Product definition for healthcare contracting: an overview of approaches to measuring hospital output with reference to the UK internal market

Neil Soderlund

Abstract

Objective – In many industrialised countries, health care third party payers are moving towards contracted provision arrangements with suppliers of hospital care. Essential to such a process is a standard approach to quantifying the care provided. This paper aims to outline the possible approaches to hospital product definition for the UK National Health Service, and recommends appropriate further research.

Methods – All published and unpublished studies on hospital output measurement in the NHS since 1980 were sought for the purposes of the review. This included both discursive and empirical work, and no exclusion criteria were applied. Most empirical reports on this topic, however, come from the United States. Consequently, the published reports since 1980 from the USA, accessed from the Medline and Healthplan CD-ROM databases, have also been included in the overview.

Conclusions – Where data are sufficient, the true casemix approach offers advantages over other methods of output measurement. In the UK NHS, two systems – diagnosis-related groups (DRGs) and healthcare resource groups (HRGs) – are the only casemix measures that have achieved any significant degree of attention. DRGs have been extensively evaluated internationally, and explain variations in resource use in the UK slightly better than do HRGs. As a local product, HRGs can be more easily adapted to the specific needs of the NHS internal market, however, and will thus probably emerge as a better measure for the UK in the long term. In both cases, locally derived cost weights are unavailable, and their development constitutes a major requirement for use in contracting. Adaptations for long stay and outpatient hospital episodes would enhance the usefulness of hospital casemix systems in the NHS. Existing approaches, such as specialty based classifications, are neither standardised nor predictive of resource use, and would be better replaced by casemix systems. Other countries facing similar choices between casemix measurement approaches will need to consider the “trade off” between the adaptability of locally derived systems on the one hand and the low cost, rapidly accessible results, and availability of international comparative data of an imported approach on the other.

(J Epidemiol Community Health 1994;48:224–231)

An interest in defining and measuring the contribution that hospitals make to society is neither new nor unique to one country. E A Codman’s original discussion of concepts of hospital output dates from 1914, and virtually all developed countries having gone through, or are currently going through, a process of experimentation with different output measurement systems. The 1980s in the USA were characterised by a shift towards prospective payment on the basis of diagnosis based casemix groupings, firstly by the Medicare programme, and subsequently by many other third party payers. Similar changes have occurred, or are occurring, in Canada, Australia, Spain, Portugal, and Ireland, to name but a few countries. Such research does not necessarily apply across national borders, however; differences in health care financing, data collection, and clinical practice necessitate substantial adaptation of methods. Nevertheless, most research comes from the United States, and some assumptions of transferability will have to be made by other health care systems until local research experience catches up.

In the United Kingdom, the establishment of an internal market within the publicly provided National Health Service (NHS) has provided considerable stimulus to initiatives to apply output measurement systems locally. Purchasers need to be sure of exactly what they are getting from hospitals before they can choose between them. Providers need to measure their output before they can estimate realistic prices that reflect local costs, but are competitive. They also need to assess internal efficiency, and manage accordingly. At the national level, monitoring the performance of the NHS as a whole requires quantification of the national output, and knowledge of how it is...
distributed. The NHS Management Executive has responded by setting up the National Casemix Office (NCMO) and the National Steering Group on Coding (NSGC), with the aims of designing a uniform product definition system, and costing those products in a standardised way. 

Data collection systems already in place in the UK are sufficient for most of the information needs of adequate hospital output measurement, and the NHS seems to be currently well placed to adapt local and international experience to suit the needs of the internal market. This paper aims to review the suitability of existing systems of product definition for use in the UK internal market. It is also hoped that this paper will assist researchers and policy makers elsewhere in the world in adapting what are largely North American systems to their own conditions.

Approaches to hospital product definition

Few would oppose the view that the principle purpose of hospitals is to bring about a change in health status of individual patients admitted to the institution, and hence also of the communities from which they come. There are other long-welldescribed hospital products also worthy of mention, such as the outpatient health gains, teaching and research, and amenity consumption goods such as hotel services. This paper will discuss the measurement of inpatient and outpatient health products only.

The inpatient health product

Measuring actual changes in health status, or health related quality of life, must remain the gold standard for hospital product quantification. For the foreseeable future, however, data gathering constraints will prevent use of such methods, and hospital output quantification will imply using a variety of process or proxy measures. It is essentially in the choice of which of these process measures to combine that hospital product definition systems differ. Hornbrook7 provides a useful conceptual background to the measurement of hospital output. He divides the hospital product into three facets – volume, “casemix”, and quality.

The volume aspect is easiest to measure, although fairly controversial in the NHS, where the “finished consultant episode” measure is particularly difficult to use for contracting purposes, as any number of episodes can be recorded during one hospital admission. Fuller discussions of this issue are given elsewhere.8,9 Quality issues have also attracted increased attention recently, with a spate of conferences and the launch of a UK journal devoted to the topic. Interpretations of the term still range widely, however, from the consumerist, all-encompassing “total quality management” approach, to circumscribed measures of medical outcome in particular diseases. This paper will concentrate on what Hornbrook has broadly described as “casemix” – that is, the disease features, and consequent health care requirements of a hospital inpatient population. Three approaches to this aspect of hospital output can be identified from published reports – the service mix approach, the true casemix approach, and the facility mix approach.10

Service mix implies measurement of the medical services administered during a patient episode – that is, the number of x rays, number and complexity of surgical procedures, etc. On the other hand, casemix implies detailing features of the patient concerned, for example, diagnosis, severity, and age. The facility mix approach classifies patients according to the capabilities of the institution concerned, for example, whether or not intensive care unit facilities are available.

FACILITY MIX APPROACHES

While this approach was prominent and influential 20 years ago (see, for example, Berry’s work),11 it has largely disappeared from US published reports and is also becoming rarer elsewhere, although it has been used in some fairly recent UK studies.12 Its immediate appeal lies in the fact that it is most easily derived from routine information and does not require individual patient data. Problems with the validity of the facility mix approach are fairly obvious. Firstly, it makes no measurements at the individual patient level, but instead attributes an institutional average to the entire inpatient population. Since no account is taken of the utilisation level of facilities, an unused facility (for example, an operating theatre) and a continuously used facility would impart the same weight to patient episodes in the respective institutions. The unused one would incur lower variable costs, and consequently, reimbursement of hospitals on this basis would encourage inefficient use of capital. Furthermore, there is no system of measurement in common use, and if there were, it would not support contracting at a level sufficiently detailed to encourage efficient provider behaviour.

SERVICE MIX APPROACHES

Like facility mix measures, service mix classification systems are concerned with the treatment process, from which is inferred a change in the health status. They do have significance at the individual patient level, and are derived from actual activity rather than potential for activity. Some measures, such as the Resource Based Relative Value Scale (RBRVS) system for estimating physician contribution to hospital product, are highly specific and correlate well with resource input.13-15 They are complicated to derive, however, and incorporate components that are highly value-laden, for example physician skill and stress associated with certain procedures. Others, such as the British United Provident Association (BUPA) schedule used in the UK are far cruder, and are based principally on time taken to perform procedures.

The service mix approach is most useful for
measuring short term technical efficiency of hospitals and is consequently of appeal primarily to hospital managers. Clearly, however, any amount of services rendered to a patient will only bring about a health improvement if they are appropriate for their particular diagnosis, age, sex, etc. Furthermore, there are usually diminishing returns for each increment in amount of services provided, implying the existence of an optimal level of treatment for every patient, above which overall efficiency would be impaired. If physicians or hospitals are reimbursed on the basis of the service mix that they provide to their patients, inappropriate interventions may be encouraged. Empirical evidence for this phenomenon is suggested by work on supplier induced demand in health care generally, and illustrated, for example, by the high procedure rates in some service mix based reimbursement systems.20

If contracts are to reflect output at the individual patient level, service mix measures do provide one of the options available to providers and purchasers. Procedural data in the form of the Office of Population Censuses and Surveys (OPCS)-4 codes are available for all inpatient episodes in the UK, but data on most diagnostic and medical therapeutic activity are not. A more detailed coding system, Read codes, would be able to support a higher degree of service mix specification, but are not widely used by hospitals. Supply side measurement, however, will always run the danger of implicitly assuming that the application of health technology is universally beneficial, and health care purchasers have been justifiably wary of reimbursing providers on this basis.

TRUE CASEMIX APPROACHES

Casemix measures are distinguished by their incorporation of information about patient illness. A considerable amount of work has gone into developing, adapting and testing casemix measurement systems globally and within the UK in the past 10 years or so.20-24 Most local research has examined their ability to explain resource use variation as measured by length of stay, although use as “iso-prognosis” groups has been suggested for some measures. Casemix measures are derived from a variety of patient specific data, including diagnosis, procedures performed, age, sex, socioeconomic status, dependency level, past treatment, diagnostic complexity, illness severity, and response to therapy. The degree to which different items of information are included by various proprietary measures largely determines their areas of usefulness.

Diagnosis related groups (DRGs)

DRGs, developed by Bob Fetter and colleagues at Yale University,25 were the first casemix measure to achieve widespread use, and are also currently the most extensively used.3 In their country of origin, the USA, they are used for determining levels of prospective payment under Medicare and other third party payers. They have also been adapted for use in Australia26 and a number of European countries, including Great Britain, where they were recommended for in-hospital management under the Resource Management Initiative.30-31 DRGs classify into just under 500 groups virtually all types of acute hospital inpatient on the basis of routinely recorded diagnostic and procedural information. They were constructed originally on the basis of variations in length of hospital stay, but have subsequently been refined using more detailed patient resource use information.

The main advantages of using DRGs are their widespread usage (and hence the potential for international comparisons), and the long period over which they have been evaluated and revised. An English version of DRGs has been developed by CASPE HKS,32 which enables their derivation from ICD-9 diagnostic and OPCS-4 procedure codes routinely collected as part of the Körner minimum inpatient dataset. A number of studies in the UK have shown the feasibility of their use, and their ability to explain some variation in the length of stay.33-35 DRGs also form the basis of the Thorn-EMI cost accounting computer software package, which is used by approximately 100 institutions in the UK.36 No UK derived version of DRG хочется пока не существует, however, and this significantly impairs their credibility within the internal market.

Healthcare Resource Groups (HRGs)

With the more widespread use of DRGs in the NHS, it became apparent that they performed poorly in some areas where there were differences in diagnostic terminology and/or clinical practice between Britain and the United States. For this reason, a system of British casemix categories called HRGs have been developed by the National Casemix Office,37 primarily by a process of clinician consensus conferencing. This aimed to make them more clinically coherent than DRGs. In other respects, however, it seems that HRGs are similar to DRGs, both in structure and performance. At this stage, little practical use has been made of HRGs. There are a relatively large number of HRGs (>500), and this number could probably be reduced quite substantially without much loss of statistical performance. Preliminary investigations suggest that they can classify approximately 85% of general hospital episodes in the NHS (P Anthony, personal communication, 1992). DRGs still explain more of the variation in costs and length of stay than do HRGs in their current form (N Söderlund, unpublished data, 1993), but version two of HRGs (due in mid-1994) are likely to be substantially improved in terms of resource homogeneity. Since HRGs require procedural data in the form of OPCS-4 codes, they cannot be used outside of the United Kingdom at this stage.

Other diagnosis based casemix measures

There are a number of other diagnosis based
Severity measures

Severity of illness may be viewed as a combination of the level of bodily dysfunction and the degree of dependence on medical treatment to reverse the illness related decline in health. While there is a theoretical basis for supposing a relationship between illness severity and resource use, empirical work from the US suggests that it is relatively weak except in a few diagnostic categories.37-39 There remains an interest in their use as iso-prognosis groupings, however.

Disease Staging:40 is a casemix system developed in the United States which aims to incorporate to a greater degree than DRGs the severity of a given disease in the classification system. It organises patients into approximately 400 diagnostic categories which are then stratified into ordinal severity stages. Disease Staging, in its routine data derivable form, measures primarily the number and type of secondary diagnosis codes, and hence is really more of a diagnostic complexity than a true severity score. Nevertheless, its ability to predict adverse outcomes is better than that of DRGs, suggesting that diagnostic complexity is a reasonable proxy for severity. Studies from the United States have indicated that Disease Staging is useful for comparing casemix differences between institutions, independently of the actual resources utilised in the treatment of patients.41 That is to say, while at a purely statistical level it is not as good a predictor of resource use as DRGs are, it does not confound treatment intensity with diagnostic casemix measurement, and therefore arguably acts as purer indicator of "need for care". This feature has led to suggestions that it may be most appropriate for use by purchasers.42 The information needed to derive Disease Staging categories can be obtained from routine NHS data, with approximately 95% of records being assigned to a Disease Staging category. Between 2 and 8% of records from different regions were assigned to non-specific groups, and it is a reasonably good predictor of both length of stay and mortality.43 It may be less useful than has been the case in the US, however, because of relatively lower levels of coding of secondary diagnoses in the UK.

The Severity of Illness Index (SOI):44 and its computerised successor, the computerised severity index (CSI)45 was developed by researchers at Johns Hopkins University in an attempt to adjust DRGs for illness severity and, in so doing, better predict patient resource needs and outcome.44 On hospital discharge, each patient is assigned a severity score on a four level scale, which is used as a stratification of their DRG group or as an independent severity measure. The SOI index comprises seven dimensions of casemix, namely, the stage of the principal diagnosis at admission, complications, pre-existing health problems, dependency, non-operating room procedures, rate of response to therapy, and residual symptoms at discharge. Data currently available on a routine basis in the UK are insufficient to derive SOI or CSI scores, however, and they have not been tested in the NHS.

The Acute Physiology and Chronic Health Evaluation system (APACHE II):46 takes the approach of the SOI index one step further and incorporates biophysical measurements together with measures of chronic health status to measure severity. It is designed to adjust for outcome in intensive care units, however,46 and is an insensitive measure of hospital product at the lower end of the severity scale.47 Furthermore, many of the measurements required by the index are often not performed on patients not in intensive care. Casemix measures that have narrowed their focus down to the extent that the APACHE system has are clearly unsuitable for the generic hospital product. Nevertheless, it has recently been quite extensively tested in the UK,48 and offers potential for quality measurement in some specialised areas of contracting, such as intensive care and trauma services.

Two other casemix systems which incorporate aspects of patient severity/complexity, Medisgroups49 and refined DRGs, have undergone preliminary investigation in the NHS. US studies indicate that Medisgroups offer few advantages over systems already mentioned; however,49 and local evaluation of refined DRGs suggests that secondary diagnosis coding in the NHS is insufficient to yield useful results.50

Specialty groupings

These are the easiest of the casemix measures to obtain, and can be taken directly from routine data. There are, however, problems with their use. Firstly, there is no uniformly applied system of specialties, so comparisons between hospitals are difficult. Whereas a large hospital might assign most medical patients to subspeciality departments, smaller hospitals would tend to admit patients with the same diagnoses to general medical wards. Furthermore, here is little intuitive basis for assuming resource homogeneity simply on the basis of...
pathology in the same organ system (for example, ENT or orthopaedic surgery), or membership of the same client group (for example, paediatrics or geriatrics). Preliminary empirical evidence suggests that specialty classification explains between 9 and 14% of variation in the length of hospital stay and only about 4% of the variation in costs, whereas HRGs and DRGs explain between 15 and 20% of variation in both length of stay and costs (N Söderlund, unpublished data, 1993). Finally, rigid product definition along specialty lines might discourage interspecialty collaboration, and thus hinder overall efficiency. Nevertheless, what little casemix adjustment is currently being used for contracting seems to be mainly of this type.

Dependency based scores
Interest in casemix systems that incorporate dependency measures stems primarily from moves towards prospective payment for nursing homes in the United States. A number of systems have been developed and refined over the last 10 years or so, and a useful overview of them is presented by Fries.51 The most extensively tested system is probably Resource Utilisation Groups (RUGs).52 This combines a five stage hierarchical ranking classification from low to high estimated resource use, depending on specific diagnostic and therapeutic features together with an Activities of Daily Living (ADL) score. Other similar systems include the Maryland Nursing Home Reimbursement System,53 the Minnesota Department of Health System,54 and the Management Minutes System.55 Although not designed for use in acute hospitals, these may have something to offer casemix measurement in the UK because of the relatively high proportion of long stay patients occupying acute NHS beds. Addition of some index of dependency might thus significantly improve the product defining ability of any casemix system applied in the UK. Unfortunately, the central input variable to such system, an ADL score, is not recorded in routine NHS data, and addition of dependency measures to any product definition system would thus involve collecting additional data from patient notes.

Outpatient health gains
Compared with the inpatient hospital product, there has been relatively little work on characterising the output of outpatient departments, or measuring the inputs. This is partly because of their relatively low total costs, and partly because of a belief that outpatient activity and output is intrinsically more homogenous than inpatient activity. With the increasing use of day-case surgical procedures, however, the distinction between the two is becoming more blurred, and there is a strong case for combining inpatient and outpatient product definition systems.

Attempts to define outpatient products have centred largely around diagnostic and procedure groupings, similar to hospital inpatient casemix groupings. The best known of these is the Ambulatory Visit Groups (AVGs) system developed by the Yale group responsible for DRGs.56 It divides outpatient episodes into 19 major diagnostic categories which are subdivided into a total of 571 AVGs. Ambulatory Patient Groupings (APGs) are similar to AVGs, but grouping is initially dependent on procedure, rather than diagnosis. Many other ambulatory casemix classification systems have been developed, and useful overviews have been done by Hutchinson57 and Arbitman.58 There has been some research comparing AVGs with DRGs for day-case procedures in the UK,59 but their use for outpatient classification has been limited by lack of diagnostic information. It is notable that their use in the United States in research applications and prospective payment formulas has been much less extensive than inpatient casemix classification systems,60 possibly because of the high cost of information gathering relative to the cost of treatment episodes being measured. NHS hospitals do not currently record diagnostic and procedural information routinely for outpatients. Read codes61 have been suggested for use in hospital ambulatory care. They tend to specify to a degree far more detailed than is needed for product pricing, however, and are not directly compatible with any of the current ambulatory grouper programmes. The UK National Casemix Office plans to develop an ambulatory casemix system, possibly based on APGs, in the near future.

Attaching costs to casemix groups
For casemix systems to be applied to the contracting process, broadly indicative cost estimates per category are needed to translate a diagnostic episode information into pounds and pence. In the US, and elsewhere, this has been done by calculating relative value weights from large cost databases.62 There are essentially two approaches to acquiring this data for the NHS and these will be discussed in relation to both HRGs and DRGs.

The first is to use weights from other countries, particularly the United States. US weights are derived from large routine data sets, the Maryland Hospital Cost Review Database being the best known. They can be divided into those based on charges and those based on actual costs, with each approach having advantages and disadvantages.63-65 The main problems with using US weights in the UK are their failure to include physician costs and their paucity of patients in day-case and long-stay care categories. In addition, it is likely that differences in clinical practice, and practices of patient cross-subsidisation in US hospitals give rise to additional problems of transfer. Since HRGs are an entirely indigenous product, no US derived weights are available, and translation of US ICD-9CM diagnostic codes to local equivalents would be necessary before US costs databases could be applied to HRGs.

The second approach is to derive weights...
from local cost information. Classically, two approaches to costing are described. The top down (or step down) approach, allocates total costs via cost centres such as pathology, radiology, etc., to homogenous groups of patients.

The bottom up approach, costs actual patient episodes and aggregates them to obtain a group estimate. The former is best exemplified by the Yale DRG cost allocation model, which has been used to derive cost weights in a number of countries using US proportional allocation data. The UK National Steering Group on Costing (NSGC) has recommended a predominately bottom up approach for HRG costing, with top down reconciliation at the specialty level. Two possible problems with their approach need to be highlighted. Firstly, any bottom up costing exercise requires considerable resource input over an extended period if reliable estimates are to be achieved, and this is likely to deter many hospitals from participating in the absence of significant immediate perceived benefits. Very few countries have attempted bottom up costing of the degree proposed by the NSGC, and the proposal might be overly ambitious, especially in view of the newness of cost accounting methods in the NHS. Secondly, pilot projects sponsored by the NSGC have opted for cost collection by “local condition grouping” rather than by standardised casemix measure, and data may prove difficult to combine and translate into HRG weights.

Some adaptation of US weights, possibly using local length of stay data and limited existing cost data from resource management sites, might be a useful stopgap until sufficient local cost data is collected.

Conclusions

Within the NHS internal market, there is some urgency in applying a uniform system for defining the hospital product if it is to function as a market at all. Even if the “market” approach to distributing health care funds, was dispensed with, however, there are still many benefits to be gained from having a system to measure hospital output. Clearly, what is immediately feasible, and what is optimal, are two very different entities. Possible short and long term approaches for the NHS are suggested.

In the short term, purchasers and providers need to be encouraged, possibly by means of the offer of centrally funded support staff, to experiment with using HRGs instead of specialty groupings in contracts. Some form of indicative cost weights, possibly in the form of a US/UK hybrid would need to be distributed to facilitate this exercise. Early use of casemix methods is also likely to spur improvements in data coding, which will in turn increase acceptance of the results generated by the casemix approach.

In the long term, assuming the continuation of the internal market in its current form, a strong case needs to be made for a universally applied prospectively determined payment system based on diagnostic casemix categories that cover acute, long stay, and ambulatory care episodes in hospitals. Essentially, any of the well known routine data casemix measures would be suitable starting points for this exercise. Because of their continuing refinement within the UK, however, HRGs led themselves to adaptation better than other groups. A number of modifications to the existing HRG system could be suggested. Firstly, specialties not currently covered by HRGs (that is, oncology, radiotherapy, acute psychiatry, and others) should be included. Attention also needs to be given to ambulatory groupings, preferably in a joint in/outpatient classification system. These should be able to group all outpatient consultations as well as day case procedures. For such a system to be widely accepted, considerable attention needs to be paid to keeping data collection requirements, and the number of casemix groups, to a minimum.

Long-stay casemix categories would substantially improve the predictive abilities of HRGs. Long-stay patients are inextricably intermingled with the acutely ill NHS hospital population and the addition of a basic ADL score to the minimum dataset would seem to offer substantial information gains. A common system of measuring similar patient episodes for different types of institutions might encourage purchasers in areas with an under supply of basic hospital care to contract for long stay care in acute institutions without fear of paying excessively for it.

Improvement of the resource use and length of stay validity of HRGs should be a continuing process. One possible approach would be to collapse down similar categories in the existing system, with subsequent subdivision on the basis of factors other than medical diagnosis, for example dependency or age. Lastly, the production of up to date, good quality cost data will greatly encourage universal utilisation.

The success of any product definition drive within the NHS is likely to be sensitive to a number of exogenous factors as well. These include the availability of skilled and empowered management, both as purchasers and providers, and a more consistent and considered approach from the NHS Management Executive. Department of Health directives since 1991 have contained a myriad of conflicting instructions simultaneously freeing and over-regulating the health care market. This policy of one step forward and one back does little to encourage creative thinking on the part of those working in the service. All this has led to a market which is, in most cases, not a market at all, except in respect of its burgeoning transaction costs. While product definition is arguably a necessity under any system that emerges, confusion as to the future health service structure discourages essential long term thinking about measuring NHS hospital output.

A number of generic conclusions with regard to product definition in industrialised countries may also be drawn from the experience in the UK and elsewhere. Output classifi-
cation systems are numerous, and extensive testing of each and every product is neither feasible or advisable. In any event, most systems will require local adaptation because of variations in coding, financing arrangements, clinical practice, and intended use, so adaptability is arguably the most important consideration. Where data are sufficient, the true casemix approach offers advantages in creating incentives for efficient use of hospital resources, while at the same time rewarding providers fairly according to the type of patients they treat. DRGs have proved the most popular choice, and have a substantial international body of evaluation and adaptation behind them. Where DRGs or other US casemix systems are considered inappropriate, the development of indigenous systems may be considered. Such an approach brings with it substantial extra costs, however, not least of which are those associated with validating and costing the categories developed. The potential for international comparisons is also lost when this approach is adopted. It may take some time for a new locally derived casemix system to achieve the explanatory power of the imported equivalents it was intended to replace, although greater adaptability to local needs is likely to make up for this in the long term.

The author wishes to thank Ruairidh Milne, Alastair Gray, and James Richardson for their comments and suggestions on an early draft of this paper. Thanks also to Hugh Sanderson, Phil Anthony, Tom McGuire, and Martin McKee who provided valuable advice and/or unpublished information.


20. Barros FC, Vaughan JP, Victoria CG. Why so many caesar-


27. McKee M, Petticrew M. Disease staging – a casemix system for purchasers. London: Health Services Research Unit, London School of Hygiene and Tropical Medicine, 1992.


Product definition for healthcare contracting


