

# nuffield/york PORTFOLIO

General Editor: A. J. CULYER

1. **THE ELDERLY.** *Who cares? Who pays?*  
Alan Maynard and John C. C. Smith.

2. **MEDICAL ETHICS.** *Health service efficiency and clinical freedom.*  
Alan Williams.

3. **VALUING HUMAN LIFE IN HEALTH SERVICE POLICY.** Gavin Mooney.

4. **COMMUNITY CARE.** *Its meaning and importance.* K. G. Wright.

5. **INEQUALITIES IN HEALTH AND HEALTH CARE.** Julian Le Grand.

6. **COST EFFECTIVE ANALYSIS IN HEALTH CARE.** Michael Drummond.

7. **CHILDREN IN CARE.** *Planning without costs.* Martin Knapp.

8. **OPTION APPRAISAL IN THE NHS.** *A guide to better decision-making.*  
R. L. Akehurst and M. J. Buxton.

9. **DOES UNEMPLOYMENT KILL?**  
Hugh Gravelle.

10. **THE ETHICAL IMPERATIVE OF CLINICAL BUDGETING.** Iden Wickings and James Coles.

ALAN MAYNARD

Professor of Economics and Director of Centre for Health Economics in the University of York.

JOHN C. C. SMITH

Nuffield Provincial Hospitals Trust.

THE AUTHORS

EDITOR, A. J. CULYER

# THE elderly

Ten billion pounds of public money is currently spent annually on the care and support of the elderly. And their numbers are growing fast—official statistics put the expected rise in the next decade or so at 20 per cent for the over 75s and 50 per cent for the 85s and over.

Within the context of perhaps the major social challenge of the turn of the century, this paper examines the position of a particularly vulnerable group—the elderly infirm—and questions whether the present systems of financial support can sustain a system of care that is cost-effective and appropriate to their needs. As a definition of infirmity we take the conditions which have to be satisfied to qualify an individual for attendance allowance—those whose infirmity (physical or mental) places them in need either of prolonged or repeated care and attention throughout the day or night, or continual supervision to avoid substantial danger to themselves or others.

## Who cares?

### *Policies in action*

Government policies recognise that elderly people have differing individual needs and that their personal preferences should be met wherever possible. Thus there are several different modes of provision and support available which can be broadly classified according to place of residence, and which theoretically match the needs of a particular elderly group:

#### 1. At home

#### 2. Sheltered housing

Provided by Local Authorities (LAs) Housing Associations and some private provision

#### 3. Residential homes

Mainly provided by LAs but also voluntary bodies and a growing private sector provision

#### 4. NHS long-stay hospitals

#### 5. Private nursing homes

Both the general practitioner services and the acute hospital services have vital supporting roles but are not regarded as separate care options in themselves.

## Who pays?

The type of care likely to be most appropriate is largely governed by the individual's degree of dependancy. A spectrum of dependancy is apparent which is satisfied at one end by care in the community (at home and in sheltered housing) and at the other by institutional care in hospitals and nursing homes with residential homes falling somewhere in between. Since individual needs change with advancing years, the system also assumes that movement from one form of care to another can take place readily.

### *Matching care to needs*

There is abundant evidence about the difficulty of securing the transfer of an infirm elderly person to a more appropriate form of care. A recent study showed that 25 per cent of elderly patients on community nurse lists and 23 per cent of elderly residents in LA (Local Authority) homes had high or very high nursing care requirements (Wade, Sawyer, and Bell, 1981). On the other hand 11 per cent of long-stay patients on hospital geriatric wards had

Table 1. *Care of the elderly: sources of funds, administration, and service*

Source provision	Administering Authority	
1. Department of Health and Social Security (DHSS)		
(a) Transfer payments Retirement pension Supplementary benefit Attendance allowance Invalid care allowance	Local Social Security Offices	Support elderly at home and assist contributions to LA services, private homes, and nursing homes
(b) National Health Service (NHS) Vote	NHS Authorities	Hospitals/Nursing Homes Community Health Services Family Practitioner Services
(c) Personal Social Services (PSS) Vote	Local Authorities	Personal Social Services
2. Department of the Environment (DOE)		
(a) Rate Support Grant (RSG)	Local Authorities	Residential Homes
(b) Housing Subsidies (HS)	Housing Associations	Sheltered Housing Other LA Housing Rent Rebates, etc.
3. Local rates	Local Authorities	Augment PSS, RSG, and HS provision
4. Personal resources Private income/capital from investments, occupational pensions, etc	—	Support elderly at home, contribute to charges made for LA services, private residential and nursing homes, but NOT in NHS facilities
5. Relatives	—	As above
6. Voluntary sector	—	Residential homes and nursing homes
7. Private sector	—	Nursing homes and some residential homes

minimal nursing care requirements. It was also found that the needs of almost a quarter of residents in LA residential homes could be more appropriately catered for in a different environment. These observations related to the physically infirm. A potentially more serious problem is the growing number of Elderly Severely Mentally Infirm (ESMI) residents to be found in the community in LA residential homes and, to a lesser extent, in sheltered housing. Wade et al define the problem as

the elderly are either receiving the care they need in an unsuitable environment or they are living in an environment that is suitable but are not receiving the care they need.

The prospect of transfer to a more appropriate form of care often imposes severe pressures on individuals who, quite naturally, are extremely reluctant to face the uncertainties associated with any move from familiar surroundings. Resistance to change and obstacles to arranging transfers have led many providers to seek to widen the scope of their services so as to cope with the rising dependency of the elderly infirm and thereby postpone the need to transfer. These efforts, however praiseworthy and desirable, do little more, however, than soften the impact of the present system without tackling the fundamental issues.

Several factors contribute to the rigidity of the system: imbalances in the scales of provision, the division of

responsibility between separate and independent authorities, professional relationships and attitudes and, perhaps, inadequate investment in care. But in our view it is the complexity, incompatibility, and lack of coherent objectives in the various systems of financial support that constitute the chief barrier to effective policy implementation.

There are positive financial disincentives to discourage present providers of care from relating to each other and jointly offering what is really best for their clients. It is this enforced segmentation that leads to some clients being held in inappropriate forms of care with others, for whom it would be appropriate, being denied access to it. The chief culprit, which produces both inhumanity and inefficiency, is the present system of finance.

### *Sources of revenue*

The main sources of finance in care for the elderly infirm are set out in Table 1. Part of the NHS vote (not shown in Table 1) is allocated to NHS authorities as Joint Finance which, by agreement between health and local authorities, is used to augment local authority personal social services.

### *Opportunity costs and transfer payments*

It is easy to confuse opportunity (economic) costs and transfer payments with the result that policy options may not be clearly evaluated and resources may be misallocated.

Table 2. Costs to be included in the appraisal of alternative patterns of care for the elderly

Hospital and residential care	Community care
Capital Costs	Capital Costs (Housing)
Care Service Costs	Care Service Costs
General service costs	Personal living expenses
Personal consumption	Informal Help

Source: Wright, Cairns, and Snell (1982)

The reason for identifying *opportunity* costs is that these measure the value of resources in their most highly valued alternative use. Accordingly, if the aim is to offer efficient services it is important to ensure that the uses to which resources actually are put are more highly valued than their (best) alternative uses. This is true within the services for the elderly, within the NHS or the public sector as a whole and, at its broadest, within the entire economy. The actual prices paid for resources do not always reflect their true opportunity cost and some resources already owned by authorities (for example, geriatric beds) may have no price attached to them at all (but their use plainly has opportunity costs).

The main opportunity costs of care for the elderly are set out in Table 2.

Most of the items are reasonably self-explanatory. Capital costs are difficult to compile because it is not easy to identify the opportunity costs of using capital resources for the elderly as opposed to some other group: the NHS keeps no capital accounts. Housing capital costs must be compiled on a replacement cost basis if new building is implied; otherwise special calculations are needed to identify the opportunity cost. Care service costs are more easily measured as the running costs of the facilities and services provided. General service costs relate to maintenance of the capital stock including cleaning and portering and are unlikely to vary much with patient dependance.

Personal consumption in residential care is financed out of pension and any other resources net of payment for care and, in long-stay hospitals, by abatement of pension (leaving the patient with pocket money only). Personal living expenses can be derived, for different care options, from the Family Expenditure Survey. The costs of informal care are borne by many individuals and groups—including relatives and charitable bodies.

Transfer payments do not form part of opportunity costs: social security payments, for example, represent a change in ownership (or distribution) of a given national income. Even though transfer payments do not use up resources but merely reallocate purchasing power, they are important for two particular reasons in relation to care for the elderly. Firstly transfer payments are important sources of purchasing power for client welfare. The payments of attendance allowances may not be opportunity costs but they transfer opportunities from the taxpayer to the beneficiaries and thus enable the latter to secure access to the services needed. Secondly, the arrangements under which transfer payments are paid can encourage or inhibit the efficient attainment of policy objectives. Transfer payments create financial flows which are a means to a caring end—but only if they are devised and implemented efficiently. They help shape effective budgets which can variously promote or hinder efficient caring.

Financial flows: Expenditure

Many authorities, institutions, and individuals take part in the decision making processes that control resource allocation. The DHSS, via the Exchequer, takes money from taxpayers and national insurance contributors and distributes it according to statutory regulations to the elderly or their agents in the form of pensions, supplementary benefit, attendance allowance, and invalid care allowance. This money is used by the elderly to purchase goods and services (for example, to meet LA charges in sheltered housing, residential homes, or to defray fees charged for private residential or nursing homes). This public sector finance augments private sources—individuals (occupational pensions and savings), relatives (gifts)

and other carers (voluntary or private bodies). The public sector agencies (the NHS and LAs) also acquire access to taxpayers' money in the form of Exchequer grants to finance the provision of services and goods.

The distinction between resource allocators (those who decide what to demand), financiers (those who finance in part or in whole what is demanded) and suppliers (those who provide the care that is demanded and financed by others) is important. Many decision-makers become involved at different levels and respond in different ways to the financial incentives and disincentives they confront. This gives rise to financial impediments to an integrated and efficient system of care.

The sources of finance

Table 3 illustrates and compares the expenditure per week flowing from public funds to support an elderly infirm person in the 75+ age group with no private resources.

**LIVING AT HOME**

Under Option 1 transfer payments are the largest item of public expenditure. These include attendance allowance to which the infirm elderly are assumed to be

entitled—that is, their degree of infirmity is taken as that ordinarily accepted on medical assessment to justify entitlement to that benefit. Under this option an additional transfer payment—Invalid Care allowance at a weekly rate of

£19.70—may also be payable where there is a relative or other person able to care for the elderly invalid at home. Where payable this would raise the total public expenditure to £147.29 weekly. The balance of expenditure comes largely from items

Table 3. *Comparison of public expenditure and sources for alternative care options (£ 1982)*

	Expenditure (£ per week) under Option				
	1 (Living at home)	2 (Sheltered housing)	3 (Residen- tial homes)	4 (NHS long-stay hospitals)	5 (Private Nursing Home)
<b>TRANSFER PAYMENTS</b>					
<b>(a) Supplementary pension</b>					
Long-term scale rate	32.70	32.70	32.70	—	—
Board and lodging rate	—	—	—	—	60.00
Extension for elderly	—	—	—	—	13.30
Personal expenses	—	—	—	—	9.45
Heating	4.65	4.65	—	—	4.65
Laundry	0.75	0.75	—	—	0.75
Housing requirement	14.22	14.22	—	—	—
<b>(b) Retirement pension</b>	(32.85)	(32.85)	(32.85)	6.55	(32.85)
<b>(c) Attendance allowance</b>	26.25	26.25	—	—	26.25
Sub-Total	78.57	78.57	32.70	6.55	114.40
<b>SERVICE PROVISION</b>					
<b>(a) Local Authority</b>					
Residential home	—	—	83.82	—	—
Capital charges	—	—	7.11	—	—
Sheltered housing	—	2.50	—	—	—
Home help	23.52	23.52	—	—	—
Meals	3.50	3.50	—	—	—
Home nursing	18.00	18.00	18.00	—	—
Sub-total	46.98	49.48	108.93	—	—
<b>(b) NHS</b>					
Long-stay hospital	—	—	—	215.00	—
Family practitioner services	4.00	4.00	4.00	—	4.00
<b>TOTAL</b>	<b>127.59</b>	<b>130.09</b>	<b>145.63</b>	<b>221.55</b>	<b>118.40</b>
<i>Notes</i>					
1. Sources: DHSS—Social Security Statistics, 1982; Benefit Rates November 1982—November 1983; Health Service Costing Returns (March 1981); CIPFA—Personal Social Service Statistics (1981/82 Actuals).					
2. Details of assumptions and computations are available from the authors.					
3. Transfer payments=gross entitlement not net payment. LA Services=actual average cost.					
4. There is evidence that clients and their relatives are ignorant of their 'welfare rights'. However reduced consumer ignorance and the open-ended nature of the social security budget has obvious implications for public expenditure, implications of which the Treasury is all too aware.					

of service provided by LAs. The items included are those it seems reasonable to expect a person in this category to receive. Nonetheless many local authorities do not or cannot afford to provide services at this level to all elderly persons in this category. The LA expenditure shown is net of charges recovered from the beneficiary—transfer payments are among the personal resources against which such charges are made.

#### SHELTERED HOUSING

Under Option 2 transfer payments are again the largest item of public expenditure. Attendance allowance has been included, even though it is rare for this benefit to be awarded to persons living in sheltered

accommodation—despite the high dependency levels reported for this group. Charges for accommodation and services are raised against tenants, usually based upon fair rent assessments. In the case of LA provision such charges may not meet full operating costs, so expenditure may be higher than that estimated. Much the same applies to sheltered accommodation provided by housing associations, where differences between costs and charges may be recoverable from 'revenue deficiency' grants.

#### LA RESIDENTIAL HOMES

Under Option 3 attendance allowance is not payable to persons living in LA residential homes even

though many residents may have a qualifying infirmity. Supplementary benefit too is assessed at the scale rate corresponding to the current rate of retirement pension. In such cases charges are levied at a level which leaves the pensioners with a small amount of pocket money. Thus, under Option 3 the main source of funds is the LA. The expenditure shown is again net of charges recovered from the beneficiary (on average charges amounted to £31.71 per resident week in 1981/82).

#### LONG-STAY HOSPITALS

Under Option 4 costs are met entirely from NHS expenditure. This is clearly the most expensive form of

provision though transfer payments are at their lowest level. Neither supplementary benefit nor attendance allowance is payable and retirement pension is subject to abatement to the 'pocket money' level. It seems clear that if patients currently accommodated in NHS long-stay hospitals could be transferred to suitable nursing homes in the private sector, the level of public expenditure incurred in providing care would be appreciably lower.

### PRIVATE NURSING HOMES

In sharp contrast, the highest level of transfer payment is to be found in

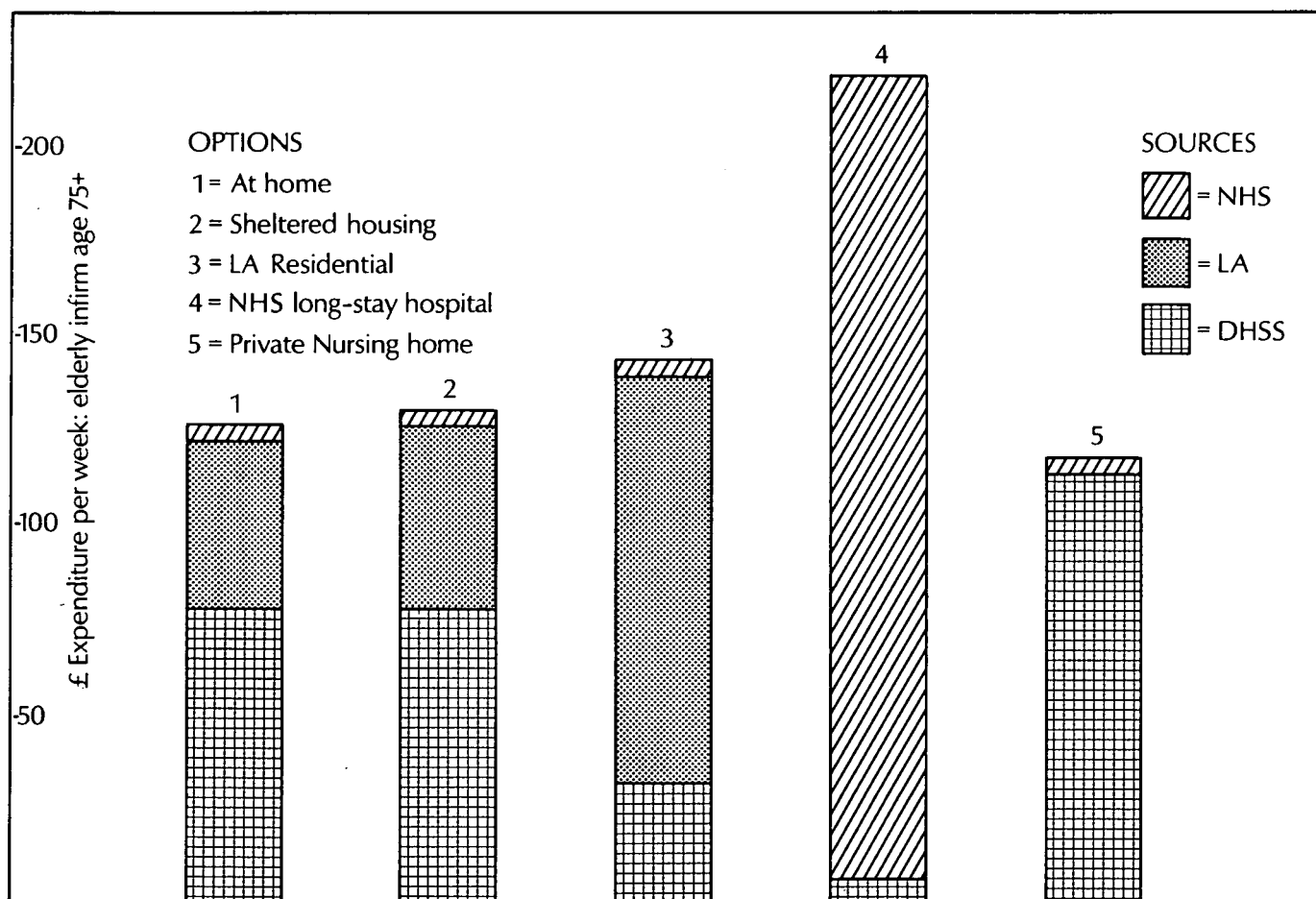
Option 5. Here the supplementary benefit is at a high level based on a locally determined board and lodging rate. Additional benefits are normally payable and so is attendance allowance. The gap between total transfer payments and private nursing home charges—about £30 to £40 per week in London and the Home Counties—is normally met from a combination of charitable donations, relatives, and other personal income. Thus although private nursing homes have tended to be regarded as open only to those with independent means and/or supportive relatives, access to them by elderly persons with personal incomes much lower than the

supplementary benefit rates is growing. Moreover, if and when personal income or capital resources becomes exhausted, many find that they are able to continue to reside in private nursing homes by taking advantage of the high level of transfer payments—particularly supplementary benefit and attendance allowance. In some cases the gap between resources and charges may be filled by grants from either the local authority or the health authority. In either case the net cost to the public authorities is clearly far lower than the cost of providing alternative accommodation in a residential home or long-stay hospital.

## Identifying the financial barriers

The information in Table 3 is shown as a histogram in Figure 1. This shows the broad relationships between the various expenditure flows under the five options.

Figure 1. *Comparison of public expenditure and sources for alternative care options (1982)*



The histogram illustrates the problem involved in securing a transfer to the mode of care that may be more appropriate to individual needs or that may be more cost-effective. Transferring a patient inevitably involves significant changes in the sources of supporting public expenditure. Each source is subject to different lines of accountability and governed by rules and regulations mostly of statutory origin.

Health and local authorities alike operate within fixed budgets. Although the allocations each receives from central government are weighted to reflect the proportion of elderly in the local

population, the resources that each can devote to services for the elderly are determined in competition with the demands on available funds from the other services provided. Decisions have necessarily to be aimed at achieving a balance across the board, having regard to the resources available, the estimated need, and a local view of priorities. Moreover, there are significant differences in the ways in which resources are channelled to health and local authorities.

### *Incompatible rules*

Local authorities can augment their total services by increasing the rate precept. They can also recover costs, at least in part, by raising charges. Health authorities do not have similar powers. Policy requires health services to be free at the point of delivery so, regardless of an individual's personal resources, health authorities have no power to impose charges on NHS patients. The principle of 'no double payment from public funds' in turn operates by withholding all transfer payments on admission to NHS hospitals, with the exception of a residual payment of 'pocket money' to those entitled to retirement pension.

The NHS thus has to meet the total cost of the services it provides from a fixed budget—none of the 'saved' expenditure on transfer payments flows back automatically to the NHS. Local authorities, on the other hand, are dealing with the marginal costs of their services. With some exceptions, the financial flow from transfer payments remains open to them to fund the balance by charges. Moreover, other personal resources can be tapped by local authorities but not by health authorities.

### *More flexible patient transfers*

The widely reported difficulties met by the NHS (hospitals and general practitioners) and by LAs (social ser-

vices) in arranging the transfer of an elderly infirm person from home to a residential home or hospital, or from residential home to hospital and *vice versa* are clearly considerably influenced by the different financial incentives. The present financial machinery can therefore be a real barrier to the operation of suitable health care policies for the elderly. For the elderly infirm the consequences are very serious. Increasing numbers can expect to be prevented from gaining access to the type of care most suited to their needs. For the providers the consequences are no less serious as infirm patients become increasingly locked into that part of the care system in which they happen to be, regardless of whether this is either the most suitable or the most cost-effective.

There are other ways in which the financing system erects barriers and influences choice. Some patients living at home, or with their relatives, seek admission to an NHS rather than a LA facility because the former is free of charge while the latter is not. So far as access to transfer payments is concerned, public sector provision is constrained whereas private sector provision is not. Thus the high level of transfer payments available is an inducement to seek places in private nursing homes. In some areas there are waiting lists for admission to such homes, in others there has been a steady growth in their number, including homes for the elderly mentally infirm.

---

## Financial barriers to policy implementation: The critical factors

Of the various financial impediments discussed, some are more critical than others. Chief among these is the inconsistent effect of transfer payments. Their role in community care and in sheltered housing is significant. Nonetheless the important contribution made by the attendance allowance to support the elderly infirm at home in the community is apparently rarely available for those whose infirmity and dependency develops when they are living in sheltered accommodation. The reason for this distinction is obscure but its impact on choice is significant. A similarly incongruous distinction occurs in LA residential homes. Here there is a statutory bar to payment of attendance allowance and supplementary benefit is geared to the level of retirement pension with no housing or board and lodging component. Inevitably this restricts LA charges to a minimum and raises the marginal burden on LA expenditure. In the NHS sector transfer payments have no impact at all on the supply of facilities; but in the private sector the wheel turns full circle and transfer payments become a major source of finance.

Of similarly critical significance is the provider's ability to recover costs by raising charges for services provided—exercised by LAs and the private sector, but denied to the NHS. In both cases charges are directed at all personal resources whether in the form of transfer payments or other personal income. The extent to which personal income other than transfer payments is effectively tapped by LAs is not readily assessable but it cannot at present be tapped at all by the NHS. Since both sets of providers are in the public sector and have a shared clientele, the differential treatment seems absurd.

# Lifting the barriers

## ... in the USA

The problems we have identified are not uniquely British. In the USA Anne Somers (1982) has said they 'result from trying to fit the new needs on the Procrustean bed of outmoded policy'. Among the basic guidelines she suggests for a new policy, emphasis is placed on the importance of achieving *neutrality* of public financing between acute and long-term care, and between institutional and home based care, and also of ensuring equal access to publicly funded benefits regardless of economic status. She advocates a new programme for long-term care which would have three principal functions:

- to ensure maximum feasible co-ordination among health care institutions, agencies and programmes involved in the care of the elderly;

- to allocate resources in the most equitable and effective manner;

- to provide comprehensive assessments, appropriate placement and cost effective case-management of individual long-term patients,

and a system that achieves a more balance blend of public and private responsibilities. Different circumstances but the analysis and objectives have a familiar ring!

## ... in the UK

*Care in the Community* addressed similar issues and exposed the Government's thinking on possible alternative options. Its emphasis, however, was on the question of facilitating transfer from hospital to community which, however appropriate for some patients, is far from obviously desirable for the highly dependent elderly infirm (especially those lacking constant family support)—in terms either of meeting their needs appropriately or of cost-effectiveness. Nor can community care always provide the security for the dependent elderly that they value highly. The Consultative Document did not consider these issues nor did it address the question of neutral financing. Indeed, though it discussed opportunity costs in relation to the possibilities for change, it did not consider at all the distorting impact of transfer payments.

Attempts to secure the integration necessary for the successful implementation of policy have hitherto pursued the 'collaboration' path. Collaboration is obviously necessary in a system that divides responsibility between so many different agencies. But its cutting edge is blunted by constant contact with the rigidity imposed by financial barriers. Legislative measures designed to reduce the division of responsibility are likely to succeed only if they can be matched by change in the financing mechanisms.

---

## Joint finance

One attempt to bridge the divide between the NHS and LAs came with Joint Finance in 1976. This was designed to pump-prime an expansion of LA services from NHS sources in the expectation that over time the NHS would not need to continue inappropriate and expensive provision. The innovation was generally if cautiously welcomed but its success is as yet unknown. It has been criticised for its limited scope, both in terms of the amount available and the purposes for which it can be used. *Care in the Community* considered widening the scope of Joint Finance as one of the options. There is little doubt that increased virement for NHS resources would provide a greater flexibility which, in some localities, could be turned to advantage. Nonetheless there must be limits to the size of a Joint Finance package. Once these are exceeded, its function as a lubricant easing the meshing of a complex allocation system will cease and leave behind yet another component to complicate and possibly add to the rigidity of the system.

Any effective reforms should aim at creating an integrated and competitive market in two key ways:

- removing the differential treatment of providers in terms of their access to finance and ability to recover costs by charges or otherwise;

- establishing a unified system of financial support for elderly infirm clients that is based upon medical assessment of need, related to individual resource capacity, and neutral to the chosen form of care.

## Providers and the market

Ideally the prices of different forms of care should reflect the opportunity costs of provision. Effective supply of care requires that providers—public and private—should have equal opportunities of access to capital and an equal

capacity to recover the running costs of provision, for example by generating revenues from the 'sale' of their services. Such competition ought also to be based upon prescribed standards to be met by all providers. Here the State has a clear role in prescribing, regulating and monitoring standards. Providers should also be open to public scrutiny through an information system showing compliance with standards and the costs/charges of provision. An integrated market would allow proper comparisons to be drawn about the quality and costs of provision and for client needs to be more flexibly matched to provision. Equal treatment of providers would remove the current segmentation of the market and ease the process by which a more suitable balance of care can become available.

The ideal is obviously difficult to achieve. Fundamental changes would be required in the public sector to place local and health authorities on a similar footing. The critical areas are, as we have seen, on the one hand the ability to recover operating costs through charges on the financial resources, both public and personal, of the client and, on the other, a common policy on means testing and subsidy and similar rules and procedures governing access to capital. The extension of charges in this sector is obviously fraught with political difficulty, particularly since the client group is highly vulnerable. The change need not, however, be as great as it seems. The costs of treatment should continue to be excluded from any charge. As regards other costs, the State already effectively imposes charges by almost wholly suppressing entitlement to transfer payments for elderly patients in long-stay hospitals. To a lesser extent, some withdrawal and limitation of entitlement also takes place in the local authority sector, but LAs may still raise charges against residual transfer payment and, more importantly, are able to take other personal resources into account.



## Clients and the market

Clients can mobilize resources from private sources (individual, family and charitable) and public sources (social security benefits). Innovations in both these sectors are needed.

In the private resource sector, it is remarkable how slow the insurance industry has been to develop cover for infirm people in old age. Making contributions to occupational pension schemes to provide income in old age is commonplace. But contributory insurance schemes that provide benefits in cash or in kind as dependency increases with age are unusual. With improving survival rates, the likelihood of survival into the eighties and beyond is increasing. But it also brings with it the problem of increasing dependency due to severe physical and mental infirmity. These risks can be actuarially assessed and suitable insurance could be offered.

Elderly people often have considerable capital resources—usually in the form of property. There are signs of growing activity in the building society, annuity and bond market to enable income to be generated from these resources in very old age. Thus the elderly can generate a cash flow from the capital they own, for example by selling the house to an institution in exchange for property rights in that dwelling until death and cash resources to buy an annuity now. There is room for the active development of this market by imaginative private institutions. Again the State has a role to play by easing some of the constraints currently applied in taxation and social benefit policies. Liberation of the resource potential in this area can, in the long run, have only beneficial effects on the level of public expenditure: mobilising private resources could save public resources and lead to a better quality of care for the elderly.

## What needs to be done

At present thousands of millions of pounds are being spent to provide care for the elderly. The market is segmented, inefficient and excessively rigid; locking many elderly people into a mode of care that is no longer appropriate to their needs. Responsibility for the care programmes is also spread across far too many institutions, public and private. There is a need to review the provision of care for the elderly, in particular:

1. to remove the barriers to ready transferability of patients between different forms of provision so that patient dependency and provision can be better matched;
2. to improve the transferability of the elderly between public and private sector institutions;
3. to improve the transformation of capital assets into cash that could finance care, and to develop an insurance market for nursing care arising from dependency in old age.

The identification of more cost-effective and more need-effective caring for elderly patients is, however, not enough. At present access to the appropriate modes of care is seriously inhibited by financial arrangements which had separate origins and have never been effectively consolidated. A piecemeal approach to particular parts of the market is inadequate. A wider view of the nature and inter-relations of NHS, LA, Social Security, and private funding is essential.

The creation of an integrated and flexible pattern of care, public and private, for the elderly will not be easy. Many interest groups will assert producer interests in an effort to defend the existing pattern of income and employment. However the resource costs of these programmes and the size of the elderly population are such that radical reform cannot be put off for long. It is high time that the clients' interests were put first—and that needs the kind of changes we have suggested.

### REFERENCES

- Department of Health and Social Security (1981a). *Growing Older*. London: HMSO, Cmnd. 8173.
- Department of Health and Social Security (1981b). *Care in the Community, A Consultative Document on Moving Resources for Care in England*. (DHSS Circular HC(81)9).
- Somers, Anne (1982). 'Long-term care for the elderly and disabled: A new priority', *New England Journal of Medicine*, 307(4), 221-6.
- Wade, B., Sawyer, L., and Bell, J. (1981). *Different Care Provision for the Elderly: Report to the DHSS by the Department of Social Science and Administration*. London School of Economics (Unpublished).
- Wright, K. G., Cairns, J. A., and Snell, M. C. (1981). *Costing Care: the Costs of Alternative Patterns of Care for the Elderly*. University of Sheffield, Social Services Monographs.

## COST SHOULD NOT BE A FACTOR IN MEDICAL CARE

*To the Editor:* Of late an increasing number of papers in this and other journals have been concerned with "cost effectiveness" of diagnostic and therapeutic procedures. Inherent in these articles is the view that choices will be predicated not only on the basis of strictly clinical considerations but also on the basis of economic considerations as they may affect the patient, the hospital, and society. It is my contention that such considerations are not germane to ethical medical practice, that they occupy space in journals that would be better occupied by substantive matter, and that they serve to orient physicians toward consideration of economics, which is not their legitimate problem. It is dangerous to introduce extraneous factors into medical decisions, since consideration of such factors may eventually lead to considerations of age, social usefulness, and other matters irrelevant to ethical practice. The example of medicine in Nazi Germany is too close to need further elucidation.<sup>1,2</sup>

It is incumbent on the physician (especially in a critical situation) to practice not "cost-effective" medicine but medicine that is as safe as possible for that patient under the particular circumstances. Optimization of survival and not optimization of cost effectiveness is the only ethical imperative. To select diagnostic tests on the basis of cost effectiveness is a deliberate statistical gamble; to use diagnostic tests in an unthinking medical fashion is poor medicine, not because of cost but because unthinking medicine is dangerous for the patient. Ethical physicians do not base their practices on the patient's ability to pay or choose diagnostic and therapeutic procedures on the basis of their cost. It may be argued that the welfare of society is threatened by escalating medical costs; indeed, that argument at first appears to introduce a dilemma. Yet a large proportion of our ills are due to smoking, heavy drinking, and overeating, and the consequences of these indulgences consume a large portion of medical-care dollars. It is unfair to deprive those who have not been overindulgent of the best medical care while allowing the overindulgence of others to consume the available money. Furthermore, our society clearly has money to spend on luxuries and baubles. A physician who changes his or her way of practicing medicine because of cost rather than purely medical considerations has indeed embarked on the "slippery slope" of compromised ethics and waffled priorities.

ERICH H. LOEWY, M.D.  
62 Elm St.

Glens Falls, NY 12801

1. Ivy AC. Nazi war crimes of a medical nature. JAMA. 1949; 139:131.
2. Alexander L. Medical science under dictatorship. N Engl J Med. 1979; 241:39-47.

Folio 2.

# nuffield/york PORTFOLIOS

Editor, A. J. Culyer

## MEDICAL ETHICS

*Health service efficiency  
and clinical freedom*

ALAN WILLIAMS

Professor of Economics, University of York

Letter from the  
New England Journal of Medicine  
vol. 302, No. 12

Health service efficiency is not just about cost. Still less is it about money. It also concerns the benefits to patients both actual and potential--benefits that are principally to do with health itself. An unfortunate division of responsibility seems to have grown up whereby costs are the business of administrators and treasurers, while benefits are the business of doctors and nurses. In this way some are cast in the role of villain and others in the role of hero or otherwise. The two sides are seldom brought into a constructive relationship with each other. Some doctors even claim that it is unethical for them to be concerned about costs. In this Folio, Alan Williams examines medical ethics and argues that a more comprehensive view of what is ethical must be adopted than is common, particularly if the clash of conflicting moral principles is to be resolved in each of the three decision-making roles of doctors: as clinicians, as practice-managers and as policy-makers. In particular he mounts a powerful case for the gross inadequacy of the common view that the prime ethical imperative in medicine is to do all possible for the patient who consults him. He suggests instead a more comprehensive ethic based on cost-benefit principles.

A. J. C.

It is argued in some quarters that the drive for greater efficiency in providing health care conflicts with the only ethical imperative which it is proper for

doctors to accept, namely, that it is the doctor's duty to do all he can for the patient in front of him, regardless of cost? I propose to argue that doctors do not actually behave in accordance with that ethical imperative. Moreover, since their actual behaviour is accepted as being reasonable by their colleagues (and by others), it cannot be regarded by them as unethical. I shall further argue that this behaviour is best seen as a crude and imperfect form of cost-benefit analysis, which could be greatly improved upon without breaking any of the principles which already seem to inform doctors' behaviour. And if it were improved upon, the provision of health care would be much more efficient (that is, the health of the community could be improved at no greater cost). I therefore conclude that not to seek to become more efficient is what is unethical, because inefficiency means needlessly worsening the health of the people that doctors are there to serve.

Let me start by distinguishing three capacities in which doctors may serve patients. They may serve patients as clinicians, as practice managers, and/or as members of management teams. In the role of clinician, on which I shall concentrate most of my attention, the doctor is advising some individual who has come to him as a trusted professional expert. In the role of practice-manager the doctor is running his own activities, possibly jointly with professional colleagues, but still at the point of delivery of professional services to his (or their) clientele. In the role of policy-maker the doctor is participating in decisions about priorities for communities larger than his own professional clientele. I shall demonstrate that these roles shade off into each other almost imperceptibly.

---

## The doctor as clinician

In the letter displayed in the panel a distinction is drawn between 'strictly clinical considerations' and 'economic considerations as they may affect the patient, the hospital, and society'. These are useful distinctions with which to begin, so let us see how far we can get with them.

It is beyond dispute that the clinician's primary ethical imperative is to do nothing that will adversely affect the health of his patient. But this immediately raises the problems: what is health? does this injunction imply that any treatment which carries an adverse risk to health is unacceptable? and is it always clear who 'the patient' is? Let me start unpicking each of these problems in turn.

---

## *Is health a 'strictly clinical' matter*

In narrow medico-technical terms health can be seen as absence of pathology, and since few of us are without some blemish or other in this respect, there is clearly some truth in the old joke about the only healthy people being those who have not yet been adequately investigated. But without going quite as far as the World Health Organization in asserting that health is a state of complete physical and mental wellbeing, it does seem more appropriate to define health in terms which relate to patients' feelings and capabilities rather than to abstruse clinical parameters which may or may not generate any adverse effects on such feelings and capabilities. If health is described in terms of such categories as physical mobility, capacity for self-care, absence of pain or distress, ability to engage in normal social interaction, (all representing quality of life), and life expectancy (representing quantity of life), are we still within the realm of 'strictly clinical considerations', or have we already stepped outside the

proper bounds of medicine and into (dare I say it?) sociology? Is it unethical for doctors to see health in these terms, because if so we are already in difficulties in judging, for instance, the severity of illness, which must generally be related to the characteristics I have described. If, on the other hand, such a definition of health is perfectly acceptable, then it must be acknowledged that it embraces many non-clinical considerations. So even at this early stage, it proves difficult to exclude what the latter terms 'extraneous' (i.e. non-clinical) factors.

Let us now suppose that a proposed course of action carries risks to health which according to good clinical practice are acceptable (i.e. the expected benefits outweigh the risks). Although it may not always be made explicit (or indeed be clear in the mind of the clinician) this risk benefit calculation has three essential elements: first, a description of alternative (prospective) states of health (including death), second, a set of probabilities attached to each, and, thirdly, a relative valuation of each health state (or, more precisely, of each time profile of health states). So a clinician making the overall judgement that the risk is acceptable is weighing the (negative) value of the probability of bad outcomes against the (positive) value of the probability of good outcomes, and we have to ask what there is in his purely clinical expertise that enables him to do that. Well, there is obviously considerable clinical expertise involved in diagnosing the underlying problem that is causing ill health, and in knowing what can be done to influence the natural history of the condition for the better. If sufficiently skilled in communication, the clinician may be able to convey to the patient precisely what the alternative prognoses mean in terms of quality and quantity of life as described earlier, and he should be able to give some indication of the respective probabilities associated with each, given the age, sex, and general condition of the patient, and any concurrent health problems the patient may have. All this is a pretty demanding task, but it is still not sufficient to enable the overall judgement to be made that the risk should be acceptable to the patient, for to do that we still need to know what are the patient's valuations of the various prognosticated health states. These valuations may not be the same as those the clinician would make, nor those of the medical profession generally, nor even those of the rest of the community of which the patient is a member, and the making of such valuations certainly does not lie within the realm of 'strictly clinical' expertise.

So it appears not necessarily to be possible from 'strictly clinical considerations' either to decide whether a patient's health is better or worse, or whether the risks associated with treatment are acceptable or not, if the role of the clinician is one of perfect agency, i.e. if he is there merely as a well-informed trusted 'adviser, giving the patient all the information he needs to make a decision, then standing ready to implement that decision once the patient has made it.

*Is the  
acceptability  
of risk a  
'strictly  
clinical'  
matter?*

Note that this is the exact opposite of the view that the patient is there to give the doctor the information that enables the doctor to make the decision, and that the patient should then stand ready to implement it (i.e. to comply). However, in the classical view, the clinician is not to impose his values on his patients, though he is, as a matter of conscience, permitted to refuse to carry out a treatment which he believes to be against the patient's interests.

But which of the patient's interests are relevant in this context? Suppose the patient's overriding concern is that he may lose his job if he has to take time off for the recommended treatment, and would therefore prefer some (less effective) treatment which avoids that risk. Should the clinician go along with this clinically inferior (and possibly more dangerous) course of action, or should he reject this economic consideration as 'extraneous' and as compromising the practice of ethical medicine? Once more whose values are to count?

---

## *Who is the patient?*

Let us go a step further and suppose now that other people's interests (and health) are also directly affected by the patient's problems and their treatment (and I do not mean the case of infectious and contagious conditions where the clinician's responsibilities are fairly clear). For instance when a clinician is considering sterilising a patient, should he take account of the effects on the patient's spouse, even if the patient decides not to do so? Or when a clinician is considering whether or not to commit a difficult psychiatric case to custodial care, should he be influenced by the effects of not doing so upon the health of other members of the patient's household? Who exactly is 'the patient' in such cases, where the treatment of person A is influenced by the effects on the health of person B (whether or not person B is one of the clinician's own patients).

My impression from countless discussions of these matters with a wide range of clinicians is that most of them do not consider it unethical to allow these economic and social consequences (both for patients themselves and for their nearest and dearest) to influence the advice they give and the treatments they carry out or organise. In other words my clinical informants weigh a wider range of costs and benefits than the purely medical ones when forming a clinical judgement, and they weigh such costs and benefits for people other than the particular patient in front of them.

---

## *Are clinicians perfect agents?*

So far this has been impeccably altruistic, but what about the clinician's own interests? According to the classical model of the clinician as a perfect agent, the clinician considers no interests but the interests of the particular patient before him. He is not interested in his own remuneration and never lets that influence his behaviour. He is not interested in his own professional standing, except to the extent that this is affected by providing the best treatment he can for each and every patient. He never lets thoughts of his wife and family

and friends, and his social life with them, influence him in the amount of time he spends with his patients or precisely when he sees or treats them. The patient's welfare dictates everything. Such clinical paragons do not exist and, if they did, they would not survive the strain for very long. Such an extreme interpretation of the dictates of ethical medical practice is clearly neither practised nor held to be a reasonable expectation.

But where then is the line to be drawn and who is to draw it? It is clearly unethical to carry out unnecessary and potentially harmful interventions simply to make money but is it still unethical if the patient is rich and wants the intervention? It is clearly unethical to refuse treatment to a patient in urgent need of something safe and effective, which you can easily do, simply because you will miss the overture at the concert you have been looking forward to as part of your wife's birthday celebration. But what if the patient is someone in slight discomfort, which may well go away anyway, and, even if it doesn't, could well be left until tomorrow without serious consequences. Sometimes the welfare of the clinician and his nearest and dearest do (and should?) intrude, so how can the widespread practice of counting these costs (and benefits), along with the others, be held to be unethical? If it is unethical, a large proportion of the medical profession is in danger of being struck off the register!

When we consider the doctor in his role as practice-manager the problem gets further compounded, for now we confront the complication that his concern (or their concern if it is a group of collaborating clinicians and their support staff that we are talking about) is now the health of all the patients for whom the practice accepts responsibility. At this level it is more common to acknowledge the fact that criteria have to be drawn up to decide whether to accept responsibility or not, and what to offer patients once responsibility has been accepted. To be 'ethical' must these criteria be purely medical? Is it unethical to allow willingness or ability to pay to count? Is it unethical to allow any aspect of the economic or social situation of the patient, or that of his, nearest or dearest, to count? Again, judging by the observed behaviour of most practices the answer to both these last two questions is: no. This means that potential patients are going to be graded in order of priority by criteria which go beyond narrow medical considerations. This in turn implies that the doctor as practice manager is making general social judgements about the value of treating one person rather than another (or sooner than another, or more extensively and expensively than another), even when the purely clinical indicators are identical. For instance, if the identical hernia stops one man from working but not another, the former is likely to get priority for treatment and in such instances it would seem very odd to argue that it is unethical for these wider (economic) costs and benefits to be weighed in the balance.

## The doctor as practice- manager

*For whom  
is the  
practice-  
manager  
responsible?*

But there is a prior ethical problem which I have so far left buried away here, namely, the judgement about whom the practice-manager feels responsible for. Some possibilities are (1) those under active treatment by the practice at the time (2) those currently known to be waiting for treatment (3) those believed to need treatment but not currently seeking it, and (4) anyone who might ever need treatment. A doctor as clinician may assert that only the patient in front of him at some point in time is relevant, but the same doctor as practice-manager has to consider how that patient got there, and why others did not. Is the priority judgement about who gets treated, and at whose expense, a social or economic one to be left to non-medical people to sort out, hence not a matter for medical ethics (but for social ethics); or is it indeed a matter of professional concern to doctors and hence within the realm of medical ethics? If it is a proper matter for medical ethics, then medical ethics must encompass many non-clinical issues including that of the cost of treatment and who should bear it. If it is not a proper matter for medical ethics, then we must also exclude many of the ethical issues we discussed earlier, and the realm of medical ethics is extremely narrow, and most doctors when establishing priorities for treatment go well beyond it, and in taking note of 'extraneous factors' will frequently undermine its narrow dictates. In this case, then, doctors must in practice necessarily be acting contrary to the dictates of medical ethics.

*What  
sacrifices  
may the  
doctor  
ethically  
impose  
on his  
patients?*

These issues present themselves in a graphic manner when we look at the way in which doctors determine the allocation of their own time and energy between the practice of proven therapy, learning, teaching, and research. So far I have implicitly regarded doctors' activity as consisting entirely of the practice of proven therapy, and hence purely concerned with the effective treatment of patients here and now. But if a doctor feels a responsibility for future patients too (though possibly only his own) then he will need to keep up, that is, learn. Given that his time and energy is limited, he will have to impose some sacrifices upon current patients in order that future patients may benefit from his improved knowledge. But a doctor may also feel a duty towards other people's patients: that is, to suffering humanity in general, and regard it as proper that current patients should also make some sacrifices for the sake of teaching the next generation of doctors, or for research into improved methods of treatment, neither of which may ever benefit these current patients personally. In general the patients who are consequently not seen by the doctor are not consulted about this, nor are they even aware of it, for it will not be obvious to them that they are being deprived of treatment because the doctor in question so manages his practice that he reserves some of his time and energy for these other pursuits. I do not myself see anything unethical in this, nor apparently does the medical profession, nor do those in authority who do know all about it, so I conclude that it cannot be unethical for a doctor to decide to do less than he might for the patients before him in order that other people might benefit from those sacrifices.

But why should this be limited to one scarce resource, the doctor's own time and energy? It must be equally true of all the other scarce resources, like nurses' time, drugs, the use of buildings and equipment, all of which, if used for one patient, deprive others of their use. Whether or not the patient in question pays directly for these service costs (and they seldom do) the ethical problem is there, and it does not seem to me to be distinguishable from the use of doctors' time, which we just concluded was a relevant cost which could ethically be compared with the benefits when making decisions as to whether it should be used for one class of patients rather than another.

We have already come a long way from the assertion that it is the doctor's duty to do all he can for the patient in front of him, regardless of cost. But we still have the doctor acting (as he would see it) in a purely professional role. In none of the foregoing situations would a doctor see himself as engaged in political activity, or in social engineering, or as behaving as an economic agent, though I detect each of those elements in one or other of the examples I have given earlier. For instance, the decision whether or not to accept responsibility for someone's health is a political decision, in the sense that it is both influenced by, and influences, the structure of power and authority in that society, as does the judgement as to whose values and welfare are to count in any decision about resource use. Similarly, the relative weight given to medical, social and economic considerations is a proper matter for social policy and, in taking responsibility for determining that balance or seeking to influence it, doctors are engaged in social engineering. Finally, in seeking to allocate between competing ends their own scarce time and energy (and the other resources of their practice) doctors are acting as archetypal economic agents, whether their objective is the maximisation of their own welfare or that of some wider community.

But a doctor may decide to play a more conscious political role, which he may still regard as little more than an extension of his responsibilities as a clinician and practice manager. For instance he may feel that his practice, or his specialty, or the people in his area, are getting too small a share of the available resources, and he may decide to play the role of advocate of their interests. Elsewhere I have commented upon this in the following terms:

If the sufferers from some improvable condition are relatively inarticulate and illorganised, the clinicians who treat them may be their best advocates in the political process. There are, however, immense dangers in this which must be recognised honestly. Firstly, political advocacy is a scant respecter of truth. It thrives on presenting a persuasive case, the immediate effectiveness of which can often be improved by biased selection of facts and unscrupulous exploitation of people's fears. It is a role which combines uneasily with the detached, scientific and even-handed ideal of the impeccably trustworthy clinician. Secondly, political advocacy of the interests of one's patients easily becomes inextricably intertwined with advocacy of one's own interests. . . . Clinicians' own interests, be they in promotion prospects, good relationships with colleagues, or hopes of political power, may well intrude, and will, from time to time, prove decisive. Finally, the role of political advocate must be played in a context in which it is accepted that one's position as an 'advocate' implies that someone else occupies the role of judge and/or jury. This has significant

implications for the role of clinicians in management teams, because if they attempt . . . to play both roles simultaneously I think the system will become corrupted, because I do not believe that political advocacy of a cause is consistent with the detachment required to take wider considerations into account when participating in a collective decision about resource allocation for the community at large . . . . And in that broader model of cost-benefit analysis, they will represent only one of many relevant disciplines and sources of information, and not necessarily a preeminent one, or one with 'reserved powers'. I believe that part of the present difficulty in getting the planning and evaluation of health care into a logical, well-informed, understandable and effectively participative framework is confusion over the nature of the calculus that is appropriate for each of the three classes of decisions I have distinguished. Some clinicians seem to believe that because they are expert at the first level, and that the second and third levels are merely there to support them, their claims should be beyond challenge at those higher levels . . . . " (1)

## The doctor as policy- maker



Unfortunately the assertion that the contributions of these other disciplines are unethical (that is, that they clash with clinical freedom) is one way of resisting their challenge, and this is another reason for exposing the very imprecise and muddled nature of this ethical claim. Only if this ambiguity is acknowledged can serious and open political discussion take place, in which the ethical dilemmas (which we all face equally) can be laid bare and discussed reasonably. It is not a realm of discourse in which any of us can claim a monopoly, nor one in which any of us has a hot line to God.

## Doctors and health economists

To conclude I would like to turn my attention more directly to health economics itself, and what its ethical pretensions are. The fundamental purpose of health economics is to test whether the value of what we get from any activity is greater than the value of what we sacrifice in order to carry it out. If the benefits do outweigh the costs that activity is efficient. If they do not, it isn't. Notice that costs are here equated with sacrifices, and sacrifices are simply benefits forgone. So our guiding principle is the entirely 'ethical' one of ensuring that the benefits gained outweigh the benefits forgone.

I do not claim that we have any magic formula for doing this, any more than it follows from a doctor's desire to relieve pain and extend life that he has found the elixir either. The difficulties (both intellectual and practical) are considerable and not likely to be resolved quickly. What I do claim is that the economists' injunction to count costs as well as benefits, and to take both into account when making decisions (even clinical decisions), cannot be dismissed as unethical, unless you believe it to be ethical for someone to say that if one person stands to benefit (by no matter how small an amount), then there is no limit to the sacrifices that others may properly be called upon to bear as a consequence. This is what regarding cost as 'not germane to ethical medical practice' means, and in my vocabulary the term to be applied to someone who acts without regard to costs is not 'ethical', but 'fanatical'. Fanaticism is just as antipathetic to good clinical practice as it is good to economic practice, which to why I believe that these ethical issues need to be explored and discussed by all the various parties involved in managing the health services and trying to make the health services more efficient, so that they work together with mutual understanding and respect. Unlike the correspondent to the New England Journal of Medicine I do not see this as embarking upon 'the slippery slope of compromised ethics and waffled priorities', but rather as embarking on a long uphill struggle to make some inroads into the enormous problems facing suffering humanity in the field of health and health care.

Published by the  
Nuffield Provincial  
Hospitals Trust

Editorial Board  
Sir Edgar Williams,  
CB CBE DSO DL  
Sir Andrew Watt Kay,  
MD CHM DSc FRSE  
Gordon McLachlan,  
CBE

BCA Graphics  
Burgess & Son Ltd  
Oxfordshire

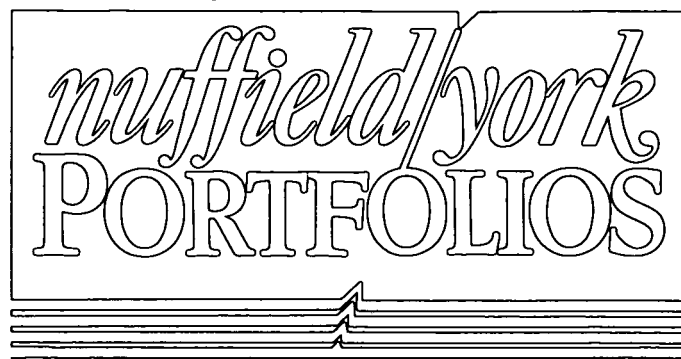
### REFERENCE

1. ALAN WILLIAMS (1979) 'The costs and benefits of surgery' in J. Leumley and J. Craven (eds) Surgical Review I, pp. 8-10. London: Pitman Medical.

Is life beyond price? Is life of infinite worth? Each question has previously been answered in the affirmative and each of these mutually contradictory answers is wrong, as Gavin Mooney shows in this folio. So far as the efficiency and fairness of resource allocation in the health services is concerned the relevant question is not so much to do with the inherent value of additional years of life. This is just as well since economists (like doctors and health service administrators) are neither better nor worse equipped to answer such ultimate questions than anyone else. Instead the relative question is the relevant one: is it better to use resources to gain life-years that would otherwise be lost or to achieve some other worthwhile objective? In this context Gavin Mooney shows that the question can be sensibly answered and, indeed, is frequently answered (though not always sensibly). He indicates the lines along which sensible answers may be sought and the kind of further questions that decision-makers should take into account in tackling decisions that involve the prospect of saving or losing lives.

A. J. C.

Folio 3.



Editor, A. J. Culyer

# VALUING HUMAN LIFE

in health service policy

Gavin Mooney

Professor of Health Economics,  
University of Aberdeen

One often comes across statements about the 'pricelessness' of human life and suggestions that there is something wrong, and indeed perhaps unethical, about attempts by economists to place a monetary value on it. For example, Muir Gray (1979) states that the value of life 'cannot be expressed in monetary terms'. Again an article in the British Medical Journal by Logan, Klein, and Ashley (1971) suggested 'we all know that cost-benefit analysis stops at age 65'—the implication being that the value placed in such analyses on human life beyond that age is zero.

Given the training of health care personnel and the ethics of medicine it is not surprising that the idea of placing human life on the measuring rod of money is anathema to many. Given the training of health economists it is no more surprising that they argue that it is not a question of liking the idea or not: it is, as a matter of fact, actually done already. The point is to do it well rather than badly and to be aware as precisely as we can be of what it is that one is assuming. Indeed one of the most important roles for economists in health care is making the values that are used explicit. This is not simply for the sake of being consistent about values but because, by being explicit, as we will indicate below, one is more likely to end up with a more efficient health care system. Greater efficiency means more health with available resources. Thus being explicit about the values placed on human life may result in more lives being saved than would otherwise be the case. Indeed an alternative title for this folio might have been 'Health Economics Saves Lives'.

# How lives are valued now

The essence of the argument of economists is very simple in principle. Since resources are scarce, the value of everything, even human life, is less than infinite. We all take risks in our day to day life, albeit to obtain some benefit or other—for example, driving to Pittodrie (home of Aberdeen Football Club) or the green grocer—from which we can infer that we do not attempt wholly to eliminate the risk of death, which would be our goal if life itself were infinitely valuable. In the NHS one has to make choices about the way in which available resources are used. Resources in life saving activities are never used up to the point where the mortality risk is zero, nor to the point where the return in terms of mortality risk reduction is zero for some additional amount spent. The NHS does not spend as much as it might on saving life before considering other benefits of health care such as reduced pain. It 'trades-off' mortality reductions against less pain. Given limited resources for health care one has to choose what benefits to provide—and what benefits to forego.

It is evident that the NHS does not allocate resources to life-saving in chronic renal failure without thought for other non-life saving programmes. If it did the cost of extending life would be very high. However, Buxton and West (1975) indicate that the provision of hospital dialysis for a cohort of 1000 patients over a 20 year period involves, on average, a cost of about £3500 per year of life extended. Now while ideally it is the cost per extra life saved which is pertinent here rather than the average, in this case the marginal and average costs may not be very different. While this form of care has been extended in recent years it remains true that not all who could benefit from dialysis do so. For example, it was recently suggested (Deitch, 1983) that 'the renal dialysis programme needs at least another £30m each year, to be as successful as it possibly could be'. Some finite value is evidently being placed on the lives of those suffering from chronic renal failure.

It is possible to show that what is paid to save life or the value of life in different circumstances and under different methods varies markedly. The following table is based on Mooney (1977).

The idea that we have to place finite values on human life is simple and inescapable in principle, provided the case for efficiency (as spelt out below) is accepted. In practice we run into problems and it is to these that most of this folio is devoted.

---

## The case for efficiency

The case for efficiency rests on the following 'rules'.

1. We should only do those things whose advantages (benefits) outweigh the disadvantages (costs).
2. We should not do or should stop doing those things whose disadvantages (costs) outweigh the advantages (benefits).
3. In deciding whether to expand or contract a service we should apply the principles of 1 and 2 above to the expansion or contraction.

Certainly if we are concerned with equity in health care—the issue of fair distribution of availability, access, risk exposure, and so on—we may want to moderate any over-zealous pursuit of efficiency. But otherwise (and perhaps even with equity in the picture) it is difficult to sustain objections to these

TABLE 1

<u>Source</u>	<u>Cost of Life</u>	<u>Comments</u>
Government decision not to introduce child-proof containers for drugs	£1000	In 1971 Government refused to introduce child-proof containers for drugs on grounds of expense; net cost per child's life estimated at £1000 (Gould, 1971, see text).
Motorway driving	£94000	Given optimum motorway speed, the price of petrol, and the value of time then the implied value of life was estimated to be £94,000. (Ghosh, Lees and Seal, 1975).
Proposals for improved safety of trawlers	£1 million	Estimate that cost per death averted would be about £1m. (Sinclair et al, 1972)
Change in building regulations following collapse of Ronan Point high-rise flats	£20 million	From the report of enquiry following the collapse of Ronan Point, a high-rise block of flats in London, Sinclair Marstrand and Newick (1972) estimated the cost of raising safety standards and the resultant fall in the risk of such occurrences. From these figures they showed that the cost per life saved would be in excess of £20m.
Questionnaire on valuation of mortality risk reduction	£3 million	Jones Lee (1976) used a questionnaire to elicit from individuals how much they would be prepared to pay to reduce their risk of death. (More detail in text).
Department of Transport	£151000	Based on the 'human capital approach' (see text).

simple rules (which also happen to be the basic rules of cost-benefit analysis) and therefore to the promotion of efficiency.

Many health service personnel harbour considerable misunderstandings about both cost-benefit analysis and efficiency. Simply, and it is hoped straightforwardly, economists seek to pursue two separate but related concepts of efficiency: (i) 'cost-effectiveness (CE)' and (ii) 'cost-benefit (CB)'. With CE concern is with how to achieve a given objective. Having decided that varicose veins should be treated, how best can they be treated? How can we best spend the £Am already designated for breast cancer screening? These are questions to which cost-effectiveness analysis (CEA) can be applied. The analysis seeks to determine the least cost solution for a particular objective or how to maximise beneficial outcomes in a particular programme or service for a given fixed budget.

Cost-benefit analysis (CBA) is more concerned with the worthwhileness of policies and options. Should heart transplant surgery be conducted? Which women (if any) should be screened for breast cancer, and how often? These are allocative efficiency questions to which CBA can be applied. This form of analysis seeks to assist choices on whether and to what extent different services should be provided. All economic analysis is concerned with assisting decisions so as to maximise the benefit to society from available resources. While this is the aim of CBA, CEA also contributes to this end in the sense that if it is possible to meet a particular objective equally well in terms of effectiveness but at less cost in terms of resource use then this should be done. In this way resources can be released and devoted to the pursuit of some other desirable objective.

The principles of CBA and CEA are very simple. The practice is more difficult especially in the case of CBA where ideally all costs and all benefits arising from the implementation of a particular policy should be placed on the measuring rod of money. Often it is at this point that the doubts about the potential use of CBA appear justified. How can we place monetary values on pain, on suffering, on death? Can the 'opportunity costs' for the caring daughter looking after the ageing mother in the community be measured in money terms?

There are problems here but they are problems that in no way undermine the basic principles of CBA spelt out above. It is important to see CBA and efficiency as an attitude of mind, as a way of thinking, rather than as is too often the case, a set of blunt tools, blunted by lack of practical application and not designed for the job anyway.

The difficulties that are highlighted by these forms of clear thinking are not special in CEA and CBA. They are inherent in the problem rather than in the method of tackling it. However decisions are currently made on, for example, renal dialysis policy, the problems inherent in reaching them were no different from those inherent in the economic approach. Some of them may have been overlooked or even studiously ignored. But ostrich attitudes do not make problems go away and explicit analysis can make the problems clearer and hence one's solution more complete.

Two of the central problems with decisions that affect the possibility of death are related to (i) principles of valuation per se and (ii) the practice of valuation. In fact these are related. Problems of the former kind stem from the uncertainty surrounding the answer to the question: whose values--consumers', doctors', health care professionals', politicians'--are related to which decisions in health care? Can the preferences of consumers apply in health care to the extent that they do in many other markets? Should they? And even if we can resolve these issues can we then derive actual money values? It is to these questions, in the specific context of valuing life, that we now turn.

---

## Deriving values for life

How monetary values for health outputs are derived is in part dependent on what underlying value system is used. Because most of the research by economists on this question has been done in the context of valuation of life, it is on this that this folio concentrates. (For a detailed review of this research see Mooney, 1977). The techniques suitably amended are, on the whole, applicable to other outputs of health care as well. Three types of approaches to valuation are outlined. These are based first on individual consumers' preferences, secondly, a 'human capital' or 'lost output' approach, and thirdly, the implied values method.

# Consumers' valuations of mortality risk reduction

If it is held that the consumer or the potential consumer of health care is the person best able to judge his own interest, then it is that individual from whom the relevant values to be attached to health care outputs should be elicited. At the point of consumption the individual's preferences have at least some relevance in his demand for health care—whether or not to seek aid from a GP, whether or not to attend for outpatient treatment or to enter hospital as an inpatient. For example, in the case of a minor accident an individual frequently has the choice of self-treatment, attending his GP or going to an Accident and Emergency Department. Clearly the expression of such preferences affects the utilisation of health services.

It is less clear however that the consumers' preferences influence to any great extent the overall deployment of health care resources. Demand in health care is in many instances influenced by supply and such supply can often be determined without any direct or even indirect consideration of consumers' preferences. Should more resources be devoted to care of the elderly? What level of supply of kidney machines is socially optimal?

In the planning processes of health care, despite various attempts to democratise the health service, it is less than certain that the value judgements of the general public are being reflected in resource deployment. Nevertheless, if the consumer-orientated approach is adopted then the question to be asked in deciding how to value the output of health care is that of Schelling (1968): 'What is it worth to the people who stand to benefit from it?'

Now for the great majority of individuals in society the prospect of death is not a certainty in any period of time—except of course in the long run! Consequently the question to be posed to the individuals comprising society is not: What would you be prepared to pay to avoid death? It is rather: What would you be prepared to pay to reduce your risk of death from  $X$  to  $X-Y$ , where both  $X$  and  $Y$  are probabilities that are very much less than 1 (and where the amount individuals would be prepared to pay is likely to vary with the size of  $X$ )? In addition similar questions would be posed about the amount other people would be prepared to pay to reduce the risk of death for those individuals at risk. Risks of morbidity could be approached in like vein.

---

## Types of risk involved

With this approach Mishan (1971) identified four components of risk, two direct and two indirect. The former are (i) direct risks that are voluntarily assumed by individuals; and (ii) direct involuntary risks (an example of this is the death arising among the frail or elderly as a result of regular supersonic flights over inhabited areas). The indirect risks arise from the concern of people for the risks faced by other individuals. These have two components: concern because of (iii) financial repercussions (for example in the case of dependents of individuals at risk); and (iv) the psychic loss—loss of companionship, or the simple fact of a compassionate regard for the welfare of others.

This viewpoint clearly has some merit and various economists (Schelling, 1968; Mishan, 1971; and Jones-Lee, 1976) have explored its implications in some detail. One way of attempting to discover how much individuals would be prepared to pay to reduce the risk of death is to study their behaviour in situations in which a risk of death is present but where some action on the part of the

individual can reduce this risk. Melinek (1974) for example uses the behaviour of individuals with regard to crossing a road by the surface (shorter riskier route) or by a pedestrian subway (longer safer route) to indicate the trade-off involved between time and risk. Given a monetary value of time in such circumstances Melinek calculates a value of life of nearly £70,000. Ghosh, Lees, and Seal (1975) combine the factors of speed, cost of travel, and accident risk on motorways to indicate the trade-offs involved in this situation, estimating thereby a value of life of over £90,000 (See table).

Such situations however are relatively few and are difficult to generalise from. For example, individuals' attitudes towards risk of death may vary for different situations (e.g. road accident mortality risk as compared with cancer mortality risk), for different levels of risk, or for family circumstances. The amount they would be prepared to pay to reduce their risk of death by say 0.1 per cent in these different circumstances may also vary. Consequently it is unlikely that the value of life will be constant in all circumstances even for the same individual.

Perhaps even more important is the question of perception (which may be erroneous) of risk. Clearly the basis of individual behaviour in risk situations is the individual's perception of the risks involved. While there is now good quantified evidence on the mortality risk associated with smoking, the individual smoker will decide whether or not to continue to smoke on the basis of his enjoyment of smoking and his perception of the risk. Whether individuals' perceptions of risk should be the basis of public sector decision-making is a difficult issue. While it is possible to defend consumers' preferences in the valuation of benefits, if the individual misperceives the mortality risk reduction it would seem wrong to base health care planning on consumer ignorance given that better informed judgements may be available.

The issue can be made clearer by examining the motives of a group of mothers who press for the introduction of a crossing patrol at a busy road to make it safer for their children to get to school. Such pressure depends on the risk they perceive, the reduction in risk which they perceive the crossing patrol could provide, and the value they place on the reduction in risk and the cost of the patrol. Given that the cost of the crossing patrol will be borne through rates and taxes by a much wider group than just the mothers, it can be argued that while their values are relevant to the decision on whether or not to introduce a crossing patrol, their risk perceptions may not be and their cost perceptions are quite likely to be biased.

One way round the problems of perception and of lack of suitable behavioural areas to study is to set up hypothetical risk situations and to elicit, by means of a questionnaire, information from individuals about their likely behaviour. Jones-Lee (1976) has developed such a questionnaire which involves a trade-off between safety and wealth in a hypothetical situation in which the individual makes choices between two airlines which differ only in their safety records and their fares (See table). These two forms of consumer orientated approach are still at the experimental stage and it may be some time yet before they yield operational values. The approach is not without its drawbacks both because of the lack of appropriate areas for studying individuals' behaviour in risk situations and the problems involved in obtaining accurate responses in simulated hypothetical risk situations. Yet, it represents an important development in valuation procedures. Given the high values of life which emerge from this approach, if it were implemented, it could result in quite wide ranging changes in health care values and hence decision-making. At the very least, economists have substantially elucidated the practical and theoretical problems that follow if one decides to adopt the consumer preferences viewpoint to valuing lives.

# The lost output or 'human capital' method

Physicians, as well as others, have quite frequently used the second approach, which is related to the idea that an important component of the value of life is the amount of useful working time that is lost when someone dies. Put more formally this approach suggests that a life's value can be measured in terms of the future expected life-time earnings of the individual concerned, adjusted to allow for working life expectancy, participation rates in the labour force, and various other factors. The value of life or, more accurately in this context, of livelihood is then obtained by discounting these future earnings to their present value as is usual in public and private investment decisions. For this reason, economists term it the 'human capital' approach. The extension of this approach to the valuation of morbidity is self evident.

This methodology, which dates at least from Sir William Petty (1699), is also the one most frequently found in cost benefit studies in health care. In some instances the valuation is made with consumption being deducted from output. For the purposes of considering investment in health this is erroneous—because if an individual is saved from death, the individual remains a member of society and any benefit to the individual (for example, the benefit he gets from his consumption) is therefore a benefit to the society of which he is a member.

Various adjustments to the basic methodology have been suggested. Dawson (1971), in the context of road accidents, includes a somewhat arbitrary sum to cover pain, suffering, and grief—his so called 'subjective' costs. Weisbrod (1962) argues that such costs are likely to follow the pattern of the discounted earnings, i.e. relatively low for children, high for young adults, and relatively low for older adults.

The methodology is appealing in its inherent simplicity and its relative ease of applicability—monetary values are quite easily obtained. It is based on the assumption that it is Gross National Product (GNP) that is to be maximised and other considerations (despite the efforts of Dawson, Weisbrod, and others) play a secondary role. As such it can be argued, given the GNP based value judgement, that it provides minimum values of life. It is the method used as a basis for the Department of Transport's accident cost calculation (see table).

Problems arise of course with non-producers in society who would have at best a zero value under this valuation process and even in some instances a negative value. It thus becomes difficult to accommodate the elderly and housewives within this scheme (although the problem of the latter group can be overcome by imputing a wage rate for housewives' services).

An intelligent use of the approach can be defended if the GNP based value system can be defended at the same time. While it would be difficult to mount a watertight defence for using this approach alone, given that one of the objectives of health care is to return individuals to productive health, then the methodology can be justified—provided the values emerging are used as minimal values. Some recent evidence (Bergstrom, 1982) suggests that the values based on consumers' willingness to pay will be higher than those emerging from the human capital approach. If this turns out generally to be the case, we again have a reason for using the human capital figures as minimal values.



# The implied values method

What resource scarcity means in practice is, as has been seen, that within the existing decision making processes of the health service by implication finite values are already being attached to the various outputs involved. Decisions are already being made which imply that not more than X pounds or Y pounds should be spent on saving a life thereby implying that the value of life is less than X pounds or Y pounds. In other cases decisions imply that Z pounds should be spent on the saving of a life thereby implying that life in this instance is worth at least Z pounds. Decisions are similarly, and more frequently, being made on treatments for non fatal illness.

Seldom are these implied values made explicit. Decisions tend not to be made in such a way that those involved are aware of the values which they are implicitly placing on lives saved. But it is possible to examine decisions retrospectively and make explicit the values used. For example Leach (1972) claims that in the UK in the 1960s one implied value of life was less than £50 because a test on pregnant women to prevent still births at a cost of £50 per still birth prevented was infrequently used.

Again, as indicated in the table, it has been claimed by Gould (1971) that the DHSS in 1971 valued the life of a child at less than £1000. The example was based on the refusal of the DHSS on apparent grounds of cost to agree at that time to the introduction of child-proof containers for medicines. On this issue it is possible to indicate (i) the size of the existing problem i.e. the numbers of children poisoned each year by taking drugs; (ii) the effect in reduced mortality and morbidity and in cost savings to the NHS (e.g. reduced hospitalisation) if containers were made child-proof; and (iii) the cost of the child-proofing.

While there are of course other factors to be considered this information can be used as an example of how to derive implied values. Since the DHSS indicated that the proposal was too costly, this implies that they considered that the benefits were less than the costs. Thus  $C_{cp} > V_1 + V_2 + C_s$  i.e. the costs of child proofing ( $C_{cp}$ ) were greater than the benefits of reduced mortality ( $V_1$ ) and morbidity ( $V_2$ ) and the savings in cost to the NHS ( $C_s$ ).

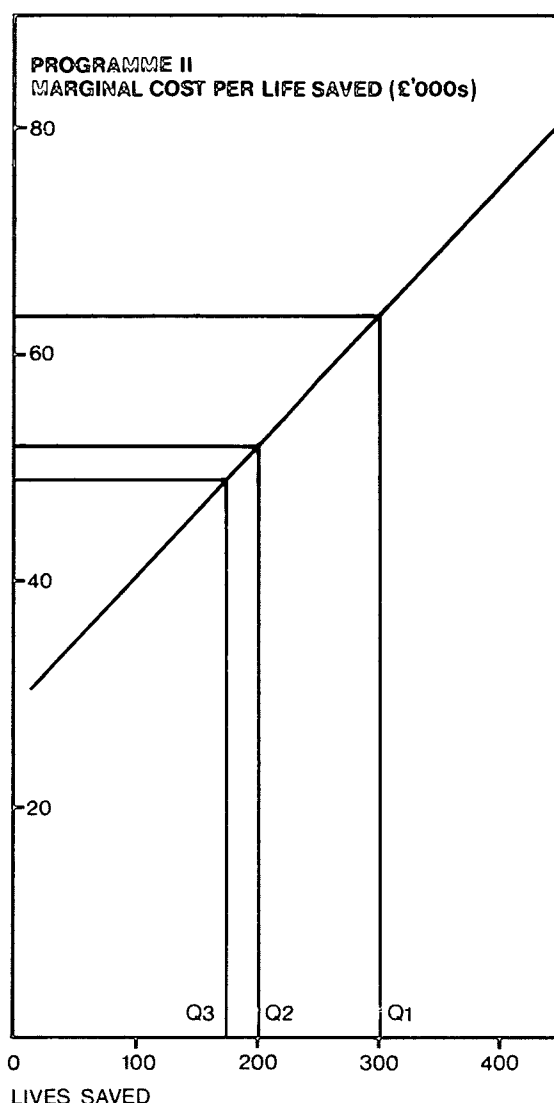
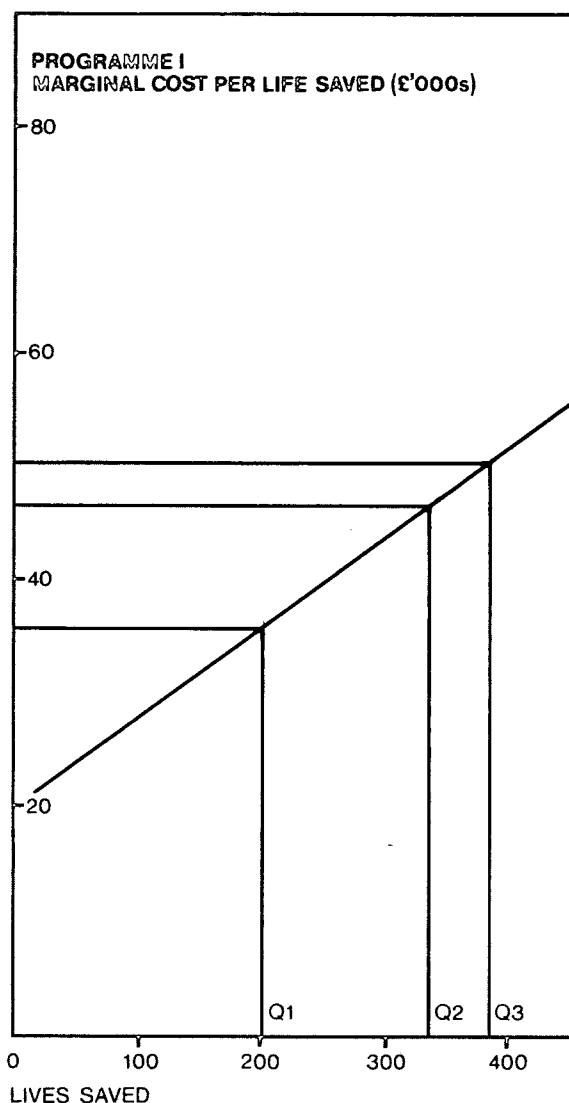
One estimate at the time (Gould, 1971) suggested that the relevant figures were of the order of £500,000 for  $C_{cp}$ , £480,000 for  $C_s$  and lives saved were about twenty. If savings in morbidity were valued at zero then the inequality above becomes  $£500,000 > 20 \text{ lives} + 0 + £480,000$  or  $£20,000 > 20 \text{ lives}$ , implying that a child's life was worth on average less than £1000.

While no claim is made as to the accuracy of these figures, the example serves to show a simple method of deriving an implied value of life.

At the other end of the spectrum, and quoting from a United States example, the American Cancer Society once endorsed a protocol of six sequential tests of stool for occult blood for screening for colonic cancer. In a study by Neuhauser and Lewicki (1975) it is shown that given the incremental costs and benefits of the sixth test (i.e. the additional cost and cases detected by a sixth test as compared with only five), the marginal (that is, extra) cost per case of colonic cancer detected by that test is \$47m.

But why bother to study such decisions retrospectively to determine these implied values? There are at least three reasons. First, there is the issue of equity. For like lives saved the ethos of the NHS would suggest that like resources should be expended at the margin. Secondly, and more important in planning terms, there is the issue of consistency and efficiency. Consistency is not necessarily a virtue in itself. It becomes a virtue in terms of efficiency if it leads to a higher level of output for the same resources. If an analysis of implied values reveals that, for the same type of output, the implied value per additional unit of output is higher in one case than another, then this in turn implies over-investment in the former area relative to the latter. By redeploying funds from the former area to the latter, the trade-off in output

**Marginal Cost Curves in Two Life Saving Programmes**  
(Hypothetical Example)



will be such that the output lost in the former area will be more than offset by increased output in the latter and an overall net gain in the productivity of the health service will be obtained. This can be shown graphically as above.

In programme I, assuming initially that 200 lives are being saved (i.e. the level of supply is  $Q_1$ ), the cost of saving the 'marginal' life is £35,000; in II, where again the level of supply is  $Q_1$ , it is £65,000 with 300 lives being saved, and total lives saved equal 500. If II were cut back to  $Q_2$ , a saving of 200 lives would be achieved and hence 100 lives fewer saved. The savings would be  $100 \times \frac{1}{2} (£65,000 + £52,200)$  which equals £5.9m. If this £5.9m were spent on programme I, then an additional 145 lives could be saved—and without any additional expenditure overall. The most efficient solution of all—assuming that all lives are equally valuable—is where the marginal cost of saving a life in programme I equals that in programme II. This will occur at  $Q_3$  when both programmes have a marginal cost per life saved of £50,000 with programme I saving 380 lives and programme II 167 lives, and total lives saved equal 547. Any further switch of resources between these programmes would then result in a fall in the number of lives saved in total.

The third reason relates directly to the question of valuation approaches. If a number of analyses can be made of past decisions regarding life saving

projects in health care, it is anticipated that a range of values would emerge, distributed about some mean. Once this mean was obtained it would then be possible to use this value prospectively in assessing future proposals involving life saving on the assumption that such a value had been already expressed by public decision makers. Of course there are good reasons why the value of life should not be the same across the whole health service—different policies will result in different saving in life expectancy, the quality of life may also vary, etc. Nonetheless, if such a mean value were used and departures from it had to be justified then it is very likely more lives would be saved and health service resources used more efficiently.

This approach has the very real merit of being easy to apply once the research has been done into past decisions and the mean implied values established. However, it is not immediately clear how the mean values established could be updated through time.

The approach inevitably reflects the existing decision making processes in the health service which are a mixture of government, parliamentary, health authority, medical profession, health care professions generally and patient preferences. Indeed it underwrites the existing processes. Consequently it is only if these are seen to be the correct way of defining the underlying value system of the NHS that the approach is justified. In the short run at least—which may turn out to be very lengthy in practice—it is unlikely that the basis of these decision processes will be radically altered.

---

## Conclusion

Perhaps the main conclusion of this folio must be that the issues surrounding valuation in health care are fraught with difficulty. In part this is due in no small measure to the complex and often seemingly conflicting value system or systems on which the health service is based and operates. In part it is due to the technical difficulties involved in placing actual values on the outputs of the health service. The question of which value system to use involves a very important moral judgement in itself.

Thus the questions raised earlier on underlying values becomes paramount. To what extent should the individual consumer have his value emphasised in decisions about the delivery of health care? As regards the value of life, should the underlying values be those of patients (actual or prospective?), the general public, politicians, administrators, clinicians? For example, is the objective of treatment of breast cancer to maximise life years saved or improve the quality of life? And when these conflict, who should decide on their relative importance: the doctors or the women?

It is not for the economics profession (any more than for the medical profession) to determine whose values should or should not be adopted. Economic analysis can indicate the possible implications of different value systems and suggest, perhaps, the decision contexts in which one may be more suitable than

another. Once a value system is chosen, it can indicate how, within that framework, the philosophy and tools of economic analysis can assist in ensuring more efficient and effective planning of health care resources.

One thing, however, is clear: there can be no eschewing the problems of valuation. Insofar as there may be an appearance of avoiding them within the existing health service decision making processes, it is only because values remain implicit, perhaps inconsistent and even, on occasion, irrational. These valuation problems cannot be escaped. Nor are they. Whatever value system is deemed politically, morally, and socially valid for the health service--and that begs questions which economists cannot resolve--the tools of economic analysis can be brought to bear to determine what monetary values should be placed on health outputs.

The debate on the underlying value system has in reality existed since the inception of the health service. It has not been commonly expressed in such terms. Yet one of the most fundamental questions underlying the stresses, strains, and conflicts over the health service is not related to whether we can place a money value on human life. There is no debate here: we do. The question is rather: whose values are appropriate?

The trouble with the more comfortable, ostrich like, alternative is that this, and the other issues pointed out here, are buried. The economic approach cannot yield 'final solutions'. What it can do is, first, expose those elements of any solution that may be controversial so that full discussion among relevant persons can proceed and second, provide some operational techniques for yielding quantitative results once the basic point of departure as regards the source(s) of values has been established. Consequently, economics tends to be opposed not by those who seek efficiency and equity in the NHS but by those who prefer their decisions to be unchallenged--even unchallengeable: by those who have reason to fear the consequences of having their practice assessed or who simply prefer the quiet life. So economics is not merely the handmaiden of efficiency and equity--she even serves the noble cause of the open society!

## REFERENCES

- Bergstrom, T.C. (1982). 'Is a man's life worth more than his human capital?', in The Value of Life and Safety, (ed) M. W. Jones-Lee. (Amsterdam: North Holland).
- Dawson, R. F. F. (1971). Current Costs of Road Accidents in Great Britain. (Crowthorne: Road Research Laboratory).
- Deitch, R. (1983). 'Commentary from Westminster', The Lancet, 2, 1316.
- Ghosh, D., Lees, D., and Seal, W. (1975). 'Optimal motorway speed and some valuations of time and life', The Manchester School of Economics and Social Studies, 43, 134-43.
- Gould, D. (1971). 'A groundling's notebook', New Scientist and Science Journal, 51, 217.
- Jones-Lee, M. W. (1976). The Value of Life: An Economic Analysis. (London: Martin Robertson).
- Leach, G. (1972). The Biocrats. (Harmondsworth: Penguin).
- Logan, R. F. L., Klein, R. E., and Ashley, J. S. A. (1971). 'Effective management of health', British Medical Journal, 2, 519-21.
- Melinek, S. J. (1974). 'A method of evaluating human life for economic purposes', Accident Analysis and Prevention, 6, 103-14.
- Mishan, E. J. (1971). Cost-Benefit Analysis. (London: George Allen and Unwin).
- Mooney, G. H. (1977). The Valuation of Human Life. (Basingstoke: Macmillan).
- Muir Gray, J. A. (1979). 'Choosing priorities', Journal of Medical Ethics, 5, 73-5.
- Neuhauser, D. and Lewicki, A. M. (1975). 'What do we gain from the sixth stool guaiac?' The New England Journal of Medicine, 293, 226-8.
- Petty, Sir W. (1699). Political Arithmetic or a Discourse Concerning the Extent and Value of Lands, People, Buildings, etc. (London: Robert Clavel).
- Schelling, T. C. (1968). 'The life you save may be your own'; in Problems in Public Expenditure Analysis, (ed). S. B. Chase, (Washington: Brookings Institute).
- Sinclair, T. C., Marstrand, P., and Newick, P. (1972). Human Life and Safety in Relation to Technical Change. (Brighton: Science Policy Research Unit, University of Sussex).
- Weisbrod, B. A. (1962). Economics of Public Health. (London: Oxford University Press).

Published by the  
Nuffield Provincial Hospitals Trust

Editorial Board  
Sir Edgar Williams, CB, CBE, DSO, DL  
Sir Andrew Watt Kay, MD, CHM, DSc, FRCS,  
Gordon McLachlan, CBE

BCA Graphics  
Printed in Great Britain by  
Burgess & Son Ltd, Abingdon

# nuffield/york PORTFOLIOS

Editor, A. J. Culyer

For more than ten years now successive governments have made the development of community care for the elderly, for mentally ill and for mentally handicapped persons a major priority of their health and social care policies. (DHSS 1971, 1975, 1977.) Yet the pace of achievement of this objective has been slow. The present government is committed to accelerating the replacement of hospital care and is exhorting local health and local government authorities to take up this enormous challenge. However, the development of community care is fraught with problems of efficient and equitable provision and both centrally and locally based initiatives must tackle these problems if the declared objectives are to be achieved.

## EDITORIAL

What is community care? For whom is it most suitable? How can it be most comprehensively costed? Is the health of patients having various kinds of handicap and disability furthered more effectively in the community or in institutions? Such questions lie at the heart of decisions at every level that have to do with determining the ideal balance between community and institutional care. That they are frequently answered unsatisfactorily, or not at all, is mainly due to the complexity and variety of the needs of the patients in question, to the great variety of services that may in principle be supplied, and to the immense difficulty of measuring the outcomes in terms of patient benefit that may accrue under the various modes. In this folio Ken Wright systematically explores these issues and shows that many of the techniques needed to resolve these thorny issues already exist. What is really needed is their application in a system that enables the efficient and humane balance both for particular groups and for individual clients to be identified and that also minimises administrative and financial barriers to the implementation of the right policies.

A. J. C.

# Community Care

## ITS MEANING AND IMPORTANCE

**K. G. WRIGHT** Senior Fellow  
and Director of the DHSS Health Economics  
Programme, University of York

There is little agreement on a definition of community care. One useful definition is Alan Walker's (1982) 'help and support to individuals, including children, people with disabilities, and elderly people in non-institutional settings'. Even with this definition it may be difficult to classify certain facilities, for example, whether residential homes for the elderly are institutional or community care. Some homes may not follow

any sort of institutional pattern as they integrate residents into the local community and make many of their facilities available to the local elderly population. Others may be run on very institutional lines and the pattern of care offered may not be easily distinguished from hospital-based care. Bearing these difficulties in mind, 'community care' is here taken to mean care outside hospital although it may include people who receive outpatient, day care or short spells of inpatient care but who for most of their time are cared for outside the hospital.

The major reason for the development of community care is the general belief that

care in the community is the best way of helping many people who are unable to care most of the time for themselves and that it gives them opportunities to lead as normal a life as possible given their disabilities. However, once one gets beyond this general belief there are many uncertainties about the detail of community care and health and social services face two immense problems:

how to allocate people between a large range of alternative forms of care within the community;

how to re-allocate the resources freed from the run down of hospitals to the community care services.

The main purpose of this folio is to examine two principal aspects of community care which may be boldly classified as efficiency and equity. Efficiency means achieving the greatest benefit or return from a given set of resources. Equity aspects are concerned with the ways in which the costs and benefits of alternative policies are distributed amongst the population.

## EFFICIENCY AND COMMUNITY CARE

### COSTS AND BENEFITS OF COMMUNITY CARE: A VICIOUS CIRCLE OF IGNORANCE

Efficiency is defined as the achievement of the greatest benefit from a given set of resources or the achievement of a

given benefit from the minimum set of resources. This section is therefore concerned with the measurement of costs and benefits of community care and attempts to define standards of care and to set cost constraints. By and large, the community care services seek to ensure that people who are unable to care for themselves in particular ways receive help to achieve a satisfactory standard of living. That standard will depend very much on the resources that society is willing to devote to the affected groups. Setting standards of care and establishing resource constraints are beset by problems because it is difficult to define precisely what standards of care are to be expected from the services provided. This results in a vicious circle of not knowing what resource constraint to set because no one knows what type or quality of care will be achieved and not setting standards of care because it is not known what these will cost.

The evaluation of community and hospital care helps to set a boundary to the community care sector by determining the limits to its efficiency. A starting point is the commonly held notion that the per client costs of community care increase with disability because, as a person suffers increasing disability, more services have to be delivered to maintain that person's standard of living. Thus, for less disabled groups of people, the costs of community care will be below hospital costs while for severely disabled groups the costs of community care will be higher than hospital costs. The idea is demonstrated in a grossly simplified form in figure 1.

The curve CHC represents the annual

FIGURE 1

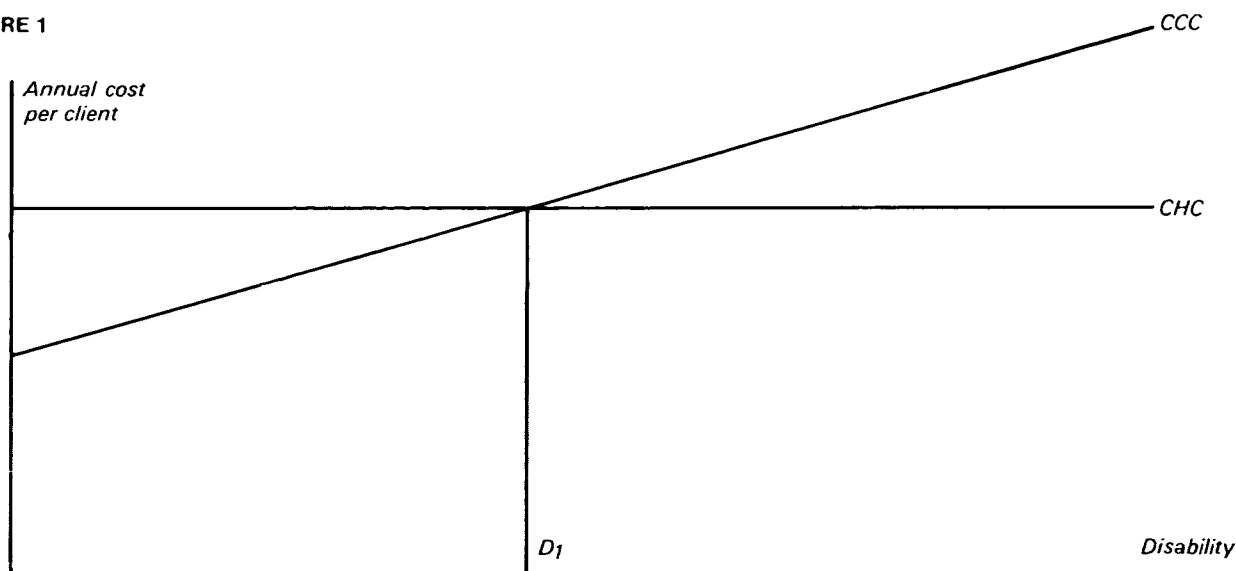
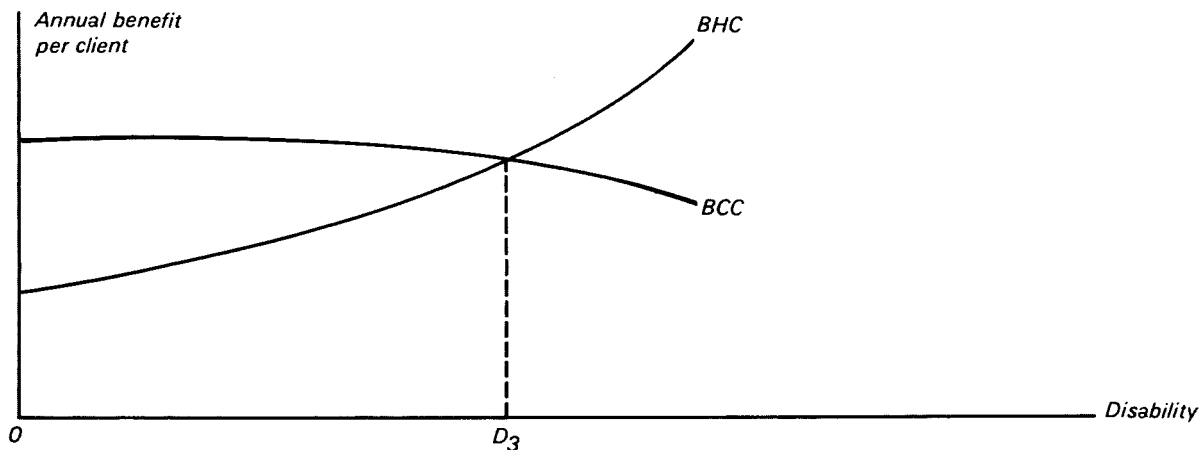


FIGURE 2

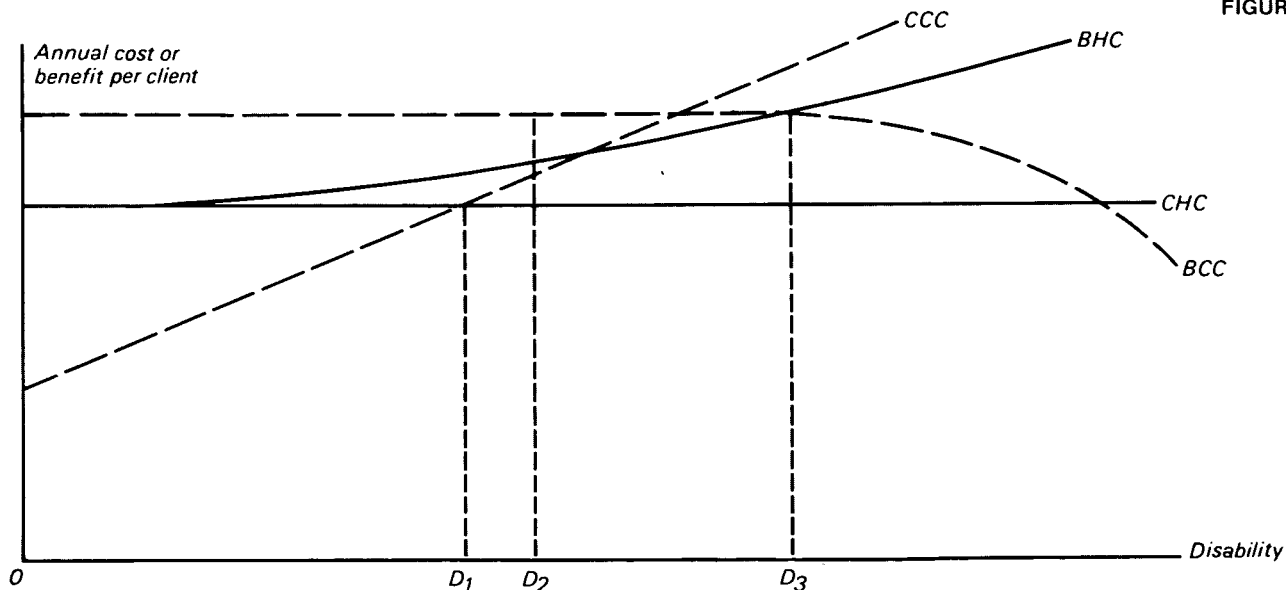


average cost per patient year in hospital care. The curve CCC represents the annual average cost per client, both dependent on the extent of disability in the community population. In practice the CCC curve is compiled by surveying the relevant population in the community, classifying the respondents into the appropriate, pre-determined categories, summing up all the costs of care received in a year by each person and completing the average annual cost per client by dividing the total cost of care per category by the number of people in that category. This allows one to identify points such as  $D_1$  in figure 1 where for disability categories between 0 and  $D_1$  community care is the less costly alternative and for categories beyond  $D_1$  hospital care is less costly.

But costs are only one side of the

argument. If we are to talk about the efficiency of the two alternatives it is also necessary to know the relative benefits they produce. Measuring benefits of care either in hospital or in the community is very difficult and we know very little about them at present. However, it is possible to speculate about these benefits so as to help push the efficiency argument on a little further. For illustrative purposes, assume that the relative benefit curves take the shape and position as illustrated in figure 2. In this example the per client benefit curves follow a different path from the per client cost curves. BHC is the hospital curve and this increases with disability, suggesting that hospital care is more beneficial to the highly disabled groups. BCC is the community care benefit curve

FIGURE 3





and is above the hospital benefit curve for all levels of disability less than  $D_3$ .

If we now put cost and benefit curves together (assuming for the time being that we can make costs and benefits commensurable) we arrive at the position set out in figure 3.

This diagram illustrates the importance of considering costs and benefits together.  $D_1$  is the point from figure 1 where the cost of the two alternatives is equal ( $BHC=CCC$ ) but in this example the benefits of community care ( $BCC$ ) are greater than the benefits of hospital care ( $BHC$ ). Thus, although the costs are equal, community care is the more efficient alternative. Conversely at  $D_3$ , from figure 2, the benefits of the two alternatives are equal ( $BHC=BCC$ ) but the costs of community care ( $CCC$ ) are greater than the costs of hospital care ( $CHC$ ) and therefore hospital care is the more efficient alternative. In effect, to define the limit of community care, one is searching for a point such as  $D_2$  where the difference between the costs and benefits is equal for both alternatives. For people with a degree of disability up to  $OD_2$  community care is the more efficient alternative beyond  $OD_2$  hospital care is more efficient.

The information requirements to complete this analysis are formidable. The next task of this section is to examine the information requirements for costs and benefits and relate these to different groups of people requiring care in the community to determine ways of moving from a theoretical model to practical policy-making.

## ***COST INFORMATION***

### **CHOICE OF COSTING BASIS**

The first decision to be made about costs is the basis on which they are to be measured. There are usually two choices. One of these relates to financial costs where resources are valued by the costs being imposed on the public purse. In this case costs are measured by the cash paid for the use of a resource. The second is opportunity costs which are based on the costing of all resources not just those which are paid for out of public money. Opportunity costs have been defined (Maynard and Smith, 1983 p.3) as

the value of resources in their most highly valued alternative use. Accordingly, if the aim is to offer efficient services it is important to ensure that the uses to which resources actually are put are more highly valued than their (best) alternative use. The valuation of these resources thus depends on the answers to the following three questions:

what resources are used in the provision of this service?

to what other use, if any, would these resources be put if they were not being used in the provision of this service?

what payment has to be made to attract or maintain these resource in the provision of this service?

In many cases the payments made for goods or labour used in the provision of community care services accurately reflect the payment needed to keep these resources in their existing use. However, there are instances when this is not true. For example, it is possible that some labour being used now would be unemployed if it were released from providing community care. Thus the economic cost of that resource is nil, even though a wage is presently being paid for its use. In other cases, commodities purchased include indirect taxation which is a payment to the Exchequer and not to the resource. Costing resources for efficiency purposes requires great care and skill in interpreting whether the price paid for a resource reflects the opportunities sacrificed by keeping it in its existing use and, in cases where some discrepancy occurs, in calculating the required adjustment to real resource costs.

### **THE COST OF COMMUNITY CARE**

There are, in summary, four main categories of cost for community care. These are summarised in table 1.

Setting the sources of finance against the main sets of resources provides a useful way of illustrating the costing dilemma. Take, for example, someone who lives in, or will move into, a dwelling place of their own. The financial basis of costing this accommodation will take into account the level of housing subsidy and housing benefit paid because these are the charges on the public purse.

However, the opportunity cost is the (annuitised) market value of the house irrespective of ownership. Similarly, the

**Table 1. Costs of community care**

<u>Cost category</u>	<u>Source of finance</u>
Accommodation: LA residential home, hostel, voluntary or private home or group home, private dwelling	LA rates and grants Voluntary funds Housing subsidies Housing benefit
Personal consumption: food, clothing, energy	Social Security Pensions Private income
Statutory services	NHS budget LA rates and grants Family Practitioner budget
Informal help	Voluntary donations Private income

financial basis would not take heed of the sacrifices made by people who have to provide care at home for a relative, whereas the opportunity cost basis would take into account the loss of earnings or leisure time sacrificed by such a person. Thus, the two different bases can give quite different estimates of the costs of community care. Since this paper is concerned with the efficiency and equity of community care it will generally use the opportunity cost basis.

#### THE COSTS OF HOSPITAL CARE

The costs of hospital care are more straightforward than those of community care since all the costs are incurred by the public sector. These costs are summarised in table 2 below.

The major problem with estimating hospital costs is the valuation of capital. Consider the depopulation of psychiatric hospitals. As the number of patients discharged to community care increases, buildings and land become available for other public use or for sale. In some cases, re-use or sale will be easier than others: a district with two medium sized psychiatric hospitals

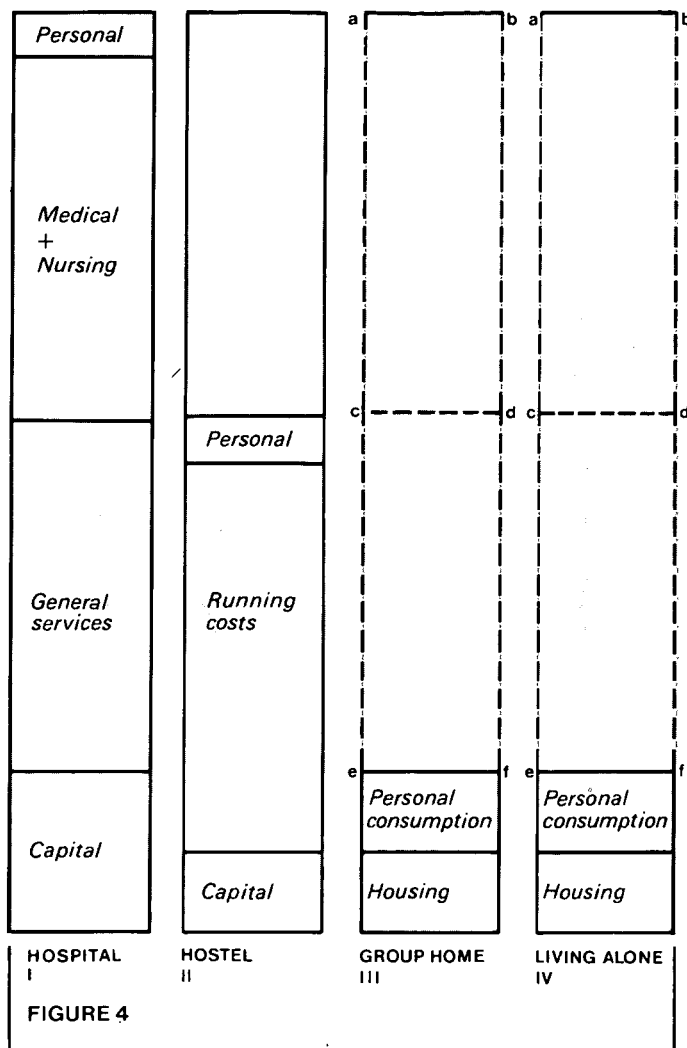
**Table 2. Hospital costs**

Capital	Building Land Fixtures and fittings
Direct patient care costs	Nursing services Medical services Para-medical services
General service costs	Catering laundry Maintenance Administration
Personal consumption	Patient's 'pocket-money' Allowance

may be able to concentrate all hospital services on one site and re-use or sell off the other buildings and site. If a district has one huge hospital it may be very difficult to change the use of the accommodation freed. In geriatric care the problem is reversed because it is likely that demographic pressure will increase the demand for long-term hospital care and in this case the capital costs will be based on providing new accommodation or converting suitable existing accommodation to specialised geriatric care. Capital costs are likely to vary considerably from one district to another according to local circumstances (for example costs of labour materials and land vary from locality to locality). There is no single answer to the question 'what are the costs?' It all depends on the local circumstances and the type of decision that is under consideration.

#### COMPARING COST OF HOSPITAL AND COMMUNITY CARE

Figure 1 sets out one set of community care costs against hospital costs. The items in table 1 suggest that community care costs depend on the type of accommodation people are occupying. Standardising for equal personal consumption (because different levels add an unnecessary complication) community care costs will vary in their composition according to the type of accommodation being occupied. People who live alone may be consuming quite valuable housing accommodation and be in receipt of several domiciliary, day-care or other community-based services but receive very little help of an informal nature. People living with others may consume a small amount of housing accommodation and a small amount of community services, but use a considerable amount of informal help. People living in hostels or residential homes may receive all their help from the staff of these establishments. Thus figure 1 could be complicated by drawing different cost curves to describe different situations. An example of the variety of alternatives is set out in figure 4. This figure indicates the relationship between hospital costs and other types of facility. Block I shows the division of hospital costs as set out in table 2. Block II represents hostel costs for a hostel which provides twenty-four hour care for all residents. Block III represents costs of living in a group home and Block IV the costs of keeping



someone in single-person accommodation. However, if compared with the items listed in table 1, it will be noticed that two items are missing: community care service costs and informal care costs. The reason for this is that these two sets of costs will vary according to the degree of help needed by the people being cared for in these different types of accommodation. It is highly probable that the need for help will increase as a person's disability increases. Thus for mildly disabled people the two sets of costs will not be very much above EF in both Blocks III and IV. As disability increases the costs will approach CD or, in extreme cases, AB. This illustrates the major problem in allocating people to different forms of care and can be summarised in the question of how much of the area CDEF is it worth paying to avoid having to keep someone in full-time residential care or how much of the area ABEF is it worth paying to avoid having to keep someone in full-time, long-stay hospital care? Indeed, if it was the general belief that community care was to be preferred to hospital care, one might be prepared to spend beyond the limit

represented by AB to keep someone out of hospital.

The levels of cost represented by the lines EF and CD highlight the old argument that community care is cheaper than institutional care. The dotted lines EF and AB show that this depends on the costs of the help that has to be given to maintain the person in the community.

However, we must not leave this question of costs without reference to the informal care system. It has long been generally recognised that community care is a partnership between statutory services, voluntary societies and an informal network of caring relatives and friends. In filling in the dotted areas such as CDEF in figure 4, the costs of the statutory services are fairly easy to calculate but the costs of other caring sectors are not. Some people would even argue that we should not cost the non-statutory sectors because the help is freely given. However, there is growing evidence that caring for a relative is often very expensive for some people, for example the person who has to give up work and loses not only the wages but also all the pension rights and the career development that go with continuous employment. It is not, however, only people who are forced to withdraw from employment who are faced with costs of providing care. Some people lose enormous amounts of their own free time or are faced with caring for two households at a time with consequent conflicts of loyalty over the amount of attention being paid to one at the neglect of the other. No-one has successfully developed a method for valuing these essential contributions to community care, but their importance is being increasingly recognised in policies designed to care for carers as much as for the disabled person.

## THE MEASUREMENT OF COST BENEFIT

### THE DEFINITION OF BENEFITS

The benefits of public services such as health and personal social services are usually defined as the success of these services in achieving their stated objectives. Thus, if a set of domiciliary services is designed to maintain a satisfactory home environment, the

benefits of the services will be measured in terms of, for example, cleanliness and warmth of the dwellings used. The implication of this definition is that before any measurement can be undertaken we need to be clear about the objectives which services are trying to achieve. Generally it is useful to borrow from work in the community care of the elderly to set out some principal objectives of community services which probably apply to all groups suffering from chronically disabling conditions. These have been set down by Challis (1981) as the provision of the following desirable characteristics:

1. Nurturance (that is, meeting the most basic needs for comfort and security);
2. Compensation for disabilities;
3. Maintenance or improvement of independence;
4. Maintenance or improvement of morale;
5. Maintenance or improvement of social integration;
6. Maintenance or improvement of family relationships; and
7. Community development (that is, the fostering or encouragement of a system of social relationships from which help and assistance is more readily available for needy members of the community).

#### BENEFIT MEASURES IN PRACTICE

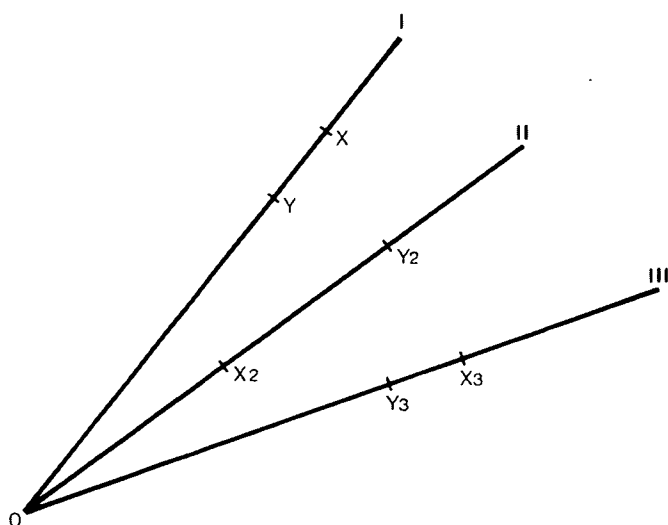
There has been considerable work on methods of measuring these benefits. There is a large number of questionnaires and measurement scales for use by anyone who cares to attempt an exercise in benefit measurement (Davies and Knapp,

1981). In addition, routine assessment by professional workers on the progress of the people they are caring for provides a further source of information because one finds in practice that those assessments frequently identify the same sort of issue as listed in 1-7 above. The main problem is to decide which of the battery of tests gives the best result for the benefit one is trying to measure.

For example, if one is concerned with the measurement of independence, (that is, how well people are able to perform the major activities of daily living) and to evaluate how well different phases of treatment, care and training might improve or maintain patients' abilities to care for themselves, there is a generous array of tried and tested scales from which to choose. However, each one is usually devised for a specific purpose and is designed to yield information on the status of health of people in different circumstances. Thus, some scales contain very few items, for example, able to bath, able to walk outdoors, able to dress, feed, wash hands and face, get in and out of bed, and so on. Others may take one component of that scale and split it into several separate steps. For example, feeding may be split into being able to eat with guidance with a spoon; to eat without guidance; to use a knife and fork if food is cut up; to eat without any help whatsoever.

The degree of disaggregation depends on the purposes for which the scale is to be used. If for example information is needed on large populations, such as census information on people in long-stay hospitals, the general categories of ability to feed, wash and dress will suffice. If, however, staff are involved

FIGURE 5



For objective I policy 1 achieves more than policy 2.

For objective II policy 2 achieves more than policy 1.

For objective III policy 1 achieves more than policy 2.

In this example we have no way of determining which is the more effective policy.

in training or rehabilitating people who do not have, or have lost, basic social skills and they wish to monitor progress over short periods of time, then a much more disaggregated scale will be useful.

## PROBLEMS OF BENEFIT MANAGEMENT

There are three inter-related, long-standing problems of benefit measurement which have severely hampered the evaluation of the efficiency of community care. The first of these is exceedingly complex and concerns the fact that policies may have differential effects on more than one objective. For example the provision of day care facilities may be good for preserving family relationships but may lower the morale of the recipient who prefers to remain at home all day. It is always difficult to assess alternative policies when they have such differential effects. The problem is illustrated in figure 5 where the lines I, II and III represent benefits along three policy objectives, the 'X' marks the achievement of policy 1 on each objective and 'Y' the achievement of policy 2 on each objective. Thus policy 1 achieves  $OX_1$  on objective I,  $OX_2$  on II and  $OX_3$  on III whereas policy 2's achievement is marked at  $OY_1$ ,  $OY_2$  and  $OY_3$ . Unless one can measure each dimension in the same terms as we have no a priori rule which says which policy gives the highest benefit. It is the same as comparing apples, pears, and bananas without a common measuring rod like weight or money value.

The second problem is that very little is known about the benefits of individual community care services or the benefits on individuals of different packages of community care services. This is particularly true where one wishes to alter services 'at the margin'. For example, what would be the effects of providing an extra two or three hours per week of home help to a housebound, disabled person living alone, or what is the effect of three days instead of two days attendance per week at a training centre for mentally handicapped persons. Very few community care services have been evaluated either singly or in combination in terms of the total or marginal benefits they produce. This means that it is exceedingly difficult to set benefits against costs as set out in figure 3 and explains why a considerable amount of the discussion about appropriate place of care for people with chronic illness has focused too narrowly

on the relative costs of care instead of the relative efficiencies of care (1).

The third problem is the very practical one of assembling data on an individual which give an indication of how well he or she is progressing. People who regularly attend training or rehabilitation centres are likely to have goals set for the acquisition of education, occupational and social skills and their progress will be monitored in terms of their achievement of these goals. The failure to follow progress and to spot a potential crisis is likely to occur for people who have only sporadic contact with the services, especially chronically ill or infirm people who live alone or who live with relatives who do not know how to ask for help from the community care services (2).

To summarise, the development of benefit measures will in the main be left to specific projects, but this should not prevent the routine collection of data about patient progress or the testing of new methods of delivering care. Information about the relative costs and benefits of different forms of care will also come from the evaluation of new or existing schemes. Without more information, especially on the benefits side, many authorities are going to be working in the dark and will have great difficulty in deciding the appropriate placement of people who are presently being cared for in hospital. In the meantime, considerable reliance has to be placed on professional judgement and recent developments in the organisation of community care have attempted to relate the expected benefits of different forms of care to cost constraints. These innovations present possibilities for assessing the efficiency of community care schemes which involve radical changes in budgeting arrangements. Before considering them it is necessary to see how the existing organisational structure and budgeting systems require amendment to encourage efficient decision-making.

# PAYING FOR COMMUNITY CARE

## GENERAL OVERVIEW

The number of organisations involved in community care, the different methods of

FIGURE 6

Social Security	NHS	Family Practitioner Committees	County Councils	District Councils	Voluntary Societies
1	2	3	4	5	6
Pensions Supplementary benefits Sickness, invalidity unemployment disability benefits Constant Attendance Allowance	<u>Hospitals</u> (including inpatient, outpatient, day facilities) <u>Community Services</u> e.g. home nursing health visiting	General Practitioner Dental Services Ophthalmic Services	<u>Social Services</u> Social Work Residential Homes and Hostels Training and Occupation Centres Domiciliary Services <u>Education</u>	Housing	Visiting Services Day Centres Group Homes Cash grants

financing each one and the development of new methods of controlling expenditure or resources use such as cash limits, local authority grant limitations and manpower targets all make the financing, delivery and receipt of community care exceedingly complex. There are six main organisations involved in the delivery of community care as shown in figure 6. For a person seeking help this is a bewildering organisational network, although its saving grace is that considerable assistance in negotiating the way through this maze may be available locally through a General Practitioner, Health Visitor, or Social Worker.

Much more serious is the fact that each of these organisations is separately financed, is working to its own budgeting criteria or financial penalties for over-spending, and has to determine many different priorities amongst competing activities. Health authorities, for example, are trying to determine priorities between acute and chronic care services, between inpatient and outpatient facilities, between old, mentally ill, or mentally handicapped people. County councils are trying to meet competing demands between different departments such as social services and education, as well as between different sections of one department, for example, casework, residential care, and domiciliary services. Given strong pressures to maintain control over expenditure it is little wonder that each organisation is reluctant to accept increased demands coming from another, for example that local authorities have been reluctant to

accept patients discharged from hospital or hospitals to accept patients from residential homes and hostels. With each organisation seeking to minimise the load on its own budget the major beneficiaries are likely to be tax and rate payers. The main victims will be the people in need of help.

This complicated system of finance affects the local authorities more than other organisations so far as community care is concerned. They are financed in the main from local rates and central government grants. The level of local rates is a very sensitive political issue and failure to keep down the rate poundage is often viewed as adversely affecting electoral success for the governing party; failure to keep within government expenditure targets will bring loss of grant which has to be made up from local rates. Thus expenditure by a local authority to provide more care in the community thereby saving actual or planned expenditure in the NHS will sometimes be seen as inefficient spending when in reality, taking the view across all organisations, it is being efficient. In other words, the use of separate crude controls encourages local authorities to block discharges, encourage admission to hospital, and appear laggardly in the development of facilities in the community, or generally to minimise expenditure and the delivery of services.

The problem has long since been recognised, but it has proved difficult to establish a mechanism by which to bring about the desired result of compensating local authorities for providing

the required mix of facilities. Recently, joint finance (3) was chosen as the most useful method especially by 'earmarking' or 'top-slicing' part of the allocation for the sole purpose of assisting the discharge of people from long-stay hospital units. Although the money made available is but a small proportion of the total needed to achieve the plans set out over ten years ago, it is a step in the right direction provided that the finance is spent wisely.

#### **PUBLIC FINANCE OR COMMUNITY EFFORT**

---

It would be a mistake to concentrate too much on the interorganisational financing problems. If community care is to be developed it will need extra resources to achieve its objectives. Where are these resources to be found: from the public purse or by voluntary effort? This is where the equity or fairness aspects of community care become so important. It could be argued that helping others, particularly members of one's own family, is a community or familial duty. One should not need to ask for help. State help is reserved for those unfortunate enough not to have a family. It has already been argued on efficiency grounds that this may not be a good policy because small amounts of help from the state services may free someone to carry out their normal employment or other duties. It might also be argued that providing state help allows people who care about others, but do not themselves possess the practical skills, to pay through the taxation system for someone who has the required skills to carry out this work. Of course, subscriptions to a voluntary society would achieve the same purpose. In addition, subscriptions to voluntary societies also enable people to donate money to that group of people they wish to see helped, whereas taxation is lumped together and need not be spent in the way people feel is correct. With voluntary effort and informal care, however, there is the problem that the duties of caring tend to fall, by default, on a willing few. With taxation it is likely that people will subscribe to services according to their abilities to pay but will have only indirect control, through the electoral or political pressure group system, over the way in which money is spent.

There is a danger, in the present trend of placing more responsibility for care on families or voluntary societies, that help is allocated to people

according to their abilities to cause problems for the statutory services: the allocation of help by so called 'decibel level'. There is therefore a need to cut through some of the rhetoric, which states that care is the family's responsibility, so as to focus thinking on the effects that are likely to impact on the unfortunate families concerned and also on the attitudes and behaviour of those people who do care but who lack an efficient method of expressing their concern in resource terms.

#### **PLANNING AND BUDGETING FOR COMMUNITY CARE**

---

As resources become available for the development of community care services the tension between the need to co-ordinate service delivery across different organisations and the need for financial stewardship will be heightened. The easing of this tension depends upon finding ways of allowing people who are responsible for patient care to have more responsibility and accountability for the services they are prescribing. At the moment there is no client-centred organisation and the placement of people in different forms of care and the monitoring of progress is, at best, likely to fall on a group of professionals and, at worse, be nobody's concern at all. It is then a question of whether or not there is one person or one group of people to whom it can be said 'here are the people who need to be cared for in the community, here is your budget, and this is a set of prices or costs of services. Match the services available to your clients within this budget constraint'.

The best example of the working of this idea is the Kent Community Care project, which has given social workers working with elderly people in the community a budget to ensure that these people remain as long as is practically possible in their own homes. The budget is fixed on two-thirds of the cost of residential care and on the whole the caseworkers involved have been very successful in running an efficient community care service for the elderly (Challis and Davies, 1981). Applied to a wider set of clients and with a population of more severely disabled elderly people, the budget could be related to hospital costs, and local authority and health services would be included (with their costs) in the battery of resources at the disposal of

caseworkers (as in the Kent Scheme) or other groups of professionals.

Such experiments need not be restricted to giving budgets to professionals. There may well be a case for giving more cash allowances to families or other principal helpers with a view to their buying in the services that they need. This is essentially an extension of the principles underlying constant attendance allowance but the criteria may not be as strict nor the weekly payment so large. It nonetheless follows the view that more help in the form of cash may give people the flexibility to tailor services to their own requirements. Bosanquet (1983) has suggested that mentally handicapped persons be discharged from hospital to a 'guardian' who receives a weekly allowance equal to average hospital cost per patient week and uses this budget to ensure the care and training of the person in his or her charge.

The common feature of these schemes is that they alter the present pattern of budget-holding. Instead of budgets being held by directors or heads of services, they are held by case managers or guardians or heads of household who are accountable not only for the money spent and the amounts of service delivered, but also for the achievement of welfare-related objectives for the people in their care. Such new ideas need careful implementation and evaluation to see who is the most efficient budget holder, what budget constraints should be set, how services used should be charged for and how the security and welfare of the person being cared for is to be monitored. Although such a change appears difficult, it must be remembered that the existing system of fragmented responsibility and an accountability based solely on financial stewardship has not been singularly successful in achieving the desired progress towards community care.

#### CHANGING THE PATTERNS OF COMMUNITY CARE

The development of community care has reached an exciting and challenging phase. After many years of frustration there is now some encouragement to try out new methods of organisation, management, and service delivery. Although the present separation of organisational and financial responsibilities does not give incentives to provide the appropriate patterns of care for the people involved, it should be

possible with a little imagination to surmount some barriers to inter-organisation co-operation and to switch the forms of accountability from expenditure and costs to:

1. identifying the most cost-effective forms of care according to the individual circumstances of mentally and/or physically disabled people;

2. monitoring the performance of teams or individuals concerned with the welfare of specific groups in terms of achieving objectives of care as well as in the approved disbursement of resources;

3. devolving budgeting responsibility to teams, individuals or clients;

4. caring for carers by ensuring a variety of well-publicized methods to ease the strain and devising better ways of protecting employment and pension rights; and

5. stimulating voluntary activity.

These changes in focus would direct managers at all levels to the main objective of community care—the welfare of the people in their care.

#### NOTES

1. The problem of this narrow focus in the provision of alternative forms of care for the elderly has been argued at length in chapter 6 of Wright, Cairns, and Snell (1981).

2. A good example is given in Wheatley (1980).

3. Joint finance is an attempt to bridge the divide between local authorities and the national health service. It is money budgeted to NHS authorities to initiate schemes in local authorities or voluntary agencies where non-NHS provision will save resource use in the NHS. A specific part of the Joint Financing Budget is earmarked for assisting the discharge of patients from long-stay hospitals into the community.

#### REFERENCES

1. BOSANQUET, N. (1983). 'Choose or we lose' The Guardian, 27 July 1983.
2. CHALLIS, D. J. (1981). 'The Measurement of Outcome in Social Care for the elderly', J. Soc. Pol., pp. 179-208.
3. CHALLIS, D. J., and DAVIES, B. (1980). 'A new approach to community care for the elderly', Br. J. Soc. Work, 10, pp. 1-18.
4. DAVIES, B., and KNAPP, M. (1981). Old Peoples's Homes and the Production of Welfare (Routledge and Kegan Paul).
5. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1971). Better Services for the Mentally Handicapped, Cmnd 4683 (London: HMSO).
6. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1975). Better Services for the Mentally Ill, Cmnd 6633 (London: HMSO).
7. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1977). The Way Forward. (London: HMSO).
8. MAYNARD, A., and SMITH, J. C. (1983). The Elderly, Who Cares? Who Pays? (Nuffield/York Portfolios No. 1).
9. WALKER, A. (ed.) (1982). Community Care, the Family, the State and Social Policy (Basil Blackwell and Martin Robertson).
10. WHEATLEY, V. (1980). 'Relative stress', Community Care, 24.
11. WRIGHT, K. G., CAIRNS, J. A., and SNELL, M. (1981). 'Costing care', Social Services Monograph: Research in Practice. (Sheffield University).





Although notions of equality pervaded the thinking of the founding fathers of the NHS, it was a good twenty years before the geographical distribution of resources received serious political attention and not until 1980 with the publication of Sir Douglas Black's report did interpersonal (or social class) inequalities in health itself gain public attention. This folio by Julian Le Grand examines the evidence, its reliability, significance and remediability. It also provides what for many readers will be a novel approach to understanding the reasons for inequalities that can be observed: an approach that treats an individual's health as a *capital stock* that can be invested and which is subject to depreciation. This

yields new insights into the *causes* of inequality in health which, even if only partly right, must have important implications for any policies designed to rectify the grosser inequalities. A.J.C.

---

# Inequalities in health and health care

---

**JULIAN LE GRAND**

*Lecturer, London School of Economics*

---

The Government . . . want to ensure that in future every man, woman, and child can rely on getting all the advice and treatment and care which they may need in matters of personal health: that what they get shall be the best medical and other facilities available; that their getting these shall not depend on whether they can pay for them, or any other factor irrelevant to the real need. *A National Health Service.*

---

The 1944 White Paper announcing the introduction of a comprehensive health service, from which the above quotation is taken, stated as its chief aim that 'everybody in the country . . . should have an equal opportunity to benefit from . . . medical and allied services'. (*A National*

*Health Service*, p.47). More recently the Resource Allocation Working Party (RAWP) gave as its principal objective the provision of 'equal opportunity of access to health care for people at equal risk' (DHSS, 1976, p.7). The Royal Commission on the National Health Service

included among a list of seven objectives 'equality of entitlement to health services... without respect to age, social class, sex, race, or religion' and 'equality of access to these services' (1979, p.9). Dr David Owen, one time Minister of State for Health, described the 'central task of the National Health Service' as 'the ultimate abolition of the present inequalities in health provision and care' (1976, p.60); subsequently, David Ennals, added the aim of narrowing 'the gap in health standards between different social classes' (Black, 1980, p.1).

Equality of some kind has thus been a major concern for the National Health Service ever since its inception. Recently this was underlined by the publication in 1980 of the Black Report on *Inequalities in Health*. This was the product of a working party set up in 1977 by David

Ennals, then Secretary of State for Social Services, chaired by Sir Douglas Black, then Chief Scientist at the DHSS, and including two other distinguished members: Professor Jerry Morris, of the London School of Hygiene and Tropical Medicine and Professor Peter Townsend, then of the University of Essex. It collected within one volume most of the available evidence relating to health inequality and, on the basis of that evidence, concluded that massive inequalities persisted despite thirty years of national health care provision.

This folio reviews some of the issues raised by the Black Report from an economic perspective. What is the evidence for health inequality? Can it be relied upon? If so, what are the explanations for the phenomenon? And what, if anything, can be done about it?

## THE EXTENT OF INEQUALITY

### *Inequalities in mortality*

The Black Report reviewed the evidence concerning health inequalities in three areas: mortality, morbidity and the use of health services. We begin with mortality. Figure 1 summarises the differences between social classes in the mortality rates for men and women for four age ranges. It is clear that there are substantial differences between Social Class V (unskilled manual workers and their families) and Social Class I (higher level professionals and their families). The significance of these differences can be better appreciated if it is realized that they imply, among other things:

- children born into unskilled manual workers' families are four more times likely to die in their first year of life than those born into professional families;
- boys in unskilled workers' families are twice as likely

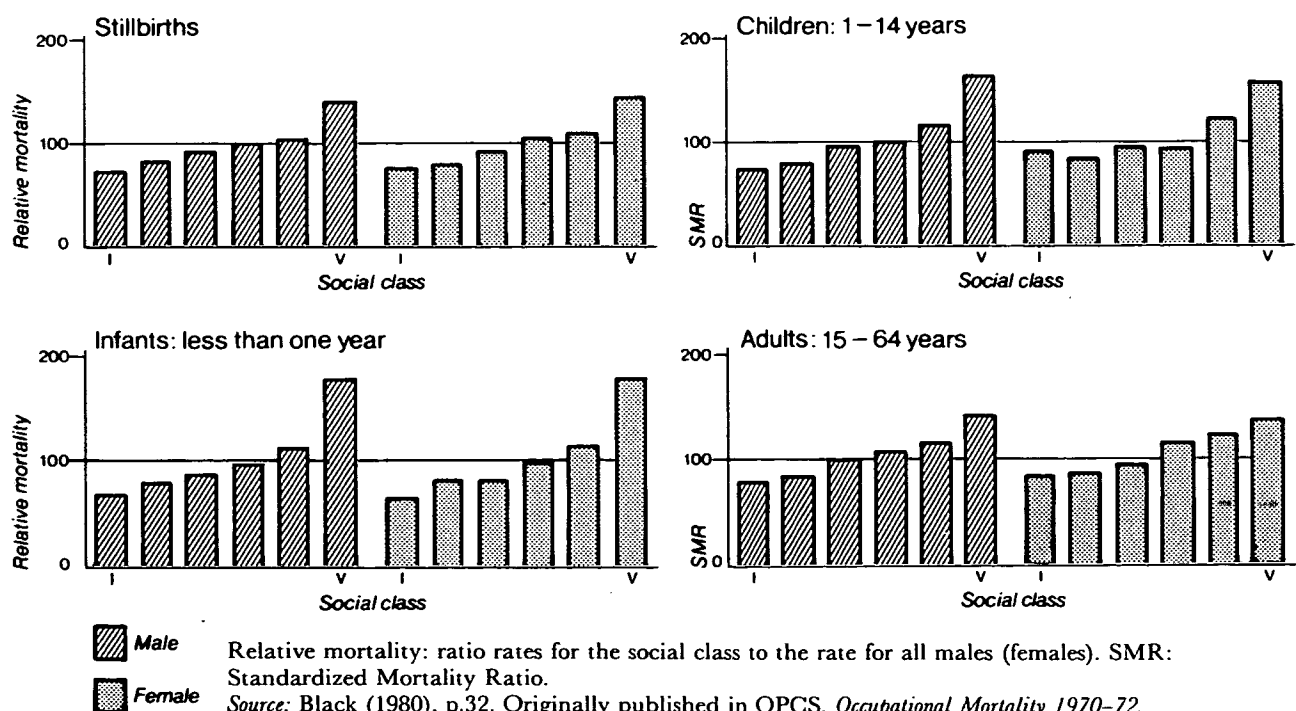
to die between 1 and 14 years of age as their contemporaries in professional families; girls, over one-and-a-half times;

● unskilled male workers are two-and-a-half times as likely to die between the ages of 15 and 44 as professionals;

● someone born to professional parents, if he or she remains in that class, is likely to live over five years longer than if he or she had been born into an unskilled manual household (Black, 1980, pp.33-9). (Figure 1)

Moreover, these differences seem to have got worse rather than better. Table 1 shows the standardized mortality ratios or SMR (see box) for adult males in the different social classes for various years between 1930 and 1972. It can be seen that the gap between Social Classes I and V actually *increased* between 1930-32 and 1959-63, since when it has declined slightly.

FIGURE 1. MORTALITY BY SOCIAL CLASS, AGE, AND SEX



**TABLE 1. Mortality by Social Class, 1930–1972**

Men aged 15–65, England and Wales

Social Class	Standardized mortality ratios <sup>a</sup>			
	1930–32	1949–53	1959–63	1970–72
I Professional	90	86	76	77
II Managerial	94	92	81	81
III Skilled manual and non-manual	97	101	100	104
IV Partly skilled	102	104	103	114
V Unskilled	111	118	143	137

<sup>a</sup> See box

Source: Black (1980), p.65, Table 3.1

## Standardized Mortality Ratio

The standardized mortality ratio (SMR) standardizes mortality rates for differences in the age structure of the groups concerned. It is calculated by estimating the deaths that would have been expected in a particular group if the mortality rates for that group at each age had been the same as the average for the population as a whole, and then expressing these as a ratio to actual deaths  $\times 100$ . Thus an SMR of 80 implies that actual deaths in the group were 80 per cent of those which would have been expected if each of the group's members had exactly the same chance of dying at his or her current age as the population as a whole.

Now care has to be taken in interpreting numbers such as these. Like is not being fully compared with like. The classifications of the different groups changed somewhat over that time and, more importantly, the size of the groups changed. Between 1930 and 1970, for instance, the proportion of the population in Social Class I approximately doubled, while that in Social Class V dropped by a third (Black, 1980, p.90). So, even if the gap between the classes has not diminished, the fact that the relatively disadvantaged group (in terms of health) forms a smaller proportion of the population while the relatively advantaged group forms a higher proportion indicates that there may nonetheless have been some improvement.

Differences in mortality exist not only between classes but also between geographical areas. Table 2 shows SMRs for various Regions within England and Wales. If mortality rates are an indication of health, the healthiest part of the country seems to be the South (East and West); the unhealthiest the East Midlands.

## Inequalities in morbidity

Data on morbidity differences show a pattern that is similar, although less stark, to that for mortality. The most commonly used indicator is self reported illness by socio-economic group, taken from the General Household Survey or GHS (see box). GHS respondents are asked

**TABLE 2.**  
*Regional Variations in Mortality*

Standard Region	SMR: Standardized for Age	SMR: Standardized for Age and Class
Northern, Yorkshire and Humberside	113	113
North West	106	105
East Midlands	116	116
West Midlands	96	94
East Anglia	105	104
South East	90	90
South West	93	93
Wales I	114	117
Wales II	110	113
England and Wales	100	100

Source: Black (1980), p.26, Table 2.2

# The General Household Survey and morbidity

The General Household Survey is a continuous multi-purpose survey of some 14,000 households (32,000 people). Reports are published annually by the Office of Population Censuses and Surveys (OPCS). The questions about morbidity on which the conclusions in the text are based are taken from the 1981 Survey (OPCS, 1983). These were, for chronic sickness, 'do you have any long-standing illness, disability or infirmity? Does [it] limit your activities in any way?', and, for acute sickness, 'during the two weeks ending yesterday, did you have to cut down on any of the things you usually do because of ... some ... illness or injury?'

questions concerning recent periods of restricted activity due to illness (taken as a measure of acute sickness) and any long-standing ailments they may have (taken as a measure of chronic sickness): for details see box above.

The top socio-economic groups (such as professionals, employers, and managers) generally report less of both types of illness than the bottom groups (such as semi- and unskilled manual workers). For instance, in 1981, 9 per cent of males and 12 per cent of females in professionals' families reported some recent restricted activity due to acute sickness compared with 12 per cent of males and 17 per cent of females in unskilled manual workers' families. Also, those in the top group (of either sex) had less than

half the number of restricted activity days per year of that for those in the bottom group. For chronic sickness, the differences are yet more marked: 10 per cent of professionals and their families (male and female) reported limited long-standing illness, compared with 19 per cent (males) and 31 per cent (females) in unskilled manual worker's families. Although, interestingly, this pattern does not appear consistently for all age groups (in particular, the children of professional families are reported as having *more* acute sickness than those in unskilled families), nonetheless the overall picture is clear: the lower the social group, the greater the morbidity.

## Inequalities in use of health services

Finally, inequalities in the use of health services. Much of the available evidence is summarized in the Black Report (1980, Ch.4). This indicates that, on average, the working class undertake more GP consultations, use hospital outpatient departments more, and form a higher proportion of hospital inpatients (relative to their proportion in the population) than do the middle class. However, preventive services, such as family planning or ante-natal care, are used much more by the non-manual classes than the manual ones.

The fact that, on most measures, the working class have

worse health than the middle class and that they use curative medical services such as hospitals and GPs more, raises the question as to whether their relatively greater use matches their relatively greater need. Regrettably, a recent review of the relevant studies (Le Grand, 1982, Ch.3) concluded that this did not appear to be the case. Table 3 summarizes the evidence from one of the studies concerned (Le Grand, 1978). The first column shows the average NHS expenditure per person for four socio-economic groups expressed as a percentage of the mean; it is apparent that there is no distinct gradient. However, it

TABLE 3. *Public Expenditure on Health Care by Socio-Economic Group*

All Persons, England and Wales 1972

<i>Socio-economic group</i>	<i>Expenditure per person: percentage of mean</i>	<i>Expenditure per person reporting illness: percentage of mean</i>
Professionals employers and managers	94	120
Intermediate and junior non-manual	104	114
Skilled manual	92	97
Semi- and unskilled manual	114	85
Mean (£)	18.1	103.2

Source: Le Grand (1982), p.26

is the second column that is important from the point of view of the relationship between use and need; for it shows the NHS expenditure per person *ill* (as measured by self-reported illness). Here there is an obvious trend, with the top groups (professionals, employers, and managers)

receiving 40 per cent as much expenditure per person ill as the bottom group (semi- and unskilled manual workers). If these calculations are correct, they suggest that the National Health Service is still some way from providing 'equal treatment for equal need'.

## EXPLANATIONS FOR INEQUALITY

The evidence outlined in the previous section immediately raises the question 'why?' Why is it that, despite thirty years of welfare state, inequalities in health and in the use

of health services (relative to need) persist? It is a simple question to ask but a complex one to answer, as we shall see.

### *Explaining inequality*

The Black Report listed four different types of explanations for inequality in health: artefact explanations, theories of natural or social selection, materialist explanations and cultural/behavioural explanations. The 'artefact' explanations emphasize the statistical difficulties of making comparisons between groups over time; in particular, it is argued that the apparent failure of inequalities in health to diminish over time is simply a consequence of failing to compare like with like. Selection theories hold that health inequalities are the outcome of a process of social mobility. Healthy individuals tend to move up the social scale, while unhealthy ones find it difficult to work, have to take an undemanding occupation (if they work at all) and hence gravitate to the lower social groups. On this reasoning, it is not the socio-economic position of individuals in a particular group that affects their health; rather it is their health that determines their socio-economic position.

In contrast, materialist explanations emphasize the difference in material constraints that characterize the social classes as the major determinant of differences in health. Individuals born into the lower social groups, or who, for whatever reason, end up there, are materially deprived relative to those in the higher groups. As a result, they are more likely to live in cramped dwellings and polluted areas, less able to afford nutritious food or appropriate clothing and, through the educational deficiencies that often accompany material deprivation, more likely to have to work in occupations that are in one way or another hazardous to health.

Behavioural explanations focus on the role of the individual. On this view, an individual's state of health is

a product or a by-product of his or her free choice. Building on the fact that those who smoke, eat the wrong sort of food, drink to excess, or take too little exercise are far more likely to suffer ill-health than their more moderate counterparts, those who hold this view argue that ill-health is largely the result of irresponsible or poorly informed choices. The reason why the lower social groups have greater incidence of ill-health than higher ones is therefore simply because they have a greater proportion of irresponsible and/or ill-educated people.

Which of these explanations is correct? The answer undoubtedly is that they all are—to some extent. We have seen earlier that the apparent failure of mortality differentials to narrow over time is indeed due in part to like not being compared with like; hence the artefact explanation for that particular phenomenon at least has some credibility. However, no amount of statistical juggling will alter the fact that class differences do exist—that, despite thirty-five years of the welfare state, people in the higher social groups report less illness and live longer than those in the lower ones. Also, since some people are socially mobile (although much fewer than popular mythology would have us believe: see Goldthorpe, 1980) and since it is plausible to suppose health is a determinant of mobility, it is quite probable that at least part of the link between socio-economic position and health is due to social selection. However, it is unlikely to be the whole story, for, as we have seen, inequalities in adults' health (at least as measured by mortality) are mirrored in inequalities in children's health—which could not be explained by this hypothesis.

### *Health as an economically chosen 'stock' of capital*

To economists, the distinction between the third and fourth type of explanation—the behavioural and the materialistic—seems false. In health, as in virtually all other areas of human activity people make choices under constraints; hence health outcomes are a product of *both* individual's choices *and* the constraints under which they make those choices. This view was first articulated formally by Grossman (1972); and understanding of his ideas is essential fully to appreciate the argument and its implications for the explanation of health inequality.

Grossman's basic argument may be sketched as follows. Epidemiological measures of health such as life expectancy and the average length of life suggest that it might be useful to see health as *stock* of capital, analogous to

other forms of 'human' or 'non-human' capital (education or physical assets). This stock will 'depreciate' over time because of the natural process of ageing, and sometimes because of the onset of unavoidable illness. However, individuals can also affect the depreciation process by their own actions. They can partly or even (for a time) wholly offset depreciation, regardless of its source, by appropriate 'investment' in health-promoting activities; taking exercise, for instance, or where appropriate, receiving medical treatment. On the darker side, they can also increase the rate of depreciation by engaging in 'health-consuming' activities, such as smoking or working in physically demanding occupations. Individuals' states of health at any one time will therefore depend on their

initial stocks of health and on the extent to which they have invested in, or consumed, that stock. Their decisions concerning health-investment and consumption will, in turn, depend on a number of factors including the benefits they perceive from health-investment (such as an improved quality of life and an increased ability to work) the costs of health-investment (such as the price of nutritious food or of obtaining medical care) and the benefits and costs of health-consuming activities (such as the pleasures of smoking and the price of cigarettes).

This formulation can now be applied to the problem of explaining inequalities in health. There are a number of reasons why we might expect this process to lead to differences in the health 'stocks' of rich and poor. To begin with, because of poor maternal nutrition and other environmental factors, individuals born into the lower social groups are likely to have a smaller initial health stock than those born into higher ones. Second, poorer individuals will engage in less health-investment. This is in part because it often costs them more: even using the 'free' National Health Service can cost the poor more than the rich (see box). It may also be because they perceive themselves as receiving less benefit. Educational deprivation means they know less about the health advantages of, for instance, a balanced diet, or of preventative medical services. Moreover, in some cases such as curative medical

care, they may actually receive less benefit. Doctors generally communicate better with their middle-class patients than their working-class ones; they also spend longer with them (Le Grand, 1982, pp.33-4).

Finally, because they have so few other resources, individuals from poorer backgrounds may be forced to consume their health stocks faster than those from wealthier ones. While the latter often have educational skills and perhaps inherited wealth, the former have only one income-producing 'asset': their health stock. As a result, to obtain the wherewithal to live they have to use up that asset at a greater rate than those more fortunately placed.

This 'economic' theory of health inequalities seems plausible and, moreover, can be used to generate testable hypotheses (Muurinen and Le Grand, 1983). However, it has to be admitted that as yet it has not yet been adequately tested—a feature that it shares with most of the other theories outlined here.

In fact, empirical evidence concerning the determinants of inequalities in health is sketchy. Such work as exists (reviewed in Le Grand, 1982, pp.41-5) indicates as relatively *important* diet and air-pollution; as relatively *unimportant*, housing conditions and—interestingly—medical care. However, the results are controversial, and much work still needs to be done in this area.

## The differential costs of using the NHS

The fact that the poor may face a higher price of health investment can be illustrated with respect to the medical care provided under the National Health Service. At first sight, it might appear that in Britain at least, most medical care does not have a 'price'; for it is provided free (or largely so) under the National Health Service. However, this ignores an important component of the cost of engaging in any activity, including medical treatment—that of time. To obtain medical treatment involves spending time in at least three ways: time spent travelling to the medical facility concerned, time spent waiting for treatment once arrived, and time spent on actual treatment. The first two of these at least are likely to be greater for the poor than for the rich. Time spent travelling will be greater because they are more reliant on public transport; moreover, per mile it costs more (at the margin). Also the poor are likely to have further to travel to obtain comparable medical care, for the areas in which they live are poorly endowed with medical facilities (Noyce, Snaith, and Trickey, 1974). Time spent waiting is likely to be longer, since they have fewer telephones and so cannot easily make appointments; also the reliance on public transport makes it difficult to arrive and depart precisely when one wishes.

Moreover, not only may the poor lose more time using the NHS than the higher paid, but the cost of each hour thus lost may be greater, particularly for those in work. Many professionals and managers are paid an annual salary and hence do not lose income directly for time spent visiting the doctor during working hours, whereas workers in manual occupations, paid by the day or the hour, many have to forego their pay for the time taken off. Cartwright (1964) found that during a spell in hospital only 20 per cent of working-class heads of household received full wages compared with nearly 80 per cent of middle-class ones. Indeed, over half the working-class heads of household received no money at all from their employers, compared with just 12 per cent of the middle-class.

Against this, it must be added that, if they *do* lose money, middle class workers are likely to lose more, as their wage rates are generally higher. But even then, the actual sacrifice involved may be less, for each pound of loss is likely to create a greater sacrifice for those on lower incomes than for those on higher ones.

# WHAT CAN BE DONE?

It is difficult to formulate policies to reduce inequalities in health and health care until we know more about their causes; and, as we have seen, as yet much of the evidence is not there. More investigation into the determinants of inequality in health is essential; and that perhaps is the most important implication of the Black Report.

However, we can use the economic analysis outlined

above, in conjunction with such evidence as does exist, to draw out some implications for policy. What is needed are, on the one hand, policies designed to affect the *health-investment* decisions of the poor (raising the benefits of health-investment or lowering its costs), and on the other, policies aimed at influencing the poor's *health-consumption* (lowering its benefits or raising its costs).

## ***Health-investment policies***

Policies that would raise the *benefits* (perceived and actual) to the poor of investing in their health include improvements in health education and changes in the training of medical personnel so as to increase the latter's awareness of the health difficulties faced by poorer groups. The former would directly benefit the poor's health; the latter would increase the likelihood of receiving appropriate help when they fell ill. But, although undoubtedly there is room for improvement in these areas, it would be unwise to place too much faith on them. A recent review (Gatherer, *et al.*, 1979) of the available evidence found that most health education programmes were noted more for their general ineffectiveness than anything else (the exception were those directed at immunization in less developed countries—not of immediate relevance to British health policy); and class attitudes are usually well entrenched by the time that most medical training is undertaken.

A more promising avenue is to lower the *costs* of health-investment to the poor. One way of doing this is to relocate health service facilities away from middle class and into working class areas. We have seen that distance is an important determinant of health service use; hence anything that reduces the distance the poor have to travel

is likely to increase their use of the service and therefore their levels of health investment.

Following the RAWP report (DHSS, 1976) a creditable attempt was made to relocate health service facilities away from the medically over-endowed and wealthy south to the under-endowed and poor north and east. However, this did little to redistribute resources *within* the regions; moreover, partly because of financial constraints and partly because the potential losers, being articulate, educated and middle-class, were able to mobilize effective opposition, even the regional redistribution has now run out of steam. It is an urgent priority therefore to revitalize this process, and to extend it to cover intra-regional disparities as well as inter-regional ones.

Another important factor affecting the costs of health-investment to the poor is the price of nutritious food. In this context, the recent (and on-going) phasing out of subsidies to the school meals service is of particular concern; for this has the effect of raising rather than lowering the cost of nutritious food and, moreover, doing so at a crucial age for health development. Subsidizing the diet of poor children through the schools system is a neat and efficacious method of lowering the costs of health-investment; it needs to be expanded, not contracted.

## ***Health-consumption policies***

The simplest way of reducing people's health-consumption activities is to raise the prices of those activities relative to those for healthier ones. An obvious example is increase the taxes on cigarettes and alcohol so as to raise their prices. However, care has to be taken with such policies if the aim is specifically to help the poor. For if the activity concerned is something on which the poor spend a fair proportion of their budget the effect may be to reduce the money they have available to spend on other things, including those that promote their health (such as nutritious foods). Taxes on cigarettes, for example, are already regressive (in that they take a larger proportion of poor people's income than that of the rich); if the effect of increasing those taxes was not significantly to reduce consumption by the poor but instead to reduce the cash available for other purposes, then the results might be quite the reverse of those intended.

The solution is to couple any increase in the relative price of health-consuming activities with measures designed to increase the incomes of the poor. Any extra revenue raised, for instance, through increased alcohol and cigarette taxes could be used to increase social security

payments or, less controversially, child allowances. Moreover, a policy of raising the income of the poor would improve their health in other ways. It would reduce some of the barriers they face in obtaining medical care, and it would contribute to improvements in their housing and in general living conditions. In short, perhaps the most effective way of *both* reducing consumption by the poor that is harmful to health and increasing their health investment is to raise their incomes.

In fact there is a more general point here. Many of the factors identified as possible determinants of inequalities in health and health care are not under the direct control of the National Health Service. Rather, they stem from basic social and economic inequalities: differences in income and in education; differences in working conditions; the class divisions that render different groups in society mistrustful of one another. Although all the facts are not yet in, the conclusion seems almost inescapable that inequality in health and health care reflect in large part inequality in society and that they will never be *fully* eliminated so long as that wider inequality persists.

# SUMMARY

Despite thirty-five years of the welfare state, inequalities in health and health care are still a major feature of British society. The poor die younger, report more ill-health and use health services less relative to need than their wealthier compatriots. A simple economic model of individuals making health 'investment' and 'consumption' decisions under constraints provides some insight into the reasons for this. The poor derive less benefit from, and face higher costs of, health-investment than the rich; they also derive (relatively) more benefit from health-consumption. Policies should be aimed at changing these relative benefits and costs. These include:

- continuation of the RAWP reallocation of health service resources away from wealthier regions to poorer ones;
- extension of RAWP to cover intra-regional disparities;
- expansion of the provision and subsidy of nutritious food to the children of poor families through the school meals service; and, *most importantly*;
- increases in the incomes of the poor.

## REFERENCES

- BLACK, D. (1980). *Inequalities in Health*. Report of a research working party, chaired by Sir Douglas Black. London: Department of Health and Social Security.
- CARTWRIGHT, A. (1964). *Human Relations and Hospital Care*. Institute of Community Studies Series. London: Routledge and Kegan Paul.
- DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1976). *Sharing Resources for Health in England*. London: HMSO.
- GATHERER, A. *et al.* (1979). *Is Health Education Effective?* Monograph No 2. London: Health Education Council.
- GOLDTHORPE, J. H. (1980). *Social Mobility and Class Structure in Modern Britain*. Oxford: Oxford University Press.
- GROSSMAN, M. (1972). *The Demand for Health*. New York, NBER.
- LE GRAND, J. (1978). 'The distribution of public expenditure: the case of health care'. *Economica*, **45**, 125-42.
- LE GRAND, J. (1982). *The Strategy of Equality*. London: Allen and Unwin.
- MUURINEN, J. M. AND LE GRAND, J. *The Economic Analysis of Inequalities in Health*. Mimeo, London School of Economics.
- A National Health Service* (1944). London: HMSO, Cmnd. 6502.
- NOYCE, J., SNAITH, A. AND TRICKEY, A. J. (1974). 'Regional variations in the allocation of financial resources to the community health services'. *The Lancet*, **7857**, 554-7.
- OFFICE OF POPULATION CENSUSES AND SURVEYS (1983). *General Household Survey, 1981*. London: HMSO.
- OWEN, DAVID (1976). *In Sickness and in Health*. London: Quartet.
- ROYAL COMMISSION ON THE NATIONAL HEALTH SERVICE (1979). *Report*. London: HMSO, Cmnd 7615.



# EDITORIAL

Folio 6.

Editor: A. J. Culyer



Economic appraisals of health services meet with three characteristic adverse reactions from health service professionals: suspicion bordering on hostility, enthusiasm bordering on the 'let's do it ourselves', and plain indifference. The first of these, if it derives from a feeling that entrenched habits of thought and practice should be protected, is scarcely worthy. But if it derives from methodological and ethical concerns it is clearly worth serious consideration. The second reaction can bring cost-benefit and cost-effectiveness principles into disrepute by misapplication. Neither economists nor medical specialists should proceed without the other in this terrain. Michael Drummond's folio, particularly if taken in conjunction with Alan Williams' folio 2, is designed to allay fears, to give insight with the cost-benefit way of thinking, and to help people who may be

involved in economic appraisals to do them. If his exemplary exposition and apposite illustrations can also whip up enthusiasm among the indifferent, then all three problems will have been tackled successfully in this paper—to the greater benefit of both the health professionals and those whom they serve: their patients.  
A.J.C.

# Cost effective analysis in health care

**MICHAEL  
DRUMMOND**

*Senior Lecturer in Economics,  
Health Services Management Centre,  
University of Birmingham*

Those who make resource allocation decisions about health care, be they planners, health authority officers and members, or clinicians, currently face a number of challenges. First, the monies made available by government, although fairly constant in real terms, appear to be increasingly stretched in the face of competing demands for their use. Second, these demands are themselves fuelled by rapid technological change which shows no sign of abating. Third, authorities have still to respond to national priorities in the development of services for the elderly, mentally handicapped, or mentally ill and satisfy, by way of the consultative process, the wishes of the community at large and special interest or 'pressure' groups. Finally, authorities, in their decision-making procedures, need to take note of several DHSS initiatives concerning value for money in the use of health service resources. The new guidance on option appraisal (DHSS, 1981) requires that proposals for capital schemes be set out in an explicit and logical framework. The Report of the NHS Management Inquiry team (1983) expresses concerns about the inability of management to grapple with the issue of clinical resource use and argues that doctors should become more involved in the management process, 'consistent with clinical freedom for clinical practice'.

Particular government proposals may come and go but one thing is clear: tough choices need to be made. Moreover they need to be made in a way that is defensible when judged from a wider perspective. How can the decision-maker come to a view on issues such as: should an authority develop a community-oriented service for the elderly mentally infirm; by what method should long-term domiciliary oxygen therapy be provided to chronic bronchitics; should an authority obtain a body scanner; should pregnant women be offered a screening test for serum alpha-fetoprotein level in an effort to reduce the numbers of infants born with open spina bifida?

Resource allocation decisions in health care have to take account of a complex interplay of social, economic, and political factors. However, for those decision-makers wanting some guidelines in the wilderness, this folio discusses a method of evaluating health service options that is consistent with the notion that resources should be deployed so as to maximize the total benefits to the community. The cost benefit approach has

been used quite extensively to evaluate the efficiency or cost-effectiveness of alternative health programmes or clinical procedures (Drummond; 1980; Warner and Luce; 1982). The growth in this literature is most marked in clinical journals and it is more likely these days that requests for service developments will be accompanied by some statement about 'cost effectiveness', perhaps backed up by some analysis.

The references cited above outline the methodology of cost benefit and cost effectiveness analysis, with the aim of helping those researchers undertaking such analyses to improve the quality of their work.

However, the emphasis here will be on using the way of thinking embodied in the approach to assist in decisions that health service planners, clinicians, managers, and authority members have to make. First, the basic principles will be introduced by reference to a number of examples. Then a simple checklist of questions to ask about any proposal for committing health service resources will be presented.

# SOME EXAMPLES

*A clinician wants the Authority to sanction purchases of a new machine called an oxygen concentrator, which he claims is a more cost effective method of delivering oxygen therapy in the home to chronic bronchitics.*

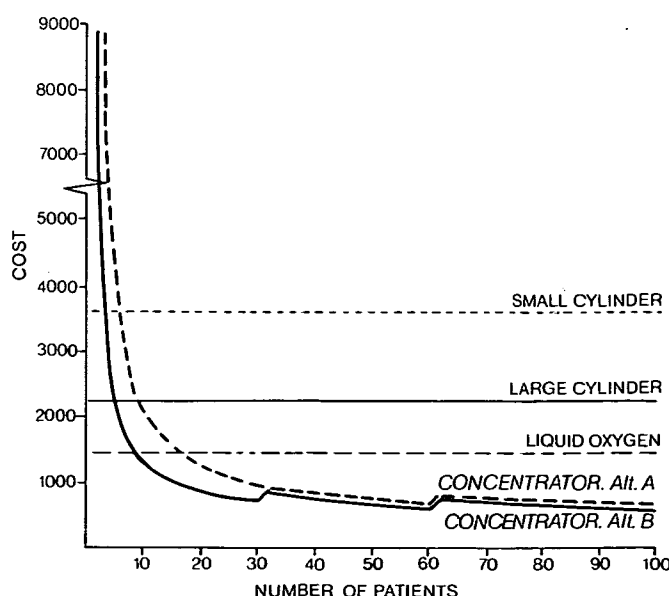
This kind of issue arises frequently due to rapid technological change in the medical equipment industry and many of the choices have been analysed within a cost effectiveness framework, including that outlined above (Lowson, Drummond and Bishop, 1981). The first issue to be clear about when beginning such a study is the range of choices (or alternatives) to assess. It can be quickly established that the oxygen concentrator (an electrically-driven machine which extracts oxygen from air) is a potential replacement for oxygen cylinders in the provision of long-term domiciliary oxygen (up to 15 hours a day) for chronic bronchitics. There is also a third method of oxygen supply, liquid oxygen delivered in bottles to the home.

Note that although it is quite legitimate to compare the three methods identified, there is always a broader choice: whether one should provide this kind of service at all given the other potential uses of the resources available. (After all, the costs of health care treatments or services should not be viewed merely as money expenditures but as potential benefits that are being foregone, because the resources can be diverted to other uses. Economists often refer to the **opportunity cost** of a resource: the benefits that would be derived from using it in its best alternative use.) Hence the logic of comparing health service options in terms of their benefits and costs. A further implication of the opportunity cost concept is that costs can be incurred without spending money; for example, in using hospital space that could be used for other activities.

In this particular case let us assume that the service would be offered in some shape or form, so the evaluation centres on the relative costs and effects of the three methods. (Both large and small oxygen cylinders were considered, making four options in total.) The effectiveness of the different methods of oxygen production could be assumed to be the same (a view based on previous medical research), so interest centres on the relative costs. A major difference between the concentrator and the other methods is that it requires capital outlays on the purchase of the machines and provision of workshop facilities. Therefore the evaluation needs to take account of the fact that, under the concentrator option, more costs would be incurred sooner rather

than later. It is usually argued that, as individuals or as a community, we are not indifferent to the timing of resource outlays. Usually a public sector discount rate (which in the UK is advised by the Treasury) is used to make comparisons between costs and benefits occurring in different points in time. In the study by Lowson, Drummond and Bishop such a rate was used to annuitize the capital outlays over the assumed useful life of the equipment (or buildings). The calculation is rather like that a bank would perform in estimating annual loan repayments. The

Figure 1. Cost per patient per annum for all methods of providing oxygen. (Source: Lowson, Drummond and Bishop, 1981).



only difference is that a bank would use the market rate of interest, which is likely to be different from the public sector discount rate. (The practical significance of this point will be returned to later. However, those requiring more discussion of the concepts involved and the calculations performed should consult Drummond (1980)).

The capital and revenue costs of the alternative methods of oxygen production were thus compared (Figure 1). These included, for the oxygen cylinders, purchase and delivery of the oxygen from local pharmacists and, for the oxygen concentrators, electricity, and routine maintenance charges. It can be seen that for all but small numbers of patients, where the set up costs per patient are large, concentrators are the cheapest method. This result was fairly independent of whether the Authority con-

cerned already had workshop facilities available (Alternative B) or whether extra facilities had to be provided (Alternative A on Figure 1).

Even this relatively simple choice raises some broader issues. Under the concentrator option, for example, patients might bear a higher proportion of the costs (in electricity charges) unless they were in receipt of supplementary benefit or over the age of retirement. Also, at the time of the study, oxygen cylinders could be prescribed freely by general practitioners, whereas purchase of a concentrator would have to be funded out of a cash-limited medical equipment budget. This position has changed somewhat, and concentrator rental is now a possibility, but it is worth noting that on occasions what is cost effective for society as a whole may not necessarily be the cheapest option for the Health Authority.

*Community pressure groups argue that the Authority should purchase a CT scanner. Although there is some worry about the cost implications, it is pointed out that there will be cost savings in that investigations previously performed will no longer be necessary.*

The last example illustrated how technological advances can be both beneficial for patients and save resources. However, more generally the rapid diffusion of new medical technologies has been a cause for concern (Banta, Behney and Willems, 1981). This is particularly true of the diagnostic technologies such as the CT (computer-assisted tomography) scanner. Although many district hospitals now have a scanner, there is still debate over the appropriate use of the technology; also a number of new imaging technologies, such as nuclear magnetic resonance imaging (NMRI) and positron emission tomography (PET), are in the developmental stage.

Pioneering cost effectiveness analyses of the CT scanner in the UK have illustrated many of the difficulties in evaluating new

diagnostic technologies (Thomson, 1977; Bartlett, *et al.*, 1978), but they also provide a basis for applying cost-benefit thinking to the problem. The central thesis of the early studies was that although the impact on patient outcome of scanners was unknown, the concerns over costs were to some extent unwarranted, as savings in resources would be generated. For example, Thomson (1977) argued that scanning of the head would avert the diagnostic costs and hospitalization associated with the current procedures (e.g. encephalograms, isotope studies, and air studies).

Whilst history has shown that other investigations are indeed averted, it is worthwhile highlighting a number of issues pertinent to an authority's decision to purchase such a machine.

### BOX 1

#### THE IMPORTANCE OF LENGTH OF LIFE AND THE DISCOUNT RATE IN ESTIMATING THE ANNUAL COST OF CAPITAL EQUIPMENT

If an item of equipment (such as a body scanner) has a purchase price of £140,000, one might be tempted to assume that this represents an equivalent annual cost of £14,000 over 10 years. However, this ignores the fact that we are not indifferent to the **timing** of resource outlays. The UK Treasury currently argues that a **public sector discount rate** of 5 per cent be used to compare costs occurring at different points in time. In our example the discount rate can be used (rather like a bank interest rate) to calculate the annual amount in each of 10 years (i.e. annuity) that would be equivalent to £140,000 now.

Assuming the annuity is **in arrears** (i.e. paid at the end of each year), the calculation is to find the amount  $x$  in each of  $n$  years which, at discount rate  $r$ , would be equivalent to our capital outlay  $k$ . This is given by the formula:

$$k = \frac{x}{1+r} + \frac{x}{(1+r)^2} + \frac{x}{(1+r)^3} + \dots + \frac{x}{(1+r)^n}$$

For our example:

$$140,000 = \frac{x}{1+0.05} + \frac{x}{(1+0.05)^2} + \frac{x}{(1+0.05)^3} + \dots + \frac{x}{(1+0.05)^{10}}$$

$$140,000 = x \div (\text{annuity factor for 10 years at 5 per cent}).$$

Using tables to obtain the annuity factor (the amount in brackets), it can be established that

$$x = £18,130$$

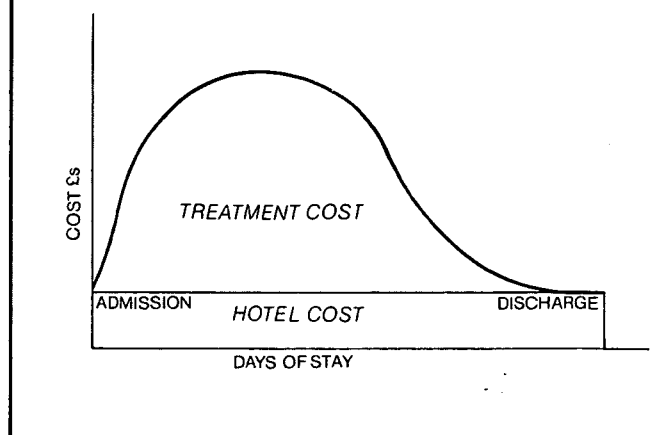
If a useful clinical life of 5 years, rather than 10, were assumed, the equivalent annual cost of the scanner would be £32,336, a far cry from the £14,000 originally suggested.

First of all, although the new technology may give equivalent (or better) diagnostic information than current methods, can this information be used to improve the outcome for the patient? There may be cases where better diagnostic information is obtained but no effective medical interventions exist. (Effective medical interventions can be defined as those treatments that do more good than harm). A particularly worrying aspect of new diagnostic technologies is that once they are available the clinical indications for their use may expand. This is especially true of the relatively non-invasive techniques like CT scanning; the unpleasant nature of some of the procedures it replaced would make any reasonable person think twice about how important the diagnostic information was for clinical management of the patient. Non-invasive techniques may not be subjected to the same close scrutiny.

Drawing on some of the lessons from the first case study, it is also worthwhile taking a closer look at the costs and potential savings from purchase of the scanner. Thomson argues that the original purchase price of the scanner, £140,000, can be assumed to represent £14,000 per year over 10 years. (A slightly more sophisticated version, available from year 2 and costing £210,000, would thus represent £21,000 per annum.) This amount is then set against the savings from averting other investigations and from avoiding the hospitalization of patients. (The scans can often be performed on an out-patient basis.) It is clear that if a more pessimistic assumption were made about the useful clinical life of the equipment (say, 5 years rather than 10) and the currently advised public sector discount rate (5 per cent) were applied, the equivalent annual cost would be higher; in fact it would be around £32,000, not £14,000 (see Box 1).

The other major point relates to the interpretation of the resource savings from reductions in hospital length of stay or reductions in hospital admissions. Often an average daily hospital cost is used as a starting point for the estimation. First, it should be noted that often the later days of a patient's stay are less resource intensive than the earlier days. (See Figure 2.) Therefore the average daily rate may overestimate these savings at the margin. Second, one has to consider the alternative uses to

Figure 2. The variation of hospital cost with hospital in-patient stay. (Source: Drummond, 1980).



which the freed resources (in this case staffed available beds) will be put. Of course, if the beds were filled by other patients, expenditure might actually increase following purchase of the scanner. On the other hand there may be benefits to the additional patients treated if the treatments they were given were effective. In the example being cited, Thomson uses an average daily cost to estimate the resource savings from investigating patients as out-patients rather than as in-patients, but halves this figure as 50 per cent of such patients would probably not have been investigated at all in pre-scanner days. Furthermore, he acknowledges that no beds were closed down as an immediate effect of the installation of the scanner.

Overall the chances are that purchase of a scanner will put up costs. Therefore in order to estimate its worth in economic terms one would need to obtain some estimate of the benefits, in improved patient outcome or some other valued output such as reassurance to patients, in order to compare with the extra costs.

*A new surgeon appointed in your District wants to perform a number of minor elective operations as day cases. He claims that this would give equivalent clinical results and would have the additional advantage that more cases could be processed with more convenience to patients.*

In discussing the last case study the question of establishing the effectiveness of medical interventions arose. It should be clear that, with the possible exception of one or two simple situations, assessment of effectiveness is an important component of a cost effectiveness analysis. (This also means that cost effectiveness analysis is usually a multidisciplinary activity involving epidemiologists and clinical researchers, as well as economists.)

In addressing the issue set out above, effectiveness is clearly important; how would this be assessed? The correct way would be to undertake a controlled clinical evaluation of the new procedure on a prospective basis. That is, patients would be assessed to see whether, in principle, they could be given the new treatment or the old and then asked whether they were willing to participate in the study. (Ethical principles dictate that they should always have the right to opt out.) Patients would then be allocated, preferably by a randomization procedure, to the new therapy or the old. (The random allocation procedure minimizes the chances that the groups will differ in any respect other than the therapy given.) The two groups of patients would then be followed up to assess their outcome (say, in terms of complications of surgery and recurrences). There are further technical issues in the design of such a study, in particular the gathering of

a large enough sample of patients to show any differences in outcome that may exist, but the basic principle is as discussed above.

The particular case outlined above has been investigated in the UK (Russell, *et al.*, 1977) and there have been other examples of the combination of clinical and economic evaluation in the fields of short-stay surgery (Waller, *et al.*, 1978), physiotherapy (Forster and Frost, 1980), community care for psychiatric illness (Weisbrod, *et al.*, 1980), and prevention of pulmonary embolism in high risk surgical patients (Hull, *et al.*, 1982). Because of the time and effort involved in undertaking such studies, these are still few in relation to the number of medical therapies in existence. Therefore, those undertaking cost-effectiveness studies, and those proposing or evaluating new clinical developments, often need to piece together evidence from a number of sources, as Culyer and Maynard (1981) discovered when they came to examine the case for cimetidine, a new drug for peptic ulcer. Thus, when judging a cost effectiveness study, or a proposal for health service developments, it is crucial to enquire after the quality of the medical evidence.

The day-case surgery example, and the study by Russell, *et al.*, (1977) raise two other important issues. First, it was found that

## BOX 2

### COST EXTERNAL TO THE DECISION-MAKER—THE WIDE RANGE OF COSTS RELEVANT TO AN EVALUATION OF A NEW COMMUNITY-BASED PROGRAMME FOR MENTAL ILLNESS PATIENTS

Weisbrod, *et al.* (1980) wished to compare the costs of a new community-based mental illness programme with the conventional hospital-orientated alternative. This involved estimation of the costs borne by a number of public agencies and the family. The results, in cost per patient (12 months after admission) were as follows (in US dollars)

	<i>Conventional hospital-orientated programme</i>	<i>Community-based programme</i>
<b>Direct treatment costs</b> (borne by the main care agency)	3138	4798
<b>Indirect treatment costs</b>		
Other hospital care	1744	646
Sheltered workshops	91	870
Other community agencies (e.g. social services, visiting nurse service)	285	310
Medical practitioners	22	12
<b>Law enforcement costs</b>	409	350
<b>Patient maintenance costs</b> (borne by government agencies)	1487	1035
<b>Family burden costs</b> (e.g. lost earnings due to patient)	120	72
<b>TOTAL</b>	7296	8093

It can be seen that the annual costs of care are \$1660 more (per patient) to the main care agency. However, when a wider range of costs is considered this cost difference falls to less than \$800. It is this extra cost that has to be weighed against any extra benefits of the new programme.

day-case patients used more community services, either from professionals such as district nurses or in more informal help provided by the family. It turned out that these costs were not large enough to reverse the decision that might be made from a consideration of hospital resources alone. However, they should be noted all the same; all too frequently health care planning decisions are made without adequate consideration of the extra burden they may impose on patients, families, and local authority services. (See Box 2 for an example of the range of costs that can be estimated.)

Second, the authors of the day-case surgery study stressed that the economic advantages from instituting the service would depend on the uses to which freed resources would be put. They

investigated two possibilities: maintaining the existing level of service and closing an existing 5-bedded ward, or expanding surgical services without the need to build a further 5-bedded ward. Depending on the circumstances facing a particular authority one might want to consider other options. For example, it may be that theatre time is the main constraint and that the beds freed might be used to relieve pressure on non-surgical waiting lists. Such a redeployment might require more discussion and negotiation, of course, as the clinician in our example was presumably thinking in terms of treating more surgical cases. Nevertheless, it is clear that day-case surgery would offer the potential for increased efficiency in the use of beds.

*General practitioners in your District feel that more long-stay beds should be provided for the elderly mentally infirm. However, the consultant psychogeriatrician feels that the District should develop a community-orientated service, based on community psychiatric nurse visiting and day care.*

Stress was laid, in the last example, on the importance of obtaining assessments of the effectiveness of health care treatments or programmes, through controlled study. In addressing the issue outlined above, one would ideally like information on the effectiveness of community-orientated care for this client group, to compare with that of hospital care for similar patients. There have indeed been randomised controlled trials of commu-

nity care versus hospital orientated care for patients with psychiatric illness (Weisbrod, *et al.*, 1980, Fenton, *et al.*, 1982) but, compared to the surgical example mentioned earlier, such trials would be more difficult to mount in this case. For example, in an average District there may be small numbers of suitable patients referred in a given year and no vacant long-stay beds available for use in the study, unless one were willing to consider

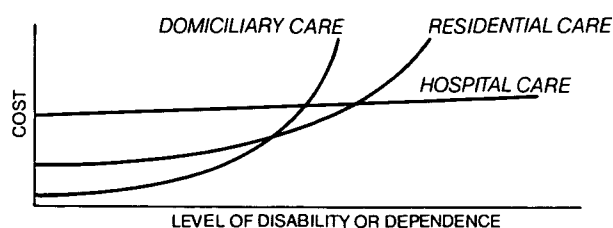
allocating (on a random basis) some existing patients to be discharged from hospital. Also, compared to the surgical example, there is much more likely to be differing opinions on the appropriate measure of output or 'success' of treatment for the elderly mentally infirm. To a certain extent, those who argue for a community-orientated service may put a higher value on maintaining the independence of the patient in his own home; on the other hand, those who argue for a more institutional-based service may put a higher value on reducing the risk of injury to the patient and the peace of mind of relatives.

Given these problems, what does the cost benefit approach have to offer? First, the resource costs of particular categories of care could be estimated; that is, what would it cost the Authority (in capital and revenue expenditure) to provide the beds being requested, and what could be provided with the same funds in a community-orientated programme, in terms of day-care places and community psychiatric nurse visits?

Second, it might be possible to investigate the likely service requirements (and hence cost) of patients of varying levels of dependency in the alternative modes of care. This approach has been used by Wright, *et al.*, (1981) in a study of care of the elderly in general, comparing domiciliary care with care in local authority residential accommodation and NHS hospital. Such a comparison should include the resource inputs from local authority and social services personnel and the family, particularly for the community care option, as in Box 2. In general one might expect the costs of care to rise by dependency for all forms of care, with the costs of domiciliary care rising more steeply (See Figure 3). The reason for this is that institutional forms of care involve more regimentation of patients and are therefore less sensitive to changes in dependency levels, although not completely independent of them.

Third, although it may be difficult to address the benefit question directly, it may be possible to obtain professional assessments of which patients may be better off, or at least no worse off, if they were moved from one form of care to another. This approach has been used by Fordyce, *et al.*, (1981) in planning for balance of care of the elderly in the Grampian Region of Scotland. Their general conclusion was that, because of

Figure 3. Alternative forms of care for the elderly: the variation of cost with level of disability. (Source: Drummond, 1980).



the relatively low costs of community services, there would be advantages in moving towards a more community-orientated service, in that with the given budget more patients would be provided with care that the professionals considered appropriate. In short, it was considered that a shift in the balance of care to incorporate more community services would give better 'value for money'. In some studies the valuation of health outputs has been approached in a more formal manner. For example, in a recent study of the economics of neonatal intensive care, Boyle, *et al.*, (1983) obtained assessments, from a sample of residents of Hamilton, Ontario, of the relative value of being in various health states. These assessments were used to estimate the quality of life-years gained as a result of launching a regional neonatal intensive care service; the quality of life is particularly relevant since not all the infants survive with good health.

However, in our example of care of the elderly mentally infirm, the cost benefit approach might be used in a less formal manner, to explore the preference of different carers (and where possible, patients too!) for various costed options.

*Your District Medical Officer argues that the District should adopt a more co-ordinated approach in its screening for antenatal detection of neural-tube defects (such as spina bifida). This may involve some equipment purchases (such as ultrasound machines) but, more importantly, requires that medical time be devoted to the follow-up of screening test results.*

The economic evaluation of preventive measures, such as population screening programmes, raises a number of methodological challenges. This particular issue, screening of the unborn foetus in an attempt to prevent the birth of handicapped children by termination of the pregnancy, was first investigated from an economic perspective in the UK by Hagard, Carter and Milne (1976). They compared the costs of mounting a screening programme, consisting of serum alpha-fetoprotein (AFP) screening (a blood test) plus follow-up of 'positives' by ultrasonography and amniocentesis (extraction and analysis of the amniotic fluid), with the extra costs of caring for handicapped children should these be born. These extra costs, over and above those of caring for a normal child, included medical care costs in immediate hospital treatment and in long-term care, the costs of special education and loss of income through handicapped survivors, or those caring for them, not being able to remain in employment. Hagard, *et al.*, considered that 'on economic grounds, screening may be worthwhile only in populations in which the incidence of spina bifida is high'. There were, however, a number of uncertainties regarding many of the

key variables, such as the accuracy of the screening tests and the compliance of patients and health care professionals with the programme. To these could be added the ethical controversies surrounding the termination of pregnancies.

The debate surrounding serum AFP screening continues, and services have developed in different ways in different locations. Whether or not a pregnant woman is offered the test may depend on her GP, or where she is booked for antenatal care. There is also geographical variation in the availability of resources for accurately establishing gestational age (which helps in interpretation of serum AFP level) and for adequate follow-up of screening test results (essential to interpretation of serum AFP level). In a recent paper, Roberts, *et al.*, (1983) argue that 'the establishment of a regional or national programme on grounds of clinical efficacy alone may be premature. The decision would seem to hinge principally around a careful consideration of the economic issues'. Therefore, how could the approach outlined in this folio be used in tackling this problem?

As in the previous examples, clarification of the options is important. Is it the intention to offer AFP screening to all

pregnant women? Will a publicity campaign be mounted to make pregnant women and the relevant health care professionals more aware of the service? Will ultra-sonography be used to establish gestational age accurately? What level of serum AFP will be chosen as that delineating a 'positive' test result? What kind of follow-up, in terms of confirmatory tests and counselling, will be made available to those with positive test results?

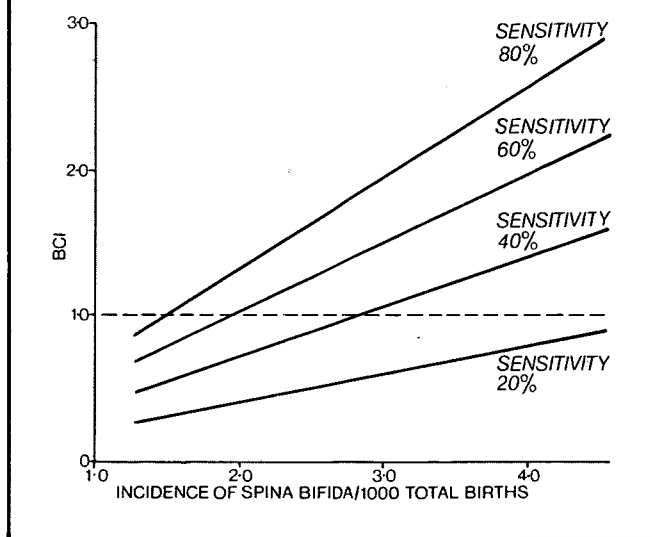
As Glass (1979) has pointed out, the decisions made about most of these factors are likely to affect both costs and benefits of the screening policy. For example, a publicity campaign may increase the proportion of the population being screened but would also require extra resources. Economists argue that one should judge component parts of the programme in terms of their **marginal** costs and benefits. That is, the relevant question to ask is 'what would be the **extra** benefits and **extra** costs of adding the publicity campaign to the basic programme?' (The same logic applies to the addition to the programme of ultrasonography to check gestational age in women who are 'unsure of dates'.) A general feature of most screening programmes is that the marginal costs rise as attempts are made to detect a higher and higher percentage of all cases of disease in the population.

Another important area for economic analysis would be to clarify the nature of the economic costs and benefits from the screening policy. The analysis by Hagard, *et al.* concentrated on the resource component of the economic benefits; that is, the extra resource costs of mounting the programme and the resource savings it would generate. However, there are other less tangible benefits from a screening policy, such as the value to parents of avoiding the birth of a handicapped child and the reassurance from a true negative test. (This broader notion of economic benefit has been explored in a more recent study by Henderson (1982)). On the other hand there may be intangible costs, such as the anxiety generated by false positive serum AFP tests. (All screening tests have false positives and false negatives associated with them. The relative proportions of these are in part determined by the 'cut-off' level chosen to delineate a positive test result. A level that is conservative with respect to false negatives—that is, indicating absence of disease when it is indeed present—will yield a high number of false positives, and *vice versa*.)

In addition there are other intangible costs, such as the unpleasantness and risk associated with some of the confirmatory tests (especially amniocentesis) and the 'psychic' costs associated with abortion. It may be extremely difficult to obtain money estimates of the intangible costs and benefits, but they should certainly enter into one's 'cost benefit thinking' about the problem, alongside the resource items. It is certainly incorrect to view economic evaluation as being concerned only with those items that can be readily expressed in money terms, although one contribution of economic analysis would be to estimate whether or not the investment in serum AFP screening would result in extra health service resources being consumed.

Finally, another contribution of analysis would be estimate the impact, on costs and benefits, of variability in a number of key factors. For this issue these might include the incidence of neural tube defects in the population, the accuracy of the screening test,

Figure 4. Variation of benefit-cost index showing screening test sensitivity and incidence of spina bifida.  
(Source: Hagard, *et al.*, 1976).



and the compliance with the programme (including termination of affected foetuses). Many economic evaluations incorporate such an approach, known as **sensitivity analysis**, where factors thought to be important to the study result are systematically varied in order to quantify their impact. (The approach can also be used to estimate the impact on results of various value judgements, such as the choice of discount rate).

A sensitivity analysis can be used to pinpoint those factors which require further study, since their accurate estimation is crucial to the balance of costs and benefits. Hagard, *et al.* showed how the ratio of benefits to costs (as defined in their study) might vary with incidence of spina bifida in the population and the sensitivity of the screening test (one measure of the ability of the test to detect disease if indeed it is present). (See Figure 4.) Later work on this topic (Roberts, *et al.*, 1983) has provided estimates of test sensitivity and a number of other factors about which earlier studies made assumptions, such as the compliance rate.

Apart from pinpointing areas for further study, sensitivity analysis can provide more information to decision-makers about the impact of factors for which costs or ethical considerations prohibit further study. It can also provide a basis for questioning some of the assumptions in a proposal for service developments. (Often those making proposals are likely to err on the side of being optimistic rather than pessimistic. It is up to those appraising such proposals to consider the impact on costs and benefits of things not turning out quite as was planned.)

This final case study, in common with many choices in health care, highlights the importance of judgement on the part of decision-makers even when more analysis is carried out. Economic evaluations are intended to be aids to the thought process, not substitutes for it.

## Some guidelines in the wilderness

The examples discussed above should have given some insights into the practice and potential of the cost benefit approach. In this final section the methodological principles embodied in the approach are consolidated in the form of a series of questions health service decision-makers might ask of proposals that come before them. (By implication they might also serve as guidelines for those who put together proposals.)



## Consideration of alternatives

What is the main justification for the proposed service; what would be the consequence of doing nothing at all?

Does the proposal contain an explicit comparison of alternative treatments or programmes, or is the implicit alternative the existing service provision?

If a completely new alternative treatment or programme is proposed:

- is it adequately described?
- why was this particular option chosen?
- were other options rejected, if so why?

## Assessment of cost and benefits

In evaluating the proposed service against alternatives, what range of costs is considered? Does this include:

- capital as well as revenue costs?
- costs other than those resulting in money expenditure (e.g. the opportunity cost of space denied other uses)?
- costs outside the immediate department where the service will be provided?
- costs on parties other than the Health Authority (e.g. patients, local authorities)?

What is known about the effectiveness of the health treatments or programmes discussed in the proposal?

- have these been evaluated by a randomized controlled trial or similar method?
- are there plans to monitor the effectiveness of any new procedures; if so, how?

Would costs and benefits be substantially different if the proposed service provision were of a different scale? That is:

- if the service provision could be larger, what would be added and what would be the additional costs and benefits?
- if the service provision had to be smaller, what aspects of it would be cut and what would be the reductions in costs and benefits?

Is it claimed that the proposal will be partly self-funding, in that savings will be generated? If so, what specific actions need to be taken to realise such savings (e.g. closing hospital wards or NHS institutions) and what are the likely resource costs associated with taking these actions?

## Other important issues

Does the proposal acknowledge any differences in the timing of costs and benefits between the alternatives assessed? If so, how is this dealt with in the proposal?

What are the main sources of uncertainty surrounding the proposal (e.g. in the effectiveness of new medical procedures, or in expected revenue costs or savings)?

- what happens to costs and benefits if the analysis is re-worked using more pessimistic assumptions?
- what could be done, perhaps at a slight increase in cost, to reduce uncertainty (e.g. by additional information gathering)?

The above checklist does not comprise a comprehensive range of questions. It is the intention that the questions themselves, and the responses they solicit, will suggest further questions pertinent to the evaluation of choices in the use of health service resources.

## REFERENCES

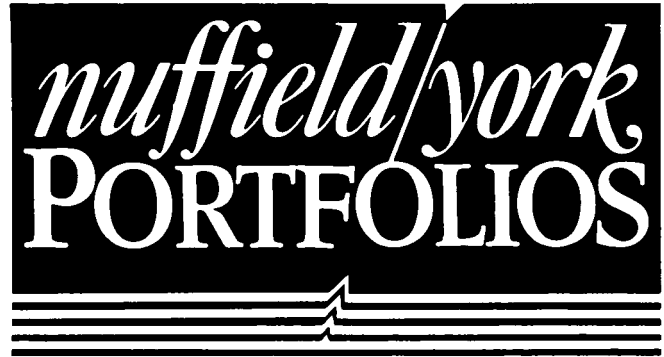
- Banta, H. D., Behney, C. J., and Willems, J. S. (1981) *Toward Rational Technology in Medicine*. New York: Springer and Co.
- Bartlett, J. R., Neil-Dwyer, G., Banham, J. M. M. and Cruickshank, D. G. (1978) 'Evaluating cost-effectiveness of diagnostic equipment: the brain scanner case; *Brit. Med. J.* 2, 815-20.
- Boyle, M. H., Torrance, G. W., Sinclair, J. C., and Horwood, S. P. (1983). 'Economic evaluation of neonatal intensive care of very-low-birth-weight infants; *New Eng. J. Med.* 308, 1330-7.
- Culyer, A. J., and Maynard, A. K. (1981) 'Cost-effectiveness of duodenal ulcer treatment; *Soc. Sci. & Med.* 15c, 3-11.
- Department of Health and Social Security (1981) *Health services management: Health building procedures*. HN(81)30. London: DHSS.
- Drummond, M. F. (1980) *Principles of Economic Appraisal in Health Care*. Oxford: Oxford Medical Publications.
- Drummond, M. F. (1981) *Studies in Economic Appraisal in Health Care*. Oxford: Oxford Medical Publications.
- Fenton, F. R., Tessier, L., Contandriopoulos, A-P., Nguyer, H., and Struening, E. L. (1982) 'A comparative trial of home and hospital psychiatric treatment: financial costs; *Can. J. Psych.* 27(3), 177-87.
- Fordyce, J. D., Mooney, G. H., and Russell, E. M. (1981) 'Economic analysis in health care. 2. An application to care of the elderly; *Hlth. Bull.* 39, 29-38.
- Henderson, J. B. (1982) 'An economic appraisal of the benefits of screening for open spina bifida; *Soc. Sci. & Med.* 16, 545-60.
- Forster, D. P., and Frost, C. E. B. (1982) 'Cost-effectiveness study of out-patient physiotherapy after medial meniscectomy; *Brit. Med. J.* 1, 485-7.
- Hagard, S., Carter, F., and Milne, R. G. (1976) 'Screening for spina bifida cystica: a cost-benefit analysis; *Brit. J. Prev. & Soc. Med.* 30, 40-53.
- Glass, N. (1979) 'Economic consequences for the public sector of a spina bifida screening programme', in Social Services and Employment Sub-Committee of the Expenditure Committee. *Minutes of evidence*, Appendix C. London: HMSO.
- Hull, R. D., Hirsh, J., Sackett, D. L., and Stoddart, G. L. (1982) Cost effectiveness of primary and secondary prevention of fatal pulmonary embolism in high-risk surgical patients. *Can. Med. Ass. J.* 127, 990-5.
- Lowson, K. V., Drummond, M. F., and Bishop, J. M. (1981) 'Costing new services: long-term domiciliary oxygen therapy; *Lancet*, i, 1146-9.
- NHS Management Inquiry (1983) *Report* (leader of Inquiry: R. Griffiths). London: DHSS, October, 1983.
- Roberts, C. J., Elder, G. H., Laurence, K. M., Woodhead, J. S., et al. (1983) 'The efficacy of a serum screening service for neural-tube defects: the South Wales experience; *Lancet*, i, 1315-18.
- Russell, I. T., Devlin, H. B., Fell, M., Glass, N. J., and Newell, D. J. (1977) 'Day-case surgery for hernias and haemorrhoids: a clinical, social and economic evaluation; *Lancet*, i, 844-7.
- Thomson, J. L. G. (1977) 'Cost-effectiveness of an EMI brain scanner: a review of 2 year experience; *Health Trends*, 9, 16-19.
- Waller, J., Adler, M., Creese, A., and Thorne, S. (1978) *Early discharge from hospital for patients with hernia or varicose veins*. London: DHSS, HMSO.
- Warner, K. E., and Luce, B. R. (1982) *Cost-benefit and cost-effectiveness analysis in Health care*. Ann Arbor: Health Administration Press.
- Weisbrod, B. A., Test, M. A., and Stein, L. I. (1980) 'Alternative to mental hospital treatment: economic cost-benefit analysis; *Arch. Gen. Psych.* 37, 400-5.
- Wright, K. G., Cairns, J. A., and Snell, M. C. (1981). *Costing Care*. Social Services Monographs: Research in Practice. Sheffield: University of Sheffield Joint Unit for Social Services Research.



# EDITORIAL

Folio 7. ■

Editor: A. J. Culyer



MARTIN KNAPP

*Lecturer in Economics and Senior Research Fellow*

*Personal Social Services Research Unit*

*University of Kent at Canterbury*

A.J.C.

---

# CHILDREN IN CARE

## *Planning without costs*

---

The House of Commons Social Services Committee Report *Children in Care* (1984) drew attention to two very durable features of policy-making in the personal social services: the influence of costs on both policy and practice and the dearth of reliable cost data. Despite taking a great deal of oral and written evidence on the costs of care, however, the Report, like so many others, ultimately relied upon poor quality data or completely ignored the cost implications of the legal, organisational and professional changes that it recommended.

Such uncritical reliance on poor quality cost indicators—or the total neglect of costs—is disappointing, surprising, and dangerous. Surprise and disappointment immediately follow from the observation that policy decisions with such an important cost dimension are rarely informed by reliable cost information. Many of the recent changes in the direction and balance of child care, for example the greater emphasis on prevention and the moves away from residential provision, were themselves strongly influenced by cost considerations. The dangers of policy-making in a cost vacuum are many and varied. They include the over-zealous pursuit of foster care without adequate provision for the additional support services and the higher boarding out fees that

will inevitably be needed, the neglect of the knock-on effects of decisions by one agency, department, or individual on others, and accusations of inefficiency by elected members and auditors in contexts where compensation and not castigation is more appropriate.

This folio examines the implications of present trends in child care provision. In particular, what are the likely costs of child care going to be in the near future? How adequate is present information for decision making? If it is inadequate, in what ways can it be supplemented to the benefit of children in care? In what ways should (reliable) cost information be integrated into the policy making process?

*Table 1. Children in Care 1971-1982*

Year	1971	1977	1980	1981	1982
Number in care (000)	87	102	100	97	93
Number in care per 1000 population under 18	6.4	7.6	7.8	7.7	7.5

Source: DHSS reports over the period.

*Table 2. Manner of Accommodation of Children in Care, 1977-82*

Year	1977	1980	1981	1982
Boarded out (%)	33.6	36.8	38.7	41.5
Local authority homes (%)	33.3	30.2	28.7	26.8
Voluntary sector homes (%)	4.2	3.8	3.4	2.9
Under charge of parent, guardian, relative or friend (%)	17.8	18.5	18.8	18.5
Other accommodation (%) *	11.1	10.6	10.3	10.3

\* Includes lodgings, residential employment, and boarding and special schools, homes and hostels for the handicapped, and youth treatment centres.

Source: DHSS (1984), table 5.

## *The child care population*

There are just over 90,000 children in the care of local authority social services departments in England and Wales. This represents approximately 1 child in 130. Numbers in care have fallen by 10 per cent from their 1977 peak, mainly due to the fall in child population rather than the fall in the 'in care' rate (Table 1). Although the numbers in care change only slightly from year to year there is in fact a high rate of turnover in the child care population. 38,000 children enter care each year, about one third of them by order of the court following care proceedings (to protect them from neglect and abuse) or criminal and matrimonial proceedings. The way in which present (DHSS) statistics are compiled and tabulated does not make it possible to calculate the lengths of time spent in care. These can range from a few hours to 18 years, but the majority of children are in care for only a few weeks at most.

One in five of the children in care are living with their natural parents, a guardian, relative or friend. Of the others, slightly more than half are boarded out with foster parents. The remainder live in a variety of residential establishments. Figures for the last few years reveal the marked trend away from residential provision in favour of boarding out (Table 2). These country-wide averages hide a number of marked and interesting variations between local authorities. Warwickshire, for example, boards out over two-thirds of the children in its care in contrast to some of the London boroughs which board out less than one-third. All local authorities have nevertheless raised their boarding out proportions in the last five years.

## *The changing balance of care*

Over a longer period, and especially the last fifteen years, the balance of child care has altered quite dramatically. There has been an increased reluctance to take children into care. There has also been an increased reluctance to keep children in care without achieving some form of 'permanent' placement. Related to these two developments have been three others: a new move away from reliance on *residential care* as the central pillar of local authority provision; the development of *community-based alternatives* to custodial sentences for young offenders, and a re-assessment in many authorities of the *respective roles of the voluntary and private sectors*. Between them these trends have given rise to new forms and arrangements of child care and the range of options now facing a social worker when making a placement decision is, at least in principle, considerably wider than ever before. Among the new services to develop out of these changes of attitude and philosophy have been the variety of special family placement schemes, professional fostering, family centres, increasingly specialised forms of residential provision, and intermediate treatment.

Each of these policy changes and each of these services has its attendant resource implications. Indeed, many of them have developed specifically because they seemed to offer local authorities and other agencies substantial cost savings. In other cases the potential for expenditure reductions was used as a convenient justification for altering the balance of provision in accordance with what were felt to be 'professional principles', which in most cases meant an assumption of improved well-being for child and family. What is far from clear, however, is whether the resource implications of these and other changes in the balance and arrangement of child care were and are fully understood. The evidence to date strongly suggests not. Indeed, as is argued below, the accumulating evidence also suggests that some of the beliefs underlying these policy changes were themselves wrong.

## *What's wrong with the present data*

There are two regular annual sources of cost information for child care services. The Chartered Institute of

*Table 3. Partial or Direct Costs of Child Care 1981-82*

Manner of accommodation	Expenditure (£m)	Average cost per child per week (£)
Boarded out	58.6	29.52
Local authority community homes *	276.7	201.01
Registered voluntary homes	21.3	160.51
Other accommodation **	13.7	49.07
Total	370.3	98.11

\* Includes some specialised residential resources (community homes with education, observation and assessment centres, adolescent units) which are generally considerably more expensive than the majority of local authority community homes.

\*\* See footnote to table 2.

Source: DHSS (1984), table A5.

Public Finance and Accountancy (CIPFA) produces detailed costings for all English and Welsh authorities for a range of services, and the DHSS produces summary figures (based on a different set of collections) which are laid before Parliament. The latest DHSS figures are reproduced in Table 3. They suggest that the cost per child week in local authority community homes is almost seven times as large as the cost of boarding out, and that the cost of residential care in the voluntary sector is noticeably lower than the cost in the public sector. Unfortunately, these cost data, upon which reliance has been placed, are inadequate and potentially highly misleading for policy purposes.

The easiest way to demonstrate this is to look at the Short Report (House of Commons Social Services Committee, 1984), and particularly to its minutes of evidence. In their examination of witnesses from the DHSS, members of the Committee raised a number of questions about costs. These, and the answers offered by DHSS representatives, are described (but not quoted *verbatim*) in Box 1. What is clear from the answers is that the information currently available can shed little or no light on most of these issues. Yet, these are precisely the kinds of issues which are frequently raised by elected representatives and faced by social services professionals in local authorities.

## **Box 1. Cost issues raised during the examination of DHSS witnesses, during November 1982, by the House of Commons Social Services Committee**

### **Why and to what extent do different local authorities pay different boarding out rates? What effect does this have on recruitment? (Q86)**

*Answer:* Some of the variation, the DHSS witnesses argued, is due to different 'types' of foster parent—those providing 'loving affection' for children, often very short term, and those 'professional' foster parents remunerated specifically for the job of caring for children with particular problems. There was no mention of the effects on boarding out payments of constraints on the supply of foster parents, or different levels and types of demand by local authorities.

### **What is the cost differential between residential care and boarding out services? (Q88, 89)**

*Answer:* The cost of residential care was cited as £9000 per child per year compared with £1300 for boarding out, but this 'does not include the costs of the placements and support services'. The witnesses agreed with the Committee that the differential is massive, but emphasised in a later supplementary note that more accurate estimates of the relative costs are needed.

### **What incentive should and does this cost differential have for the relative expansion of boarding out services? (Q89-91)**

*Answer:* If matching is done properly foster care is better than residential. The cost differential provides an incentive to expand foster care and it has certainly accelerated its development in recent years. However, 'Ministers... would not want to see fostering developed at a rate which is much faster simply because it saves money'.

### **What are the 'hidden' costs of child care—particularly residential care and boarding out—not covered by the published statistics? (Q88, 103-107)**

*Answer:* The hidden costs include field social workers, other support staff, overhead administration and office staff. Local authorities vary widely in recording administration costs. There is no accurate information on these.

### **Why has the real cost of both residential care and boarding out risen in recent years? (Q93)**

*Answer:* The available statistical data cannot provide an answer. There are complications in introducing a constant price series to deflate costs for different years to a common base. 'It is our impression, looking at the children as we go round the homes, that they are more difficult and disturbed. We do not collect statistics that quantify their degree of difficultness; it is quite hard to do'.

### **What are the cost differences between voluntary and local authority residential homes, and what accounts for them? (Q95-102)**

*Answer:* 'We can only speculate as we have no survey information'. The lower voluntary home costs could be due to: exclusion of capital costs from their figures, different child characteristics, differences in efficiency, lower salaries and higher occupancy rates. However, the fundamental problem in making a valid comparison of the relative costs of local authority and voluntary homes lies in deriving the costs per child for similar types of establishments providing similar types of care to children with similar difficulties'.

### **What effect does the bed occupancy rate have on the cost of residential care? (Q92)**

*Answer:* Occupancy rates are levelling out now. The DHSS witnesses also pointed to a need for research and local assessment of the reasons for cost variations between homes.

### **How successful has intermediate treatment been? Is it not only a better but also a cheaper way of dealing with children with problems? (Q65-69)**

*Answer:* It has successfully attracted resources. There has been no proper research on its success in terms of the effects on children. 'Our impression is that certainly it is at least as good as the alternatives of detention centre and borstal. It is certainly cheaper than they are. We happen to think it is better.'

The cost issues that arose during the examination of witnesses by the Social Services Committee, and other related issues, illustrate how the presently available cost statistics—and the ways in which they are employed—are deficient in a number of respects.

- First, and perhaps foremost, many policy choices are made without reference to costs.
- Second, the available figures include only the obvious or direct costs of care services. For some services and some children, the *hidden costs* are far greater than these direct costs.
- The figures do not properly compare like with like. Children presently placed in local authority and other residential homes differ in a number of ways from children boarded out (not least in the 'difficulties' they display), and these differences need to be taken into account.
- The quoted cost measures are simply averages and therefore gloss over marked variations between children, between establishments, between sectors, and between areas of the country.
- In addition to these deficiencies, the statistics are often employed without reference to the outcomes of care. While this is often viewed as a different issue, it can be dangerously misleading to separate cost from outcome questions—for there could be no better recipe for the indiscriminate and inefficient pruning of services.
- Finally, and as a direct corollary, the presently available information provides only a 'snapshot' picture of child care, and does not allow consideration of the important dynamic element in care and planning.

---

## Cost vacuum

The introduction of cost information into the planning process has been the subject of much criticism. Child care services, it has been argued, are the preserve of the social worker or the student of social policy, and should not be the testing ground for economic theories or cannon fodder for central government fiscal policies. Costs, in short, are held to be anathema to social care. The denigration, rather than applause, which greets new attempts to impose a degree of 'cost effectiveness' upon care policies is a common feature of current trends. This denigration stems in part from a feeling that services as indubitably and inherently 'desirable' as

child care should be above the vicissitudes of national economic welfare. Nobody would deny the importance of child care services, but it would be dangerous and foolish to argue that they are beyond economic analysis. Few activities are costless. Resources are scarce and allocating them in one way immediately implies the rejection of an alternative allocation. The present economic climate has heightened our awareness of scarcity. If we want to make the best use of scarce resources, if we want to deploy available child care services in such a way as to maximise their effectiveness or to distribute them in accordance with agreed criteria of fairness or justice, then we need to take a long and careful look at the cost implications of policy plans (Knapp, 1984).

---

## Hidden costs

### *Residential and foster placements*

The figures in Table 3 are simply the direct costs of care borne by social services departments. They are the expenditures directly attributable to residential homes (including fees paid to voluntary and private homes for the accommodation of children in the care of local authorities) and boarding out allowances paid to foster parents. They ignore the costs associated with field social work support, the recruitment and support of foster parents, the range of nonresidential services received by children in residential or foster homes (psychiatry, intermediate treatment, child guidance,

playgroups, occupational and speech therapy, and so on), and the variety of administrative and overhead costs. Many of these costs will fall to agencies other than the social services department.

Easily the most important of the hidden cost elements for the majority of children in care is the time spent by the child's field social worker. One recent study looked at the field social work input for children in residential care. This input averaged three hours per child per week, and included visits to the residential home and the child's natural family, contact and consultation with agencies and individuals with other responsibilities for the child, travelling time, and desk-based administration. Around this average there were marked variations, reflecting differences in the needs of children, characteristics of the homes, the extent to which other

**Table 4. Direct and Hidden Costs of Boarding Out**

Length of care period (days)	Direct placement cost (£)	Social worker cost (£)	All other costs (£)	Cost per child per week (£)	Hidden cost as percentage of direct cost %
0-7	30	88	20	138	360
8-30	31	29	16	76	145
31-91	27	12	8	48	74
92-365	27	12	7	47	70
Total	29	37	14	80	176

Source: Knapp, Bryson and Lewis (1984), table 12.

agencies were involved and the distance between the social worker's office and the residential placement. At the time of the study (June 1983) this field work input to residential care was costing at least £25 per child per week. Similar cost calculations were made for other hidden elements of care; these included services provided by the social services department, the education department, the National Health Service and the Probation Service. Together, these hidden cost elements were equivalent to approximately 13 per cent of the directly observed residential costs. It should be noted that these calculations are all for children who have been in their residential placements for at least three months, and these children generally impose lower demands on the non-residential sector (Knapp, 1984a).

The hidden cost element for children boarded out is generally larger than for children in residential placements. Field social worker involvement is greater, and a larger number of other social services department staff and other agencies will be involved. The only data currently available on these hidden costs of boarding out are not immediately comparable to the figures on the hidden costs of residential care given above, for the former refer to a cohort of children recently admitted into care, and who are not therefore as settled as the residential group. The hidden costs of boarding out range from a very small proportion of the direct costs (the payments to foster parents) to a large multiple (Table 4). Hidden costs are on average 176 per cent of

the direct costs, and can be as much as 360 per cent for those children in care for only a few days. Not included in these costs of family placements are the costs of recruiting, preparing, training, and matching foster parents with children in care, nor the costs of foster placement breakdown or foster parent turnover. These can often be considerable. Nor do they include hidden foster family costs. The boarding-out payment may or may not accurately reflect the amount that foster families require if they are to continue offering homes to children in care. If the payment is greater than the amount required then there exists a potential for cost savings by local authorities. It is more likely, however, that the boarding-out payment is below the adequate perceived compensation required by foster families. This can be only speculative, but it is supported by the excess demand for family placements in virtually all areas of the country. If the boarding-out payment is 'insufficient' for some families already providing care they will eventually give up fostering. If more boarding-out is planned, the level of payment will have to rise, so that even if it is the 'right' amount currently, it must understate the amount required if all children whom it is intended to board out are actually to be boarded out.

### *Intermediate treatment and custody*

A similar problem of hidden cost elements arises in the comparison of services for juvenile offenders. There is a strong commitment to community-based programmes, most of them labelled *intermediate treatment* (IT), as alternatives to custody in detention centres or youth treatment centres (the former borstals), and in some cases as alternatives to residential placements. Central government encouragement for the growth of IT has recently taken the form of short-term subsidies, but local authority providers cannot ignore the longer term and hidden cost elements. A study of one IT programme in Kent found that the directly observed costs were £83 per child week for IT and £161 for detention centres, but that the adjustment for anomalies and for hidden elements of cost altered these figures substantially (Knapp, 1984b). The costs were adjusted to cover only equivalent activities (for example, the detention centre

### **Box 2. Average cost per child week (£) for intermediate treatment and detention centres, 1980/81**

Cost Definition	IT	DC
Directly observed ('accounting') costs	82.81	160.93
adjusted for anomalies and to include costs to direct provider (IT centre or detention centre) previously excluded	96.53	154.14
adjusted to include hidden costs borne by social services department, probation service and education department	115.64	154.14
adjusted to include family costs and value of lost employment for those in custody	121.31	162.75
Final cost as percentage of directly observed cost	146%	101%
Average length of sentence*	110 days	47 days
Average cost per child per sentence	1906.30	1092.75

Source: Knapp (1984b)

\* The *Criminal Justice Act*, 1982 has since reduced the average length of a detention centre sentence.

costs included an administrative element for which there was no available data equivalent for IT) and to include items, such as vehicle usage and staff accommodation in the prison service, which were under valued or ignored in the accounts. This, however, is not enough. Further adjustment for the costs of supervision of IT youngsters during the day, their social workers' time, and the special tutors employed by the education department provides a more accurate indication of total public sector costs. To these can be added the costs borne by families (which are removed, in part, during periods of custody) and the value of the lost output of those youngsters sentenced to detention centres.

Together these adjustments alter the cost per child week to £121 for IT and £163 for detention (Box 2). There is thus a considerable narrowing of the cost differential as a result of these adjustments. Furthermore, the cost per sentence is *higher* for intermediate treatment than for detention. This contradicts the view

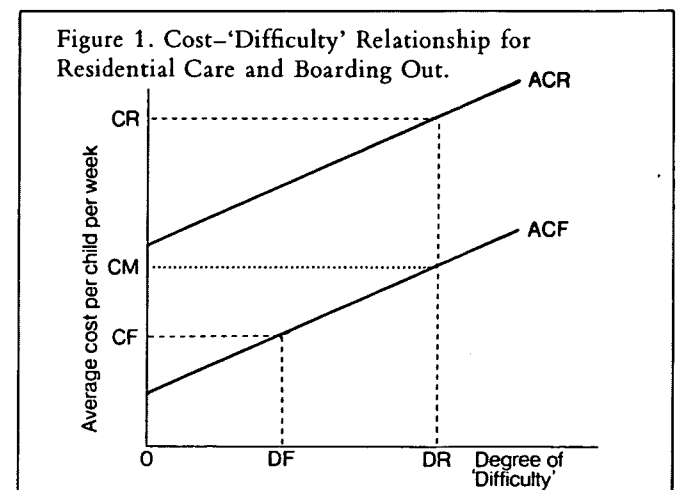
of the DHSS representatives questioned by the House of Commons Social Services Committee (1984, volume II, question 65). It is important to separate two issues here. First, the cost per child day for IT is much higher than accounting figures for the service presently indicate. This needs to be recognised in the planning of this expanding form of provision. Second, the cost per child day is lower for IT than for detention and the cost per sentence (excluding the 'post-sentence' supervision period) is higher. This reversal of the cost ranking is the result of the very different lengths of time spent in the two alternatives, a difference which has recently widened with the implementation of the *Criminal Justice Act, 1982* (reducing detention centre sentences to 21 days). This clearly begs questions as to the relative effectiveness of different sentence lengths and of IT as an alternative to custody. The available evidence in Britain, unfortunately, remains equivocal (Rutter and Giller, 1983).

## Fair comparisons

Many comparisons of the costs of residential care and boarding out fall into the trap of assuming that like is being compared with like. However, it is quite clear that the characteristics of children accommodated in residential homes are different from the characteristics of children boarded out. Children in residential care are typically more difficult than children boarded out. The meaning of the term 'more difficult' is not easy to specify exactly, but it is clear that children with physical and mental handicaps, with delinquent tendencies or backgrounds, older children and those with certain emotional or behavioural characteristics (aggression, hyperactivity, propensity to self mutilation, and so on) are much less likely to be boarded out than to be in residential care. Furthermore, the average cost of care will be positively associated with the 'degree of difficulty' displayed by children because of the greater demands that they make on staff and on the families in which they board. If both of these arguments are true—that 'more difficult' children are likely to be found in residential care and, other things being equal, generate higher costs—then the resource implications of trying to board out a higher proportion of children will be very different from those suggested by a simple comparison of present average cost figures.

These two arguments can be summarized with the help of Figure 1 which demonstrates some of the cost implications of a shifting placement pattern. The two lines drawn in the figure represent the average costs of care in a residential setting (ACR) and a foster home (ACF) for each level of 'difficulty'. Each line is drawn to represent average cost as an increasing function of difficulty. For simplicity it has been assumed that the cost of residential care is greater than the cost of foster care at all levels of difficulty. The 'average degrees of difficulty' of children currently accommodated in residential and foster homes are denoted as DR and DF respectively. Associated with them will be the average costs. It is these two average cost figures (denoted CR and CF) that are quoted by the DHSS and by local authorities, although of course without including the hidden cost elements. If a child of difficulty DR (the average amongst the residential population) is successfully boarded out, the cost to the authority will not be the present average boarding out cost (CF), but will be the larger amount CM. The savings to a local authority

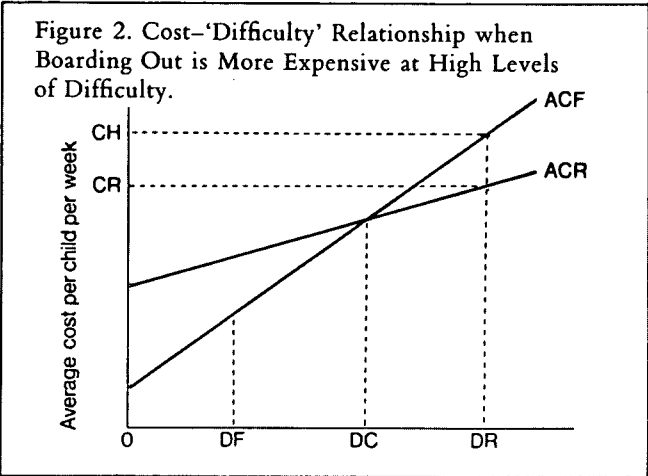
of boarding out a larger proportion of children in care, therefore, will be exaggerated by the simple (mean) average cost figures prepared for children currently in care (even if these costs include the omitted services discussed earlier). Whenever a child of difficulty between DF and DR is moved from a residential to a boarding out placement, there will be an increase in the (mean) average cost of *both* services (although a fall in total care expenditure).



The assumption that residential care is always more expensive than boarding out can be relaxed without altering the thrust of the argument. Indeed, it strengthens it. Assume that above a certain level of difficulty (DC in Figure 2) the average cost of residential care exceeds that of boarding out. This might be because in order to achieve a certain degree of 'success', foster placements need a great deal of social worker and peer-group support, because boarding-out payments have to be so much higher to attract foster parents, because recruitment costs may escalate, or because children placed outside the authority impose high 'visiting costs' upon social services departments. In this case, moving 'very difficult' children from residential to family settings will actually raise average cost from CR to CH for these children and also raise the total cost of the child care system.

These two figures can explain in large measure the real cost inflation experienced by both residential and boarding out services in the last few years. As an increasingly large proportion of children in the care of

the local authorities has been boarded out (rising from 33 per cent in 1976/77 to 44 per cent today), so the average difficulty of these children has gone up, and so has the cost (see Table 5). Social services departments have had to pay higher boarding out allowances and provide a greater degree of support from fostering officers and field social workers. This means higher direct and hidden costs of foster care and higher recruitment and training costs. The alternative is a higher breakdown rate. Residential care is thus left with the more difficult children, raising its own average cost. Residential care will become more conspicuous to elected members in budget-conscious authorities and residential staff will face a harder residential task at the same time. This will in turn mean a higher rate of staff turnover from an already volatile sector.



This argument is not suggesting that the relative expansion of boarding out services is in any sense 'wrong'. What it suggests is that the comparative cost figures presently available exaggerate the savings to be achieved by family placements. The danger is not that too many children will be boarded out, on the contrary, but that local authorities' attempts to move towards higher boarding out proportions will be under-funded. This would follow from the under estimation of both the higher boarding out allowances and the fieldwork support that will be needed. The implications in terms of higher breakdown rates are obvious.

Table 5. Cost Inflation and the Changing Balance of Care					
31st March	Financial Year	Boarded Out Proportion*	Cost Index**	Residential Care*** Proportion*	Cost Index**
1977		33.6		37.5	
	1977/78		100		100
1978		34.5		36.0	
	1978/79		112		108
1979		35.9		34.6	
	1979/80		113		114
1980		36.8		34.0	
	1980/81		115		123
1981		38.7		32.1	
	1981/82		125		129
1982		41.5		29.7	

\* Proportion of all children in care  
 \*\* 1977/78=100  
 \*\*\* Includes children in local authority and registered voluntary homes.

Source: Computed from DHSS (1984), tables 5 and 11.

## Misleading averages

The need to compare like with like, or more generally to qualify or standardize cost variations for differences in the characteristics of children and other factors, is not confined to the comparison of residential and boarding-out services. One common error is to assume that the public, voluntary, and private sectors are sufficiently alike to enable a direct comparison between costs or charges across sectors. Statistics collected in a recent national survey of children's homes reveal that the average cost per child week during 1982/83 in local authority 'ordinary' community homes was £169, compared with the cost of £125 in the voluntary sector, and a weekly charge of £115 in the private sector (Knapp and Smith, 1985). It would be quite wrong to assume that such cost differences reflect differences in the efficiency with which the three sectors achieve their child care objectives. Other information collected in the survey reveals marked differences between homes in different sectors which may account in part, if not perhaps in full, for these costs differences. These include differences in scale, the percentage of places occupied, the use of self-contained living units, transport facilities and differences in the characteristics of children (age, length of time in home, previous residence, handicap and behavioural difficulties displayed).

Even within sectors, however, it is necessary to avoid making misleading comparisons between (say) homes and between local authorities. There was much variability around the average public sector cost of £169 per child week, and it is important to understand why this variability occurs and (where appropriate) to act upon it. Variations between homes in the average cost of care and, by extension, variations between authorities may be attributed to a number of influences. For example, one might expect larger homes to enjoy economies of scale. Lower unit costs are more likely to be found in homes with a higher percentage of available places occupied throughout the year. Homes accommodating more difficult or disturbed children are more likely to need higher staffing levels and will therefore be more expensive. London authorities face a different labour market from authorities elsewhere and this will raise the price they have to pay for residential staff. Among the other factors that can be expected to raise the unit costs of residential care are the provision of day care services, the provision of education on the premises and the employment of residential staff for peripatetic care tasks. None of these cost raising influences reflects an 'inefficiency' in the provision of care services by individual homes, although clearly some of them are within the control of local authorities (at least in the long run) and therefore could reflect on the organisational or planning efficiency of authorities. The important point, however, is not to make comparisons between bare cost figures without first considering the possible reasons for differences between them. Cost-raising factors beyond the control of local authorities or homes should clearly be put to one side before pronouncing on the relative efficiency of homes or different care providers. On the other hand, cost-raising factors beyond their control do have immediate ramifications for the pursuit of efficiency, and understanding their influences will help policy makers plan the resource implications of any policy changes they may be contemplating.

# How to take account of costs

The role of cost information in the formulation of child care policy must be neither exaggerated nor devalued. The sensible integration of cost information into the

decision-making process requires improvement in present data and greater awareness of the utility and the limitations of particular information.

Elements in this greater awareness should include the following:

- The indirect or 'hidden' costs of child care are often larger than the directly observed costs. For children boarded out these hidden costs can be as much as four times as great as the boarding out payment. Ignoring them might easily mean the under-funding of any policy of foster care expansion. A full commitment is needed to the expansion of support services and the necessary increases in boarding out payments.
- The current trend towards a higher boarding out proportion will induce real cost inflation in both the residential and the boarding out sectors. This is not an inefficiency, rather the necessary consequence of boarding out 'more difficult' children.
- There is a similar tendency to 'under-cost' intermediate treatment programmes. On a *per diem* basis IT is cheaper than either residential care or custody, but it can be more expensive per sentence or period in care. There is a widespread belief that IT is the better care option for many youngsters, but there are still no research answers to the question of cost effectiveness.
- Cost comparisons between services, between local authorities and between sectors are complicated by marked and pervasive differences in the services offered and the characteristics of children receiving them. If like cannot be directly compared with like, then explicit account needs to be taken of those differences between care units or agencies which impinge upon cost.
- Some of the observed differences in the cost of providing apparently identical services will be attributable to differences in the efficiency with which social care agencies carry out their responsibilities. However, there are a number of factors, some of them beyond the immediate or even long term control of agencies, which will be partly responsible for the observed variations. Care agencies should not be castigated for inefficiency if the cost-raising factors are beyond their control.

## REFERENCES

- DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1984). *Children in Care in England and Wales March 1982*. (London: HMSO).
- HOUSE OF COMMONS SOCIAL SERVICES COMMITTEE (1984). *Children in Care*, HCP 360 (Session 1983-84) (London: HMSO).
- KNAPP, M. R. J. (1984). *The Economics of Social Care*. (London: Macmillan).
- KNAPP, M. R. J. (1984a). 'Non-residential services for children in residential care: the Mid Glamorgan study', PSSRU Discussion Paper 321. University of Kent at Canterbury.
- KNAPP, M. R. J. (1984b). 'The relative costs of intermediate treatment and custody', *Home Office Research Bulletin*, 18. (London: Home Office).
- KNAPP, M. R. J., BRYSON, D., AND LEWIS, J. (1984). 'The comprehensive costing of child care: the Suffolk cohort study', PSSRU Discussion Paper 355. University of Kent at Canterbury.
- KNAPP, M. R. J. AND SMITH, J. (1985). 'The costs of residential child care: explaining variations in the public sector', *Policy and Politics*, 13(2).
- RUTTER, M. AND GILLER, H. (1983). *Juvenile Delinquency: Trends and Perspectives*. (Harmondsworth: Penguin).



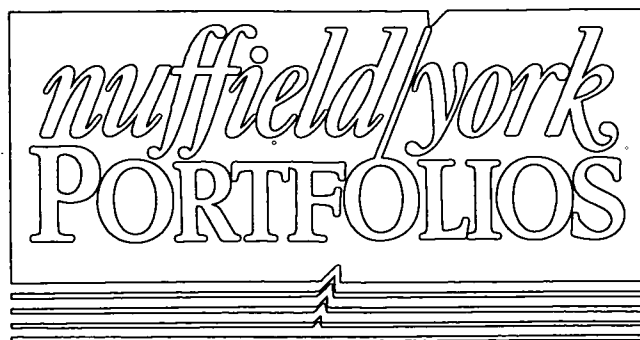
# EDITORIAL

Folio 8. ■

Editor: A. J. Culyer

In financial year 1984/5 some £724m has been allocated to NHS capital spending. Is it too much, too little, or just right? In large part, the answer depends on political judgements about the benefits expected and the opportunities forgone. But whatever the level, the sum allocated has to be spent wisely—not just in terms of public accountability but also in recognition of the fact that today's capital spending will determine the pattern of services for many years to come—witness our inheritance of the Victorian era. This folio recounts the recent development of option appraisal in the NHS as applied to capital schemes, offering suggestions as to how best to conduct such appraisals and exploring the implications for DHSS, Regions, and Districts.

A.J.C.



**R. L. AKEHURST**

*Senior Fellow at the University of York*

AND

**M. J. BUXTON**

*Senior Research Fellow at Brunel University*

## A guide to **OPTION APPRAISAL** in the NHS

Capital spending in the Health Service is an investment with long-term consequences (witness our inheritance of Victorian buildings and 'hatted' facilities) so it is not surprising that there have been central government pressures to subject NHS hospital developments to the sort of appraisal required for investments elsewhere in the public sector and commonly used in the private sector. A series of working groups have produced reports recommending and elaborating appraisal processes. In 1979 a joint DHSS and Treasury review of capital spending (1) recommended the adoption of *Investment Appraisal* to bring the NHS into line with other parts of the public sector. At about the same time the joint NHS/DHSS Standing Group on Planning identified a gap in the planning system between the identification of broad strategic priorities and detailed decisions about particular capital developments. A procedure called *Planning Appraisal* (2) was proposed. Following the change of Government in 1979, a Rayner style review was set up to examine the process by which capital spending decisions were made. The conclusion of the small team was that the decision-making process needed improvement and it recommended a procedure entitled *Option Appraisal* (3). The team stressed the need to generate and examine systematically a range of options for development rather than simply assessing one or two options in great detail.

## BOX 1

### Parliament and Option Appraisal

'This is the third successive year in which Committees of Public Accounts have thought it necessary to report on defects in control of hospital building. . . . The Department told Your Committee that the projects under examination this year had gone sadly wrong, and some others had gone wrong too.'

[*Third Report from the Committee of Public Accounts, Session 1974-75, HC 374, HMSO.*]

'We consider that whenever a publicly financed project has gone badly wrong, as the Liverpool Teaching Hospital project has done, the reasons should be carefully examined and responsibility for deficiencies allocated, and that appropriate action should be taken whenever the facts warrant it.'

[*Ninth Report from the Committee of Public Accounts, Session 1976-77, HC 532, HMSO.*]

'...during the last year we have concentrated in the Department on action emphasised by the Committee to ensure that the NHS as a whole has benefitted from the lessons of the Liverpool experience. . . . Lastly . . . we have completed a thorough-going review during the year of the use of capital in the NHS and have circulated this to health authorities. In brief, the review has been concerned with getting better value for money from capital investment. What we are after is a much more rigorous approach to project appraisal, focussing on identifying and costing alternative ways of achieving the service objectives and on a much more systematic comparison of the full costs, both capital and revenue, and the service benefits of the available options.'

[Permanent Secretary of the DHSS in evidence to the PAC, in *Eleventh Report from the Committee of Public Accounts, Session 1979-80, HC 498, HMSO.*]

'...the position simply is that a hospital has been built and it is not possible to open it properly because the money is not there to do it and this is the case in eight other hospitals to a lesser degree.'

[MP taking evidence at the PAC in *Seventeenth Report of the Committee of Public Accounts, Session 1981-82, HC 375, HMSO.*]

'Examination in the two regional health authorities studied suggested that, despite the reference to investment appraisal in the CAPRICODE procedures for control of capital projects, appraisals were not in practice a regular part of financial control. Nor did there appear to be any monitoring of the use made of investment appraisal at area, regional or DHSS level. . . . new procedures will require health authorities to prepare an explicit appraisal of alternative building options and place a more definite responsibility on health authorities to evaluate future running costs as part of the planning of new projects.'

[Memorandum from the Comptroller and Auditor General in *Twenty-eighth Report from the Committee of Public Accounts, Session 1981-82, HC 417, HMSO.*]

Apart from differences of emphasis, the recommendations of these three groups were conceptually very similar. They offer frameworks within which to apply the general cost-benefit approach to major capital developments in the health service. Unlike the previous proposals, however, Option Appraisal was made *mandatory* and it was intended to be integrated fully into a revision of the formal Capricode procedures governing major capital developments. Health Notice HN(81)30 required appraisal for all submissions to the DHSS for permission to spend over £5m capital. This requirement was considerably broadened when HN(82)34 made it clear that *all* capital developments, including those not requiring Departmental approval, were to be subject to the same procedures. It is the responsibility of the Department's auditors to see that this is so.

In addition to these pressures to encourage better decision-

making, there has also been an important impetus for clearer public accountability stemming from the investigations of the Public Accounts Committee into a number of capital planning decisions which had seemed with hindsight to make little sense (4). There was particular concern at the totally inadequate documentation of the reasons why decisions had been made, so that subsequently it was impossible to judge whether they had been sensible in the light of the assumptions that were reasonable at the time. To check whether the planning 'disasters' that came to their attention were the exceptions or the tip of an iceberg the Committee asked its executive arm (the Exchequer and Audit Department) to look at routine decisions made in a couple of NHS Regions. The conclusion of the investigating team was that a lack of record and absence of systematic appraisal were normal.

Thus, allied concerns for better decisions and more account-

## BOX 2

1. Identify the service objectives of development.
2. Identify a wide range of options that might be expected to achieve the service objectives ('the long list').
3. Apply criteria consistently so as to reduce the number of options to be considered in detail to a manageable size (the 'short list').
4. Analyse the cost and benefits of each option in the 'short list'.
5. Examine the sensitivity of the results.
6. Examine the robustness of the results.
7. Present the results of the appraisal so as to highlight the nature of the decision to be made.

When appropriate, LOOP BACK, AMEND and REITERATE earlier steps.

ability lie behind the current requirement for Option Appraisal of developments. Some officers and Authorities have greeted this advent with horror; others have seen it as a threat; some are wholly uncomprehending. In this folio we shall try to show that Option Appraisal, properly understood and applied, is a means of liberating the imagination, of encouraging multi-disciplinary cooperation, of promoting efficient and sensitive planning, and of making the basis of decisions less opaque to fellow professionals, authority members, CHCs and the public at large.

## WHAT APPRAISAL ENTAILS

HN(81)30 and various guidance documents set out the basic elements of an appraisal. Whilst they do not put forward frameworks that are identical in every detail, their broad thrust is similar. The basic framework is set out in Box 2.

The sequential ordering of the various stages in Box 2 oversimplifies the dynamics of the practical process. The stages become rather blurred in practice and discoveries made in later stages may lead to a reassessment of judgements made earlier. For example detailed work on a particular set of options may lead to the realization that a further option exists, combining the characteristics and advantages of several of the existing options. Nonetheless it is important to keep a clear view of the *logical* sequence in which issues fall so that in looking back and reconsidering issues that have already been provisionally settled, the overall logical structure is maintained.

## IDENTIFYING OBJECTIVES

This crucial starting point is one of the weakest parts of current appraisals. Officers find it difficult to define objectives in terms of the appropriate *service for patients*—rather than in terms of particular *inputs*. Thus the objectives need to be expressed in terms of the desired nature, quality, and pattern of care. When officers can be persuaded to define their objectives in such terms, new alternatives for solving problems that deserve formal consideration are often identified. For example, defining objectives in terms of total care for the elderly may lead to

non institutional options, while defining them in terms of correcting deficiencies in geriatric accommodation will not.

In part, these difficulties reflect weaknesses elsewhere in the planning process. In particular, strategic objectives frequently have been presented in terms of a concrete solution rather than a less constrictive and more general statement. Thus in some strategic plans a too narrowly specified proposal makes it hard to identify the basic objectives in ways that assist subsequent, more careful, consideration.

This problem should diminish as the overall quality of appraisal continues to improve and lessons learnt filter up to the strategic level, so that the one provides a sensible context for the other.

## IDENTIFYING OPTIONS

One clear lesson learnt from early attempts at appraisal is the importance of identifying an appropriate set of options to be considered in detail. This is a major difficulty in the complex political reality of service planning. Official guidance material starts largely at the stage at which a short list of options has somehow emerged. But *getting to that stage* is at least as important and as difficult as anything which follows.

Successful generation of options requires a willingness to take steps to break away from whatever ideas may exist already about the best solution. Some sort of '*brainstorming*' process is required, together with a willingness to reassess options to be included in the light of ideas that emerge subsequently. A careful definition of objectives nearly always leads, and use of imagination may well lead, to the identification of at least one previously unconsidered option. Similarly, when options are being discussed, important clues may be given to underlying but hitherto unarticulated objectives, which may then need to be incorporated.

Additionally, it is essential to make a point of including, at least at the early stages, options which are strongly favoured by influential groups such as CHCs, relevant consultants, and groups of authority members, or options to which there is some existing commitment. Even if such options are regarded as unsuitable or infeasible they should be included if only to demonstrate clearly and fairly that this is the case. Otherwise, the appraisal is left open to the criticism that important options have been omitted and the danger arises of future delay when eventually they do have to be considered formally.

## MAKING A SHORT LIST

The preceding process will usually produce a far larger number of options than can be considered in detail. A satisfactory process for narrowing down to a short-list is therefore essential. There are several criteria that can help 'deselect' options at this early stage.

1. The application of a carefully considered (and justified) set of binding constraints which options must satisfy may rule out many. For example, site configuration or available building space or policies adopted by Region or DHSS may rule out some options. More commonly, constraints reflect a concern that some minimum standard of attainment should be achieved, for example, percentages of geriatric beds located with general acute beds, or maximum travelling times for particular patients. The main point to bear in mind is that using constraints in this way allows no possibility of trade-offs between the achievement of objectives.

2. Some options may be identified as being typical of a sub-set so that only one representative option need be appraised initially. Only if it becomes the preferred option need the detailed variants be considered. For example, acute rationalizations usually permit a large number of possible permutations of use of existing hospital sites. Many permutations will be quite similar to one another and representative options may first be chosen for detailed appraisal.

3. A selection process which goes through a preliminary and coarse cost-benefit comparison may also help. This differs from the detailed appraisal to be shortly discussed only in terms of detail, refinement and time input. Thus a list of say twelve significant options might be assessed in terms of very broad estimates of their costs and benefits and the most promising appraised in greater depth. Questions of trade-offs between objectives can be avoided at this earlier stage by use of the notion of 'dominance' and eliminating only those options clearly inferior on some grounds, and no better on any others.

At present, the process by which the number of options considered is reduced is often neither explicit nor consistent. Certain considerations are applied to rule out one option (which is disliked for some unexpressed reason) but the same considerations are not mentioned in the context of the favoured options.

**Consultation.** The short-listing process constrains the range of likely final recommendations and at this early stage some consultation is desirable. It is sensible to put the issues to the relevant Authority at the stage at which objectives have been defined and options for detailed appraisal short-listed. Such consultation has two major advantages. First, it involves Authority members, allowing them to express their views at a time when they still may contribute to the process, and it thus helps to ensure that members will not at some late stage introduce a new objective or ask for a new option to be considered in detail. Secondly, it formalizes the status of the appraisal, making it easier to approach relevant outside officials, for example the District Valuer or a Local Authority Personal Social Services Department.

Discussion in public is not without its own problems. It may fuel fears about closures or changes of use that eventually prove groundless; it may raise expectations about new developments that are subsequently dashed; indeed it may introduce an unnecessary degree of uncertainty amongst staff about their future; it may affect negotiations about land/property acquisitions or disposals. All of this, in turn, may lengthen the decision-making process.

These anxieties may recede as option appraisal becomes more familiar and the nature of the exercise more widely understood. But, even if this were not the case, strong arguments for an open process of consultation remain. Option appraisal is not merely an internal technical exercise, but a political process that should reflect society's values and priorities.

## ASSESSING COSTS AND BENEFITS

Official guidance material referring to appraisal contains ambiguities about the range of costs and benefits that should be considered. There is even doubt about which items should be treated as costs and which as benefits. HN(81)30 recommends that all effects (positive or negative) that can be measured in money terms should be grouped together as costs, while important effects which cannot be so measured should be grouped together as 'social benefits'. This is not a very helpful division. First, lumping together entities preceded by pound signs may be misleading if some pounds are regarded as more important than others. For example, a Health Authority might regard costs falling on its own budget as more important than travel costs to people visiting hospital patients. Second, and much more fundamentally, such a grouping ignores the crucial effect of budgets and the role that costs can have in indicating to a health authority what it has to give up in order to develop a particular service. Authorities have to *choose* between developments and the costs to the authority determine which combination can be undertaken. In other words the costs indicate what must be forgone elsewhere in the authority. Which projects *should* be chosen is a matter of both cost and benefit and this judgement is best made with all those considerations that do not constitute authority costs grouped together.

**Costing.** There is considerable variability in the ease with which officers undertaking appraisals are able to obtain appropriate costing information. This reflects the widely differing attitudes and capabilities of the professional officers and their staffs who provide data. Anyone undertaking an option appraisal is heavily reliant on a variety of professional inputs for cost information.

Capital costs cause problems, for example, because judgements about the relative cost of different sites may be required at an early stage before detailed information is available from which to accurately estimate 'on-costs'. In such circumstances works professionals are often reluctant to provide even rough estimates. Nevertheless, such judgements have to be made and it is better that they are made by the relevant professional. They will be better made if he is reasonably conversant with what is going on at that stage of an appraisal. Moreover, as professionals become more attuned to the appraisal approach, they become more confident in entering its spirit.

There are genuine difficulties in estimating costs in options involving refurbishment and upgrading. It may be impossible to make even rough guesses at costs without detailed work on site. A common response to this problem is to estimate the cost of providing new accommodation on the basis of Departmental Cost Allowances (DCAs) and then to use the Needleman Formula to calculate the maximum acceptable upgrading cost. This is done on the grounds that were the maximum upgrading cost exceeded we would build new anyway. The approach is potentially misleading. Actual upgrading costs may in reality be well above the maximum shown as 'allowable' on the

### BOX 3

## Glossary of some of the Official Terminology

**APPROVAL IN PRINCIPLE (AIP)**—the new first stage approval introduced by the DHSS (in HN(81)30) for major capital schemes. This approval is to be sought at an early stage in the planning process and should include an explicit option appraisal to show which option is preferred, and why.

**CAPRICODE**—the Code for handling NHS Capital Projects. It consists of a number of Health Procedure Notes governing the planning and processing of individual health building developments. It was first consolidated into a single code in 1967 and last revised in 1974.

**DEPARTMENTAL COST ALLOWANCES**—a set of cost guides for hospital buildings based on the functional content. These regularly updated guides effectively determine the basic permissible capital cost for a development. 'On-costs' relating to the costs imposed by the communications and services between departments, external work and drainage, and reflecting the particular site conditions are then added.

**NEEDLEMAN FORMULA**—for estimating the maximum proportion of the full new building cost that should be spent on a shorter-life upgrading. This formula is based solely on relative capital costs.

**REVENUE CONSEQUENCES OF CAPITAL SCHEMES (RCCS)**—an estimate of the running costs of a development. It particularly refers to the estimating process, based on the '45 Hospital Formula' once used by the DHSS to estimate an appropriate allowance for the running costs of a major new development when these were funded directly. The formula was based on the actual running costs of a sample of 45 existing hospitals—mainly full-scale district general hospitals. Although no longer used or calculated by the DHSS, an updated version is produced by Shropshire Health Authority.

**EXETER FORMULA**—a more sophisticated basis for estimating revenue costs. This regression based formula was developed at Exeter University based on the cost returns of all large acute or partly acute hospitals.

Needleman formula yet the latter figure might tip the balance and commit the Authority to an option that relies on upgrading. At the very least such figures should be presented with a very clear description of their derivation and an indication of their likely margin of error.

Over a period of time the costs of running the service usually far exceed the initial capital costs. Despite this, the thought and effort being put into estimating differences in revenue costs between options is disappointing. Too often calculations are unquestioningly based on a standard calculation, such as the old DHSS forty-five hospital RCCS formula as updated by Shropshire Health Authority (6) or the Exeter formula (7), with little thought to their appropriateness to the specific local circumstances. These formulae are derived from observation of the running costs of a national sample of hospitals with a particular size range. They were not, for example, designed for the purpose of estimating the costs of relatively small additions to existing hospitals or of building small community units. Sometimes it is simply assumed that additional facilities will cost the same as those already existing locally. Frequently, the assumption is that, as all options deliver the same level of service, they will all cost the same. This is despite common and good reasons for assuming otherwise. For example, split site working, different configurations of facilities, or varying availability of key staff, may each lead to differences between options that offer the same overall capacity. The error may be compounded if the assumptions built into the benefit estimation about the level of service to be provided differ from those

used for simplicity's sake on the cost side. This can all too easily occur if the finance staff responsible for revenue costs are not fully involved in the whole appraisal process. Again, integration and involvement are important components of good appraisal practice.

**Benefits.** Identifying differences in non-cost aspects between options is crucial. Otherwise the exercise degenerates into simple cost-minimization. Because of the shortage of genuine outcome measures, benefit estimation tends to throw up a hotch-potch of measures and professional judgements which somehow have to be combined together. It is worthwhile considering some rules for ensuring that this part of the appraisal is done as systematically as possible and that absence of ideal information does not serve as an excuse for either ignoring the benefit side altogether or giving it no more than a token nod.

The first step in identifying appropriate benefit measures is to use the set of objectives laid down at the first stage of the appraisal. In this way a broad strategic objective such as 'improve quality and quantity of services to the elderly and maintain them as far as possible in their own homes' might lead to a variety of criteria for comparison. Among them might be number of inpatient beds, number of day places, expected 'throughput', the quality of the hospital environment provided (measured for example by space standards and suitability for purpose) and the number of elderly who could continue to live at home with appropriate support.

The major requirement for systematic benefit analysis is the devising of measures to indicate the degree of achievement of the stated service criteria expected to be attained by each option. Space standards, or travelling distances, or numbers of beds exemplify benefits that have 'natural' units. Unfortunately some important criteria have no such natural units of measurement: no one has yet produced a standard scale for 'quality of medical care'. For these aspects, professional judgements about the relative performance of options must suffice, and some explanation of the basis of these judgements needs to be recorded.

In addition to those criteria which follow directly from the objectives there may be others which are relevant. For example, relative ease of staffing is a consideration of importance in comparing options even though it forms no part of the main objective. It is, therefore, valuable to consider explicitly whether there are valid criteria for comparison in addition to those that follow directly from the stated service objectives.

The most disturbing aspect of benefit estimation in the Health Service is that no one seems to regard it as their particular responsibility. It is extremely difficult to obtain advice on, for example, quality of care of patients in different options that is truly disinterested—that is independent of professional temptations to show themselves in the best possible light. It is disappointing that Community Physicians have proved so reluctant to contribute actively to the development of this crucial territory. Until this void is filled, it will continue to be difficult to judge the recurring issue of how much extra it is worth paying for a 'medically superior' option.

Once all the relevant and practically obtainable benefits have been measured in their own units or presented in terms of professional judgements, they need to be brought together so that they can be compared with one another and collectively be compared with the costs of the options. A common approach, successful and illuminating when well used, is to give each option a score on each of the benefit dimensions (say from 1 to 10) and to weight the dimensions one against the other, thereby indicating their relative importance, to yield an aggregate benefit measure. Appraisal teams have found this process challenging and illuminating and, if well presented, it is of considerable value in making clear to those receiving appraisals just what is at issue.

However, the job must be done carefully. All the *main* dimensions of benefit must be included, their definitions must avoid double-counting, the interaction of weighting and scoring must be carefully examined to ensure that implied rates of trade-off are as intended and, as with so many other aspects of appraisal, the sensitivity of the analysis to changes in weights and scores must be established. Technique must be seen as an aid to clear thinking—not a substitute for it.

Benefit measurement is an element of appraisal that raises issues of consultation in its widest sense. Values and weights may reflect trade-offs between, for example, accessibility and quality of care, staff and patient convenience, or quality of care and political constraints. These are not issues to be decided solely by Health Service professionals let alone small caucuses of them. There is a need for the active involvement of a variety of interested groups including, of course, Authority members and CHCs.

---

## SENSITIVITY AND ROBUSTNESS

---

HN(81)30 uses the term 'uncertainty' to encompass both of these concepts. By 'sensitivity' is meant the degree of change brought about in the relative assessment of options by changes

in the assumptions and judgements that have been made in the appraisal in estimating the balance of costs and benefits. Relevant questions are whether changes within a plausible range of error of cost estimation, or small changes in the relative ranking of particular benefits, might alter the choice. This requires that all doubtful estimates be presented as a range, say 'low' 'expected' and 'high'. Thorough sensitivity analysis is rarely done, despite its inclusion in HN(81)30 as one of the stages. This is unfortunate because sensitivity analysis is helpful in a number of ways. First, it indicates where additional thought and measurement effort should be directed. If a choice is particularly sensitive, say, to the relative importance of accessibility for patients and visitors and the number of beds provided, then it is appropriate that these issues should be discussed at greater length than otherwise. The obverse of this is that time should not be wasted on those issues not affecting the choice. Second, if between approval and the starting of a project further information arises concerning some of the assumptions made, a well conducted sensitivity analysis will indicate, without the need for extensive reworking of large parts of the exercise, whether the difference is enough to change the choice of preferred option and, if so, which one will move into 'first place'.

Analysis of robustness is similar to sensitivity analysis except that it concerns itself with judgements external to the appraisal and the subsequent service development, rather than internal to it. It concerns questions like: is the choice particularly sensitive to assumptions about population or caseload? What would happen if the birthrate changed suddenly? What if a particular medical technology were to change significantly? What if priorities for resources for particular care groups were to change? Would the building being proposed be of value for any other group if priorities do change?

---

## PRESENTATION

---

Sometimes analysis reveals a clearly superior option: an option that is both the most beneficial and the cheapest and whose superiority is both insensitive and robust. More commonly the appraisal will not indicate an unambiguous choice: often the more beneficial options will also be more expensive, or the benefit ranking will be highly sensitive to the weights put on different benefit criteria, or the preferred option will not be at all robust. In such cases the final choice will be a matter of judgement, and the results of the appraisal need to be presented in such a way that the nature of the remaining judgements is highlighted. The crucial choices and trade-offs involved should be explicit. Few current appraisals do this adequately, nor do they demonstrate clearly why one option in the end is preferred over its rivals. Appraisal is not a substitute for judgement—but it should clearly indicate the important judgements that have to be made.

---

## APPRAISAL IMPOSES OBLIGATIONS

---

Even though the process is mandatory, the *spirit* of option appraisal will survive only if it can be seen to be a genuinely useful technique. The spirit of appraisal must therefore not be at odds with the environment within which it is used. Its rational principles must not be devalued by arbitrariness. Its logical consequences must not be frustrated by unnecessary constraints in the NHS 'management system'. It imposes obligations on everyone involved in NHS decision-making.

## AT DHSS

The promises of speedier handling of major submissions and the concentration on major issues of policy that appraisals raise need to be kept. The more complete information thus provided should not be used as a reason for going into details about the way local judgements were made. At the moment DHSS progress in approving submissions is rather slow relative to the aspirations expressed when HN(81)30 was published, but this is part of a learning process on both sides as appraisals get referred back to Regions for more information (8).

Appraisal also obliges the Department to desist from imposing a variety of 'arbitrary' norms or maxima on Authorities. These may be cost norms such as Departmental Cost Allowances (DCAs) or norms of provision measured in terms of capacity or throughput relative to a population. While there is a need for rules of thumb at both District and higher levels, strict insistence on such norms or maxima imposes unnecessary additional constraints when a choice is being properly appraised. Provided the case for a development is clearly made, the Department should relax such constraints when an appraisal justifies it in particular local circumstances. DHSS should also publicly make it clear that it is willing to do so. Good appraisals will also make ministerial intervention on purely political grounds more difficult by indicating, in a public document, the costs of going for inferior solutions to a local problem. If one of the major claims for the cost-benefit approach is that it helps clear thinking, another is that it makes life harder for those who prefer behind-the-scenes manoeuvring.

A further problem that still needs to be alleviated by a clear lead from the DHSS, with Treasury backing, concerns the allocation of capital and revenue into separate budgets. The formal requirements in the appraisal process are that both capital and revenue costs should be brought together and discounted. The rationale for this depends on certain assumptions which are inconsistent with the position in most Regions. In particular these include the assumption that capital and revenue can be traded or exchanged, that there is no accounting distinction between capital and revenue sums occurring in any one year, and that current (capital) funds can be bought at the cost of future revenue and vice versa. Although the position is gradually changing, the scope for exchange between the two expenditure heads of capital and revenue is seen to be very limited from the perspective of those within the NHS at District level. The Department needs to make clear its willingness to facilitate interchange between the two budgets: otherwise the consideration of the present value of capital and revenue together is irrelevant.

## AT REGION

Similar obligations are implied at Regional level: abstaining from the imposition of arbitrary norms, avoiding interference in local detail rather than simply exercising regional responsibility for financial allocation and showing willingness to make budgetary allocations as flexible and rational as possible. In too many Regions, capital is still a 'free' good obtained at the expense of other Districts rather than at the expense of revenue. Indeed, in Regions which still pay some or all of the revenue consequences of capital spending, a successful capital bid will increase future revenue availability rather than reduce it. The appraisal approach will not be used as intended if the allocation procedures give no incentive to a District to try honestly and openly to minimise the combined revenue and capital costs of a scheme.

Finally Regions have a duty to reconsider the nature of their role in the appraisal of projects, so that Districts see it as a

process to help *them* decide how to meet *their own problems* and to accept responsibility for the agreed solution. Appraisal of capital schemes needs to be seen as a District activity, albeit an activity for which Region has a supply of certain professional skills and over which Region must play a monitoring, overseeing, and advisory role.

## AT DISTRICT

Appraisal has to be taken as an integral part of the decision-making process at District. It must be seen as leading to the taking of a decision, rather than, as is still often the case, a procedure that is made to justify a decision *already taken*.

Appraisal must be opened up to consultation from a range of skills and viewpoints, not least from Authority members and the CHC, so that it ceases to be seen as a merely technical exercise or a mysterious part of the professional officers' alchemy and becomes openly regarded as a way of focusing on important value judgements and classifying issues of wide interest to members of the local community.

---

## A FUTURE FOR APPRAISAL

---

Some Health Authorities have been alarmed by the implication of HN(82)34 that *all* capital schemes must be appraised in the manner prescribed in HN(81)30. Some Regions may follow the letter of that Notice and insist that Districts evaluate *in detail* all schemes, even those involving small capital sums. There is a real danger that effort will be diverted from carrying out appraisals of significant developments, in the spirit of the requirements, to creating an appearance of satisfying the letter of the requirements in every case, for the benefit of Region or the auditors. It can never have been the intention of the DHSS to insist on *detailed* appraisal of all small developments. Authorities must be given some discretion as to how they display their scarce analytical resources. The best approach might be to use the appraisal framework to structure 'back-of-the-envelope' consideration of small schemes with the process of reasoning being briefly recorded and placed on file. It is much to be hoped that Authorities—particularly larger ones at District level and certainly those at Regional level—will develop the necessary pool of resource skills so that appraisal is not only enthusiastically promulgated throughout the Service but implemented in an imaginative and competent fashion.

Unfortunately the focus of the requirements is entirely on *capital*. The developments that might benefit most from appraisal are not necessarily those involving the largest capital sums or indeed capital at all. Some schemes that are relatively capital inexpensive may have major cost implications for a local service as a whole: service developments may benefit even more from formal appraisal than those involving capital.

If appraisal is to be used intelligently and imaginatively, and not be subject to cynicism and abuse, a much wider range of people need to understand the process. Obviously those officers responsible need to fully understand the technique and to be able to carry their less well informed colleagues with them. The Chief Officers above them have to appreciate how to use appraisals appropriately to guide their decisions and recommendations. Officers lower down the system have to understand how best to provide detailed inputs to the analysis. At Region, officers have to be able to assess the appraisals of others as well as conduct their own. Authority members have to receive both intermediate and final forms of appraisal reports and CHC members have a role in effectively representing views which might otherwise not be registered. Consul-



tants, both in their roles as medical experts and service providers and in their role of potential or actual member of District or Unit management team have important roles to play.

So far most education in the art of appraisal has been aimed at the officers directly involved. Despite their importance, other groups have largely been neglected. A consequence of this is that appraisals have been inappropriately received and handled by Chief Officers and Authority members and have sometimes been misunderstood by CHCs. As a result some officers have become very frustrated because other major actors around them have not entered constructively into the appraisal process. Disillusionment and scepticism can quickly set in.

It is clear from the good examples now existing that option appraisal can be used as a forceful and illuminating technique to assist in decision-making in the NHS. There are nonetheless difficulties still to be overcome before the full potential of the

approach can be realised. In particular those practices which engender cynicism and encourage mere 'window dressing' appraisals need to be eliminated. Arbitrary constraints and politically motivated overturning of appraisal results have to be avoided. Education in appraisal roles has to be more widespread. Finally, the fiction that the process does not consume significant management and planning resources has to be exposed for what it is, particularly in the light of current attacks on health service administration as wasteful. Appraisal certainly takes time—but major issues deserve such formal attention. It is not unreasonable to spend management effort in ensuring that the Health Service resources are wisely and fruitfully spent. This naturally requires judgement about how appraisal resources are employed. But if wiser and more productive spending decisions result—and sensitive use of Option Appraisal will be a major step towards this goal—the Public Accounts Committee will have served a useful purpose in acting as the catalyst to a necessary change.

#### BOX 4

### Key Points for Good Practice

1. Identify objectives as ends rather than means.
2. 'Brainstorm' on options to break away from the shackles of past history and the limitations of imagination.
3. Apply consistently clear criteria for the reduction of a long list of options to a short list.
4. Consult interested parties early in the process.
5. Avoid unthinking use of formulæ in estimating costs.
6. Adopt measures to indicate degree of achievements of objectives and use them to compare options.
7. Relate any scoring and weighting matrix used to firm measures or descriptions of services so that the values implied can be challenged.
8. Explicitly consider questions of sensitivity and robustness.
9. Present results so that they highlight key issues in the decision.
10. Higher tiers should avoid arbitrary constraints which conflict with the spirit of appraisal.

#### REFERENCES

1. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1979). *Review of Health Capital*. DHSS.
2. STANDING GROUP ON PLANNING (1979). Draft Addition to the *Blue Book*. DHSS, mimeo.
3. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1981). *National Health Service, Review of Capricode: Building Strategy*. DHSS.
4. For examples see Box. 1.
5. See 1–3 above and DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1982). *Appraisal of Development Options in the National Health Service*. DHSS; H. M. TREASURY (1982) *Investment Appraisal in the Public Sector*. H. M. Treasury.
6. SHROPSHIRE HEALTH AUTHORITY (1984). (Financial Information Project) *RCCS Hospital Sample Standard Data*. Shropshire Health Authority.
7. BAILEY, T. C. AND ASHFORD J. R. (1984). Specialty Costs in English Hospitals—A Statistical Approach Based Upon a Costs Component Model. *Journal of the Operational Research Society* 35(3), 247–56.
8. Fourth Report from the Social Services Committee, Session 1983–84. (1984) *Public Expenditure on the Social Services*. HC 395. HMSO.



Current high levels of unemployment have led to renewed interest in the suggestion, first investigated in the 1930s, that unemployment has serious adverse effects on health. If the suggestion is correct, then it could be important for debates about macroeconomic policy, because it implies that social costs of unemployment are greater than is commonly assumed. The conclusion drawn by some researchers, that unemployment kills is obviously an emotive one and for this reason it is essential that it be examined dispassionately. The relationship between health and unemployment appears to be primarily medical or epidemiological, yet models of behaviour affecting



health, developed by health economists, and the various procedures used by economists to test these and other models, can help to evaluate the often rather confusing evidence. This paper outlines the results of some of the studies which have claimed to find a link between unemployment and health, subjects them to a critical review, and concludes that, while unemployment is a very serious problem, there is little evidence for the view that it is, in addition, a killer.

A.J.C.

# DOES UNEMPLOYMENT KILL?

**HUGH GRAVELLE**  
*Queen Mary College, University of London*

Studies of the effects of unemployment on health have been of three main types: aggregate time series, aggregate cross-section, and disaggregated (focusing on individuals). In aggregate time series analyses the relationship over time between the ill-health (usually mortality rates) of whole communities and the percentage of the population who are unemployed is examined via *multiple regression*. This is a statistical technique for summarising the linear correlation between one variable (in this case health) and a number of other variables. The advantage of the technique is that it enables one to assess the strength of the correlation between mortality and unemployment after allowing for any possible correlation between mortality and other variables. Thus the separate association between mortality and unemployment can be measured.

This approach has been used in a large number of studies by the American sociologist M. Harvey Brenner of Johns Hopkins University. Brenner argues

that mortality rates in England and Wales over 1936–1976 depend on four factors (Brenner, 1979):

(a) the level of national income, reflecting improvements in nutrition, sanitation, education, and medical technology;

(b) the unemployment rate. Since the effects of unemployment may take a number of years to emerge, mortality rates are hypothesised to depend on current and past unemployment rates;

(c) rapid economic growth, which is argued to be associated with instability and stress for particular groups in the society;

(d) government welfare expenditures, which are meant to alleviate economic distress and hence may also improve health.

TABLE 1  
*Estimates of some social costs  
of unemployment (USA 1970).*

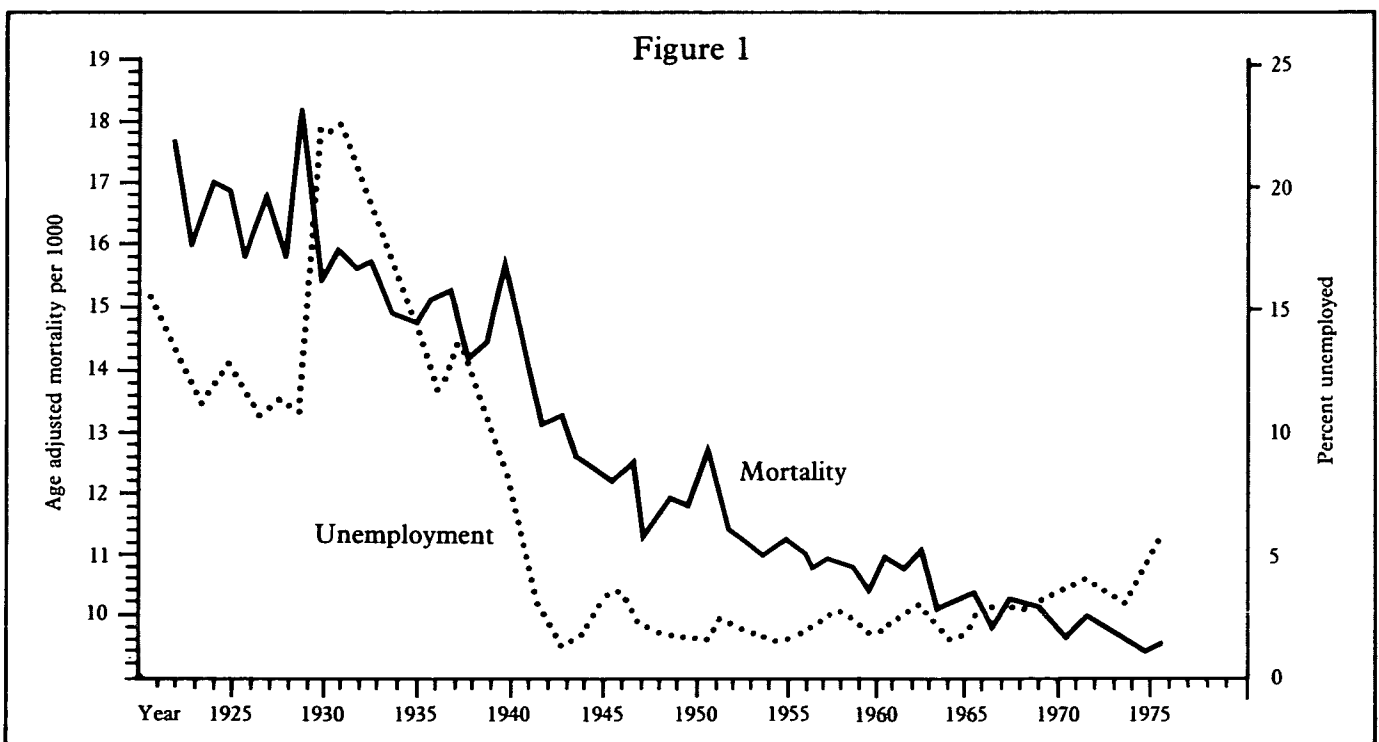
	Increase resulting from 1% higher unemployment over six-year period.
Total mortality	36,890
Cardiovascular mortality	20,240
Cirrhosis of liver mortality	495
Suicide	920
Homicide	648
State mental hospitals (first admissions)	4,227
State prison admissions	3,340

Source: Brenner (1977). Estimates based on US population in 1970.

When age adjusted and age specific mortality rates were regressed on these variables the results appeared to provide impressive support for Brenner's arguments. In addition to the expected negative relationship between income and mortality, unemployment was significantly and positively associated with mortality rates at all ages.

Similar studies have found the same statistically significant relationship between unemployment and mortality in the U.S., Australia, and Scotland and for specific causes of death (such as heart disease, suicide, homicide, and cirrhosis) (Brenner, 1971; Brenner, 1977, Brenner and Mooney, 1982; Bunn, 1979). Mortality is, of course, a rather crude measure of health, but statistical analyses using first admissions to mental hospitals and the volume of prescribing as an indication of population health yield very similar results.

However, time series studies by other researchers have often come to rather different conclusions, even though they have followed the same kind of statistical modelling and used similar data. Gravelle, Hutchinson, and Stern (1981), for example, repeated Brenner's analysis of England and Wales. They found that the association between unemployment and mortality was not statistically



significant when the time period used was extended to 1922–1976 or was split into two at 1951. One reason why this may be so can be seen from figure 1.

The age adjusted mortality rate shows a long-term, nearly linear, downward trend over the whole period 1922–1976. The dominant feature of the unemployment series over the period 1936–1976 is the very large fall in unemployment in the early 1940s (from 9.7 percent of the insured population in early 1940 to under 1 percent in 1942). Any statistical analysis which includes this fall in unemployment is almost bound to find a positive association between unemployment and mortality. The longer the period analysed the less the relative effect of the large fall in unemployment. The analysis over 1922–1976 rather than 1936–1976 shows a much reduced association. Similarly, analysis of 1951–1976 (a period over which unemployment was historically very low) shows no significant correlation at all.

McAvinchey (1984) studied the relationship between mortality at different ages for men and women and the

unemployment rates over the previous ten years in post-war Scotland. He found that for women there was a significant relationship between unemployment and mortality in eight out of twelve age groups but in only two cases was it positive. In the remaining six age groups it was *negative*: higher unemployment was associated with *lower* mortality. For men the association was positive in two age groups and negative in two, the remaining eight showing no significant correlation. Forbes and McGregor (1984) using almost identical data found that the relationship between mortality from all causes and unemployment in Scotland was rarely significant and when significant was negative. The association between ischaemic heart disease mortality and unemployment was rather more marked but was very sensitive to the length of lag on unemployment considered. IHD mortality was positively linked to unemployment rates over the previous five years but when unemployment over the previous ten years was considered the relationship became negative!

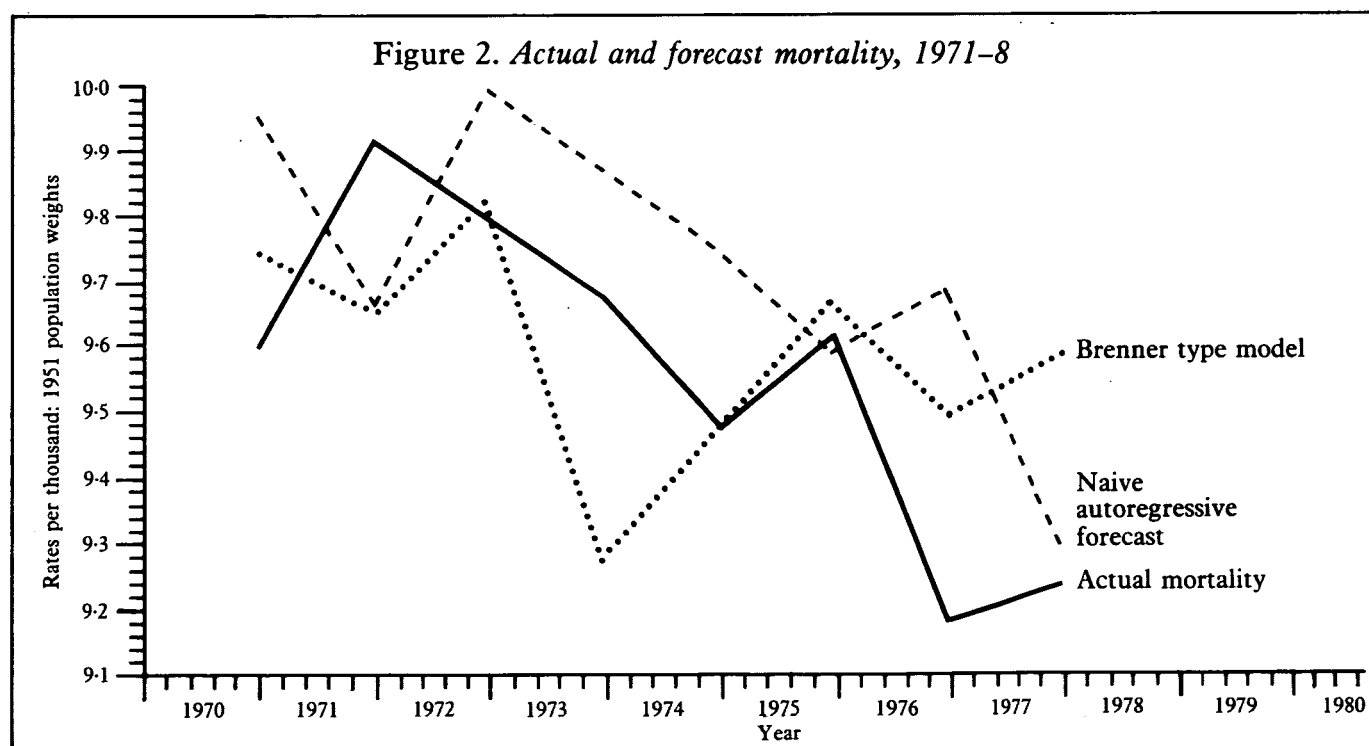
## Lies, damned lies and time series results

There are many pitfalls in the interpretation of time series. Spurious correlations between time series are frequently observed. For example, Hendry (1981) has reported a variable which performs at least as well as the money supply in 'explaining' the rate of inflation: cumulated annual rainfall. He goes on to emphasise a statistical model's *ability to forecast* as a test of its validity.

Thus Gravelle, Hutchinson, and Stern (1981) examined the forecasting performance of the Brenner model. They first estimated the model over the period 1936–1970 and obtained the usual good fit and the expected positive relationship between unemployment and mortality. Using the estimated relationship and the actual values of unemployment, income and so on for 1971 to 1978, they then forecast the mortality rates for 1971 to 1978. The model forecast badly both by reference to actual mortality

and by comparison with the simple alternative hypothesis that mortality in the current year is proportional to that in the previous year (see figure 2).

Such forecast tests require long runs of data, since the forecast tests must be made with data from outside the period over which the model is estimated. This presents two problems. First, information is often available only for relatively short periods. The Scottish investigators, for example, were restricted to 23 post war observations once account was taken of the need to include lags of up to ten years on unemployment. Given the number of variables hypothesised to affect mortality and which should, therefore, be included in the regression this can easily make forecasting tests impossible because all the data have had to be used to estimate the model. The second difficulty arises from the fact that it is unlikely that



the relationship will remain unchanged for long periods. Unless one also has a theory of how the relationship between unemployment and mortality changes over time

and can allow for this in the statistical modelling, the estimated relationships are likely to forecast badly.

---

## Correlation not cause

Even if these purely statistical problems were unimportant and the statistical association between mortality and unemployment were robust, it is not proven that mortality is determined (even in part) by unemployment: *correlation is not proof of causation*. The correlation between two variables may be purely accidental or it may reflect the influence on each of some third variable. Eyer (1977) suggested just such an explanation for the correlation of mortality and unemployment. He argued that mortality in modern industrial societies is largely a result of stress and that the aetiological evidence suggests that there is very short lag between increased stress and higher mortality rates. Mortality thus tends to increase during booms in economic activity because of stress induced by overwork and fragmentation of community through migration. Since economic cycles are fairly regular, booms (and associ-

ated higher mortality rates) tend to follow depressions and periods of higher unemployment with a lag of 2–3 years. This, he suggested, explained the correlation between current mortality rates (due to the current level of economic activity) and unemployment two or three years previously. Unemployment was conceded to be important in explaining suicide and homicide, but these account for only a small proportion of the total variation in the overall mortality rate. Eyer emphasised the importance of specific causes of death rather than overall mortality. He suggested that, for example, deaths from accidents increase as the level of output rises over the cycle because more people are employed and exposed to the risks of accidents, which are in turn heightened by the increased pressure of work.

---

## Other influences on health

Many variables affecting health have been suggested by different authors. In addition to unemployment, these include education (measured, say, by years of schooling or by number of teachers or educational expenditure), housing conditions, alcohol, tobacco, sugar, health service expenditure, doctor/population ratio, per capita income, water purity, adequacy of excreta disposal, climate, air pollution, and water hardness. Such variables must be allowed for in a study of the link between unemployment and health.

Unfortunately, it will not usually be adequate merely to include additional explanatory variables in the regression analysis along with unemployment. Many of the variables which plausibly influence health are highly correlated. People with lower incomes are usually less well educated, smoke and drink more, are more likely to be unemployed,

to live in poor housing, and so on. This means that it will be difficult to isolate the separate effects of different variables on health, since they all tend to vary together. It will not be possible to determine the effect of particular *individual* variables, only their *combined* effect. It will therefore not be possible to determine the optimum mix of policies to reduce mortality and morbidity. Although the costs of say increasing the income of the unemployed and improving housing conditions could be calculated, the separate marginal benefits of these policies in health terms cannot be estimated.

A high degree of correlation amongst the explanatory variables also means that estimated effects are not robust. Even minor changes in the data used, such as the length of period over which a regression is estimated, can lead to dramatic changes in statistical results.

---

## Sins of omission

If the explanatory variables were not correlated with each other the dangers of omitting some of them from the statistical analysis would not be too serious. The estimated size of the effect of unemployment on health would still be unbiased: that is on average the estimated effect would equal the true effect. However, the estimate would have a larger margin of error because of the omission of other explanatory variables, so that there would be a tendency to conclude the estimated effect was not statistically significant from zero even when there was a positive relationship between unemployment and mortality.

Unfortunately, when the explanatory variables are correlated the implications of omitting some of them are much more serious. To see why this is so, consider again Brenner's 1936–1976 study. As already noted, the

salient feature of the unemployment series over this period is the very large drop in unemployment in the early 1940s. Greaves and Hollingsworth (1966) suggested that diet has a major influence on health. The war years saw a significant improvement in diet, particularly for the poorer sections of the population, *as well as* a fall in unemployment. In 1936–7 expenditure on food by the poorest group of households in the National Food Survey was 65 per cent of the average food expenditure but by 1954–5 it had increased to 96% of the average. At the same time there were significant improvements in the quality of food following the introduction of free school meals and free school milk, wartime flour milling regulations and the addition of calcium carbonate to flour. Thus the level of unemployment and the adequacy of diet are strongly and inversely correlated over the 1936–1976 period. Since

there is no measure of diet in Brenner's regression equation but unemployment is included, the estimated effect of unemployment on mortality will be biased upward since dietary adequacy is correlated with the unemployment rate. The fall in unemployment in the early

1940s is, in effect, also acting as a measure of the simultaneous improvement in diet and so the estimated unemployment coefficient is biased upward by the omission of a measure of dietary adequacy.

---

## Morbidity or mortality?

Data rarely correspond with the theoretical constructs. The most obvious example of this is the use of mortality rates as the measure of the health of the population. Relative mortality from particular causes is by no means perfectly correlated with relative morbidity. It could be argued that unemployment has a significant effect on mental rather than physical health. If so, mortality rates are a very poor measure of the relevant concept of health. Unfortunately, morbidity data are not readily available for many conditions. Even when available, they may be an inaccurate measure because they usually relate to the use of health services (bed days, consultations, etc.) and thus conflate the provision of services and the health of the population. There are equally serious measurement difficulties connected with the explanatory variables. There is, for example, no obvious single measure of diet, housing conditions, or the state of development of medical technology. Moreover, even the main economic variables such as unemployment or per capita income are subject to measurement problems. Unemployment is usually measured as the number of unemployed at a particular time, i.e. as a stock. However, the same stock of unemployed individuals could be compatible with a wide variety of inflow and outflow patterns each associated with different durations of unemployment for those becoming unemployed. If the number becoming unemployed and the number finding new jobs in a period were both to increase, the number recorded as unemployed at any point in time might be unchanged but those who became unemployed would spend a shorter time unemployed. If the length of time for which an individual is unemployed affects health the use only of the *number* unemployed could be seriously misleading, since a given total number unemployed at a particular time could be associated with very different average durations of unemployment.

It is possible to include flow and duration related unemployment measures in any statistical analysis, though few investigators have yet done so, but there is a further problem with the unemployment data. The recorded numbers of unemployed in Britain are based on the numbers who register as unemployed in order to claim benefits. Thus some people who are unemployed in the sense of seeking work at prevailing wage rates may not be registered as unemployed because they are not eligible for benefit. This is particularly so for married women. Similarly, changes in benefit regulations can cause entirely spurious changes in the recorded level of unemployment.

Measurement errors in the data can take two forms: (a) the measured variable can measure the true variable with a constant error or bias; (b) the measured variable on average correctly measures the true variable but any given observation may be too large or too small. The effect the first type of measurement error is to bias the estimated effects of the variable on mortality. For example, there are two alternative unemployment series available for the inter-war period in Britain, one of which is consistently and substantially lower than the other. If the lower series is used over the 1936–1976 period then the estimated effect of unemployment on mortality is larger because the same mortality reduction is being associated with a smaller fall in unemployment in the early 1940s. Similar differences in the estimated effects of income will arise depending on whether one believes that personal disposable (after tax) per capita income or per capita gross national income is the correct measure of income.

Even with the second type of error, when the measures used are correct on average, the effects of the incorrectly measured variables on mortality will be underestimated. The investigator may mistakenly conclude that the explanatory variable has only a small effect on health.

---

## What determines health?

Perhaps the major source of problems with statistical investigations of the unemployment/health relationship is the neglect of the fact that many of the variables influencing health are affected by decisions made by individuals in their roles as workers, consumers, patients, health service providers or macroeconomic policymakers. This means that the determination of the health of an individual or of a population should be seen as part of a wider process in which health, and many of the variables affecting it, are *simultaneously* determined. Unless this is allowed for in modelling the health-unemployment relationship seriously misleading conclusions may be drawn about the factors affecting health. Four examples show how pervasive this problem can be:

(i) The causal link between unemployment and health is

unlikely to be one way—from unemployment to health. Workers with poor health may stand a higher risk of being sacked and have lower probabilities of finding work if firms believe that they have lower productivity. Thus comparisons of the health of unemployed and employed *individuals* may confound the effect of health on unemployment risk and unemployment on health. The observed correlation between poor health and unemployment will overstate the effect of unemployment on health. Aggregate studies based on the health of whole populations will not be affected by this problem. Consider for example the effect of a reduction in demand in the economy which reduces the number employed. Suppose firms sack less healthy (because less productive) workers when demand falls. If unemployment had *no* effect on the health of individuals a comparison of population or average health

before and after the increase in unemployment would show no change: the less healthy would have merely changed their employment status, but there would have been no reduction in the average health of the whole population. By contrast a study of individuals would indicate that the unemployed were less healthy and thus might lead to the mistaken conclusion that unemployment affected health.

(ii) Aggregate cross-section studies will be prone to *selection bias*. Healthier individuals tend to be more mobile so that they are more likely to move from areas of high to low unemployment seeking work.

Hence low unemployment areas will have healthier populations both because of the effect of unemployment on health and because of migration. This kind of selection bias will not affect studies of individuals or time series analyses.

(iii) Individuals' consumption of health services depends both on the state of their health and on the cost of the services to them. In the NHS most services have a zero money price but do impose costs on patients in the form of time and travel. The cost of time required to consume NHS services will depend on the alternative uses of time which are forgone. The unemployed will usually have lower time costs than those in work and hence, other things being equal, will be more likely to consume health services. This may lead to biased estimates of the effect of unemployment on health if, as in some studies, health is measured by the use of health services. For example, many studies have found that unemployment is positively related to admissions to mental health facilities (Brenner,

1977). However, when direct measures of the mental health of the population are available from surveys it has been found that they are not correlated with unemployment (Catalano, Dooley, and Jackson, 1981). Thus it would appear that increases in admissions following unemployment do not reflect worsening mental health. One explanation is that the opportunity costs of seeking treatment for existing conditions is reduced. Alternatively, unemployment may make it less easy for others to tolerate or support an individual's symptoms and lead to pressure on them to seek treatment. In either case it is important we have fuller models of behaviour for the interpretation of what is observed.

(iv) The consumption of health care services tends to increase when health deteriorates. Individual patients or consumers will increase their demand on services. Policy-makers may increase provision to compensate areas with greater health 'needs' for example via explicit resource allocation formulæ such as RAWP. Thus, not only does health service expenditure affect health but health will in part influence health service expenditure. Unless this is explicitly allowed for, the beneficial effect of health care on health will be underestimated since areas or individuals with worse health will tend to consume more health care. Indeed, in extreme cases, a negative correlation between health and health care may be observed. This can lead to incorrect inferences about the effectiveness of health care and can also bias estimates of the relationship between health and other explanatory variables if these other variables are correlated with health care.

---

## How does unemployment influence health?

Many investigators fail to be sufficiently clear about the precise hypotheses they are attempting to test. Unemployment may affect health *directly* in a variety of ways:

(a) *being* unemployed may increase morbidity and mortality by inducing stress;

(b) *becoming* unemployed may impose psychological costs of adjusting to changes in consumption patterns, use of time and social relationships;

(c) *anticipating* unemployment in the future may lead to anxiety and stress.

Unemployment may also have *indirect* health consequences because it leads to lower income levels. Thus individuals may have their health affected by having a lower income, by the psychological costs of adjusting to a lower income, or by worry about the possibility of having a lower income in the future. If it is low income rather than unemployment *per se* which affects health then these consequences can be alleviated by raising unemployment and supplementary benefits or by reducing the unemployment rate. Since these alternative policies have very different economic consequences, the particular way in which unemployment affects health has considerable policy relevance.

These different hypotheses relate to different groups of individuals and require different tests. Hypotheses (a) and (b) are concerned only with the health of the unemployed, so that attempts to test them with aggregate data on the entire population (of whom the unemployed are a small

proportion) may not be successful. Not only will the health of the population be dominated by the employed majority but explanatory variables, such as income, will be population averages and thus will not relate to the unemployed. Even if attention is restricted to hypotheses of type (c), where the unemployment rate affects the health of the employed as well as the unemployed, studies which use aggregate data may not be able to produce firm conclusions. A large number of other variables influence health and these variables do not have the same value across all members of the population. Income, housing conditions, education, occupation, duration of unemployment, and consumption patterns differ widely and may have considerable effects on health, but the average health of a population may be only very weakly related to average income, education level, and so on.

Micro-level studies of individuals appear to be the most promising way of establishing the health-unemployment relationship because the averaging effect of aggregate studies is avoided and because it is possible to obtain many more observations and thus more precise estimates. However, micro-studies may also yield inconclusive results unless based on careful specification of hypotheses, behaviour modelling, and an awareness of the many other factors which can affect health. Given the complexity of the process by which health is determined it is essential to combine cross-section and longitudinal comparisons of individuals. Unfortunately, such data collection exercises are time consuming and expensive.

# Longitudinal studies

Some existing British longitudinal studies can be used to examine the unemployment/health link but they were generally not specifically designed for this purpose and so need cautious interpretation. The DHSS cohort study (Moylan, Miller, and Davies, 1984) of a sample of men who became unemployed in the autumn of 1978 provides a comparison of men who were continuously unemployed for a year with those who found work within the year. As table 2 indicates, most men did not experience a decline in health and there was no statistically significant difference between the two groups, indicating that duration of unemployment did not appear to be related to health.

The National Training Survey (Metcalf and Nickell, 1979) interviewed 50,000 individuals in 1975-6 about their previous work experience, qualifications, education, etc. and also gathered data on sickness spells of over three months duration and unemployment spells. After allowing for a large number of other factors (including age, region, race, marital status and education), it was found that the number of years between 1965-75 in which an individual had a three month sickness spell was larger the greater the number of unemployment spells. Unfortunately, since the order of sickness and unemployment spells was not recorded, the result may also reflect the effect of sickness on unemployment. In any case the estimated effect of unemployment was rather small (a 100 per cent increase in unemployment spells was associated with a 7 per cent increase in sickness spells).

The OPCS Longitudinal Study (Fox and Goldblatt, 1982) is based on about half a million individuals, starting with their 1972 Census returns and adding routine OPCS data on them (births, deaths, cancer registrations, the 1981 Census, etc.). Since the 1971 Census records an individual's economic status, it is possible to relate 1971

TABLE 2

<i>Health problems reported at:</i>	<i>All interviewed at third interview %</i>	<i>Continuously unemployed until first to third interview (i.e. for 1 year) %</i>
Neither interview	73	61
Both interviews	14	24
First interview	7	5
Third interview (after 1 year)	6	10
All	100	100
(Sample size)	(1503)	(217)

Source: Moylan, Miller, and Davies (1984).

(Based on a sample of men unemployed in autumn 1978 who were interviewed three times. Of the 1503 interviewed a third time (after 1 year) 1286 had found a job during the year).

economic status with subsequent mortality. Of men aged 15-64, those who were seeking work had a standardised mortality ratio of 130 compared with a SMR of 86 for those who were employed (see table 3). This difference may be partly due to poor health before the 1971 Census increasing the chances of being unemployed at the Census and of dying after it.

The study by the MRC/SSRC Social and Applied Psychology Unit in Sheffield (Banks and Jackson, 1982) of school leavers was designed to separate the effects of unemployment on health and health on unemployment. The mental health of a sample of school-leavers was assessed before they left school and again after they had been in the labour market for 12-18 months. After

TABLE 3  
*Economic position and mortality*

<i>Economic position</i>	<i>Males aged 15-64 observed</i>	<i>Mortality Expected</i>	<i>observed ÷ Expected</i>	<i>Females aged 15-59 observed</i>	<i>Mortality Expected</i>	<i>Observed ÷ Expected</i>
<i>Active</i>						
Employed	3021	3508.7	86	682	844.4	81
Off work, sick	211	65.3	323	48	11.1	432
Seeking work	165	126.9	130	20	25.0	80
<i>Inactive</i>						
Retired	91	59.4	153	37	26.2	141
Permanently sick	370	94.5	392	101	20.1	502
Student	26	31.5	83	16	17.5	91
Other inactive	43	41.0	105	646	605.9	107
Total	3927	3927.3	100	1550	1550.2	100

The table shows the mortality of individuals in the longitudinal study sample by their economic position in 1971. The expected number of deaths is the number which would have occurred in a group in a particular economic position if that group had experienced the average mortality for all groups with the same age and sex.

Source: Fox and Goldblatt (1982).

controlling for sex, ethnic group, and qualifications, those who were unemployed had worse mental health than those who were not. More important, there was no significant difference in initial (at school) mental health between those subsequently unemployed and those in

work, on the YOP scheme, or in further education. Thus it appears that initial mental health was not a factor in subsequent employment status: the association between unemployment and poorer mental health indicates that unemployment does affect mental health.

## Conclusion

The early sweeping claims that unemployment has a large effect on population mortality are not supported by the evidence to date. Given the many difficulties in interpreting the data it would be a rash commentator who reached firm conclusions about what the precise effects of unemployment are. Despite this some tentative judgements can be suggested:

(a) unemployment is not important in explaining variations in general population mortality;

(b) there appears to be some independent effect of unemployment on suicide;

(c) unemployment has not been shown to be a significant cause of physical ill health;

(d) unemployment probably leads to some reduction in mental health.

Any attempt to evaluate the social implications of these conclusions will be even more dubious but one more assertion can be hazarded:

(e) the health costs of unemployment are relatively insignificant compared with its most obvious and very large social cost: the lost output forgone by having workers idle rather than producing goods and services.

### REFERENCES

- BANKS, M. H., AND JACKSON, P. R. (1982). 'Unemployment and risk of minor psychiatric disorder in young people: cross-sectional and longitudinal evidence', *Psychological Medicine*, 12, 789-98.
- BRENNER, M. H. (1971). 'Economic changes and heart disease mortality'. *American Journal of Public Health*, 61, 606-11.
- BRENNER, M. H. (1977). 'Health costs and benefits of economic policy', *International Journal of Health Services*, 1, 581-623.
- BRENNER, M. H. (1979). 'Mortality and the national economy: A review, and the experience of England and Wales, 1936-1976', *The Lancet* (ii), 568-73.
- BRENNER, M. H., AND MOONEY, A. (1982). 'Economic change and sex specific cardiovascular mortality in Britain 1955-1976', *Social Science and Medicine*, 16, 431-42.
- BUNN, A. R. (1979). 'Ischaemic heart disease mortality and the business cycle in Australia', *American Journal of Public Health*, 69, 772-81.
- CATALANO, R., DOOLEY, D., AND JACKSON, R. (1981). 'Economic predictors of admissions to mental health facilities in a nonmetropolitan community', *Journal of Health and Social Behaviour*, 22, 284-97.
- EYER, J. (1977). 'Prosperity as a cause of death', *International Journal of Health Services*, 7, 125-150.
- FORBES, J. F., AND MCGREGOR, A. (1984). 'Unemployment and mortality in post-war Scotland', *Journal of Health Economics*, 3, 1-11.
- FOX, A. J., AND GOLDBLATT, P. O. (1982). *Longitudinal Study: Socio-Demographic Mortality Differences* (London: HMSO).
- GRAVELLE, H. S. E., HUTCHINSON, G., AND STERN, J. (1981). 'Mortality and unemployment: A critique of Brenner's time series analysis', *The Lancet* (ii), 675-9.
- GREAVES, J. P., AND HOLLINGSWORTH, D. F. (1966). 'Trends in food consumption in the United Kingdom', *World Review of Nutrition and Dietetics*, 6, 34-89.
- HENDRY, D. F. (1980). 'Econometrics: Alchemy or science?', *Economica*, 47, 387-406.
- MCAVINCHIE, I. D. (1984). 'Unemployment and mortality: Some aspects of the Scottish case 1950-1978', *Scottish Journal of Political Economy*, 31, 827-53.
- METCALF, D., AND NICKELL, S. J. (1979). *Notes on the incidence of unemployment and sickness spells of 3 months or more*, Working Paper no 79, (Centre for Labour Economics, London School of Economics, London).
- MOYLAN, S., MILLER, J., AND DAVIES, R. (1984). *For Richer, For Poorer? DHSS Cohort Study of Unemployed Men*. Social Research Report No. 11 (London: HMSO).



# EDITORIAL

Folio 10

Editor: A. J. Culyer

Despite the three recent major organisational changes in the National Health Service the most striking features that continue to characterise its management are the absence of variety in experimentation in alternative ways of getting things done (things are very different in the United States) and the absence of clear incentives that make for better management and patient care (apart from the innate professionalism of its staff). The NHS's history instead has been one of monolithic imposition of uniform organisation structures that offer no clear and direct rewards to those whose performance improves. This folio reports on what is the one outstanding exception to these deficiencies: some real experiments in

## muffield/york PORTFOLIOS

offering clinicians budgetary incentives to be better managers. Their importance is scarcely to be underestimated, given the uniqueness of such ordinary experiments in Britain. Iden Wickings and James Coles make the ethical case for clinical budgeting in the NHS and show how it links up with new developments in the provision of information for management at all levels. A. J. C.

# The ethical imperative of clinical budgeting

**IDEN WICKINGS**

*Deputy Director, Kings Fund College and  
Director of the CASPE Research Unit*

**JAMES COLES**

*Fellow in Case-Mix Accounting  
Kings Fund College and Associate Director  
of the CASPE Research Unit*

'Ethical' and 'imperative' are both strong words. In this paper we hope to justify their use by showing that a major opportunity to improve patient care will have been missed by any Health Authority that does not participate fully in the move towards clinical budgeting and so does not make the cultural and managerial changes that are needed to introduce this working partnership with its clinicians. The NHS should not willingly fall short of its potential and its managers have some openings that are denied to those in charge of the different systems of health care found in many other countries. There will be a substantial cost both in monetary terms and in the need for managerial effort as clinical budgeting is introduced, but significant and permanent benefits are available. This was recognised by the Griffiths Inquiry team who put the case powerfully:

'Rarely are precise managerial objectives set; there is little measurement of health output; clinical evaluation of particular practices is by no means common and economic evaluation of those practices is extremely rare. Nor can the NHS display a ready assessment of the effectiveness with which it is meeting the needs and expectations of the people it serves. Businessmen have a keen sense of how well they are looking after their customers. Whether the NHS is meeting the needs of the patient, and the community, and can prove that it is doing so, is open to question.' (1)

---

On balance, we accept the above criticism although the comparison with British businessmen may not be persuasive to everyone. One could say that for many of those health outputs that are measured, Britain is amongst the better performers internationally while the costs of health care in Britain appear to be relatively modest (2,3). British business performance overall has not always been so impressive. But the NHS could undoubtedly perform much better and due to its structural cohesiveness may have unusual opportunities for so doing. It could, as we argue, achieve better patient care outputs in quality and quantity for the same level of expenditure. That in essence is the ethical case for clinical budgeting.

---

## Ethical criteria

Even in the USA, which is one of the richest countries in the world in terms of GNP per caput, various forms of health care rationing take place and some patients find it difficult to get treatment at all (4,5). In the UK rationing tends to take different forms with long waiting lists (6), poor standards of care for the chronically ill (7), and limits placed upon the numbers given expensive treatments such as kidney transplants (8), or coronary artery bypass grafts (9). Many patients are also treated in hospital facilities that are poor or badly maintained (10).

We shall suggest later that in the clinical area, just as the Rayner scrutinies have shown elsewhere, there are expenditures that could be reduced with no disbenefit to the patients. However, in addition to their separate responsibilities that are outside the immediate purview of the statutory authorities, clinicians within the NHS also have a role as allocators of public resources. It is, of course, one of the duties of public resource allocators to

try and apply available funds where they will do most good. All senior managers have to accept the need to ration scarce resources: the clinician's time, the available beds, the distribution of skilled nurses, and so on. Clinical budgeting creates opportunities for the roles of the resource allocators to be extended.

Rationing inevitably raises ethical questions. Sometimes these concern choices that must be made between individual patients; at other times the choices are between programmes for the elderly, say, or for prevention or for acute care. The selections will often require judgements to be made between qualitative and quantitative improvements. All the above decisions will be capable of altering the service outputs in terms of lives saved, or injuries repaired or avoided, or in benefits measurable in the quality of patients' lives.

If real progress can be made in tackling some of these evaluations, then it is ethically unacceptable for doctors, general managers and the statutory Health Authorities to shy away from the difficult decisions that have to be made. The potential benefits in terms of improved patient care are too important to be lost.

But to date little quantification takes place and few decisions about resource allocation choices are based on estimates of the benefits that are being sought. It is enormously difficult to handle these issues but one can scarcely be proud that, after 37 years of the NHS, so little has been achieved in this area that Griffiths was able to record: 'little measurement of health output; clinical evaluation . . . is by no means common and economic evaluation . . . is extremely rare.'

Many decisions, although not all, will be assisted by the activity and cost information that will be provided by what

Griffiths called management budgeting. However, in this paper the term clinical budgeting is used because it is upon that particularly difficult area of health service expenditure that we are concentrating. In any event, the differences between clinical and management budgeting are more to do with the wider spread of management budgeting, and its incorporation of overhead costs, than with any differences of substance (11).

## What is clinical budgeting?

At its minimum, clinical budgeting is budgeting for each major area of clinical activity instead of simply for the functions that exist solely to support that clinical work. The functions, such as nursing, the pharmacy, radiology, and so forth now have annual budgets prepared for them in most Authorities. But these functions may be regarded as the input side of health service management.

The output of the NHS is patient care including the treatment or prevention of sickness and disability. Clinical budgeting requires Health Authorities also to have budgets related to these outputs: there will be a budget for the geriatric service, say, and for general surgery. Health Authorities will then be in a position to decide whether changing the budgets for or between these two services will change the outputs (the patient care provided) and which balance stands the best chance of optimising the 'benefits' for the community served. For example, once you have a budget for a clinical service, say general surgery, then other managerial options become available. The general surgeons may want to manage their own budget; the Health

Authority may be prepared to delegate some powers of discretion to the surgeons provided that they stay within the budget and deliver the planned surgical service. This is often referred to as the incentive: the general surgeons have an incentive to review their resource use more carefully if their own patients will benefit from any savings achieved. In this way, all parties can benefit because the Health Authority through its general manager can make real choices about the balance to be struck across all the clinical budgets, and the clinicians can be given extra discretion and thus have an incentive to use their allocated resources more efficiently in the interests of their own clinical service and its patients. In this way optimising the output of the NHS, in terms of the quality and quantity of the service provided, becomes a matter of concern to all those involved.

The different ways in which such budgets can be framed can be seen in Table 1.

## A conceptual framework

Health Authorities and clinicians need a shared analytical framework if they are to work together to make well informed decisions. The questions often sound deceptively simple:

- How much money should be spent next year on general surgery?
- What outputs should be expected from this clinical service or that?
- Is terminal care in the community cheaper, or better, or worse, and for which group of patients?
- How does our ENT department compare with others?

Table 1. Illustrative table of functional and clinical matrix\*

Clinical headings	Functional headings					Total of departments shown (£)
	X ray (£)	Pathology (£)	Nursing (£)	Pharmacy (£)	etc (£)	
Surgeon Mr A						
Patient Mrs P	7	2	30	10		49
Patient Miss Q	15	7	25	18		65
Patient Mr R	2	5	45	0		52
Other patients	720	390	3200	540		4850
Total for Mr A	744	404	3300	568		5016
Total for Mr B	890	444	3750	485		5569
Total for Mr C	960	218	3800	960		5938
Total for general surgery	2594	1055	10850	2012		16523
Total for general medicine	3000	3750	14950	3600		25300
Total for geriatrics	1500	1890	19250	2600		25240
Total for psychiatry	300	150	11000	900		12350
Total for other specialties	1000	500	5000	1000		7500
Total for hospital or group	8394	7356	61050	10113 (W)		86913

\*Data for illustration only and to represent nominal financial units.

(W) = Functional budgets and/or expenditure.

(X) = Specialty budgets and/or expenditure.

(Y) = Clinical consultant budgets and/or expenditure.

(Z) = Patient costs.

- Are the patients reasonably satisfied?
- What should we stop doing to allow something important to start?

It will be noted that objective answers to all of the above questions are likely to benefit from comparative information either within the Health Authority concerned (general surgery versus other claimants) or across Authorities (how are we doing?). Comparisons are always difficult to make fairly but in the commercial world a firm that does not learn from its rivals will soon become bankrupt. The NHS should take comparative performance equally seriously. It is, one could argue, even more important because patients' and not just companies' lives may be at stake. As long as all concerned are genuinely seeking to improve the quantity or quality of patient care provided, then in our experience the active support of the medical profession can be expected. But the commitment to patient care must be real; those whose prime aim is to control doctors or simply to cut costs will be exposed quickly and will quite rightly fail. Clinical budgeting needs true commitment to the patient or it is an indefensible bureaucratic incursion into a complex and sensitive political minefield.

## The evidence about clinical budgeting

There have now been many clinical budgeting experiments. In England these include those at Westminster Hospital in 1974 (12, 13) in East Cumbria in 1978 (14), in North Manchester (at Booth Hall Children's Hospital) (15), at Southend and Oldham (16) and in Lewisham and North Southwark. Management budgeting demonstrations, incorporating clinical budgeting, were begun in 1983 as a result of the Griffiths report. They are now in progress in over twenty Health Authorities. Thus far, there is plenty of evidence that beneficial redeployments can be made within the context of the individual clinical budget. For instance, the changes made have included the following:

- One general medical firm saved enough to employ a ward clerk.
- Mobile X-rays in ITU reduced by 88%
- Bacteriology tests reduced by 60%

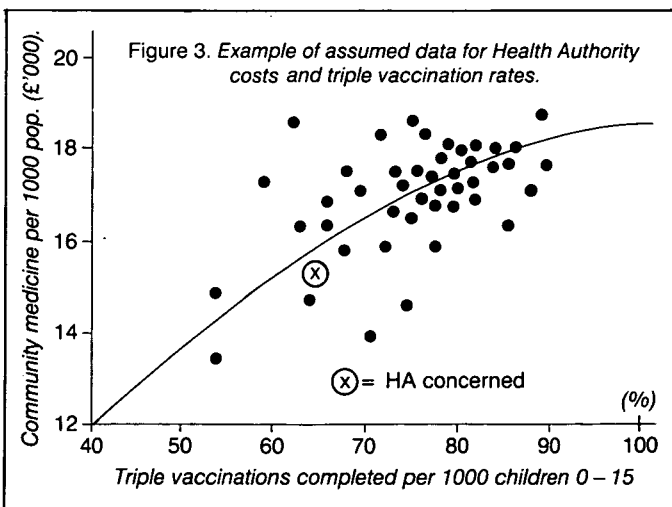
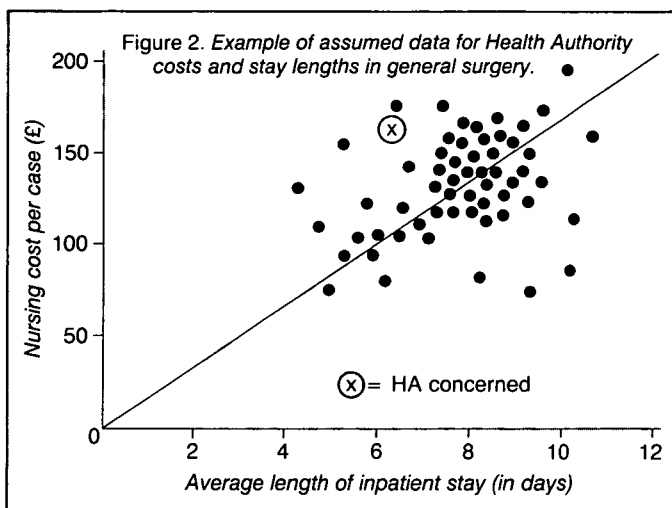
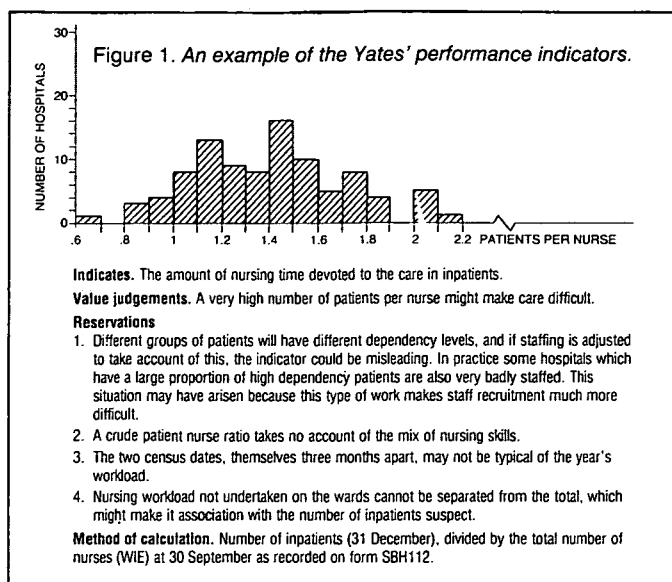
- Savings to fund a new bathroom for disabled patients
- 21% reduction in A & E radiology requests
- Radiotherapy firm funding extra staff through economies in drug use
- Valuable savings made on intravenous fluid usage
- Reduced number of beds in a paediatric unit freeing resources for other child care developments.

Encouraging results have also been reported in the USA (17), and in France and Scandinavia (18).

## Comparisons and performance indicators

An important way of making better informed health service comparisons has been developed by John Yates and Michael Davidge at the Health Services Management Centre in Birmingham (19). An example, in this case illustrating patients per nurse in mental handicap hospitals, is shown in Figure 1.

The majority of Yates' performance indicators, of which there are now 33 for most hospital specialties, show comparative scales of provision (inputs) or relate these to measures of workload (throughputs). There is a continuing need for measures of relative output, to which we shall return, but the apparent crudity of Yates' indicators is misleading. They are of great potential value to Health Authorities and their clinicians as aids to decision-making about requisite directional shifts for the local indicators. A Health Authority can decide in which direction and by what degree it wishes its own indicators to be moving and subsequently monitor the general manager's progress in achieving the chosen organisational change. Such decisions can often be implemented more easily within a clinical budgeting system. In like manner, clinicians can be made aware of their own comparative performance. When clinical budgeting has been implemented widely, it will allow the inputs and throughputs for each service, as revealed by the Yates (21), or the DHSS (22), indicators, to be related both to the treatment processes with their associated costs and various indicators of outputs. For example the types of data shown in Figures 2 and 3



should become available in reality whereas the figures at present can only incorporate assumed data.

The members, officers and clinicians of Health Authority X should have some useful questions to consider as a result of examining the two figures: why is our nursing cost per case high when our average length of inpatient stay is only

average? Is it because we are getting proportionately better clinical results or very high levels of patient satisfaction? Our medical staffing costs in the community are very low—are our vaccination levels low as a consequence?

The fact that these questions are not easy ones to answer does not exonerate responsible managers or clinicians from trying to agree upon the appropriate courses of action. After all, if an unnecessary amount of the total available resource is being spent on nursing then some benefits that could be obtained from other patterns of expenditure are being forgone.

So why should Health Authorities and their clinicians be sceptical about getting involved in these debates? In the past such issues have generally been addressed only at times of confrontation, with the parties having differing objectives. However, if the objective is shared, and there is good reason to judge that a particular action might achieve improved, or at least maintained levels of patient care, then both the clinicians and the Health Authorities should be striving to expand their activities in this field. Clinical budgeting must not be seen as merely another way of saving money. For some diagnoses an increase in diagnostic work is associated with better outcomes (20). The shared method of taking decisions that forms the basis of clinical budgeting should be seen as a tool for optimising the outputs of the NHS, in terms of service quality, quantity, or both, for any given level of resources.

## A shift in the burden of proof about outputs and outcomes

Medicine is rightly a highly regarded science and doctors are amongst the most respected members of society. Perhaps this is because they have so often to act with confidence in situations where there are no well researched protocols to follow. In this respect they are not unlike health managers who have to take decisions with little in the way of guidance that might be available from text books.

Thomas Huxley was at his most thoughtful when considering scientific and Christian values. On science he wrote in The Method of Zadig: 'Science is nothing but trained and organised common sense, differing from the latter only as a

Figure 4. Examples of opportunities for better resource usage in the clinical field in the NHS.

RESOURCE	EXAMPLE	SOURCE
ACUTE BEDS	1964 : 22.5 empty 1976 : 30% empty 1979 : 28% empty	Yates (21) 1982
ADMISSIONS	Four-fold variations in stay for comparable cases	Ferster (23) 1978
MEDICAL	<div> <div>New</div> <div>New O/Ps Session</div> <div>Old/New Ratio</div> </div> <div> <div>1970 533000</div> <div>3.72</div> <div>5.9</div> </div> <div> <div>1976 474000</div> <div>3.02</div> <div>6.8</div> </div>	SH3s
X-RAYS	24-fold variations in use for comparable cases	Ashley (24) 1972
MICROBIOLOGY	< 35% requests probably unnecessary	Spencely (25) 1979
DRUGS	1.25 tons found unwanted in S. Wales homes in one week	Olsen (26) 1974
THEATRES	26% sessions not used	Rawlinson (27) (1981)
OXYGEN	ten-fold variations in domiciliary oxygen consumption	Williams (28) 1981

veteran may differ from a raw recruit: and its method differs from those of common sense only as far as the guardsman's cut and thrust differ from the manner in which a savage wields his club.'

Science, or as Huxley would have it 'common sense', has to be allowed a greater influence upon resource usage decisions in the NHS. The Rayner scrutinies have shown unacceptable levels of resource mismanagement and waste in the non-clinical areas but there are similar questions to be raised in the clinical field as well (see Figure 4).

Common sense becomes essential when considering the very different clinical practices revealed in Figure 4. Can those working in the way that is most expensive show any evidence of better clinical results? Where they can, then in most circumstances their activities should be encouraged (at least until they inhibit the development of other clinical programmes); where they cannot, then the resource allocators should be reviewing competing opportunities to discover whether greater benefits are likely to be produced elsewhere. It is in this sense that we believe that the onus to provide sufficient evidence (the burden of proof) is shifting from the statutory Authorities and their general managers to the clinical practitioners. Pressure upon resources is increasing and yet many improvements are still needed in the NHS. As mentioned earlier, services for the chronically ill and the handicapped are often woefully inadequate; many patients who are immobilised and in pain owing to arthritic hips are waiting too long for their operations; the support given to the elderly in their homes is often insufficient to let them remain independent with dignity; and, on a

different tack, health service buildings are often poorly and inefficiently maintained. If these demonstrable needs are to be met from a constant level of finite resources or even from any foreseeable level of modest growth in real terms, then some existing practices will have to be forsaken. A careful review of the existing practices that seem not to be fully justified can be contemplated where the cost of these practices appears high. Any wider review would simply open up too many issues to be handled within a Health Authority's planning and budgeting timetable.

## A method of organising common sense

Using Huxley's concept of science makes us ponder how the application of common sense can be organised most effectively in the NHS. This is likely to require the development of some new skills and some new methods of working. Figure 5 introduces the terms 'Expected Range', 'Current Value', and 'Planned Range'. The 'Expected Range' should be regarded as the template for a particular Health Authority if it was working well at current technological standards. This can be considered in both resource and clinical terms. For instance, what costs,

Figure 5. Diabetic services in England by geographical distribution (14 Regions).

	ITEM	DISTRIBUTION	EXP. RANGE	CURRENT VALUE	PLANNED RANGE
STRUCTURE	Cons. per 100,000 population				
	Specialist clinics per week per 100,000 population				
	Combined clinic with ophthalm. per 100,000 population				
	New patients per week per 100,000 population				
PROCESS	Bed days per 100,000 population				
	Drug costs ppd				
	Path. tests ppd				
OUTCOME	Amputation				
	Avoidable blindness				

or staffing levels, or throughput figures, would we expect to see for any particular clinical specialty for this particular Health Authority given its demographic and socio-economic profile? Different members of the general manager's team will be able to research and calculate these template figures which can then be contrasted with the 'Current Value' for each item, which in future will often be available from clinical budgeting data. Sometimes, as in the example quoted for diabetics in Figure 5, other sources can be used--in this instance the data show regional differences and are taken from a report of the Medical Advisory Committee to the British Diabetic Association (29). In some cases, the Yates' indicators will be the primary source of data allowing an expected range to be calculated but in others the District Medical Officer will be researching in the library or using well known criteria such as Charlton and Holland's survey of 'Avoidable Deaths' (30). As this method of organising common sense becomes more widespread, different Health Authorities will be able to learn from one and another.

Deciding upon the 'Planned Range' needs a different approach, however, because this depends upon genuine agreement being reached with the clinicians working in the specialty concerned. It is the clinicians who will often be best placed to know the direction and speed in which their specialty is moving and they may have quite different needs and expectations from those known to the general manager and his team. Nonetheless it is the discussions preceding agreement upon the planned range, and how to get there, that will produce genuine co-operation between managers and clinicians. A more basic method of reaching such agreement has recently been tested in some clinical budgeting experiments. The method involved district managers and clinicians negotiating Planning Agreements with Clinical Teams (PACTs). In early 1985, an independent Evaluation Group chaired by Professor Buller, the previous Chief Scientist at the DHSS, concluded:

'The Evaluation Group is not aware of any other system than PACTs that offers similar interaction between managers and clinicians and notes the adoption of a generally similar format by the demonstration districts in the management budgeting

programme, at a considerably greater introductory cost . . . the Evaluation Group is unanimously of the view that in principle this PACTs centred budgeting system has all the right ingredients for improved resource management in the NHS, and it should be given the support needed to ensure its wider dissemination within the service.' (16)

---

## Why we cannot do without clinical budgeting

Health service resources grew fast during the early years of the NHS. There were understandable expectations that, ultimately, supply would outstrip demand. That is no longer a tenable expectation and some form of choice has to be made between competing options under all systems of health care in even the richest countries. This is certainly true in the United Kingdom where some Health Authorities are also experiencing real cuts in their resources in a nationally led attempt to improve overall patterns of financial distribution. The assessment of competing claims is just as necessary within Authorities as it is between Authorities, but the problems of achieving such redistributions well may be greater.

A framework for analysing the evidence that is already available, or the evidence that will become available when most of England's 192 Health Authorities have the data from clinical budgeting, has been described. It allows the very difficult decisions about trade-offs between specialties to be considered in an organised way. In general these decisions need to be joint ones between the Health Authority, through its general managers, and the appropriate clinicians. Similar agreements, even before the requisite data base has been developed nationally, have been tested and independently found to be valuable.

The realities of resource constraints upon the NHS make it ethically mandatory for Health Authorities and clinicians to work together in the way described. Not only does this allow clinicians and Authorities to share their individual perceptions of the desirable ways to go forward but it allows them to explain to each other the real constraints under which they are working. Both can then consider what can be forgone and what must be retained and reach their

decisions on an informed and collaborative basis. Once the resources to be allocated to a particular specialty have been determined then responsibility for their management can be delegated to clinicians and their teams and this can provide additional incentives for resource redistribution at that level as well. We have reported the benefits that can follow elsewhere (11), but this aspect of clinical budgeting is secondary to the more major debate that can probably only be conducted within a framework similar to the one we have described. It is at least arguable that only in such a way can the ethical responsibilities of all the many resource allocators in the NHS be fully

discharged. Only when working in such a manner can both the Health Authorities and their clinicians conscientiously seek to increase or optimise the service outputs in quality or quantity for any given level of resources.

The Griffiths report showed the way. But like so many of the previous attempts to improve NHS management, this initiative will fail unless behavioural incentives are used more widely and the accountability for making choices at the clinical, unit and Health Authority levels is truly made clear. Clinical budgeting provides only a modest beginning—but not to start at all is, quite simply, ethically unacceptable.

## REFERENCES

1. Griffiths, R. et al. (1983). NHS Management Inquiry. London: DHSS.
2. Maxwell, R. J. (1981). Health and Wealth: An International Study of Health Care Spending. USA: Lexington Books.
3. Culyer, A. J. (1976). Need and the NHS. London: DHSS.
4. Aaron, H. J. and Schwartz, W. B. (1984). The Painful Prescription. USA: The Brookings Institute.
5. Lurie, N., Ward, N. B., Shapiro, M. F. and Brook, R. H. (1984). 'Termination from Medi-Cal - Does it affect health?'. The New England Journal of Medicine, 311, 480-4.
6. Cullis, J. G., Jones, P. R. (1983). 'In-patient waiting: a discussion and policy proposal'. Brit. med. J. 11, 1483-6.
7. Malin, N., Race, D. and Jones, G. (1980). Services for the Mentally Handicapped in Britain. London: Croom Helm.
8. McCarthy, M. (1985). 'A Decision-maker's guide to renal failure' in Health Care UK 1985. (eds) Harrison, A. and Grettton, J. 84-8. London: CIPFA.
9. Coronary Artery Bypass Surgery - Consensus Statement. (1985). London: King's Fund.
10. Davies, C. (Chairman). (1983). Report of the Enquiry into Underused and Surplus Property in the NHS. London: HMSO.
11. Wickings, I. (1983). 'Consultants face the figures'. Health and Social Services Journal, 1466-8.
12. Wickings, H. I. (1977). 'Putting it together: the patient, the purse and the practice'. The Lancet, 1, 239-40.
13. Coles, J. M., Wickings, H. I. et al (1975). 'The effect of presenting management information to clinically accountable teams', Brent HD.
14. Chin, P. L., Chatterjee, J. et al (1981). 'Speciality Budgeting: a pilot study carried out in East Cumbria Health District'. Health Trends 13.
15. Lamb, S. M., David, T. J. (1985). 'Playing with fire: an experiment in clinical budgeting'. Brit. med. J. 290, 650-1.
16. Buller, A. (Chairman) (1985). Clinical Accountability, Service Planning and Evaluation - Interim Report of the Evaluation Group. London: DHSS.
17. Solomon, S. (1979). 'How one hospital broke its inflation fever.' Fortune.
18. Bally, Yvonne Clinical Budgeting in the NHS or Going Dutch (Swedish French) Discussion Paper No. 05(82 presented to the Health Economics Study Group, University of Aberdeen.
19. Yates, J. M. and Davidge, M. G. (1984). 'Can you measure performance?' Brit. med. J., 1, 1935-6.
20. Garber, A. M., Fuchs, V. R. and Silverman, J. F. (1984). 'Case mix, costs and outcomes'. The New England Journal of Medicine, 310, 1231-7.
21. Yates, John. Hospital Beds - a problem for diagnosis and management? Heinemann Medical Books.
22. Performance Indicators. National Summary and Regional Summaries for 1981. (1983). DHSS Regional Liaison Division.
23. Ferster, G. (1978). 'Estimation of notional savings in revenue expenditure by specialty at district level'. Hospital and Health Services Review, 87-90.
24. Ashley, J. S. A. et al (1972). 'How much clinical investigation?' The Lancet, 890-2.
25. Spencely, M. et al (1979). 'The clinical value of microbiological laboratory investigations', The British Society for the study of Infection. Journal of Infection, 1, 23-6.
26. Olsen, N. D. L. (1974). 'Restricted list of NHS drugs in economic benefits of restrictions on clinical freedom'. Brit. med. J., 4, 273-4.
27. Rawlinson, Dr. C. (1981). 'Operating Theatres - how many and how busy?' Health and Social Services Journal, XCI, 406-11.
28. Williams, B. T. (1981). 'Geographical variations in the supply of domiciliary oxygen'. Brit. med. J., 282, 1941-3.
29. British Diabetic Association (1984). Report from the Medical Advisory Committee.
30. Charlton, J. R. H., Hartley, R. M., Silver, R. and Holland, W. W. (1983). 'Geographical variation in mortality from conditions amenable to medical intervention in England and Wales.' 1, 691-6.

Acknowledgement is made to the Office of the Chief Scientist at the DHSS for providing grants.