaining \((1 - \epsilon)\)\% of enrollees. The plan exclusively allocates unprofitable patients \((Hs)\) to the generously reimbursed risk pool. Figure 3 assumes fully prospective payment \((\epsilon = 1)\) for the nonpooled enrollees.

For the explicit dumping model, the simulation assumes dumping costs of the form

\[
Y(t, c) = r \left[\exp\left(\frac{t}{1-t}\right) - 1\right]
\]

Selection shapes the distribution of consumers to whom the plan will most appeal. Specifically, assume that \(t\) reduces ‘travel cost’ to the nearest competitor from \(t\) to \((1 - t)\). Perfect selection \((t = 1)\) reduces the travel cost to the competitor to zero, so that all unprofitable patients go to the competitor.

The plan will only ever find it profit-maximizing to dump high risks, \(H\). Although I assume plans’ selection activities perfectly target \(H\) consumers and do not affect \(L\) demand, I
Comparing methods to stimulate quality improvement

also assume that perfect selection—dumping all high risks onto competitors—is prohibitively costly. The model incorporates this into the demand function of high-risk consumers (i.e., the probability that an H consumer will choose the plan; see [10]) so that as $t$ increases, fewer high-risk consumers find it utility-maximizing to enroll in the plan exerting selection effort $t$. For $t$ approaching 1, enrollment of $H$ approaches zero.

**Definitions of terms and model parameters**

Capitation—A fixed amount of money per enrollee per period of time (an example of full supply-side cost sharing, with very strong financial incentives for cost control and selection).

Dumping—Refusing to treat specific patient(s).

Mandatory high-risk pooling (MHRP)—The insurer must predesignate some enrollees (high risks) that will be reimbursed more generously than other patients. Because insurers can claim reimbursement of costs for patients that the insurer anticipates will be costly, there is less incentive to discriminate against such patients or to ‘cream skim’ profitable patients (see 23).

Mixed payment (risk sharing)—An intermediate level of supply-side cost sharing ($0 < s < 1$), so that insurer/providers receive partial reimbursement for care provided beyond that covered by prepayment. The insurer and the payer share some risk, with part of costs covered by prepayment and part covered by reimbursements.

Prospective payment—Prepayment of a fixed amount, an example of full supply-side cost sharing.

Quality-distortion selection—Favoring profitable patients by providing high quality for services used by such patients and low quality for services that disproportionately attract unprofitable patients.

Risk sharing—See mixed payment.

Risk selection—See selection.

Selection (risk selection)—The incentive for a provider to treat profitable patients and turn away unprofitable patients.

Supply-side cost sharing—Cost sharing between the purchaser and insurer; specifically, the fraction of cost (beyond some fixed prepayment) that is not reimbursed. 0 means all incremental costs are reimbursed. When supply-side cost sharing is 0.5, the provider receives 50 cents for each extra dollar’s worth of services provided and when supply-side cost sharing is 1, the provider does not receive any reimbursement for providing an extra dollar’s worth of services.

**Key model variables and parameters**

$q_1$ — Insurer’s chosen threshold for chronic care (the socially efficient threshold is 1)

$q_2$ — Insurer’s chosen threshold for nonchronic care (socially efficient threshold is 1)

$r$ — Fixed prepayment (capitation)

$s$ — Supply-side cost sharing

$c$ — Patient travel cost to competing insurer/provider

$L$ — Low-risk (profitable) patients/enrollees

$H$ — High-risk (chronically ill, unprofitable) patients/enrollees

$\text{Alpha}$ — Insurer/provider’s concern for patient welfare, relative to profit