UK Budgetary Systems and New Health-Care Technologies

A. McGuire, BA, MLitt, PhD
LSE Health and Social Care, LSE, and Kings College, London, UK

ABSTRACT

Objectives: This article outlines the budgetary setting within the UK health-care system.
Methods: It is argued that while prospective budgets can give rise to efficient resource allocation outcomes, this relies on the budget being set at an appropriate level and the accompanying incentive structures being efficient. The organizational structures and the interrelationships are critical. The recent history of UK National Health Service reforms and expenditure is outlined. It is suggested that until recently, although the budget system has the potential to promote efficiency, the aggregate budget allocated to the NHS has probably been too low given public expectations, technology advances, and preferences for health care.
Results: The aggregate budget is due to rise considerably over the next 5 years. While some incentive and regulatory provisions will move the budget toward an efficient allocation many microlevel incentive issues remain.
Conclusion: Whether efficient patterns of health-care allocation emerge remains open to debate, however, because the existing incentive mechanisms are not optimal.

Keywords: health-care budgets, health-care technology, pharmaceuticals.

Introduction

Introducing new health-care technologies, for example, new pharmaceutical products, is normally expensive. Either the unit cost of these new technologies is high, partly reflecting high fixed research and development costs, or the volume of uptake may be high. The budget impact can consequently be considerable. Conversely, budget constraints may inhibit the benefits to be gained from a new technology as uptake is suppressed. The issue of concern here is how a limited budget impacts on the diffusion of new technology with special reference to pharmaceuticals. It is important to emphasize that interest is focused on the pursuit of efficiency. Economists have a number of definitions of efficiency. Of importance, there is technical efficiency, ensuring that maximum output is produced from inputs; productive efficiency, ensuring that the maximum output is produced at minimum cost; and allocative efficiency, ensuring that the appropriate mix of outputs is produced. These definitions of economic efficiency differ from what may be termed budgetary efficiency. Budgetary efficiency can be thought of as a provider of health care attempting in principle to produce as many outputs as possible within a given budget. However, staying within budget is crucial. Output is of secondary concern; budgets may therefore distort economic incentives.

It is argued here that the UK health-care system gives rise to a mixture of incentive structures reflecting the pursuit of budgetary efficiency at an aggregate level and economic efficiency at the purchaser/provider level. At the individual clinical level, which is the main decision-making level with respect to the actual transformation of budgets into choices about the use of resources, it is budgets that matter and therefore expenditure levels rather than outcomes that dictate behavior. The actual budget allocation mechanism, the ring fencing of budgets, the fact that prices are set and volumes, which are generally less flexible, must adjust as the budget changes. Thus the specific regulatory characteristics of the UK system all combine to place emphasis on budgetary efficiency rather than economic efficiency. Cost levels rather than output mixes dominate. To pursue this argument we describe the various levels of expenditure operating within the UK health-care system, outlining the relevant incentive structures.

UK Health-Care System

The Department of Health and various bodies within the National Health Service (NHS) manage
the health-care budget in the United Kingdom. The NHS budget is financed centrally largely through taxation, although some charges are levied—most importantly on pharmaceutical prescriptions. Aggregate funding for the NHS has historically been determined through the Department of Health negotiating with the UK Treasury for an annual spending level. Recently this process has been adapted to reflect the devolution of responsibilities for the NHS to other bodies in Scotland, Wales, and Northern Ireland and to incorporate longer-term expenditure plans over a 3-year period. This total budget is then allocated to the various commissioning groups in the health service according to a formula reflecting the needs of the local population. These allocations are broadly based on standardized mortality rates with some further correction for expenditure equity considerations.

Recent organizational changes in England have included the formation of primary care trusts (PCTs), which have become the purchasing, termed commissioning, agents responsible for the health care of their populations. These PCTs are amalgamations of individual general practitioner practices, which will service the population on their combined patient lists. PCTs are expected to provide primary care to their populations and to enter into long-term relationships with NHS Trusts, which are largely individual hospitals providing secondary or tertiary level care, to provide any higher-level health care for their populations. PCTs are now being established throughout England and the funding allocation for 2003 to 2004 onward will flow straight from the Department of Health to the PCTs.

The current trend is for the purchasing authorities, the newly formed PCTs, to act in strategic collaboration with their local Strategic Health Authority offices, to determine purchasing priorities and strategies. These must confirm to some degree to the stipulations of national guidance on priorities. Currently national guidance covers such areas as the reduction of waiting lists and times for various secondary care services; measures aimed at reducing inequality of care; targets aimed at reducing mortality and morbidity from specific diseases, such as coronary heart disease; and measures to improve services generally in specific disease areas, such as cancer. PCTs are then meant to develop their own individual plans with specific reference to these national priorities. These local responses are likely to vary both in their content and in their implementation. There is no clear model of PCT level priority settings, which can be taken as illustrative of the general approach. Each group will respond to its own environment.

Within this general framework there are also a variety of more specific regulatory policies. Of direct concern to the pharmaceutical sector, for example, are the regulation of pharmaceutical prices and products and the regulation of treatments utilized by the NHS. All pharmaceutical products are regulated to ensure safety and efficacy before a license for their use is granted. Furthermore the profits of the pharmaceutical firms on NHS sales are regulated through the Pharmaceutical Price Regulation Scheme (PPRS). This is a voluntary scheme that operates between the Association of the British Pharmaceutical Industry (ABPI) and the Department of Health, albeit with statutory backing for the government to fine companies that do not comply with the rules. Each participating company is allowed to earn a given return on capital in the United Kingdom equal to profits from NHS sales minus allowable costs. As the scheme operates on overall profitability it is possible for companies to introduce new products at a premium price while reducing the price of older products. Pharmaceutical products also currently feature prominently in the guidance issued by the National Institute for Clinical Excellence (NICE). NICE is a national body introduced to evaluate the clinical, cost-effective, and more general patient preference based arguments for NHS funding of individual treatments and services.

**Budgetary Operations**

Having outlined the organization of the UK budget process we can consider both the conceptual basis of how any budgetary process operates, in terms of the major impact it may have on costs and volumes of health care delivered, and the incentive mechanisms that are embedded within any budgetary regime.

A budget defines the total expenditure available. Formally the budget defines the total expenditure given over to any activity, in this case health-care delivery:

$$B = \sum_i P_i Q_i$$

where $B$ is the budget, $P_i$ is the unit price of the service or product delivered $(i)$, and $Q_i$ is the quantity of service or product provided. Reimbursement mechanisms apply individually to $P_i$, to $Q_i$, or to a combination of both $(P_i Q_i)$. The various applications will be associated with different incentive
mechanisms and may affect different levels of the organization. For example, setting different fixed prices for different treatments will impact on \( P_i \) and influence the provision of individual treatments. Budgets may be set at departmental levels. Constraining \( P_i Q_i \) at the departmental level will affect the quantity of service delivered, and if \( P_i \) is correlated with the quality of service, the quality also. Global budget setting constrains the aggregate across departments

\[
\sum_i P_i Q_i.
\]

Moreover, the budget can be set retrospectively or prospectively. There is no consideration given to individual prices, volumes, or total expenditure. By their nature, retrospective regimes have no incentive powers as they reimburse fully all costs. There is no incentive for the provider to be cost conscious, nor to provide the appropriate quantity and quality of care. If costs are fully reimbursed then there is a tendency to overprovide quantity and quality of care. As a consequence most health-care systems have moved to prospective payment systems.

Under a prospective system the budget is predefined in one of three main ways. First, the overall budget may be set before any service delivery with the understanding that all population requirements are to be met. In this case a lump-sum payment is made in exchange for access to services. This is a very powerful incentive scheme with no de facto reimbursement of costs.

A second approach, cost and volume contracts, combine preset volume targets with some limit on the cost (i.e., the price, \( P_i \)) of each service provided. This involves a fixed payment to treat patients up to a given volume above which there is either no reimbursement or the next form of budget agreement is implemented. Again this is a high-powered incentive contract. Third, cost per case contracts can be struck where the costs, or prices, of individual episodes of care are specified and either the volume of each individual service is negotiated as required or there are additional agreements over workload. The incentive power of this form of contract obviously depends on the cost levels specified and the volume agreements.

In all prospective cases, although the incentive mechanisms differ in specific ways, the financial implication is that the risk is being pushed back from the third-party payer of health care to the provider of health care. This is precisely what occurs with prospective budgeting. When budgets are allocated, the provider carries the risk of running over budget; however, at the same time they are free to allocate resources as they deem fit. Of course, two problems arise. First, to stay within budget the provider may choose inefficient mixes, in economic terms, of health-care inputs to produce outputs. In terms of the definitions given above, technical and productive efficiency may override budgetary efficiency. Second, as the budget constraint is neared, activity may slow down to stay within budget. In other words, while prospective budgets can be an efficient resource allocation mechanism that provides the latitude to the provider to behave in an efficient manner, this will only occur if the appropriate incentives are maintained and the budget is set at an appropriate level for the attainment of desired levels of output.

Global Budget Aspects

As noted above, in the United Kingdom the budget operates at the global level, but also at the regional and individual trust level. Such budgets are prospective in each case with the total health-care budget

\[
\sum_i P_i Q_i
\]

being determined by central government expenditure rounds. This overall global budget approach has kept the UK health expenditure at historically low levels compared to other industrialized countries.

Currently the United Kingdom spends approximately 7% gross domestic product (GDP) on health compared to an OECD average of approximately 10%. Rates of growth of health-care expenditure have been relatively high compared to the growth rate of GDP. However, this is not unusual. Most industrial countries have higher growth rates in health-care expenditure than that of GDP.

The 2002 UK Treasury budget, which outlines tax levels as well as major expenditure commitments, indicated that by 2007 to 2008 UK health spend will rise to 9.4% of the GDP. This represents a growth rate of 7.4% per annum in real terms over
the next 5 years, taking UK health expenditure from £65 billion in 2001 to 2002 to over £105 billion in 2007 to 2008. This is an unprecedented increase in the tax funding allocated to health care. The efficiency with which these monies will be allocated and spent does, however, depend on the within-system incentives—an issue returned to later. The government has concentrated less on setting new incentive mechanisms than by decreeing that certain targets, especially waiting list targets, must be met. Instead of creating incentives to attain these targets the predominant mechanism has been the imposition of penalties if they are not met. Currently there is widespread criticism that the targets will not be met, even with the dramatic increase in the budget, there is not enough flexibility in the volume of resources to meet the target levels. Staff shortages remain problematic and productivity and activity responses have been negligible. The lack of increase in measured activity may also reflect a shift in the input mix, with more money being spent on pharmaceuticals in primary and secondary care. The number of episodes of hospital care is the main measure of NHS output, and more effective use of medicines would not lead to a change in output on this measure.

It is instructive to analyze past expenditure in a little more depth. It is possible to analyze this at an aggregate level by breaking down the individual components of

$$\sum P_i Q_i$$.

Growth in spending could be explained by either rising prices or unit costs or increasing volumes of health care delivered. UK health-care expenditure is disaggregated in Table 1. This is based on a simple decomposition of expenditure as based on a statistical identity. The volume series is defined by abstracting out the relative price series from the expenditure series, thereby embedding any measurement error in the volume series residual. The relative price series states whether prices in the health-care sector are growing at a generally faster rate than in the economy. The volume figures indicate whether output in the health-care sector is rising faster than in the economy as a whole. In general, the growth in relative prices was modest, until the 1990s. For the period since 1985, the expenditure growth can be explained by similar growth in volume delivered as in relative prices growth.

Relative price growth does not appear to be the major explanatory factor determining expenditure. On the volume side the accompanying expenditure growth does not appear to be due simply to demographic or technology-induced pressures, which are generally said to total no more than 1% per annum. In other words, historically there appears to have been real growth in the health-care budget. This growth is greater than that in the GDP, but may still have been lower than necessary to attain the desired output of health care.

The decomposition of pharmaceutical expenditure is particularly interesting. Pharmaceutical expenditure accounts for approximately 13% of total health-care expenditure and has remained relatively constant over time. In contrast the volume of pharmaceutical use, as measured by community prescription, has grown from 488.2 million in 1992 to 613.1 million in 1999, a growth of just below 80%. Recent evidence is that prescription volume is growing at a historically high rate, 5% per annum compared with around 3% per annum during most of the 1990s, possibly reflecting improved treatment guidelines in some major therapy areas [1]. Yet the growth in total prescription costs fell in 2000 to 2001, largely reflecting increase use of generic compounds. It has been suggested about 50% of this growth is attributable to growth in volume, while 50% is due to product mix changes reflecting the use of newer more expensive medicines. Indeed, aggregate price growth is negative, reflecting PPRS constraints on price increases and increasing use of generics.

Therefore, a general conclusion may be reached then that the pressure of the UK aggregate budget is felt on the volumes of health care delivered. Given a negotiated national level of health-care expenditure and negotiated prices for most categories of health-care input, for example, labor and pharmaceuticals, then the only scope for manipulation is through local pay negotiations that supplement, to a very

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Price and volume growth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real health-care expenditure growth</td>
</tr>
<tr>
<td>Expenditure growth</td>
<td>4.4</td>
</tr>
<tr>
<td>Relative prices growth</td>
<td>8.7</td>
</tr>
<tr>
<td>Volume growth</td>
<td>–4.0</td>
</tr>
</tbody>
</table>
limited degree, the national agreements or, more importantly, through the volume of inputs used in the delivery of health treatments. In short, budgets are set and prices are set, so volumes are the adjusting factor. Even with increasing budgets, as is currently being experienced by the NHS, financial difficulty may occur if output targets are set at levels that require even larger budgetary increases. This is what appears to be currently occurring within the NHS as the government sets, for example, waiting list targets, which are incompatible even with the large increases in budget recently announced.

Regional Budgetary Aspects—Over to PCTs

The aggregate health-care budget, under the current reforms of the UK NHS, will be allocated to PCTs. These are merged primary care groups that will be responsible for the provision of primary care and purchasing of health care from secondary providers for their populations. The allocation of budgets to PCTs is largely through capitation payment. That is, the population served by the PCT is estimated, weighted based on standardized mortality ratios (SMRs) and deprivation indices, such that the final capitation payment takes account of population health as proxied by SMRs and deprivation. At this level, there is little incentive structure imbedded within the budget allocation. Proposed changes in the near future will retain the capitation basis of payment but this will be weighted by volume, deprivation, and a quality-of-service index.

The major form of treatment in primary care is the prescription of drugs. Unless the individual is exempt through age, income level, and pregnancy or as they have a chronic illness, a patient charge is levied on the prescription. Approximately 75% of patients requiring prescriptions are exempt. The charge has risen steeply since 1979. Hughes and McGuire [2] examined the impact of the increase in prescription charge finding that the implied price elasticity was −0.37, which had become more elastic over time. In other words a 1% increase in the prescription charge leads to a 0.37% fall in the demand for prescriptions. Out-of-pocket costs thus do have an impact on the demand for prescription-based treatments. The primary care physician is constrained in their choice of prescription treatment through the operation of a selected list of unapproved medications—although this has not been extended recently—is further informed of their choice of prescription drug for individual disease areas by national information on the volume and branded or generic type of prescription delivered by each practice. The budgetary impact of general practitioner (GP) practice prescribing can therefore be seen immediately. However, this expenditure information is not integrated with health outcome data.

In choosing where to purchase secondary care for their patients it is likely that PCTs will opt to maintain historical referral relationships with hospitals, although it is possible that increasing information on the quality and cost of treatments provided by the secondary care providers will impact on this decision. This point is taken up later, although it should be noted here that there are financial incentives being introduced at the PCT level to increase service quality and reduce waiting times.

Unit of Delivery Budgetary Aspects

At the level of health-care delivery the PCT acts as a purchasing agent for its population, both providing in-house health care and purchasing secondary care from hospital trusts. It is anticipated that there will be a degree of competition among secondary providers at this level. Delivery of the optimal volume of health-care output requires optimal input. It is envisaged that an evolving payment mechanism will relate inputs to outputs.

An objective of any payment mechanism is to provide incentive signals to ensure optimal effort, essentially optimal quantity and quality of care, is delivered under specific conditions. In the case of health care the delivery of optimal effort is difficult as there are several informational problems, for example, over quality and cost of service, which render the contracts between purchasers and providers difficult to specify. Ideally such contracts would be complete contingent contracts, in which all conceivable contingencies are prespecified and strictly enforced. This is also particularly difficult to accomplish in the health-care sector because it is difficult to both fully observe and verify the relationship between effort and outcome.

The fundamental problem is that the health outcome is to a significant degree stochastic. That is, there is considerable uncertainty over the impact that treatments may have in each individual case. This uncertainty may be associated with individual characteristics, for example, treatment compliance, or with inherent characteristics of health-care delivery. As diagnosis and treatment become more complex then the uncertainty over outcome increases. Different cases treated in the same manner, with the same inputs, may have different outcomes; those
treated differently may have the same outcomes. As extreme examples consider that exactly the same surgical intervention may lead to either recovery or death or that myocardial infarction patients treated medically or with surgery may have similar prognosis.

In such circumstances it is difficult to specify at the individual level the relationship between health-care inputs and health-care outputs. In aggregating up to the population level this problem remains. Given that the relationship between health-care inputs, that is, the optimal combination of staff, capital, and materials, is also not well understood, the optimal manner of production and organization is also to some extent stochastic. All of which results in the specification of appropriate cost levels being extremely complex.

The design and implementation of health-care budgets is thus clearly a difficult issue. As noted previously, setting the budget to reimburse a health-care provider for its total costs, or expenditure, creates no incentive to minimize costs. Such a reimbursement mechanism will encounter technical inefficiency—the quantity of care supplied will be greater than is optimal—as well as productive inefficiency—quality may be higher than optimal. Cost-saving measures are unlikely to be introduced.

Acknowledging all of the problems outlined above, it is still the case that reimbursement must be related to the costs of production in some manner. A typical scheme reimburses costs through the following general approach:

$$\sum_i PQ_i = a + b[TC]_i$$

where \( \sum_i PQ \) is again the budget, which may now be thought of as the reimbursement revenue, \( a \) is a fixed fee, and \( b \) represents a proportion of the costs (\( TC \)) reimbursed. Two extremes exist. First, there is the cost plus fixed fee or cost-plus contract (\( b = 1 \)). The provider bears no costs. This creates low-level incentives. Alternatively, there is the fixed-price contract (\( b = 0 \)). The provider is responsible for and claimant of all cost savings. No costs are reimbursed. A fee is paid, creating a high-powered incentive. Linear contracts (where \( b \) has a slope of between 0 and 1) are called incentive contracts as they share the degree of risk taken by the provider and the reimbursed. Obviously costs must be observable and some notion of minimum cost must be paid. Difficulty in specifying the production/cost relationship is a further complication in having the optimal reimbursement level be specific.

One of the major problems in this area is that of moral hazard; agents may have better information than the principal over their performance. In particular, unobservable behavior/hidden actions may be present. A health-care provider, be it a doctor or a hospital, therefore may not be cost-minimizing in producing health care. Because the health-care production technology is difficult to specify it may be impossible for those funding to observe whether behavior is efficient. A second major problem is patient selection and cream-skimming. GP practices and hospitals may seek only to treat those patients for which they know cost is below the reimbursement level. In both types of cases it is difficult to specify optimal reimbursement to ensure that the most efficient outcome is attained.

The problem can be stated in a more general manner. The principal, in our case the commissioner cannot fully observe or verify the agents’, or health-care providers’, actions. The outcome of the agents’ actions may be well observed, but the relationship between the actions and the observed outcome are stochastic; in other words, the agents’ actions do not fully determine the observed outcome. In such circumstances the principal must design the reimbursement package in such a way that it indirectly gives the appropriate incentive to the agents to undertake actions, which are compatible with an efficient outcome. That is, the reimbursement package must move the agents to take actions that would be contracted for if their actions were observable.

Even in theory it is difficult to specify optimal incentive contracts in such circumstances. It has been shown that if reimbursement is linked to the quantity of health care delivered, quantity of health care delivered correlates positively with quality, and commissioners can observe this correlation, an optimal incentive contract will include a fixed cost reimbursement element and a variable cost element, to vary with quantity of care delivered, reimbursed at some proportion of full cost. If the quantity–quality correlation is not easily observed, for example, if there are many dimensions of quality, or the quantity–quality correlation is negative, then the optimal incentive contract is unknown [3].

In such circumstances, given no undermining of competition among hospitals through the formation of cartels or selection of patients, basing reimbursement on yardstick competition may be best. This is the basic principle behind diagnostic-related group (DRG) reimbursement and is also the underlying principle behind average cost pricing and health-related group (HRG) pricing, which is currently being implemented in the United Kingdom. The
basic idea works as follows: if total revenue for each hospital is set equal to the average cost of a sample of similar hospitals plus a lump sum transfer to cover fixed costs, this provides an incentive to each hospital to cut costs to below the average, through increasing their investment in cost-reducing technology, to generate a surplus. This brings the sample average cost down over time and this incentive continues until $p = MC$.

If the number of other identical hospitals is large there is little efficiency loss from simply setting price equal to the average across all hospitals. As stated by Schliefer [4], who first proposed the idea, considers the case when transfers are allowed and shows that even here prospective payment based on observed average cost is second best. Of course the payment can be related to similarly defined treatments, for example, DRGs, rather than operated at the aggregate hospital level. This is the basis of Medicare payments in the United States. Thus, a fixed pricing rule based on observable data can lead hospitals to efficient production. In other words, prospective pricing based on DRGs, with some additional adjusters that are not patient-specific, is close to optimal.

In the context of primary care it is harder to apply these principles. DRGs or their UK equivalent have not been developed for episodes of ambulatory care activity. Nor is it clear that there is enough measurement of health outcomes or a strong enough evidence base to enable commissioners to draw clear conclusions between activity and likely quality of health outcome. More work needs to be done developing disease-based guidelines using evidence of cost-effectiveness and of cost before a DRG equivalent can be put in place. In the meantime, an emphasis on comparing GP practice pharmaceutical expenditure with local and national averages, in the absence of evidence on the quality of outcomes, could create “silo” effects, achieving budgetary efficiency at the expense of economic efficiency.

**Individual Actor Incentives**

Below the purchaser/provider level is the myriad of decisions taken by the most important single agent within the health-care sector—the medical specialists. UK NHS hospital specialists are paid through salaries linked to national scales. These salaries are not linked to workload, although there are expectations about time commitment and caseload volume, which are implicit to their contracts and may be explicitly stated at the departmental level. Each specialist or team manages the departmental or team allocated budget within the hospital. This budget is normally decided by historical precedent, political concerns, and maneuvering by the individual specialists within the hospital. Financial incentives are therefore nonexistent and indeed budgetary incentives are largely inadequate. Indeed the incentives may be perverse in some instances; consultants who increase their workload may not be provided with extra resources and indeed may be curtailed. For example, surgical volume may be negotiated at the beginning of the financial year with the PCT purchaser and a prospective payment put in place; if the surgical team meets the negotiated target they can be asked not just to slow workload down but to stop taking patients altogether.

In choosing treatments information is directed to the commissioners and providers by a number of central bodies including the Department of Health and the NICE. Both bodies are concerned with the effectiveness and the cost of treatment. NICE recommendations must be funded by individual commissioners but their employed specialists need not follow recommendations dogmatically. It is also true that the data on both costs and effectiveness of individual treatments, while improving, are far from satisfactory. Information on costs is improving, but often the information relates to tariff prices rather than the true costs and is not joined with effectiveness data. The latter is difficult to collect, as it is difficult to measure health outcome. Budgetary management therefore tends to be dominated by expenditure flows and the requirement that hospitals maintain financial solvency. Indeed NHS hospitals are required to break even. Centrally imposed rates of return on capital restrict access to and movement of capital itself, and most importantly, after allowing for capital adjustments prices for cost per case payment must be set equal to average cost. As Propper et al. [5] note, such a regulatory framework makes the hospital vulnerable to short-run changes in income, thereby providing an incentive to break the regulatory requirement that price equal average cost. Where there are joint costs, and the apportionment of costs to any individual service is not clearly specified, it is difficult to verify that the regulation is being maintained and considerable cross-subsidization of costs is likely, all of which weakens the incentive structure at the individual departmental and specialist level. If the financial solvency of the hospital is paramount, then it is unlikely that individual incentives, including reward for good practice, will be allowed to undermine this.
Clinical Efficacy, Cost-Effectiveness, and Budgets

The general environment defined by the NHS is one of tension between central bodies such as the Department of Health and NICE giving strategic advice and the day-to-day expenditure flows being controlled by a myriad of different channels: individual clinicians, individual hospital trusts, and individual PCTs. Although the system of integrated or unified budgets mean that PCTs can move money between health-care providers to produce the most efficient mix of services to deliver health outcomes, this does not appear to be happening in practice. Moreover, once money is distributed, there is little incentive for individual actors in the system to consider the systemwide effects of their expenditure decisions. There is still little scope and less incentive to vie across budgets. The result is that systemwide efficiencies tend to be lost. Seen in aggregate, the NHS system appears to be a low cost in a budgetary sense, and some might even argue a relatively low quality system of high waiting lists for certain procedures and restricted access to some treatments, but in a budgetary efficient system, in that the global budget is not exceeded.

Whether this will change in the near future as the NHS experiences a dramatic rise in its budget and the new commissioning arrangements develop is an open question. The increased budget will be accompanied by the introduction of the HRG payment system, increased quality control through bodies such as NICE and CHI, and increased individual autonomy. Whether the centralized regulatory bodies and a hospital payment system based on averaging reimbursement will end up being compatible with individual consultant and PCT expenditure decisions remains to be seen.

As the cash gets closer to the health delivery point, inefficiencies do currently materialize. To be crude, the system of expenditure determination in the United Kingdom is volume led at the hospital level with the volume determined through negotiation between commissioners and providers, but really largely reflecting historical trends, with total payments reflecting the amount of health care to be delivered at predefined average cost. At the primary care level payment is through capitation and therefore does not even reflect volume delivery, but is constrained by the capitation payment. Given that the overall health budget is constrained at the central government level and prices set by average cost or capitation rates, then the volume of delivery is likewise constrained. At the primary care level, individual GP income to treat patients is given determined largely through the capitation payment less any costs. Given these payment systems, there is little incentive to introduce new health-care technologies or prescribe effective therapies if these result in heavy budgetary impact.

There has been a strong emphasis on constraining expenditure on pharmaceuticals, with the GPs first receiving information on the costs of their prescribing in the late 1980s. This was followed in 1990 by the introduction of indicative prescribing budgets for GP practices. All health authorities were required to introduce prescribing incentive schemes providing extra money for GP practices that under-spent their drug budgets. More recently there has been more emphasis on the quality of prescribing. In addition to the requirement to fund NICE recommendations, prescribing schemes set by PCTs typically reward GPs for improving the quality of prescribing as well as reducing cost [6]. The new government contract with GPs, which is about to be finalized, includes proposals to pay GPs additional salary, as opposed to making money available to the GP practice, for hitting evidence-based quality prescribing targets. How the new emphasis on quality, and by implication on the use of new expensive technologies when they are cost-effective, can be made compatible with the continuation of the traditional budget constraints is not yet clear.

Recent work on coronary heart disease has documented the low level of take-up of new health-care technology in this area by the United Kingdom in comparison to a number of other countries [7]. Figure 1 is representative of the findings from this study. Using linked data on Scottish patients within the NHS who have suffered a myocardial infarction, it was shown that the numbers having catheterization, angioplasty, or bypass surgery were low compared to most other countries, reflecting a late uptake of these procedures and slow growth rate after uptake. There is no reason to suppose that these findings are specific to this particular treat-
NHS appears to operate rather well with regard to aggregate level, as Barr [8] points out, the Conclusions provide an inducement to underprescribe and to reimbursed directly and that capitation payments that given the lack of budgetary incentive to pre- grow in volume and cost of prescribing in general practice is the implementation of the National Service Frameworks for Coronary Heart Disease, Mental Health, Older People and Diabetes. Coronary heart disease was the second of these NSFs to be published and it appears to have had a large impact on prescribing of lipid regulating drugs, antiplatelets, beta adrenoceptor blocking drugs and angiotensin converting enzyme inhibitors” [1]. That said, it is almost certainly the case that the increased use of statins is associated with secondary prevention, even though there is overwhelming evidence that this intervention is also effective in primary prevention and the NSF guidelines suggest using statins if coronary heart disease 10-year risk is greater than 30%. This may reflect the fact that there is increasing utilization of generic compounds to reduce costs and statins are not yet available generically. Prescribing in this area clearly suggests that given the lack of budgetary incentive to prescribe effectively, GPs require clear treatment guidelines. Even then they may be wary of the budgetary impact, given that expensive prescriptions are not reimbursed directly and that capitation payments provide an inducement to underprescribe and to gear prescription toward generic compounds.

**Conclusions**

At an aggregate level, as Barr [8] points out, the NHS appears to operate rather well with regard to some aspects of efficiency. Given that the budget is funded through general taxation the general difficulties that arise with private insurance markets are overcome. This is a conclusion recently endorsed by the Wanless Committee Report [9]. Problems relating to risk rating and exclusion, for example, through adverse selection and high probabilities of requiring treatment, simply do not arise. Public funding of health care based on taxation means that information concerned with risk rating and utilization is not necessary to the budgetary process. This is, however, a cost as well as a benefit.

In particular, there is little incentive inherent in a tax-funded system to provide information on costs and their linkage to health outcomes. As such treatment costs are not linked to either episodes of care or to the health outcome achieved. The system is largely volume-constrained, given that the aggregate budget is set centrally, that prices are largely set, and that primary sector providers, GPs, gain income through a capitation system, currently unrelated to volume of service delivered.

In spite of the introduction of unified budgets, the budgetary system still does little, or indeed nothing, to aid incentive structures, which would ensure the promotion of cost-effective treatments. Indeed the effectiveness of the system appears to be determined despite the expenditure system. In particular, the role of the individual clinician and the various treatment guidelines issued through National Service Frameworks, the National Institute for Clinical Excellence, and the Royal Colleges all support the delivery of effective treatments. The budgetary constraints act as a limitation on the volume of treatment delivered, but the mix of treatment delivery is not optimal, as financial incentives do not exist that promote the efficient delivery of care at a micro-level. Even with the increase in the budget that the NHS is currently witnessing, given the continued reliance on a budget mechanism coupled with penalties to allocate resources, it is not ensured that economic efficiency will be attained. Prospective budgets can be efficient allocative mechanisms but they have to be set at the appropriate level and be accompanied by a compatible incentive structure.

Over the recent past there has been a growing recognition that the UK health-care budget has been set at levels that are too low to attain desired levels of output as determined by the public’s expectations and preferences. This is changing. The NHS budget is set to grow dramatically. Whether this rise in aggregate budget gives rise to increased efficiency in both the mix of health-care inputs and the level of outputs delivered remains open for debate, how-
ever, as it is not clear that the appropriate incentive mechanisms are yet in place.

I thank Adrian Towse and the editors of this supplement for their helpful comments and additions.

References