Decentralization in health care: lessons from public economics

Rosella Levaggi[1] and Peter Smith[2]

1. Dipartimento di Scienze Economiche, University of Brescia, Italy
2. Centre for Health Economics, University of York, UK.

Paper Prepared for Conference on Economics and Health Policy
Centre for Health Economics, University of York
16th December 2003

DRAFT: NOT FOR QUOTATION

Abstract

One of the central interests of modern public finance theory is the implication of decentralization for the equity and efficiency of public services. It is therefore somewhat surprising that health policy in the UK and elsewhere is formulated with hardly any reference to this theory and the associated empirical evidence. The purpose of this paper is to examine the relevance of mainstream public economics for countries such as the UK, Spain and Italy that are currently decentralizing major aspects of their health system. The paper seeks to offer an economic framework for thinking about programmes of decentralization now under way.

Acknowledgements

We thank Diane Dawson and Luigi Siciliani, University of York, and Richard Saltman, WHO European Observatory, for helpful comments.

1. Introduction

The most appropriate decentralization of policy making powers is an important unresolved policy question for most health systems. At one extreme lies the English National Health Service, in which most policies are set by the central authority, and lower levels have little room for manoeuvre regarding the nature or financing of services. At the other extreme lies the United States, with a pluralistic web of purchasers and providers, and little central policy of any effectiveness.

The difficulty of commanding a health system from the centre has led many systems to explore the potential for decentralizing powers to lower levels of government. The United Kingdom has already started this process, as the systems of Wales, Scotland and Northern Ireland begin to diverge following the introduction of devolution. And policy makers are – at least in their rhetoric – beginning to promote greater decentralization of NHS powers within England [1]. Traditional NHS type systems such as Italy and Spain have devolved health system policy making and finance to regions covering populations of about 3 million people. In contrast, countries such as Norway and Portugal are currently moving towards more centralization of powers [2]. Decentralization has also been an important unresolved element of health system design in many developing countries [3, 4].

Many health systems already delegate substantial powers. In Scandinavian countries a large degree of responsibility for the health system has traditionally been vested in local government. Federal countries, such as Canada and Australia, have made provinces or states the principal locus of health policy making. Yet it is worth noting that – even in these well-established, decentralized systems – the national government often retains considerable powers of oversight and regulation, and there remain considerable tensions about where the balance of responsibility for the health system should lie [5].

Some proponents appear to view decentralization as an unambiguously virtuous ambition. Yet the ultimate logic of decentralization is that responsibility for health and health care should be devolved to the household. The manifestly dysfunctional nature of health systems that promote this principle (most notably the US) should therefore alert us to the danger of a blind pursuit of decentralization. Whilst a degree of decentralization down to some level of collective authority may indeed yield substantial gains for the health system, pursued excessively there can be no doubt that decentralization in health care leads to serious difficulties.

Economists have developed a substantial literature on the topic of decentralized public services, usually referred to under the banner of 'fiscal federalism'. This literature focuses on the optimal administrative level at which to vest powers of finance and purchasing of public services, and examines the consequences of alternative distributions of responsibilities. It therefore seems very germane to recent debates on decentralization in health care, though to date there has been few English language analyses of the implications of the fiscal federalism literature for health system design (see Petretto [6] for an exception).

This paper seeks to correct this lacuna. We first offer some brief comments on what is meant by decentralization in health care. The next section introduces the economic view of decentralization, and sets out the major economic arguments adduced in the decentralization debate. We then focus on three key issues: the diversity of services and expenditure that may arise from decentralization; the role of information in decentralization; and the coordination needs of decentralized services. The main contribution of economic models is to offer a framework for thinking about decentralization, rather than any firm policy prescriptions. The paper concludes with a discussion of what we feel are the key judgements needed to develop effective policy towards decentralization.

The discussion is at quite a broad level of generality, and is concerned mainly with the purchasing of health services. In practice, local providers, most especially hospitals, are often in the driving seat of local health services, and the purchasing function is weak. However, we believe this reality reflects a failure of local governance, and that it is the purchasing function with which local communities should be pre-eminently concerned. We leave open the question of who should provide the services. Moreover, health care is hugely diverse, in both the tasks it undertakes and the technologies it deploys. It is therefore highly likely that an organizational structure that is good for some health system tasks may be less satisfactory for others.

Throughout we refer loosely to 'local government'. This is merely shorthand for a sub-national institution that enjoys a certain amount of autonomy in setting priorities and (possibly) raising revenue, and is not intended to refer necessarily to existing local government arrangements (such as local authorities in the UK). The local government under discussion could range in size from Australian states to Finnish municipalities. Also, how the governors of the local institution are appointed is left open. However our usual assumption is that they are subject to periodic popular local elections, in contrast (say) to being appointed by a national minister.

2. What is decentralization in health care?

Decentralization in health care is difficult to define. However, in broad terms it entails the transfer of powers from a central authority (typically the national government) to more local institutions. Given the immense complexity of health and health care and the associated governance arrangements, it is possible to envisage infinite variety in the nature and strength of any decentralization, embracing considerations as diverse as political autonomy, service provision, representation, finance and legal frameworks. Saltman *et al* [7] cite four types of decentralization: delegation, deconcentration, devolution and privatization. Delegation transfers responsibility to a lower organizational level, de-concentration to a lower administrative level, devolution to a lower political level and privatization takes place when tasks are transferred from public into private ownership.

In this paper we do not dwell on these subtly different notions of decentralization, which may have radically different implications for system behaviour. A full treatment of decentralization would require commentaries from a number of perspectives, including political science, sociological and clinical. Instead we

comment from an economic perspective on just two issues that are common to all types of decentralization: transfer of finance powers and transfer of policy powers.

The extent to which local institutions are given autonomy over how they can raise and use finance is a central design decision in any decentralization policy. At one extreme, localities may be assigned a fixed budget by the national government and allowed no fiscal autonomy at all. Indeed the national government may subdivide the budget so that expenditure on certain activities is 'ring-fenced'. Local choice then becomes one of deciding a preferred pattern of services within the fixed budget.

At the other extreme, local governments may be free to use any local tax base they choose (for example, the voters of Seattle were recently asked to consider a 10 cent 'coffee tax') and to set any level of tax rates. In health care, an important autonomous source of finance may be charges for service users, which in the context of health care can be considered a tax on the sick. Although not used to any great extent in UK health care, widely varying charges are levied on social care for older people in English local government.

Similarly, local governments may at one extreme have absolute autonomy over the policies they adopt, or they may be subject to strong central controls on (say) minimum standards, at the extreme becoming mere agents for the national government. In short, it is important to distinguish between the nominal degree of decentralization and the real extent of local autonomy. In health care, national governments almost invariably insist on certain minimum standards, often in the form of a 'basic package'. Localities may then be free, at the margin, to enhance the package or alter user charges. Any variations from national norms will usually result in variations in the local tax rate.

The impact of decentralization depends very strongly on the rules governing patients' access to health services, and how local governments reimburse providers. For example, in Italy after a 1995 reform, each Region was free to decide the level of competition between private and public providers. In some cases (Lombardy) the patient became free to choose any provider, while in others competition remains almost non existent (Emilia Romagna). If patients are free to seek out care from any provider (public or private), and local governments must reimburse according to a national schedule of fees, then there may be little incentive for localities to develop local policies or engage in active purchasing with local providers, and limited scope for expenditure control. On the other hand, a requirement that patients use only 'preferred providers' sanctioned by the locality may have serious implications for patient choice and competition. It is noteworthy that many highly decentralized health systems (such as Canada) have a requirement that a core set of national benefits are 'portable' between jurisdictions.

There is thus scope for huge variations in autonomy, even with nominally similar systems. In Italy, further reforms have recently decentralized health care provision and (partially) finance to the regional level. The national government has defined the list of the minimum number of services to be provided by each regional system, the so called LEA (Livelli Essenziali di Assistenza – Minimum Treatment Levels). They define for each therapy group what has to be provided as a minimum, either to the

entire population or to some subgroups (children, old age, means tested). Each region can refine and augment the list, but the treatments defined at national level have to be provided. Parallel reforms in Spain go much further in devolving almost all policy powers to the regions [8]. The centre's role is confined mainly to arranging redistribution of financial resources between regions. Arrangements in the UK have in most respects not yet approached these levels of decentralization.

Left entirely autonomous, local governments would experience massive variations in health needs and revenues sources. Indeed, high health needs and small tax bases often coincide. Left unattended, this situation would lead to huge variations in local services and local taxes, and a flight of mobile citizens from disadvantaged areas. Therefore, national governments invariably effect a system of grants-in-aid that often constitute the major source of local government income (in England they account for about 80%). These grants are effectively a transfer from low needs, wealthy areas to high needs, poor areas, and usually seek to allow localities the opportunity to deliver some standard level of care at a standard rate of local tax and user charges [9]. More generally, any system of central government transfers to localities gives the centre considerable power to influence the pattern of local services. They are for example often used by national governments as a lever for insisting on certain minimum standards in local services, or for protecting localities from certain types of risk.

3. The public economics perspective on decentralization

Public economics is concerned with public goods and their financing. A public good is one that a competitive market alone cannot provide in line with society's wishes. We take it for granted that health and health care fall into this category (and so shall side-step the sterile debate about the role of markets in health care). The issue we wish to address is therefore the following. Given that the stewardship of the health system is a governmental responsibility, when and how should national governments share powers with more local institutions?

The principle underlying local government is that for some kinds of public goods, the benefits accrue to local residents, and there is therefore a presumption that – at least up to a point – local people should determine their nature. Economic arguments in favour of decentralizing policy making of public services to lower levels of government arise in a number of forms. They include the following:

Information: remote national governments cannot understand all the opportunities and constraints that affect the supply of local services. They may seek to impose managerial solutions that are inappropriate for local circumstances, and strike poor bargains with providers. Equally, they may not be sensitive to variations from the national norm in demand, a particularly important consideration in health care, which is vulnerable to considerable random fluctuations in demand.

Preferences: local governments can respond to local preferences and seek to design services that reflect local priorities. Local elections are the usual means of expressing such preferences, and some degree of freedom to set priorities according to local electoral choices is generally considered a pre-requisite of true local government.

Local co-ordination: many public goods (but especially health care) require local co-ordination of a variety of statutory and voluntary agencies. Information limitations mean that local governments may be best placed to secure such co-ordination.

Efficiency: because they are closer to local institutions and citizens, local managerial boards may be able to identify and root out sources of inefficiency. More generally, local people may be more prepared to become active and encourage efficient delivery of locally governed public services, especially if their local taxes finance the service.

Accountability: the notion of accountability is often poorly defined. However, for economists it is closely related to allocative efficiency, and reflects the idea that those who (individually or collectively) benefit from a good or service should bear the financial consequences [10]. Under this view, decentralization of the financing of local public goods can (properly implemented) contribute to economic efficiency.

Equity: when budgets are constrained, local governments may be better placed than national governments to ensure that resources are allocated equitably within their borders.

Innovation: autonomous local governments may be more willing and able to experiment with new modes of delivery.

Competition: if suitable comparative information is collected and disseminated, autonomous local governments may effectively compete with each other to provide efficient and effective services, through the process that has become known as 'yardstick competition' [11]. There may even be a 'market' in local governments offering different packages of services and different user charges and tax rates.

However, there are also economic arguments in favour of centralization, some of which directly contradict those just cited:

Information: the information asymmetry between locality and centre may lead to *worse* outcomes under decentralization. For example, local purchasers and providers might collude to hoodwink the centre about local spending needs. More generally, local governments might act strategically in an effort to secure more than their fair share of central resources, for example by blaming high spending on high local needs rather than inefficiency. This phenomenon is likely to be important if central financial contributions depend (say) on past local expenditure levels.

Economies of scale: there may be higher production, purchasing or managerial costs associated with decentralization. In particular, larger entities may be able to secure more favourable contracts with service providers. The monopsony power of the NHS as an employer of clinical staff, and its restraining influence on pay, has often been adduced as an argument in favour of central control.

Transaction costs: in the UK, the managerial costs associated with small administrative units have been a persistent policy preoccupation in local government. More generally, decentralization may impose higher burdens in terms of information flows or the need for local managerial expertise to design and monitor local contracts.

Spillovers: local governments may to some extent be inter-dependent. The services provided by one jurisdiction affect citizens from another. For example in health care there may be public health interventions, such as childhood vaccination programmes, that will ultimately yield benefits for the whole country. Such interdependencies (or externalities) suggest some role for a national government.

Equity: unfettered local government may lead to greatly varying services, standards, taxes, user charges and outcomes. These variations may compromise important equity objectives held at a national level, and so are a special class of spillover effect.

Macroeconomy: the actions of local governments may collectively create important adverse macroeconomic effects. This is, for example, an argument often put forward for imposing strict borrowing controls on otherwise autonomous local governments.

Competition: competition between local governments may be harmful rather than beneficial. For example, if jurisdictions compete on tax rates (rather than services) there may be widespread under-provision of public services [12]. In health care, local areas may perversely have an incentive to perform *poorly* on certain types of chronic care (say dialysis treatment) if to do so deters potentially expensive patients from settling in their locality. There may therefore be a role for a national government in harmonizing standards.

There are of course a number of additional reasons for seeking to decentralize services, such as promoting local democratic involvement and distributing political power in order to reduce the potential for corruption or despotism. Such considerations may have important implications for efficiency and effectiveness, but in this essay we shall focus only on traditional economic concerns.

We now consider three broad types of consideration that are incorporated into economic models of decentralization. The first relates to the welfare improvements associated with the increased diversity and choice that often accompanies decentralization. The second addresses the lack of information available to run services efficiently from the centre. The third concerns the potential costs that arise from fragmentation and a lack of coordination of public services.

4. Diversity and decentralization

The traditional fiscal federalism literature has focused on the extent to which decentralization allows local communities to shape local services closest to their preferences [13]. It is important to recognize that diversity could be secured consciously or unwittingly by a central authority, for example through political patronage, differential funding or numerous other devices. However, there is a general presumption that local decision-makers are better at identifying local preferences than their central counterparts, and so some form of local governance is likely to secure welfare improvements compared with a central authority.

In considering how this argument relates to health care, it is first worth considering the implications of an entirely centralized system, in which every patient's entitlement

(and therefore expenditure) is explicitly defined. This of course assumes unambiguous information about a patient's condition and the appropriate treatment. With uniform levels of efficiency throughout the system, this might result in a system close to many systems of social security, in which a 'demand led' national set of entitlements is carried out mechanically by local administrative offices, and is not far removed from what systems of social health insurance historically sought to secure (before recent reforms) [14]. A major implication of this system would be that it leads to an open-ended budget for the health system – demand cannot be predicted in advance, either at the local or national level. It also has major managerial requirements for specifying and monitoring adherence to entitlements.

In an attempt to secure expenditure control, many health systems therefore allocate prospective budgets to local administrators, and require them (to a greater or lesser extent) to meet all local demand within that budget [15]. This system has many virtues. It is in practice impossible to offer detailed epidemiological predictions of diseases and their treatment requirements. Yet, on average, the costs of delivering a given level of service to a reasonably large population can be predicted with some accuracy. Therefore offering a global budget can often offer local decision makers an opportunity to implement the large majority of national guidelines. Within their budget, they can often trade off lower than expected demand for some interventions against higher than expected demand for others, and thereby secure budgetary control to a tolerable level of accuracy.

Within such systems, central authorities often seek to circumscribe local freedom by 'ring-fencing' some part of the local budget for specific functions, or prescribing required treatments for certain conditions (as through the National Institute for Clinical Excellence in the UK). Clearly such constraints circumscribe local freedom and reduce the effective degree of decentralization.

Difficulties arise when the level of devolution is too local, when the amount of mandatory provision is too extensive, or the budgetary mechanism is faulty. Then, random fluctuations in demand can lead to massive overspends or underspends of budgets, and – without adequate risk sharing arrangements – gross inequities can arise between otherwise identical patients in different localities if local decision-makers sacrifice uniformity in the interests of meeting budgets [16]. It is for this reason that Smith [17] advocates a range of risk sharing arrangements when setting formulaic budgets for small administrative units such as general practices.

The extent to which local diversity is desired or efficient in health care – as compared to other public services – is a matter for debate. The widespread adoption of clinical guidelines and defined 'basic' packages of care suggests that many national policy makers believe that a uniform package of health care is a desirable policy objective. There is also widespread popular concern with 'postcode' rationing of health services. So the extent to which local diversity addresses policy objectives deserves careful scrutiny.

However, it is of the essence of local government that there should be variations in levels of service and tax rates between jurisdictions. In a classic paper, Tiebout [18] argued that citizens might 'vote with their feet' to settle in jurisdictions that provided

a service mix that suited their preferences. Equally, communities might choose their mix of services deliberately to attract (or deter) certain types of citizen. A corollary of this viewpoint is that communities that fail to provide attractive services will lose mobile citizens – frequently those who provide tax revenue in excess of their demand for local public expenditure.

While Tiebout's viewpoint is deliberately extreme and provocative (and perhaps more relevant to a consumerist US setting), it nevertheless offers a great deal of food for thought when considered in relation to health care. For example, if variations in health care provision or user charges (or even local taxes) emerge, will mobile citizens move to areas offering their preferred system of health care? This is unlikely to be more than a marginal consideration so far as general acute services are concerned (although employers may take the quality of local health services into account when considering relocation decisions). However, for citizens with chronic conditions, or older people with generally high health care needs, proximity to relevant services of high quality (or low levels of user charges) might be very important considerations when choosing where to settle. Whether the implied concentration of certain types of health care provision in certain locations leads to a welfare gain is a matter for conjecture, but local diversity is likely to benefit those who are able to move (and can therefore exercise choice) more than those who cannot.

A particularly interesting phenomenon arises when local governments rely on a property tax as their revenue base. Effectively, when buying a property, one secures the right to gain access to local public services as well as the intrinsic benefits of the property (and one also assumes concomitant responsibilities, in the form of local property taxes). Therefore the property price should in principle partially reflect these considerations – in other words, the expected benefits and costs of local public services might be 'capitalized' into house prices. For example, there is empirical evidence that school education is in England a valued consideration when choosing where to live, with a large impact on house prices [19]. It is therefore highly likely that – if great variations in health care provision arise – similar considerations might apply.

One does not need a system of local government for the Tiebout effect to arise. Indeed the education evidence cited above arises from variations in school quality *within* a local government, rather than between jurisdictions. Considerable variations in service standards exist even in national government programmes, such as the NHS, and one would expect some sort of Tiebout effect to be in place already. Decentralization is merely likely to make it more pronounced if local jurisdictions introduce service variations as a matter of deliberate policy, or if variations in local taxes or user charges are permitted.

In health care, whether local governments would seek to encourage (or deter) certain types of patient depends heavily on the finance regime [20]. At present, local areas (primary care trusts) in the English NHS are funded exclusively from national taxes, predominantly on the basis of population size, demography, and an index of socio-economic disadvantage [21]. The extra funding for an additional citizen will therefore be a crude age-related capitation payment, albeit with some adjustment for

general social conditions. Under this sort of funding arrangement, jurisdictions have a strong incentive to deter citizens they know to have health care expenditure needs in excess of the age-specific local average. That is, they may wish to deter citizens with chronic conditions, unhealthy lifestyles and generally poor health.

It is of course quite beyond the powers of local governments explicitly to refuse residence to such citizens. However, there are numerous indirect ways in which jurisdictions could signal that patients with chronic care needs are not a high priority, such as poor facilities, difficult access, and even poor reported outcomes. In short, patients with long term needs might become a very low priority in a system of competitive local governments.

5. Information asymmetry and decentralization

Recently research has focused on the role of information asymmetry in determining the optimal level and nature of decentralization of public services. A central theme of fiscal federalism has always been the informational advantages enjoyed by localities to understand local demand for and supply of local public goods. Formal analysis of this advantage is beginning to emerge.

Seabright [22] examines the distribution of powers between central, regional and local governments. The advantage of decentralization is that it brings electoral power closer to local people, and so may more closely align local preferences with local services. The advantage of centralization is that it permits better coordination of public goods, most notably when the choices of one locality have spillover effects for other localities. In the health domain, one particularly important spillover effect concerns the potentially negative impact of devolving choices to local government on various notions of equity, such as equity of health, equity of access, or equity of financing.

Seabright's model presumes that governments at all levels are interested in re-election, and that the probability of re-election is determined by the level of welfare enjoyed by the population. National (or regional) governments are interested only in those lower level areas that are marginal to their expected re-election (a sort of 'jurisdictional' median voter model). The existence of positive spillovers from one locality's services to another's welfare *increases* the case for centralization. However, this must be traded off against a lack of accountability in jurisdictions that are not critical to the central government's re-election. The case for centralization is also strengthened if localities have similar vulnerability to unforeseen nation-wide circumstances (such as macroeconomic shocks).

There is an implication that aggregate spending will usually be higher under centralization, because the central government takes into account the positive spillover benefits from higher spending. Centralization also increases the willingness to transfer resources from rich to poor areas, therefore benefiting disadvantaged localities. However, centralization might benefit some localities more than others, most notably the 'pivotal' electoral battle grounds. This prediction is borne out by research showing that in England national grants have been skewed to electorally important local authorities [23, 24].

Gilbert and Picard [25] assume that central governments are less well informed than local government about two crucial aspects of local services: local production costs and local preferences. They argue that – if central government had full information on production costs – then full centralization is optimal, whilst the reverse is the case if the central government had full information on local preferences (including the values attached to spillovers). Ambiguity arises when (as is usually the case) the there is imperfect information on both costs and preferences. If information on costs improves, then the scope for exploitation by local providers decreases, so central government is in a good position to exercise its prime role of accommodating spillover effects. If on the other hand information on costs is poor (or spillovers not important) then decentralization is preferred because the better knowledge of local governments about the efficiency of local providers.

Decentralization supported by central grants offers localities an incentive to act strategically in misrepresenting their true needs and preferences. Levaggi and Smith [26] give an example of the nature of the game in which the locality increases its spending beyond its preferred level in order to attract higher government grant. Barrow [27] shows how the competition between jurisdictions for a fixed central grant can induce spending in excess of efficient levels. In the same vein, Besley and Coate [28] present a model of political economy in which localities have an incentive to elect representatives with high spending preferences to national legislatures.

Finally, Laffont [29] examines an important class of problem, in which decentralization increases the probability of collusion between local purchasers and providers. This risk is especially important in health care, where there is an ever-present danger of local purchasers being 'captured' by powerful providers. A key element of his model is the bounded rationality of the centre in capturing and processing information about localities – in short, the information requirements of effective centralization may be costly. Once again, economic analysis offers no clear-cut policy prescription. The informational advantages of delegation have to be weighed against the potential efficiency costs of collusion.

Analyses of this sort emphasize the crucial role of information asymmetry in determining optimal structure of government. But, as Seabright [22] argues, 'the choice between centralised and decentralised forms of government is very sensitive, not only to variable features of the particular policies in question, but to estimates of the quantitative significance of the phenomena – such as 'accountability' – that are in the nature of things very hard to quantify'. In short, whilst we can develop useful models of the structure of the decentralization problem, it is very hard to be able to offer useful policy advice on optimal structures of government.

6. Spillovers and decentralization

The main role of central governments in the models discussed above is in ensuring that the public services accommodate any valued spillover effects that would otherwise be ignored by jurisdictions. Important examples of these effects can be found in any health care system, and are the reason for the generally high level of central intervention. They include:

- *Clinical training and research:* left to their own devices, localities would probably seek to free-ride on the training and research provided by others, leading to a chronic under-provision.
- *Public health:* given the high mobility of citizens, there is an incentive for localities to ignore actions such as health promotion that secure benefits only in the long term.
- *Inequalities:* the diversity inherent in unfettered local government may compromise nationally held equity objectives.
- *Information:* only a central authority can specify and mandate the collection of the comparative data needed for informed decision-making.
- *Macroeconomic factors:* the health system is a big segment of the economy with major implications for the nation's productivity. There may be a number of features of a decentralized system, such as inhibitions to labour mobility, that have adverse macroeconomic consequences requiring correction by the national government.

The national government has available a number of instruments for accommodating spillovers under four broad categories: centralization of services; central rules and standards; performance reporting; and financial and non-financial incentives.

The centre can indeed internalize the spillover problem by centralizing powers. It is likely that functions such as clinical training and research should be organized directly by the centre. It is difficult to envisage any circumstances in which more oblique attempts to influence system behaviour will be as effective. However for direct patient services there will always be a need for local organizations that purchase local services, and centralizing may merely mean the replacement of local democratic governance by a local administrator accountable to the centre.

More important than structural form are therefore the rules and standards imposed by the centre on local services. Whatever the governance structure, these are always likely to be extensive in health care, particularly in the domain of minimum standards of care and information provision [6]. In the UK, standards have taken the form of guidelines, such as those promulgated by NICE and the National Service Frameworks, whilst in Italy and many other countries they have taken the form of a national basic package of care. Central to the effectiveness of all such instruments are the sanctions associated with departures from the standards, and the extent to which patients are empowered to ensure they are adhered to.

Rules concerning patients' rights can also address spillover problems. For example, a guarantee of patients' mobility can reduce inequity when the provision of hospital care is not equally distributed. For highly specialized treatments, say, patients could then move to where the intervention is supplied. Some patients (those living closer to

the hospital location) will be better off than others, but the cost and quality benefits of concentrating services might outweigh the implied inequity, so long as mobility is guaranteed.

Performance reporting is becoming widespread, and one frequently cited objective is to encourage competition and reduce disparities. However, there is an open question as to what indirect incentives might be introduced by public reporting, and the optimal deployment of comparative data remains a matter for research [30]. Reporting can nevertheless contribute to democratic dialogue and perhaps help the national government learn where spillovers most need attention. Certainly, the emergence of credible OECD data that indicated that - relative to its international peers - the UK health system performed poorly on many aspects of health care was an important stimulus for the long term review of the UK system undertaken by Derek Wanless. One particularly important type of reporting is accreditation and periodic inspection of health system organizations, and properly done such activity can clearly help secure adherence to any nationally specified standards.

Financial incentives can take a number of forms. The traditional fiscal federalism literature considered three broad types of grant-in-aid: unconditional lump sum grants; unconditional matching grants; and conditional grants [9]. Each of these has very different implications for the magnitude and mix of local services, and therefore for spillovers. Hospital systems have experimented extensively with payment mechanisms. For example, English hospitals are gradually moving from a system of block grants towards diagnosis-related group (DRG) funding. The former system tends to discourage treatment, whilst the latter can stimulate treatment in excess of optimal levels. Many academic researchers therefore advocate a 'mixed' block and DRG funding system, as used in Norway.

There is some concern that payments mechanisms for local institutions are ineffective because they are not 'passed on' using equivalent incentives for front line professionals, who ultimately determine the nature of service delivered. To address this, the UK government has recently proposed an immensely ambitious reform of its contract with general practitioners, in which up to 20% of practice income will be determined by reported performance on 150 indicators of clinical quality [31]. We must await impatiently evaluation of this experiment. Increasing attention is also being paid in the UK to the use of non-financial incentives, such as 'earned autonomy' in the form of freedom from standards or less frequent inspection. Again, the effectiveness of these instruments is a matter for research.

Levaggi [32] examines the procedures used to distribute the total budget between competing services in a decentralised system. She finds that it may be most effective to offer grants dependent on providing specific services rather than use block grants and seek to protect parts of the budget. In the extreme, the creation of separate local agencies for different functions may be preferred to a single local organization. Clearly there is then a trade-off with the local coordination of the separate functions.

Thus the 'autonomy' within which any decentralized organizations operate is highly dependent on the system of rules, standards, reporting requirements and incentives within which the centre asks them to operate. In principle, the centre probably always

has enough instruments available to force localities into a particular pattern of service delivery. We would therefore argue that it needs to use these instruments with discretion, addressing legitimate spillover concerns, but equally ensuring that legitimate local freedoms are respected.

7. Concluding comments

This paper has sought to highlight some of the important themes in the fiscal federalism literature that may be relevant to policy makers seeking to identify optimal decentralization policies in health care. We have noted the multidimensional nature of the concept of decentralization, and the difficulty of securing a simple definition of what it means. There are numerous economic arguments relevant to decentralization debates, but three central issues have dominated the discussion: the implications of diversity amongst local jurisdictions; the implications of local informational advantages; and the implications of spillover effects between jurisdictions.

Diversity amongst local governments, and the associated competition, can induce both beneficial and adverse behaviour. At the very least – providing the national government makes provision of comparative data mandatory – localities will be required to account to their electorate for their performance relative to their peers, through the mechanism now known as yardstick competition. Hitherto UK governments have relied heavily on comparative performance data in health care, but have used centrally driven performance management arrangements rather than local democratic control to provide the necessary incentive mechanism. Which of these two arrangements is more effective is an open question.

The scope for competition between local jurisdictions can lead to adverse outcomes. There is a large literature that shows that the mobility of tax bases might lead to levels of local taxation that are lower than optimal, as jurisdictions 'beggar their neighbours' through tax competition. In health care, this might lead to more restricted packages of care or higher user charges than is optimal. It also creates an incentive to give services for chronically sick and elderly people a low priority. There is an important role for the national government to assure minimum standards.

Compared to their local counterparts, national governments may suffer an informational disadvantage when purchasing services. Information asymmetries come under two broad headings: service costs and local preferences. In a service as complex as health care it is very difficult for the centre to determine whether an apparently high level of local costs arises because of inefficiency or external demand factors. The argument for decentralization is that – by bringing accountability for local expenditure closer to local people – inefficiency will be discouraged.

It is not known how much variation in local preferences exists in health care. For example, it is likely that maximizing health gain is a universally held central objective of all health systems. However, it is equally reasonable to suggest that there may be considerable variation in the local weight given to issues such as access, responsiveness and equity. This being the case, there is a strong case for putting in place local governance mechanisms to solicit local preferences. However, a special concern in health care is the vulnerability of the political process to 'capture' by

interest groups (either patients or producers). This is an area that a vigilant central government should be alert to.

One of the main arguments for a strong central role in public services is the presence of important spillovers, when residents in one locality are affected by the nature of services in other jurisdictions. In health care there are clear reasons to believe that such spillovers are important. If one jurisdiction provides poor quality services, it may induce unwanted migration of chronically sick people to other jurisdictions. Variations in the availability and quality of services have obviously adverse consequences for equity. Some localities may neglect the public health and macroeconomic consequences of their services. Medical education is likely to be a national public good that would be underprovided without central coordination. These sorts of considerations provide a compelling argument for a strong central role, even in a mainly decentralized system, using minimum standards, performance reporting requirements and financial transfers.

We believe that diversity, information and spillovers are the three main considerations when discussing the optimal level of decentralization in health care. However we noted earlier that other arguments have been adduced. Scale economies are often cited as an argument for centralization. Some health system functions – such as research, technology assessment, and some public health actions – clearly benefit from economies of scale if provided at a national (or even supra-national) level. However, although decentralization requires greater use of local contracting, it is difficult to identify large economies of scale to be derived from national as opposed to local purchasing of most services. Even centralized health systems such as the 'old' NHS, required a local bureaucracy to purchase local services. Therefore we think it unlikely that transaction costs will be materially higher under decentralization.

It is also claimed that the diversity encouraged by decentralization can offer an incentive for innovation. There is scant empirical evidence to support this hypothesis, and an examination of the extremely decentralized US system offers little support for it in health care.

One final point is that the optimal degree of decentralization is likely to be different for different health system functions. Services for primary care and chronic care may have much more scope for local discretion than (say) secondary care services, and may therefore benefit more from decentralization. Yet coordination may require that health system functions are best organized locally by a single purchaser. The optimal size and operational constraints imposed on that purchaser may therefore be something of a compromise.

We have therefore argued that the appropriate level of decentralization in health care is a difficult policy judgement, involving a trade-off between a number of conflicting objectives. Public economics can inform the debate, but can offer no clear-cut recommendations. However, it is likely that an optimal system in health care is likely to combine a strong central role of oversight, standard-setting and information provision with a strong local role that allows local preferences to be expressed and promotes accountability of local services. It is difficult to see how this localism can be achieved in reality (rather than rhetorically) without a robust system of local

democracy and some degree of financial autonomy. In its avowed aim of decentralization it will be interesting to see how far the English NHS goes down the path of real local democratic control.

References

1.Department of Health (2003) *Keeping the NHS local - a new direction of travel*. London: Department of Health.

2.World Health Organization (2003), *Health Care in Transition: Country Profiles*. http://www.euro.who.int/observatory/toppage.

3.Mills, A. (1994), "Decentralization and accountability in the health sector from an international perspective: what are the choices?" *Public Administration and Development*. 14, 281-292.

4.World Bank (2003), *Decentralization and Health Care*. http://www1.worldbank.org/publicsector/decentralization/healthcare.htm (19 November 2003).

5.Lazar, H., K. Banting, R. Boadway, D. Cameron and F. St-Hilaire (2002) *Federal-provincial relations and health care: reconstructing the partnership.* Ottawa: Commission on the Future of Health Care in Canada.

6.Petretto, A. (2000), "On the cost-benefit of the regionalisation of the National Health Service"*, Economics of Governance*. 1, 213-232.

7.Saltman, R., V. Bankauskaite and K. Vrangbaek (2003), *Decentralization in health care: strategies and outcomes*. http://www.euro.who.int/observatory/Studies/20031112_1 (19 November 2003).

8.Lopez Casasnovas, G. (2001) *The Devolution Of Health Care To The Spanish Regions Reaches The End Point*. Barcelona: Center for Health and Economics, Pompeu Fabreu University.

9.King, D. (1984) *Fiscal tiers: the economics of multi-level government*. London: Allen & Unwin.

10.Barnett, R.R., R. Levaggi and P. Smith (1991), "Accountability and the poll tax: the impact on local authority budgets of the reform of local government finance in England"*, Financial Accountability and Management*. 7(4), 209-228.

11.Shleifer, A. (1985), "A Theory of Yardstick Competition"*, Rand Journal of Economics*. 16(3), 319-327.

12.Wilson, J.D. (1999), "Theories of tax competition"*, National Tax Journal*. 52(2), 269-304.

13.Oates, W. (1972) *Fiscal Federalism*. New York: Harcourt Brace Iovanovich.

14.Normand, C. and R. Busse, (2002) *Social health insurance financing*, in *Funding health care: options for Europe*, E. Mossialos, A. Dixon, J. Figueras and J. Kutzin, Editors. Open University Press: Buckingham.

15.Rice, N. and P. Smith, (2002) *Strategic resource allocation and funding decisions*, in *Funding health care: options for Europe*, E. Mossialos, A. Dixon, J. Figueras and J. Kutzin, Editors. Open University Press: Buckingham.

16.Martin, S., N. Rice and P.C. Smith (1998), "Risk and the general practitioner budget holder", *Social Science & Medicine*. 47(10), 1547-1554.

17.Smith, P.C. (2003), "Formula funding of public services: An economic analysis", *Oxford Review of Economic Policy*. 19(2), 301-322.

18.Tiebout, C. (1956), "A pure theory of local expenditure", *Journal of Political Economy*. 64(5), 416-424.

19.Leech, D. and E. Campos (2003), "Is comprehensive education really free? a case study of the effects of secondary school admission policies on house prices in one local area", *Journal of the Royal Statistical Society Series A-Statistics in Society*. 166(1), 135-154.

20.Ellis, R.P. (1998), "Creaming, skimping and dumping: provider competition on the intensive and extensive margins", *Journal of Health Rconomics*. 17(5), 537-555.

21.Department of Health (2003) *Resourca allocation: weighted capitation formula:*. London: Department of Health.

22.Seabright, P. (1996), "Accountability and decentralisation in government: an incomplete contracts approach", *European Economic Review*. 40(1), 61-89.

23.John, P. and H. Ward (2001), "Political manipulation in a majoritarian democracy: central government target of public funds to English subnational government, in space and across time", *British Journal of Politics and International Relations*. 3(3), 308-339.

24.Ward, H. and P. John (1999), "Targeting benefits for electoral gain: constituency marginality and the distribution of grants to English local authorities", *Political Studies*. 47(1), 32-52.

25.Gilbert, G. and P. Picard (1996), "Incentives and optimal size of local jurisdictions", *European Economic Review*. 40(1), 19-41.

26.Levaggi, R. and P. Smith (1994), "On the Intergovernmental Fiscal Game", *Public Finance-Finances Publiques*. 49(1), 72-86.

27.Barrow, M. (1986), "Central grants to local governments: a game theoretic approach", *Environment and Planning C: Government and Policy*. 4, 155-164.

28.Besley, T. and S. Coate (forthcoming), "Centralized versus decentralized provision of local public goods: a political economy approach", *Journal of Public Economics*.

29.Laffont, J.-J. (2000) *Incentives and political economy*. Oxford: Oxford University Press.

30.Marshall, M., P. Shekelle, H. Davies and P. Smith (2003), "Public reporting on quality: lessons from the United States and the United Kingdom", *Health Affairs*. 22(3), 134-148.

31.Department of Health (2003), *Investing in general practice: the new GMS contract*. http://www.doh.gov.uk/gmscontract/index.htm (28 October 2003).

32.Levaggi, R. and R. Zanola (forthcoming), "Flypaper effect and sluggishness: evidence from regional health expenditure in Italy", *International Tax and Public Finance*.