MEMORANDUM

No 18/2004

When quality today affects service needs tomorrow

Kari Eika

Department of Economics
University of Oslo
List of the last 10 Memoranda:

<table>
<thead>
<tr>
<th>No</th>
<th>Author(s)</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Rolf Golombek and Michael Hoel</td>
<td>Unilateral emission reductions when there are cross-country technology spillovers. 23 pp.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Alexander W. Cappelen and Bertil Tungodden</td>
<td>Rewarding effort. 31 pp.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Alexander W. Cappelen and Bertil Tungodden</td>
<td>The liberal egalitarian paradox. 15 pp.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>John K. Dagsvik and Steinar Strøm</td>
<td>Sectoral labor supply, choice restrictions and functional form. 15 pp.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Rolf Golombek and Michael Hoel</td>
<td>Climate Agreements and Technology Policy. 38 pp.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Isilda Shima</td>
<td>The shadow economy in Norway: Demand for currency approach. 23 pp.</td>
<td></td>
</tr>
<tr>
<td>09</td>
<td>Steinar Holden</td>
<td>Wage formation under low inflation. 22 pp.</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>Steinar Holden and Fredrik Wulfsberg</td>
<td>Downward Nominal Wage Rigidity in Europe. 33 pp.</td>
<td></td>
</tr>
</tbody>
</table>

A complete list of this memo-series is available in a PDF® format at: [http://www.oekonomi.uio.no/memo/](http://www.oekonomi.uio.no/memo/)
When quality today affects service needs tomorrow

Kari Eika

First version: October 2001
This version: October 2004

Abstract

Quality in human services has long term effects. Reduced quality of service today increases the service recipients’ future service needs and other social costs. This paper shows that such effects should be considered when a government designs contracts with a for-profit service provider. If the contract relies on verifiable information only, short contract periods fail to give the provider proper incentives to internalize future effects of quality. Long term contracts are problematic if the effects of quality are not over time indicated by verifiable measures. Relational contracts seem more robust to changes in the model’s informational assumptions and rely on trust to deal with hold-ups. Long term quality effects matter for the relative merits of integrated provision and contracting out, and may create adverse quality incentives if a for-profit provider has market power.

1 Introduction

One motivation within the human services is to improve service recipients’ welfare and reduce their service needs in the future. For example, the school and social
services that an intellectually disabled child receives are intended to improve the child’s future skills and social functioning, thus lowering the child’s future service dependency. A prison’s main task is to detain convicts. Still, the effect of the detention on former convicts’ criminal activity once released is also important for social welfare, both directly and through the demand it places on society’s resources, e.g. in upholding the law. In either case, it takes time for all the effects of a service to be realized.

Long term effects depend largely on the quality of service. The higher service quality is today, the lower are future service needs. In the human services it is also possible that service provision, if quality is sufficiently low, increases long term service needs and decreases recipients’ welfare. Quality, however important, is difficult for outsiders to observe and the information that can be obtained is largely unverifiable. Neither is it easy to maintain a sufficient pressure on providers for quality service through consumer choice or voice in many human service contexts.

Most of the human services are either wholly or largely supplied by a third party, usually by public authorities, although private health insurance companies are important in many countries. If the government contracts out service provision to a profit-maximizing firm, how should contracts be designed so as to give the provider incentives to internalize the long term effects?

An example is the proposed Ticket To Work program in the US. The aim of the program is to enable people to move from disability income support into active work life by offering training and job search assistance. Someone receiving disability income benefits can choose to enter into a contract with a training agency which is then rewarded if the training results in paid employment (Frank and McGuire (2003)). The quality of the job training cannot be directly observed by the government or anyone else not participating in the training. Furthermore, the information that is made available is difficult to verify in court. It is not possible to specify in a contract prior to training the appropriate ways to motivate and qualify a beneficiary. Training must be tailored to the individual trainee, and much of the information needed to do this tailoring, for example, about the client’s preferences, his problems and the job opportunities, is revealed through the training. This information is too complex to be specified in a contingent contract. At the same time, the intended effects, increased labor participation and social security savings, are realized over time.
When quality cannot be directly controlled, a static or myopic perspective may be problematic. In the formal analysis I show that short contract periods weaken incentives for quality care.

The next section describes the main characteristics of the human services that motivate the analysis. Section 3 presents the analytical framework, and derives the socially optimal quality level. This is compared to the quality level that a for-profit service provider chooses under alternative informational assumptions and different types of contracts. Relying on long term contracts is not desirable if the effects of quality are not indicated by verifiable measures over time. An alternative is to make contract renewal more probable if the authorities deem the quality level to be high. I argue that the advantages of this contract design seem more robust to changes in the model’s informational assumptions than do long term contracts.

Long term contracts are often difficult to use in practice. Section 4 discusses hold-ups and other obstacles to long term contracts. Either contracting party can in principle be held up by the other. Additionally, the client may hold-up the provider. Long contract periods also limit consumer choice, which is an important informal sanction in most markets where quality is non-verifiable. Restricting consumer choice is particularly problematic in the human services (though for some services, such as prisons, it is an integral part of the service). It may, however, prevent the provider from being held up by the client.

Other results that do not follow from the formal analysis are also briefly discussed in section 4. Inter-temporal quality effects are of importance for how efficient service provision is conceptualized and measured. It also has a bearing upon more profound organizational issues than the ones I analyze. The long term effects of a service should be taken explicitly into account when assessing the relative merits of in-house service provision versus the contracting out to an independent provider. Hart et al. (1997) outline theoretical conditions under which public or (for-profit) private production of services is to be preferred. Taking long term quality effects into account may alter their policy conclusions. Lastly, market power can give the provider incentives for low quality service. Low quality increases future service needs, thus expanding the market. This is another supplier-induced demand mechanism than those mentioned in health economics literature, and by which the clients concerned as well as society at large incur substantial welfare losses.

The main results from the analysis are summarized in section 5.
The human services

According to common usage the human services refers to the health and social services and education. I also include in this term parts of law enforcement, such as prisons and community policing. I am concerned with service contexts where the object of the service is a person. To detain a prisoner, to teach a child, to care for a help dependent elderly person is, in this sense, literally “people work”. In my analysis, I also assume that the following characteristics are significant:

- A service has long term effects. In particular, the way in which a service is provided, that is, the quality of service, affects a person’s need for the same or related services in the future. The aim of the Ticket To Work program is to reduce the need for disability income support. The responsibilities of the police are to detect and verify criminal activity (which in itself is intended to deter future crimes), and to undertake active crime prevention policies, for example, building community relations or cooperating with social workers getting troubled youth off the street. The intention of nursing home care is to assist the sick and elderly in chores they are no longer able to perform, but also to improve their ability to function.

Long term effects are not unavoidable by-products of the service itself. Their direction and strength depend on the quality of the service. Successful habilitation or rehabilitation in health and social care implies that service provision reduces future help dependency relative to giving no assistance today. However, service provision may actually contribute to higher service dependency, rather than to reduce future service needs. In the significance of such mechanisms, the human services differ from other industries. Humans respond to the provided service in ways that affect their future needs. These responses are not necessarily conscious, nor rational, and are powerful mechanisms contributing to virtuous or vicious circles with respect to personal future service dependency. My analysis does not apply to the unsuccessful outcomes of inherently risky treatments. For example, heart operations usually improve functionality, but sometimes also worsen conditions. The situations of interest for me are those where the effect on functioning depends on the type of service given.

Living in a nursing home may lead to unnecessary physical passivity so that a resident’s physical condition deteriorates further. For example, the resident
is capable of setting the table, opening the door etc., but such activities are not demanded from his or her, and perhaps not even allowed for. Similarly, extensive and “wrong” care may deprive a person of the skills she once had. Additionally, and often parallel to this, there are powerful psychological mechanisms which reinforce help dependency. The way a patient is told about a serious disease may affect the patient’s ability to cope with her health problems in the future, by either helping the patient to cope, or demoralizing or discouraging her. In social care, the perception of being help dependent may in itself create help dependency. At one extreme of the scale is so-called learned helplessness. This may occur if the care provided deprives the care recipient of control over his or her life, increases unpredictability, and if the care recipient also perceives this to be the case. Observable implications of this are passivity, and possibly over time cognitive retardation, low self-esteem, sadness, loss of aggression, immune changes and physical illness as well (Peterson et al. (1993)).

- The gains from high service quality are often substantial. The social gain may be tremendous for the individuals directly concerned, and often so are the indirect effects also, although frequently dispersed across a large number of individuals. Successful crime prevention policies that reduce homicide rates and overall crime levels serve to illustrate the importance of both effects.

- The gains from high service quality are often foregone. The attention that successful policies sometimes receive is one indication of this. These policies stand out, appearing to be exceptions to the rule rather than the standard. Moreover, there is surprisingly little comprehensive evidence available on the costs and benefits from alternative policies. The absence of such calculations may in part reflect the collective good and externality problems, which result in too little public attention to these issues. For the same reason, successful policies often require coordinated efforts. Even within closely related parts of government, coordination has proven difficult. English authorities, for example, recognize that socially productive investments in preventive services and rehabilitation of the elderly is not undertaken due to a lack of coordinated effort by the public bodies responsible for, respectively, hospital services and long term care (Audit Commission (1997)).
Norton (1992) studied a social experiment on nursing home rehabilitation in California. One group of homes were given explicit monetary incentives that rewarded improvements in the state of the residents’ health, and discharge. Another group, the control group remained with a prospectively set reimbursement rate that was equal for all residents, irrespective of health status. The reward system relied on a third party, registered nurses employed by an independent firm, evaluating residents’ health status at regular time intervals. Using a Markov model to represent changes in health, he found that residents’ health improved significantly. Moreover, more residents were discharged from the homes. Overall health expenditure went down due to shorter waiting time for entry from hospitals.

- Substantial informational problems add to the organizational difficulties. The causal mechanisms through which long term effects of quality are mediated are complex human and social processes. The processes are highly idiosyncratic, differing from individual to individual and from case to case. Much is unobserved. It is therefore difficult to make inferences about quality from observed outcomes. The complexity also makes quality unverifiable. Firstly, the appropriateness of any given action is highly context-dependent. The recipient’s needs and preferences, the characteristics of the situation, and also the capacities and talents of the provider and her relation with the recipient are all factors that must be taken into account. Writing a contract covering all eventualities, if at all possible, is too costly, cf. also Hart et al. (1997). Secondly, important quality dimensions are intangible, for example, the quality of social relations. Intangible dimensions cannot be described in a contract in a verifiable way. Therefore, quality is unverifiable even after the context-dependent characteristics are known. As a consequence, quality assessment unavoidably contains a subjective, judgmental element. Slagsvold (1997) investigates the validity of alternative methods for quality assessment in nursing homes. Her findings support this view. Quality of service is also difficult to observe for outsiders, that is, for those not participating in the service delivery. In order to fully observe quality, one must observe service delivery and be well informed about the context-specific characteristics. That quality is non-verifiable and hard-to-observe are common assumptions in the literature on the human services, see for example, Arrow (1996) and Chalkley and Malcolmson (2000).
• Often, individuals receive a human service for a long time. This is characteristic of education, training, rehabilitation, imprisonment and long term care. Many long term care recipients – individuals with physical or intellectual disabilities and nursing home patients – need assistance for as long as they live. In most cases, service provision is better described as a continuous activity rather than one that can be decomposed into a series of one-time events such as “the repair of a car”, “heart operation”, “court sentencing” etc. One implication of continuity is that improvements in quality can benefit recipients directly, those that experienced lower quality and that still receive the service, as well as new recipients (Kane and Kane (1988)).

• Lastly, there is generally limited scope for demand to influence quality. Third party purchase is not in itself a hindrance for consumer choice to be a powerful mechanism in ensuring quality, see Chalkley and Malcomson (2000) and references they quote. Also, opportunities to voice discontent can represent an important alternative pressure for quality (Hirschman (1970)). However, in some service contexts, service recipients have weak consumer sovereignty. That is, they have weak voices and limited ability to choose in an informed and rational way (Eika (2003)). Even for services that are technologically simple, they are not in a position to demand from the provider those actions that are required to realize a satisfactory quality level. To a varying degree this is the case within psychiatric care, care for the weak and demented elderly, care for intellectually disabled and services for children. Rational choice is often more difficult the higher the care needs are, for example, caused by serious mental problems. Yet another reason may be that choice is more or less excluded by the nature of the service. For example, the exit option does not exist for a prisoner because he or she cannot choose whether or not to go to jail. Choice is in general restricted within law enforcement. Enforcement of the law presupposes extensive authority, and that authority cannot be subjected to competition.

It is a challenging task to organize service provision in such a way that decision makers internalize the long term effects of quality. Other factors than those addressed in the following contribute to this, and deserve mentioning. Firstly, weak consumer sovereignty creates an incentive problem in the monitoring of service qual-
ity. Those to whom quality matters the most, the service recipients, are not able to effectively monitor service provision directly, or indirectly by monitoring the efforts of outside monitors (Eika (2003)). Secondly, the gains to society as a whole are dispersed across a large number of individuals, which give rise to both externality and collective good problems. Thirdly, it may take long time, for example several decades, before the gains from high quality or the costs of poor quality materializes. This creates externalities across generations. Myopic politicians or firm managers, hyperbolic preferences or other irrationalities would further increase time inconsistencies. Lastly, the often complex causal processes make it difficult to grasp the significance of service quality, particularly if it takes a long time before the full effect is seen. These problems are disregarded in the following. I assume that the government is benevolent (concerned with social welfare) and rational.

3 A model of service needs

The government is responsible for supplying a citizen, hereafter called a service recipient, with some service. The government secures its financing, while the service provider determines the level and quality of service. The higher the quality is, the lower are future service needs. Quality is non-verifiable and difficult to observe by outsiders.

If the government has decided to contract out service provision to a profit-maximizing firm, how should contracts be designed so as to give the producer incentives to internalize the inter-temporal effects of quality?

3.1 The framework

There are two decision makers, the government who determines the service contract and purchases the service accordingly, and the service provider. The third party, the service recipient, makes no decision. For analytical convenience the restriction on consumer choice is taken to the extreme. Rational consumer choice is either not possible or not permitted. The provider’s incentives for quality are determined

---

1This service does not have to be wanted by the service recipient. Participation in the proposed job training program is voluntary. However, detention is not preferred by prisoners. Furthermore, the government is responsible for the purchase, but may or may not finance the service.
entirely by the contractual arrangements with the purchaser, the latter possessing only imperfect and non-verifiable information about quality. Consequently, there is the problem of multitasking (Holmström and Milgrom (1991)).

Let $s_t$ be the service level, $q_t$ the service quality, $w_t$ a verifiable outcome measure, and $H_t$ service dependency in period $t$.

The government is concerned with some aspects of the service recipient’s state. For now, assume that the verifiable index $w$ is an indicator for how the government evaluates these aspects. The outcome is better the higher $w_t$ is. I later discuss the case when $w_t$ is not an adequate indicator for outcome.

The measure $w_t$ depends in part on the person’s service needs, $H_t$, and the service level, $s_t$, that he receives.

$$w_t = w(s_t, H_t) \tag{1}$$

The service level improves the immediate outcome for the service recipient (hereafter called the service outcome), $w'_s > 0$, while the effect of higher service dependency is negative, $w'_H < 0$. $w(\cdot)$ is assumed quasi-concave. In the care services $H$ is help dependency or care needs. In a prison, a prisoner is more demanding, and more likely to commit new crimes once released the higher $H_t$ is. In the Ticket To Work program $H_t$ are those circumstances that qualify the recipient for a social security disability income. The authorities require that the outcome is at least as good as $\bar{w}$. With scarce resources we can assume that this constraint is binding. $\bar{w} = w(s(H_t), H_t)$ defines

$$s_t = s(H_t) \tag{2}.$$ 

It follows that $s' = -\frac{w'_H}{w'_s} > 0$ and $s'' > 0$.

Labor is the only factor of production, but there are two types of labor input, high and low skilled, respectively, $E_t$ and $e_t$. The service level is determined only by total labor input, so $s_t = E_t + e_t$. I define service quality as the share of high skilled labor input in total labor input, $q_t = E_t/s_t \in [0, 1]$. High quality effort ($E$) is more costly for the provider than unskilled or simple effort ($e$). $p > 0$ measures the additional relative cost of $E$. Total costs per period are therefore $e_t + (1 + p) E_t$, or,

$$s_t (1 + pq_t) \tag{3}$$.
Service dependency is partly determined by exogenous factors, represented by the parameter $\theta > 0$, and partly by past service provision.

\[(4) \quad H_t = \theta + s_{t-1}h(q_{t-1}),\]

where $h' < 0$ and $h'' > 0$. For convenience, I assume that $h'(0) = -\infty$ and $h'(1) = 0$. Note that $\partial H_t / \partial q_{t-1} = s_{t-1}h' < 0$. Higher quality always reduces service dependency and thus future service requirements, but the marginal effect is weaker the higher the quality is, $\partial^2 H_t / \partial q_{t-1}^2 = s_{t-1}h'' > 0$. Lowering future service dependency is the only role of quality in the model. The effect of present service on future service needs is determined by $h(q_{t-1}) = \partial H_t / \partial s_{t-1}$. Beyond these properties, $h(\cdot)$ may take various forms depending on the individual client. As argued in section 2 it is, for example, possible that $h(0) > 0$ and $h(1) < 0$, so that future service needs increases if quality is very low.

For simplicity, only two periods are assumed, the “present” period 1, and the “future” period 2. It is the service in period 1, and specifically the choice of high skilled labor input, that is the focus of my analysis. Period 2 has only interest to the extent that the care outcome in that period matters for appropriate or actual behavior in period 1.

### 3.2 Optimal service provision

To use as a benchmark in the discussion I derive the optimal service provision. It is defined as the service level ($s$) and quality ($q$) which minimize the net present costs of fulfilling the welfare requirements $w \geq \bar{w}$, taking into account the effect of quality on future service needs (through equations (2) and (4)). Let $r$ be the discount rate. The optimal service provision is found by

\[
\min_{s_1,q_1,s_2,q_2} \left\{ s_1 (1 + pq_1) + \frac{1}{1+r} s_2 (1 + pq_2) \right\} \quad \text{given eq. (2) and (4)}.
\]

Obviously, it is optimal to set $q_2 = 0$, since $q_2 > 0$ is costly but does not increase welfare. The service level in the first period follows from the predetermined service dependency $H_1$, so $s_1$ is given by $s(H_1)$. Using equation (2) and (4), the
problem then reduces to

$$\min_{q_1} \left\{ s_1 \cdot (1 + pq_1) + \frac{1}{1 + r} s (\theta + s_1 \cdot h(q_1)) \right\}.$$  

The optimal quality level in period 1, denoted $q^*$, is uniquely determined by

$$\begin{align*}
(1 + r) p &= -s' (H_2) h'(q^*) ,
\end{align*}$$

The interpretation of (6) is straightforward. On the margin, the gain from quality care, $-s'h'$, must equal the opportunity cost of this investment, $(1 + r) p$. The more important quality is in reducing future service dependency ($-h'$ large), and the larger decrease in future care this allows for without also lowering patient welfare ($s'$ high), the larger the optimal level $q^*$ is. This is shown in the figure. Since $s'' > 0$ and $h'' > 0$, the right hand side in (6) is a decreasing function of $q$. 

Figure 1:
3.3 Long term contract with a for-profit service provider

Now assume that the provider is a for-profit firm, having contracted with the government after a bidding round. What are the incentives for that provider to give service at a high level of quality? In this section I discuss how the provider’s incentives depend on the length of the contract, and more generally, on the conditions for contract renewal.

Let $\tau$ be the probability of contract renewal, $P_t$ provider remuneration in period $t$, and $\tilde{q}$ the profitmaximizing value of $q_1$.

The incumbent provider in period 1 has expected profits over the entire time horizon equal to

$$(7) \quad \{P_1 - (1 + pq) s_1 + \frac{\tau}{1 + r} [P_2 - s_2 (H_2)]\}$$

where $H_2 = \theta + s_1 h (q_1)$.

The determination of $P_1, P_2$ and $\tau$ depends on the type of contract. The set of contracts that are feasible is limited by the information available to the contracting parties. I look at three different cases that differ in their informational assumptions as shown in table 1.

Table 1. Informational assumptions

<table>
<thead>
<tr>
<th>Case</th>
<th>Informational assumptions</th>
<th>Outcome indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>H verifiable, $w(\cdot)$ is known.</td>
<td>$w$ indicates outcome</td>
</tr>
<tr>
<td>Case 2</td>
<td>No verifiable variables (except $w$)</td>
<td>$w$ indicates outcome</td>
</tr>
<tr>
<td>Case 3</td>
<td>$w$ does not indicate outcome</td>
<td>$w$ indicates outcome</td>
</tr>
</tbody>
</table>

Note that $w$ by construction is verifiable, but does not always indicate the outcome (case 3). Furthermore, it is always possible to obtain noisy information $q$.

**Case 1** $w$ and $H$ are verifiable. $w$ indicates outcome. The government knows $w(\cdot)$.

The socially optimal solution can be realized through a long term contract, or a short term service contract in combination with a long term financial obligation. In the first case, the service provider is the same in both periods and total remuneration is determined prior to period 1. In the other case, the service provider in the first period is contractually committed to pay the period 2 service costs, $s(H_2)$, regardless of who actually provides the service in that period.

In either contract, $\tau = 1$. The provider has to pay the period 2 service costs with certainty. Maximizing (7) is then equivalent to minimizing social costs, (5).
The provider gets the entire gain from higher quality through lower costs of future provision, for which he or she is also responsible.

The latter contract type resembles a sales contract with a product warranty, for example, a car manufacturer being responsible for the repair expenses of a car that breaks down within the warranty period. The car repair problem usually can be assessed fairly easily, and the required repair can be undertaken by any authorized car repair shop. This is not a common situation in the human service contexts relevant here. Typically, the costs, \( s(H_2) \), are not possible to specify contractually. Service dependency may not be verifiable, or the \( s(\cdot) \)-function, the costs of providing adequate service given \( H_2 \), may be difficult to specify in advance, either because it is not known, or for other reasons. Case 1 is primarily included in the analysis as an illustrative contrast to case 2 and 3.

**Case 2** Only \( w \) is verifiable. \( w \) indicates outcome.

The use of a long term contract realizes the optimal solution in case 1 with far lower informational requirements than the other contract design. In case 2 the future service costs are non-contractible. A long term contract is still feasible, and yields optimal incentives as long as all effects of quality eventually show up in the verifiable measure \( w \).

The Ticket to Work program is most closely approximated by case 2. The outcome dimension with which the authorities are concerned is easily verifiable. Training of a recipient is successful if, and as long as, the social security beneficiary earns an income beyond a certain threshold level. The government can then withdraw the cash benefit. In principle, a long term contract covering the period until the person reaches the age of retirement gives the job training agency sufficient incentives. To some extent the proposed reward scheme has a long term perspective. With successful training, the training agency receives 40 per cent of the withheld cash benefit for a period of up to 5 years. Frank and McGuire (2003) note, however, that many of the social benefits (and public cost savings) can be realized over a longer period than that.

Short term contracts do in general lead to too weak quality incentives when non-contractible quality has long term effects. To illustrate in the simplest way, assume that the authorities choose the period 2 provider through a competitive bidding round at the end of period 1, and that the service provider in period 1
takes the probability of contract renewal, \( \tau < 1 \), for granted. One justification for an exogenous probability of contract renewal would be that there is a fixed cost of production in period 2, randomly determined for each potential producer at the end of period 1. The firm that realizes the lowest fixed cost wins the period 2 contract at price \( P_2 \), which is equal to the production costs of the second most efficient firm.\(^2\) Maximization of (7) then implies the f.o.c.

\[
(1 + r)p = \tau \left( \left( \frac{\partial P_2}{\partial \tilde{q}} \right) / s_1 - s' \left( H_2 \right) h' \left( \tilde{q} \right) \right).
\]

Spot contract competition weakens quality incentives in two respects. Firstly, \( \tau < 1 \). The incumbent provider may not realize the lowest fixed cost, in which case the contract is not renewed. If \( \tau = 0 \), for example because the number of competing firms is very large (and all have the same probability of realizing the lowest fixed cost), there is no incentive for providing quality. The incumbent cannot expect to keep the gain from lower future service costs.

Secondly, since the authorities allow for price competition one would expect that \( \partial P_2 / \partial q_1 < 0 \). Higher quality, by making it cheaper to provide the required service in period 2, lowers the bid from any potential provider. If the competitors are fully informed about the degree of service dependency in that period, and if quality reduces service costs equally as much regardless of who provides the service, then \( \partial P_2 / \partial q_1 = s_1 s' h' \), and there would be no incentives for \( q_1 > 0 \).

It is, however, profitable to choose \( q_1 > 0 \) if the future effects of quality are at least partly relation-specific. A hallmark of high quality human service is the ability of the provider organization and of individuals within it to learn from the interaction with its service recipients, and through that acquiring a better understanding of individual recipients and developing good working relations with recipients. As long as the workers remain with the incumbent provider, the relation specific future effects of quality are also firm specific. The reduction in the recipient’s future service needs that high quality service allows for is then partly contingent on provider continuity. Let \( \alpha \in [0, 1) \) measure the degree of relation specificity. A marginal increase in quality today would reduce future service needs by \( s_1 s' h' \) for the incumbent, and

\(^2\)It is assumed that all the firms know the realization of its own and of its competitors’ costs when they submit their bids, and that the firm realizing the lowest fixed costs always has the lowest total service costs.
by \((1 - \alpha) s_1 s'h'\) if a competitor takes over service provision in period 2. With \\
\(\partial P_2/\partial q_i = (1 - \alpha) s_1 s'h'\), equation (8) can be rewritten as \\

\[(9) \quad (1 + r)p = -\alpha \tau s' (H_2) h' (\tilde{q}).\]

Price competition reduces the incentives for quality since the future gains from 
quality (the left-hand side in (9)) is scaled down by a factor \(\alpha < 1\). Incentives for 
quality are not entirely eliminated only in those instances when there is some degree 
of firm specificity affecting marginal gains from quality, that is if \(\alpha > 0\).

The effect of price competition is here similar to that of cost plus contracts. In a cost plus contract with full compensation for higher cost, the cost gains from higher service quality would be exactly offset by reduced reimbursement resulting in no incentives for quality.

In summary, optimality requires that \(\alpha \tau = 1\). If either \(\tau = 0\) or \(\alpha = 0\), a 
for-profit provider would not invest in quality, that is \(\tilde{q} = 0\). Incentives for quality 
are eliminated, either because a competitor reaps the entire gain (\(\tau = 0\)) or because 
the authorities do through lower \(P_2\) (\(\alpha = 0\)).

Clearly, there is little economic rationale for short term spot contracts within 
this framework. In principle, strong incentives for quality are reconcilable with 
short term contracts, but it would require that the authorities used information 
about the provided quality to determine whether the contract should be renewed 
and at what terms. Since quality information is unverifiable, this would entail some 
sort of relational contracting. A long term contract, on the other hand, provides 
optimal incentives, resulting in a first best quality level, without relying on such 
unverifiable and noisy information.

This strong conclusion in favor of long term contracts rests on several strong 
and in many cases questionable assumptions. One of them is that the outcome, 
although realized only over time, is fully verifiable. In the model it means that the 
verifiable measure \(w\) is an adequate outcome indicator.

Case 3 Only \(w\) is verifiable. \(w\) does not indicate outcome.

Case 3 differs from case 1 and 2 in that the service outcome is not verifiable. Let \(W\) be the correct indicator (valuation) of the outcome, taking into account all
dimensions that the government considers relevant.

\[
W_t = W(s_t, H_t)
\]

\(W\) increases with \(s\) and decreases with \(H\). The outcome requirement \(W_t \geq \bar{W}\) is fulfilled with strict inequality. \(\bar{W} = W(s_t, S(H_t))\) defines

\[
s_t = S(H_t).
\]

\(S(\cdot)\) is a convex function similar to \(s(\cdot)\) in (2). I assume that \(S' = -\frac{W''}{W'} > s'\). (I make no assumptions about the level of \(S(H_t)\) relative to \(s(H_t)\) since it has no bearing on my analysis).

For any level of service dependency \(s(H_t)\) underestimates the social disadvantage of becoming more service dependent. \(S'\) measures the marginal increase in true social costs in period 2. These costs are higher than the costs \(s'\) which is implied by the verifiable measure \(w\). The higher cost stems partly from a non-verifiable increase in true service requirements. If a prisoner becomes more demanding (\(H_2\) increases) it is more costly to fulfill the true welfare standard (\(\bar{W}\)), for example, respecting basic human rights such as the prisoner’s legitimate need for social contact and physical activity. To the extent that it is not possible to verify the fulfillment of these requirements, they cannot be enforced through a formal contract. Similarly, former convicts committing new crimes increase law enforcement costs (police investigation, trial, and new imprisonment). If the cause of the cost increase is not fully verified, the provider cannot be held accountable. However, included in \(S'\) is also any non-verifiable welfare loss (measured in money terms) of a person becoming more service dependent that cannot be compensated through a higher service level. This welfare loss can be borne by the client herself, or by other persons, such as family members, or, for a criminal, by those hurt by the criminal offense.

The socially optimal quality level is found by optimizing (5) with (11) replacing \(s_t = s(H_t)\). This quality level, \(q^{**}\), satisfies

\[
p(1 + r) = -S'(H_2)h'(q^{**}).
\]

Since \(S' > s'\), the social gain from quality \(-S'h' > -s'h'\) for any \(q_1\). Consequently, \(q^{**} > q^{*}\). The optimal quality level in case 3 is higher than that which follows
from maximizing (7). A long term contract with a for-profit service provider realizes a quality level that is too low because only part of the costs of higher service dependency are borne by the provider. When only some of the beneficial effects of lower service dependency can be verified over time, any contracting regime that relies exclusively on court-enforceable contracts results in suboptimally low service quality.

Examples of non-verifiable long term quality effects are easy to find in the human services. Case 3 is a highly relevant case empirically. One example is rehabilitation of the elderly. The geriatric ward at the Huddiksva hospital in Sweden changed its care practices substantially in 1978 (Sundman (1990)). A new senior physician was employed that emphasized active rehabilitation of patients. The changes at the hospital and its perceived effects received national attention. The average length of stay in hospital was reduced, allowing the department to serve a larger number of citizens. The majority of the geriatric patients seemed to function better upon discharge. There were, however, indications that the new care practices did not function well for all. Though many were content, fierce criticism was voiced by some. Mortality also increased in the ward, and in the county, though the latter increase was not significant. What was the relation between care and these changes in the rate of mortality among the elderly? Sundman found it difficult to draw strong conclusions.

Patients' functioning and welfare are the result of highly complex and idiosyncratic processes. Separating out the effect of care from the effects of the illness or the injury and from other exogenous factors is difficult both at the individual level and in the aggregate. It cannot be done without subjective evaluation. The judgements made by peers, surveyors and researchers are rooted in professional and public discourse in addition to laws and regulation. The judgements provide information and insights that can be used in relational contracting, but has less relevance for formal contracts. Except for in cases of grave misconduct, suitable evaluations are not something courts can make.

High quality in the human services also has intrinsic value. Considering quality in terms of the process through which a service is given, the way a service is performed is important in and of itself, separate from any (likely) beneficial effects on future service dependency and welfare. In the care services respecting recipients' individuality and autonomy has intrinsic value. Such factors are difficult to verify,
let alone to measure. Process is equally important in law enforcement, for example in the respecting of civil rights. The intrinsic value of high quality is not explicitly taken account of in (12). It is an additional reason for taking quality considerations seriously, and since quality is non-verifiable it is also an argument against having a strong reliance on formal contracts.

3.4 Relational contracting

The alternative to long term contracts would be some sort of relational contracting where the authorities used their available information about service quality to determine if the contract is to be renewed and on what terms. Let this imperfect quality information be summarized in the belief \( q \in [0, 1] \). At the beginning of period 1 the government announces the following policy. If at the end of this period the provider is sufficiently high, \( q \geq q_{\text{min}} \), the provider is offered the period 2 contract. Conversely, if \( q < q_{\text{min}} \), the contract is not renewed. The assumption \( \partial Pr(\hat{q} \geq q_{\text{min}}; q_1) / \partial q_1 > 0 \) ensures that for the provider \( \partial \tau / \partial q_1 = \tau' > 0 \). Furthermore, \( P_2 \) must be set so that \( P_2 - s(H_2) = P_2 - s(\theta + s_1 h(\hat{q})) > 0 \), making the period 2 contract commercially attractive to the provider.

The provider maximizes profits wrt \( q_1 \), taking into account \( \tau' > 0 \) and the announced value of \( P_2 \). The 1.o.c. then becomes

\[
(13) \quad p(1 + r) - \tau' [P_2 - s(H_2)] = \tau - (s'(H_2) h'(\hat{q})).
\]

By making contract renewal contingent on available subjective information so that \( \tau' > 0 \), and allowing the provider to make extra-normal profits in period 2, the provider can be given strong incentives for high quality service. It is in principle possible to give the provider sufficiently strong incentives to realize the socially optimal quality level by letting \( \tau' > 0 \) or \( P_2 > 0 \) be sufficiently large.

The relational contract is essentially a performance pay contract. The provider gets the contract at commercially attractive terms as a reward for high quality in the previous period. Norton (1992) found that monetary incentives to improve nursing homes’ rehabilitation efforts improved residents’ health status while saving overall care costs, cf. section 2. The study is, however, not informative about the effects of the policy change for outcome dimensions apart from health status, access and discharge or for intrinsic quality. For example, there is no information on how
residents or their families evaluated the resulting care.

The performance pay relied on externally employed nurses evaluating residents’ health status at regular intervals. Discharge would also have to be approved by an independent party. This reflects the informational problems in assessing help dependency, which also increased administrative costs. For the same reason, the incentive system presupposes mutual trust and cooperation between the contracting parties since the metering of rewards unavoidably involves discretionary judgement.

An important issue that falls outside the scope of this analysis is what information the government should collect, and to what extent, that is, how to generate \( \hat{q} \). Some remarks on information costs are still required. Relational contracts can achieve the first best quality level \( q^{**} \), provided that the information costs are independent of the targeted quality level. If information costs increase the marginal social cost of quality, relational contracting could only realize a second-best quality level, which would be lower than \( q^{**} \). Furthermore, the social gain from increasing the quality level (by \( q^{**} - q^* \), if the first best is achievable) must exceed the total information costs in relational contracting. If so, relational contracting is to be preferred to a long term contract.

4 Broader perspectives

Long term contracts are problematic if the outcome cannot be indicated by a verifiable measure. A second objection to the use of long term contracts is that such contracts are difficult to implement and enforce in practice. This may result in hold-up problems. Long term contracts also interfere with consumer choice. I also argue that dynamic quality effects have implications for the understanding and measurement of efficiency in the human services and for assessing the relative merits of integrated service provision versus contracting out. Lastly, they may create adverse incentives for quality if a for-profit provider has market power.

4.1 Hold-ups

When it takes time for service quality to have an effect, the cost of providing quality is an investment. Consequently, hold-up problems may arise. Either party can in principle be held up by the other. The provider is held-up in period 2, if she chose high quality service in the first period in order to save future service costs, but the
service purchaser (the government) is able to seize the cost gains by breaking an agreement with the provider. The service purchaser is held-up in period 2 if the provider chose sub-optimally low quality level in the first period, and can renege on a promise to pay for the resulting increase in service needs. As an example of the latter, a nursing home may deliberately choose a quality level, \( q_1 < q^* \), resulting in sub-optimally high future care needs, \( s(H_2) \), if the firm believes that it can circumvent an agreement to pay for the associated increase in future care expenses. The possibility of hold-ups, by the government or by the provider, weakens provider incentives for quality.

Hold-ups are possible even when the service outcome is contractible (case 2) and the service contract is long term, because long term contracts are difficult to enforce in practice. For example, a nursing home filing for bankruptcy at the beginning of period 2 may not be able to fulfill its contractual commitments in that period. The possibility of bankruptcy therefore weakens incentives for quality.\(^3\) Replacement costs aggravate this hold-up problem by making bankruptcy more attractive. In period 2 it is cheaper for the authorities to renegotiate the service agreement than replace the incumbent provider. The provider can be compensated for the high costs (caused by low quality) and may even make extra profits because of the competitive advantage that the replacement costs represents. Replacements costs can be substantial, even when replacing a low quality provider, for example because the provider owns important capital equipment. A nursing home provider may be the only owner of a building that is a suitable location for a nursing home in an area. Such a situation is particularly likely if the contract horizon is long.

To the extent that these capital investments are specific, having higher value in this usage than in an alternative, and are important for service quality, quality incentives may be reduced by the reverse hold-up problem, which is that of the service purchaser holding up the provider. For the service purchaser it can be difficult to make credible any long term commitment not to hold-up the provider.

---

\(^3\) A financial guarantee from the provider could restore incentives. However, that amount would have to be sufficiently large so that the provider, for a given bankruptcy probability, would not find it profitable to deliver a quality so low that future service costs \( s(H_2) > \text{financial guarantee} \). The larger the required amount, the greater the gains are from quality \( (\partial s_2/\partial q_1 = -s'h') \) since a fall in \( q \) would result in a large increase in \( s(H_2) \). Credit market imperfections and incomplete information about \( s(\theta + s_1h(\cdot)) \) limit the feasibility of such arrangements.
This is particularly so when the purchaser is the government. The authorities have multiple roles, including that of the law maker and the judicial power. They cannot credibly commit not to change the regulatory environment, particularly not on behalf of future policy makers (Levy and Spiller (1996)). For many of the human services, the time horizon is long. It can last the length of whole life times. Therefore, the problem of long term commitment can be significant.

Furthermore, the nature of these services – that it is “people-work” – creates some additional hold-problems for the provider. High quality is an investment in people, improving the functioning or the capabilities of the recipients that the service contract concerns. For example, enabling a social disability beneficiary to qualify for a paid job increases the probability of employment for that recipient only. Consequently, the quality investments are specific to the recipients contracted for. The investments are inalienable to the recipients, it is not an investment owned by the provider.

This inalienability limits the extent to which formal contracts can be used to ensure that the provider is rewarded for high quality service. Firstly, it can be argued that it is morally (and politically) problematic to write long term contracts concerning the “performance” of individuals, literally attaching contractual commitments to people. Is it acceptable to reward (punish) a prison if a former convict is law-abiding (commits new crimes) during his remaining life time? It could also be practically difficult because the recipient, or others, could act strategically given such clauses. A former convict, threatening to commit new crimes, could blackmail the prison, or the prison management could kill or cripple him or her to avoid receiving fines due to future criminal activity.

Secondly, long term contracts would in some cases violate basic individual liberties that are protected by law. For example, a disabled person in need of assistance over a long period of time (perhaps his or her whole lifetime) would not be able move to another place if that would necessitate a change in service provider.

Thirdly, this interference with consumer choice is also problematic for efficiency reasons, particularly when service outcomes cannot be indicated by verifiable measures over time. Trade-offs between consumer choice and long terms contracts are discussed more in the next subsection.

Long term contracts are difficult for any contracting problem of some complexity. The alternative is relational contracting, or in the framework of section
3.4, short term contracts with contingent contract renewal. Such contracts provide greater flexibility in the face of non-contractible outcomes and in the use of subjective information. If the provider is to take into account non-contractible outcomes, some sort of relational contracting is required.

Hold-up problems are potentially more important when the outcome is not fully contractible (case 3), for example, when a care provider chooses not to comply with an informal agreement about a minimum quality level. This contributes to higher future costs. However, for the same reason that the government cannot take the provider to court for contractual non-compliance, there might be costs on a political level for the government from punishing the provider. Even if the government knows that costs have increased because quality has been substandard and not because of other factors for which the provider has no responsibility (θ), the government cannot credibly convey this subjective information to the public. Consequently, the authorities may prefer not to punish the provider, but renew the contract and compensate the provider for the cost increase.

Relational contracting does not eliminate potential hold-up problems, but relies explicitly on mutual trust to deal with them. Since ∂τ/∂q_1 > 0 is an informal rule, and not legally binding one, the opportunity for the government to hold up the provider appears to be at least as great as with a long term contract. However, relational contracts can also be used to reduce directly the risk to the provider of hold-up by the government. With long term contracts, the rewards of high quality take time. How long of a time it takes depends on how quickly service needs, s (H_t), respond to changes in the quality of service. With relational contracts, rewards can also be based on informal information about present quality, and not only on s (H_t). Thus rewards to the provider can be made more compressed in time, reducing the requirements on long term commitments.

4.2 Consumer choice

My analysis demonstrates that as long as the outcome over time is fully verifiable, a long term contract leads to socially optimal quality incentives, even when service quality itself is unverifiable. The lack of consumer choice that this contract implies is then not a problem. However, the service recipient may be better informed about the appropriateness of the actions taken by the provider, that is, in judging quality,
while also finding it difficult to convince the provider about this. Not having the possibility to switch provider gives the recipient little strength in voicing complaints. Of course, if the outcome is not verifiable, the lack of consumer choice is much more problematic. Consumer choice is the most basic market mechanism for sanctioning low quality, and it is often the only effective sanction in markets where quality is unverifiable.

Even in those settings where rational consumer choice is limited, most service recipients can to some extent, if given the opportunity, assess unverifiable quality dimensions in some way or another, either personally or through representatives (for example, family members). It is probably for this reason that the Ticket To Work program allows service recipients to choose (from a list of qualified job training agencies) the provider with whom to write a contract, and to terminate the agreed training program on short notice if dissatisfied with the service. The effect in many cases is probably the intended one with consumer choice leading to a competitive pressure that improves the quality of training. In recent years, however, the composition of disability income recipients in the U.S. has changed, with an increasing number of persons with mental problems becoming beneficiaries (Frank and McGuire (2003)). For many of these, job training is also social training for work life, as in having to get up in the morning, to commit themselves to do assigned work tasks, or to accept the authority of others in the workplace. Such training is not always pleasant, and trainees can at times lose their motivation. The opportunity for the service recipient to end the contract on short notice, though intended to enhance the quality of training, is also in some respects problematic. In the presence of hyperbolic discounting, or other forms of irrationalities, it reduces the training agencies’ incentives to effectively acquaint some recipients to the demands of a regular work life. A training agency placing much effort into this risks losing its investment through the recipient terminating the contract.

This tension between the benefits of long term contracts on the one hand, and consumer choice and flexibility on the other, is also an issue in health insurance literature. Mutually binding longer term contracts reduce the problem of adverse selection by consumers or providers, and longer contract periods also improve on inter-temporal insurance. The resulting lock-in of consumers, however, mitigates competition when important product dimensions are not contractible. For a discussion of this see Cutler and Zeckhauser (2000). Inter-temporal quality effects are an
additional reason for having a longer term perspective in health insurance. Short contract periods reduce the insurer’s incentive for preventive care since frequent recontracting facilitates the readjustment of premiums to new information about risks. The risk of losing a customer, in which case the cost savings from preventive care accrues to another provider, also reduces incentives for such efforts. In the real world, contract periods are often quite short. In the US, the contract period is usually for one year, and very long term contracts do not exist.

4.3 Service efficiency and its measurement

Attention to the inter-temporal effects of service quality may also contribute to a richer understanding of service efficiency and how to measure it. When calculating the social efficiency of a prison, comparing output to costs, one must include in its output not only the detention of the prisoners, preventing escape and fulfilling welfare requirements, but also the prison’s success or failure in rehabilitating prisoners. Likewise, a nursing home must be judged not only by how efficiently it can care for a group of patients, but also by its ability to improve or maintain their future functioning. However, the effects on future service needs (or other types of social costs) are seldom taken into account. It is, for example, generally ignored in the literature on nursing home efficiency, see Rosko et al. (1995) and other studies referred to therein. As a consequence, the social efficiency of high quality service providers tends to be underestimated.

4.4 Should public services be contracted out?

Hart et al. (1997) deduce theoretical conditions for when either private or public production of services is to be preferred. The model is designed to fit prisons in particular, but the authors also find it useful for analyzing other traditional public services, such as schools and health care.

Hard to observe and non-contractible inter-temporal quality effects are highly probable in detention institutions. If a prison is badly run, it is likely to contribute to increased social problems and hostility among the prisoners. One might say that the prisoners’ social capital, which presumably is generally low already and vulnerable to external influences, deteriorates further. As a result more resources are required to run the prison (if service standards are to be maintained) or alternatively, qual-
ity, in terms of security and prisoners’ living conditions, is lowered even more. Hart et al. (1997) claim to “show that our model can be used to organize, if not resolve, the debate over prison privatization.” In my view, their model is not general enough for that purpose. Taking into account these potentially important inter-temporal quality effects may alter policy conclusions. For example, Hart et al. (1997) in reviewing quality standards for prisons find that contractual incompleteness is particularly important for rules regarding the use of force and the quality of personnel. They conclude that the case against privatization is stronger for maximum security prisons than for private half-way houses and youth correctional facilities. However, the effects of low quality (in the respects pointed to by Hart et al. (1997)), by building down the prisoners’ social capital, may be equally or more important in youth correctional facilities, and possibly also in half-way houses than in high security prisons.

4.5 Provider market power

Inter-temporal effects of quality may actually give incentives for low quality if the service supplier has market power. Sufficiently low quality (by definition) expands future service needs. For example, in a market for “prison services” low quality, to the extent it reduces the probability of rehabilitation, expands the future market as long as the authorities maintain service standards and adapt capacity to changes in the volume of prison sentencing. Knowing this, a provider with market power that sets the service price as a mark-up over costs can increase his or her period 2 profits by having low quality in period 1. I am not aware of prison procurement contracts that give the supplier a bonus if a released prisoner avoids being sentenced for committing new crimes, or that possibly increase the bonus as well if this former convict actively works for reducing criminality in general. (Writing such a contract would of course have its difficulties.) In the absence of such contractual clauses, however, it would be in the prison firm’s interest that a former prisoner continues

4In both youth correctional facilities and in half-way houses prisoners’ social capital may be more easily influenced by the quality of the service (e.g. the use of force and the quality of the personnel) than is the case for prisoners in maximum security prisons. In addition to this, the lowering of prisoners’ social capital has social costs in terms of increasing future criminality. Such costs are likely to be higher the younger the criminal is.
his criminal career and that he recruits more people into criminal activity etc., thus increasing the probability of him or others being imprisoned in the future.

Provider induced demand is well-known in health economics literature (see McGuire (2000)). To increase his or her own income the physician, who is better informed than the patient and any third-party payer, manipulates the patient to increase her demand. The physician induces an increase in perceived care needs. When the service provided today affects future service needs, another form of supplier-induced demand is possible. In this case real care needs increases. Furthermore, it is an induced demand mechanism by which the clients concerned as well as the wider society incur substantial welfare losses.

5 Conclusion

The issue of quality is at the core of any comprehensive analysis of the human services. I have emphasized three characteristics of service quality and of the context within which the service takes place. Firstly, the quality of the service affects the needs for that service or related services in the future, and is often of crucial importance for the welfare of those affected. Secondly, those directly affected by the service have limited ability or opportunity to demand quality according to standards, either through voice or through choice of provider. Thirdly, quality is unverifiable and can only be imperfectly observed by outsiders. The last two points imply that monitoring by outsiders is often the only way to create external pressure for quality, but that this is difficult.

A premise in my analysis has been that the main dimensions of quality are contained in its inter-temporal effects. When this is the case, and the outcome can be verified, sufficiently long term contracts can make the provider internalize the incentives for quality. However, long term contracts are problematic for a number of reasons, and this limit their practical relevance. Also, important outcome dimensions cannot be verified in many cases. An alternative to long term contracts is relational contracting, where rewards and contract renewal are made contingent on subjective quality information. I have argued that the advantages of this contract design are more robust to changes in the model’s informational assumptions than are long term contracts.
References


