Resource allocation and purchasing in the health sector: the English experience

Peter C Smith [a]

Abstract The United Kingdom of Great Britain and Northern Ireland has extensive experience in allocating health service funds to regions and localities using funding formulae. This paper focuses on England. Special attention is given to recent policy concerns to reduce avoidable health inequalities by broadening the remit of the resource allocation formulae. The paper also examines the issues that arise when seeking to allocate funds to very small organizational units, such as general practices. The English example is relevant to less-developed health systems, especially for those governments seeking to decentralize, to improve accountability and to promote equity.

Bulletin of the World Health Organization 2008;86:884–888.

Introduction

Most health services are geographically specific and so a central policy issue in many countries is how national (often tax-based) funds should be allocated to localities. Rather than relying on arbitrary methods of solving this "resource allocation" problem, such as historical precedent or political patronage, many health systems are seeking to place greater emphasis on the use of systematic funding formulae.[1] This trend has been given added impetus by the increased decentralization of health services worldwide. If decentralization is to be effective, national governments need to be seen to be treating different localities even-handedly.

The use of systematic formulae for allocating funds offers the best prospect of satisfying equity criteria. It is becoming increasingly attractive due to improvements in the scope and timeliness of data sources measuring the inputs, activities and outcomes of health services. Even in countries where information resources have historically been poor, there is an increasing demand to allocate resources systematically and fairly, in line with policy intentions, and emerging data resources may now make this feasible. This paper uses the English experience of allocating funds using formulae as a case study that may be useful particularly for settings where information resources are limited.

National Health Service

Established in 1948, the National Health Service (NHS) of the United Kingdom of Great Britain and Northern Ireland claims to be the world's largest publicly-funded health service. It delivers more than 87% of the country's health care and spends about 8.4% of the country's gross domestic product. Each country of the United Kingdom manages its own NHS but this paper concentrates on England only.

The English NHS is managed by the Department of Health and is administered locally by 152 geographically defined health authorities, known as primary care trusts (PCTs), with average populations of about 400 000. PCTs are almost completely reliant on financial allocations from the national government to fund their activities.

All citizens register with a general practitioner (GP) of their choice. With the exception of emergency treatment, patients can gain access to NHS hospital care only if they are referred by their GP. The GPs therefore act as gatekeepers to hospital and community care and prescription medicines. The costs of all local health care are met from within the local PCT's fixed budget, as set by the national ministry. If hospital referrals or other aspects of local clinical practice imply expenditure in excess of the local budget, then some sort of rationing takes place. This may take the form of a delay in treatment or a refusal to prescribe certain medicines. Patients can at any stage seek private care although, in practice, this accounts for only a small proportion of health care in the United Kingdom.

The national government allocates the overall Department of Health budget in its annual public expenditure negotiations. The Department of Health sets the cash-limited budget available for allocation to PCTs. In the financial year 2006–2007, this amounted to £64.310 billion.[2] These funds are then distributed to PCTs to finance hospital and community health-care services as well as all prescribing, primary care and health promotion. While formulae are applied to all the categories of expenditure, this paper focuses mainly on the acute sector, which accounts for 66% of all PCT expenditure.

Weighted distribution

The starting point for any discussion of formula funding in the United Kingdom's health system is the recommendation in 1976 of the Resource Allocation Working Party (RAWP)

[a] Centre for Health Economics, University of York, York YO10 5DD, England.
Correspondence to Peter C Smith (e-mail: pcs1@york.ac.uk).

Peter C Smith

for allocating NHS funds to English regions.[3] At that time, funds were allocated mainly according to historical precedent, leading to a very large bias in favour of London and the south-east of England.[4] This imbalance had become politically unsustainable.

The objective of the RAWP approach was to set budgetary targets for the 14 regional health authorities, each covering populations of about 4 million. The services in question included hospital inpatient and outpatient care, and some community care, but not primary care or prescribing. First, the health services were disaggregated into a small number of disease categories, corresponding to specialities based loosely on WHO International Classification of Disease chapter headings.[5] RAWP then recommended that, in each speciality expenditure, targets for health authorities should be based on:

- population size;
- an adjustment for demographic characteristics (age and sex adjustments specific to each speciality);
- a further weighting for additional clinical need (as measured by local standardized mortality rates for the specific specialities, assumed to be proxies for morbidity);
- an adjustment for variations in the input prices of local services.

This approach gave rise to the notion of "weighted capitation", the principles of which are still in force in the NHS although the methods of quantifying the different elements have changed (Box 1).

The total health services budget for each region was the sum of its disease-specific budgets. The budgetary targets implied by the RAWP recommendations were phased in gradually over a period of 15 years. They took formula funding to a new level of intellectual coherence and sophistication, and have been highly influential internationally. They remained in force until 1990, by which time most regions were spending very close to their expenditure targets.

However, increasing pressure developed in the 1980s to place NHS resource allocation on a more empirically sound basis.[6,7] As new data became available, RAWP was therefore superseded by a series of more complex empirical formulae, of which one example is the York formula intro-

duced in 1995. This disaggregated the population by age, and then applied a further "needs" adjustment based on the five variables shown in Box 2. The results were derived from extensive econometric analysis of small area NHS utilization data.[7,8]

Reducing avoidable inequalities

Traditional NHS capitation payments are based on current patterns of expected utilization under the assumption that the system is currently meeting the needs of the population. They therefore perpetuate any inequity implied by the existence of unmet need. For example, ethnic minority groups systematically appear to receive less care for circulatory diseases than people of a European ethnic origin.[9]

The Labour government elected in May 1997 established a public health agenda, with the objective of "improving the health of everyone, especially the worst off".[10,11] This commitment to reducing health inequalities resulted in a reappraisal of the capitation criterion. For the first time, allocations in the financial year 2001–02 contained a "health inequalities adjustment" of £130 million targeted at those PCTs judged to be making the biggest contribution to avoidable health inequalities. While this seems a small amount (about 0.5% of total expenditure), it marked

a major departure from conventional resource allocation approaches.

Source of inequalities

Avoidable health inequalities come from three broad sources: variations in (i) the quality of health services; (ii) access to health services; and (iii) factors outside the direct control of the health system, such as wealth, lifestyle, genetic and environmental considerations. There is considerable evidence that many populations suffering poor health outcomes suffer on all three counts: they use poor quality services, to which they have relative difficulty securing access, and they also suffer multiple "external" disadvantages. Each issue has quite different implications for policy and resource allocation.

Poor quality services for disadvantaged populations are, in principle, a management issue. The right amount of money is being spent on such populations but it is not being spent wisely, perhaps because the organization of local services is poor or some providers are substandard. The policy implication is that the quality defects should be rectified by (possibly radical) managerial action, but that resource constraints are not the principal source of the problem. Therefore, there is no need for any major change to the funding system. Rather, attention should be directed at securing better management of resources in deprived areas.

Box 1. **Summary of the RAWP recommendations**

Per capita need was calculated by first disaggregating the population by age and sex. The different expected health-care utilization of each demographic group was approximated using the national average per capita hospital bed utilization. These were in turn adjusted by a series of standardized mortality ratios (SMR). The SMR is defined as the number of *observed* deaths in an area as a percentage of the *expected* deaths in the area, given its demographic profile. It was used by RAWP as an index of an area's relative morbidity and therefore as a proxy for medical need over and above demographic considerations.

RAWP also broke down health care into a small number of broad categories of conditions and the index of relative need for care for each category was determined by applying the condition-specific SMR to the population of an area. This process generated a notional total use of bed days by the population in an area, assuming utilization conformed to the national average, after adjusting for local need, as indicated by the SMRs. Algebraically, the equation can be represented as follows:

$$RA_i = \sum_j SMR_{ij} \left(\sum_k BEDS_{jk}\, POP_{ik} \right)$$

where RA_i is the financial allocation to area i; SMR_{ij} is the SMR of condition j in area i; $BEDS_{jk}$ is the national number of bed days required by age/sex group k diagnosed with condition j; and POP_{ik} is the population in area i in age/sex group k.

The final stage was to apply an "area cost adjustment" to all budgets to reflect the large variations in input prices, especially pay, among the regions.

RAWP, Resource Allocation Working Party.

> Box 2. **The five variables used as the basis of the York acute model of formula funding**
>
> - Standardized limiting long standing illness ratio (aged less than 75 years)
> - Standardized mortality ratio (aged less than 75 years)
> - Proportion of economically active who are unemployed
> - Proportion of pensionable age living alone
> - Proportion of dependants in single-carer households.

Poor access for disadvantaged populations implies that they are not receiving some services to which the remainder of the population secures access. Traditional resource allocation formulae do not capture any elements of unmet need.

Poor life chances among disadvantaged populations pose the most fundamental challenge to the health system. Health inequalities can arise from lifetime exposure to numerous sources, such as genetic, environmental, income, lifestyle, welfare service and health utilization variations. However, for capitation purposes, it is necessary to identify the specific potential contribution of health care to health improvement. If the NHS is to tackle these health inequalities, it will need to target the vulnerable populations in a way that it has not done before. This might entail offering such populations preferential access to NHS services, e.g. accelerated access to surgery or provision of therapies that are not available to all. In short, addressing health inequalities may require abandoning the principle of equal access for equal clinical need, in favour of equal access for some concept of equal social need.

Adjusting for inequalities

Designing a funding formula for the health inequalities adjustment requires the resolution of the following issues: (i) identification of effective health-care interventions designed to reduce the health inequality; (ii) identification of disadvantaged groups at which the intervention will be directed; (iii) identification of the areas where such groups live; (iv) allocation of resources according to the group composition of an area; and (v) ensuring that the resources are spent appropriately on disadvantaged groups and the necessary interventions.

In its first year of operation, however, the distribution of the health inequalities adjustment was based simply on the magnitude of an area's "avoidable mortality". This is defined as the number of years of life lost under the age of 75 over a three-year period, where diagnosis of death is in certain broad categories deemed to be "avoidable". This preliminary index was chosen because of the ready availability of mortality data and its plausible link to the concerns of health inequality policy. However, it is clearly very broad. For example, should

all years of life lost be counted equally? Why use age 75 as the benchmark? Are the chosen diagnoses the most appropriate? Is current mortality (backward-looking) a suitable index of current need for inequality interventions (which are forward-looking)? How should migration between PCTs be accommodated? These questions reflect the same sort of issues that troubled commentators on the original RAWP formulae, and led to their eventual replacement with more evidence-based indices such as the York formula (Box 2). The health inequalities adjustment was withdrawn in 2003. However, with renewed political concern about health inequalities, there is considerable pressure for the reintroduction of some equivalent formula.

GP budgets

In 1991, the government began an experiment with a system of allocating budgets directly to those GPs that elected to become "fundholders". These budgets were carved out from the local health authority budget and were intended to cover expenditure on a wide range of routine non-emergency inpatient procedures and prescribing expenditure. Expenditure on the remaining aspects of care (predominantly emergency care) and for expenditure by non-fundholding practices remained the responsibility of the health authority (predecessor of the PCT). Fundholders could not retain any annual surplus for private income, but could use funds to provide additional services for patients. The typical size of a fundholding practice was 8 000 patients.

There is some evidence that fundholding moderated increases in health-care expenditure. However, finding a satisfactory method of calculating budgets for fundholders was a central requirement that was never satisfactorily resolved.[12] Fundholding was abandoned in 1998, by which time more than 50% of general practices had become fundholders. However,

this initiative has recently been reintroduced in a different form known as practice-based commissioning. This introduces indicative budgets for all general practices (it is effectively compulsory) and covers most aspects of health care. Implementation has been slow and many details of budget-setting and rewards and sanctions for practices have yet to be resolved.[13]

In the early years of the fundholding scheme, the predominant approach was to base budgets on the practice's previous expenditure levels. However, this approach introduced perverse incentives to maintain high spending levels, and was therefore clearly unsustainable. As a result, some local health authorities began to base GP budgets on the York formula, which was originally designed to allocate funds to much larger populations. According to Martin et al.,[14] the York formula is inadequate for small general practices (typically 8 000 patients) where there is a one-in-three chance of a divergence in expenditure of more than 10%. The formula is vulnerable to random fluctuations and too much expenditure variation is left unexplained, thereby introducing much higher levels of budget uncertainty.

Elements of risk

Budgetary risk is the propensity for actual expenditure to vary from predicted expenditure. It can be interpreted and formalized in several ways, but is often measured by statistical measures of dispersion, such as the variance of outturn expenditure. If demand for health services is independent of the budgetary regime, the aggregate of budgetary risk in the system as a whole cannot be altered. But how it is shared between various parties can be profoundly altered by the chosen organizational structure and payment mechanisms.

Risk is therefore important under any system of formula funding. However, in strict capitation systems (with

no cost sharing), it becomes especially important. Capitation mechanisms usually seek to reflect the expected expenditure on a population with certain measured characteristics if some standard set of policies is applied. They effectively create an insurance risk pool, and the budget holder's role is to ensure that the expenditure made from the risk pool is in line with some agreed set of policies. Risk is concentrated on the budget holder because a fixed budget is allocated to manage uncertain demand and unit costs.

It is possible to partition variability in health service expenditure into four broad components: (i) an element that is due to individual characteristics (e.g. age and sex) that are captured by the funding formula that has been chosen to allocate funds; (ii) an element that is due to individual characteristics that are *not* captured by the formula (e.g. smoking status or the presence of diabetes); (iii) an element that is due to the practices of local providers (e.g. local clinical practice); and (iv) an element that is totally random, caused by the unpredictable incidence and severity of individual health-care need.

Governments will of course seek out a formula that explains as much individual variation as possible, however, it will never be remotely practical to measure all relevant observable individual characteristics for all citizens. The use of local area characteristics (such as mortality rates) as a proxy for individual characteristics circumvents some of this problem, but can never offer a perfect substitute for unmeasured individual characteristics. Variations in clinical practice reflect illegitimate influences on expenditure variation and so should be ignored in a formula. Random variation is inherently unknowable. Therefore, all expenditure that is not related to individual characteristics will be perceived by the local organization to be expenditure "risk", which arises from weaknesses in the formula, clinical practice variation and random variability.

Individual capitation methods

The concern about budgetary risk in calculating budgets for small units such as general practices has led the Department of Health to commission new research exploring the feasibility of introducing "person-based resource allocation" methods. In particular, a research project is exploring the feasibility of attaching diagnosis-based capitation payments to every individual, along the lines of the systems developed for Medicare in the United States of America (USA) and some other systems of competitive social health insurance.[15]

The essence of the approach is to develop individual capitation payments for age and sex, and to adjust these for any unambiguous and universally recorded measure of disadvantage (such as welfare receipt) and for any verifiable indicator of previous disease diagnosis. In contrast to the traditional uses of such approaches, the interest in England is not in creating a "portable" capitation amount that is paid to the insurer chosen by the citizen, but to secure more accurate indicators of health services expenditure at the level of small (GP) populations.

The researchers have tested several schemes from the USA that use indicators of health status arising from previous encounters with the health system. These include the Hierarchical Condition Categories model introduced into Medicare in 2004.[16] Such systems seek to incorporate increasingly refined indicators of diagnosis to be used in patient classification, including pharmaceutical use and ambulatory care, as well as hospital care. Their major technical contribution is to collapse a wealth of diagnostic information derived from patient encounters into a manageable number of patient classifications.

Diagnosis-based capitation methods offer greatly increased predictive power over other methods, explaining up to 17% of an individual's future health-care expenditure. This means that when individuals are aggregated into purchaser risk pools, such as PCTs or general practices, the precision of the purchaser's budget is very much higher than under cruder area-based capitation methods. The major limitations of the methods are their considerable information requirements, the risk that they will stimulate unnecessary health-care use, the incentives they offer GPs to "cream skim" patients for whom expected expenditure is less than the capitation payment, and the risk of underestimating unmet need (a problem shared with most empirical methods). Researchers are currently seeking to set out these trade-offs for Department of Health policy-makers.

Conclusion

In countries such as England, the continued refinement and development of funding formulae will be essential if the principle of public funding of health services is to continue to enjoy widespread political and public support. The introduction of formula funding is also becoming increasingly urgent in many less-developed health systems, especially those pursuing a policy of decentralization. If localities do not believe that the national funds are fairly allocated, it is unlikely that any decentralization policy will be sustainable. Of course, data availability often limits the scope of any formulaic approach. However, the English experience demonstrates that substantial progress in developing formulae can be made even with limited information. And without such formulae, there is little prospect of promoting accountability and equity within the health system. ■

Résumé

Allocation des ressources et achats dans le secteur de la santé : l'expérience anglaise

Le Royaume-Uni de Grande Bretagne et l'Irlande du Nord disposent d'une expérience étendue dans l'allocation de fonds destinés aux services de santé aux régions et aux localités à l'aide de formules de financement. Le présent article se concentre sur le cas de l'Angleterre. Il s'intéresse plus particulièrement au souci politique récemment exprimé de réduire les inégalités évitables en termes de santé en élargissant le champ d'application des formules de financement. Il examine également les problèmes qui apparaissent lorsque l'on tente d'allouer des fonds à des unités organisationnelles très petites, telles que les cabinets de médecine générale. L'exemple anglais intéresse les systèmes de santé moins développés, en particulier ceux que les gouvernements cherchent à décentraliser, en vue d'une responsabilisation et d'une équité plus grandes.

Resumen

Asignación y adquisición de recursos en el sector sanitario: experiencia de Inglaterra

El Reino Unido de Gran Bretaña e Irlanda del Norte tiene una vasta experiencia de uso de fórmulas de financiación en la asignación de fondos para servicios de salud a regiones y localidades. Este artículo se centra en el caso de Inglaterra. Se presta especial atención al reciente interés que en materia de políticas suscita la posibilidad de reducir las desigualdades evitables ampliando el alcance de las fórmulas de asignación de recursos. En el artículo se analizan también las cuestiones que se plantean cuando se intenta asignar fondos a unidades organizacionales muy pequeñas, como los consultorios generales. El ejemplo de Inglaterra reviste interés para otros sistemas de salud menos desarrollados, especialmente para los gobiernos que aspiran a descentralizar la atención, mejorar la rendición de cuentas y fomentar la equidad.

ملخص

تخصيص الموارد والمشتريات في القطاع الصحي: التجربة الإنكليزية

تتمتَّع المملكة المتحدة لبريطانيا العظمى وأيرلندا الشمالية بخبرات واسعة في مجال تخصيص الاعتمادات المالية للخدمات الصحية بالأقاليم والمحليات باستخدام صيغ تمويل مختلفة. وتركِّز هذه الورقة على إنكلترا، وتولي اهتماماً خاصاً للشواغل التي ظهرت مؤخراً حول السياسات، في ما يتعلق بتقليص مظاهر الجور التي يمكن تلافيها في مجال الصحة،وذلك بتوسيع نطاق صيغ إرسال الموارد المالية إلى الوحدات المختلفة. وتتناول الورقة أيضاً القضايا التي تنشأ عند طلب تخصيص اعتمادات مالية لوحدات ذات حجم تنظيمي صغير جداً، كوحدات الممارسة الطبية العامة. ويُعَدُّ المثال الإنكليزي مناسباً للنُظُم الصحية الأقل نمواً، ولاسيَّما لدى الحكومات التي تسعى إلى تطبيق اللامركزية وتحسين عملية المساءلة، وتعزيز جوانب المساواة.

References

1. Smith P. *Formula funding of public services*, London: Routledge; 2006.
2. *2006-07 and 2007-08 Primary Care Trust initial revenue resource limits*, London: Department of Health; 2006.
3. *Sharing Resources for Health in England: report of the Resource Allocation Working Party*. London: Department of Health and Social Security; 1976.
4. Mays N, Bevan G. *Resource allocation in the health service*. London: Bedford Square Press; 1987.
5. *International Classification of Diseases*. Geneva: WHO; 2008. Available from: http://www.who.int/classifications/icd/en/ [accessed on 6 October 2008].
6. Smith PC, Rice N, Carr-Hill R. Capitation funding in the public sector. *J R Stat Soc Ser A Stat Soc* 2002;164:217-57. doi:10.1111/1467-985X.00200
7. Carr-Hill RA, Sheldon TA, Smith P, Martin S, Peacock S, Hardman G. Allocating resources to health authorities: development of methods for small area analysis of use of inpatient services. *BMJ* 1994;309:1046-9. PMID:7950737
8. Smith P, Sheldon TA, Carr-Hill RA, Martin S, Peacock S, Hardman G. Allocating resources to health authorities: results and policy implications of small-area analysis of use of inpatient services. *BMJ* 1994;309:1050-4. PMID:7950738
9. Martin S, Rice N, Smith P. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *J Health Econ* 2008;27:826-42. doi:10.1016/j.jhealeco.2007.12.002 PMID:18261812
10. Acheson D. *Independent inquiry into inequalities in health*. London: HMSO;1998.
11. Department of Health. *Saving lives: our healthier nation*. London: The Stationery Office; 1999.
12. Audit Commission. *What the doctor ordered: a study of GP fundholders in England and Wales*. London: The Stationery Office; 1996.
13. *Putting commissioning into practice*. London: Audit Commission; 2007.
14. Martin S, Rice N, Smith P. Risk and the general practitioner budget holder. *Soc Sci Med* 1998;47:1547-54. doi:10.1016/S0277-9536(98)00231-7 PMID:9823050
15. Van de Ven WPMM, Beck F, Buchner D, Chernichovsky L, Gardiol A, Holly A. Risk adjustment and risk selection in the sickness fund insurance market in five European countries. *Health Policy* 2003;65:75-98. doi:10.1016/S0168-8510(02)00118-5 PMID:12818747
16. Pope GC, Kautter J, Ellis R, Ash AS, Ayanian J, Iezzoni L, et al. Risk adjustment of medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev* 2004;25:119-41. PMID:15493448