Casemix classification systems

Robert B Fetter

Robert B Fetter is Professor Emeritus, Yale University.

Abstract

The idea of using casemix classification to manage hospital services is not new, but has been limited by available technology. It was not until after the introduction of Medicare in the United States in 1965 that serious attempts were made to measure hospital production in order to contain spiralling costs. This resulted in a system of casemix classification known as diagnosis related groups (DRGs). This paper traces the development of DRGs and their evolution from the initial version to the All Patient Refined DRGs developed in 1991.

Introduction

Casemix classification was first suggested by Florence Nightingale in 1852. She stated that a system for categorising cases needed to be developed in order to analyse the costs and benefits of the treatment given patients with differing illnesses. She believed that it was necessary to classify cases so that they could be grouped according to illness, and treatment patterns studied. Unfortunately, the technical capabilities of the time precluded the kind of system she had in mind.

It was not until the early 1900s that practical attempts were made to further this notion. Dr Eugene Codman (1914, p 491) of the Massachusetts General Hospital and the Harvard Medical School advanced the notion in 1914 that:

> *... the whole hospital problem rests on this one question: What happens to the cases? We must formulate some method of hospital report showing as nearly as possible what are the results of the treatment obtained at different institutions.*

In his book on diseases of the shoulder, Dr Codman (1934) included inside the front cover a matrix with pathologies on one axis and treatment procedures on the other. In each cell, therefore, was a collection of records which could be examined to better

understand the treatment process, including its cost and efficacy. He believed that each case needed to be followed up in order to demonstrate whether or not the treatment applied had had the desired result. He asked physicians to record their expectations on the patient record. As might be imagined, Dr Codman was not popular in the Boston medical community. The idea was not followed up with any vigour, except by Dr Codman himself. His ideas led to his resignation from the staff of the Massachusetts General Hospital and the founding of the Codman Hospital in Boston.

In 1965 Professor Martin Feldstein (1965) demonstrated the central role of casemix in explaining differences in hospital cost behaviour, although his model was too simplistic to be of any practical utility. Advances in technology and skyrocketing health care costs following the introduction of Medicare in 1965 provided the means and the necessity to recommence such efforts.

The casemix approach to managing hospital services differs from the traditional approach in that it focuses on the final product of the hospital – the bundle of goods and services provided to a patient with a particular illness – rather than on the individual services as ends in themselves. This allows the effective utilisation of services to be taken into account, as well as the efficiency of their production (see Figure 1).

During the 1970s the cost of health care in the United States grew at a rate of more than 15% per year, more than twice the rate of general inflation. Spending on health services as a proportion of gross national product was 7.4% in 1970 and rose to 10.7% in 1983 and to more than 15% in the 1990s. The United States leads in per capita expenditures on health care, the rate being 40% higher than that of Canada, 60% higher than Sweden, 90% higher than Norway, and 120% higher than Finland and Australia.

At the same time, there are wide disparities in these expenditures within the United States, as well as different rates of service, and very large, and, during this period, unexplained differences in hospital costs. No statistics of public health indicated that our population was better off than those of other western countries. At Yale University we addressed the task of developing a means by which the output of the health care sector could be measured and evaluated. While we began this effort in 1967, it was not until 1975 that we received explicit funding for this work from the Social Security Administration. As the agency responsible for the Medicare program, their concern was to help develop a system for paying hospitals based on an explicit product or case, each case type representing a distinctive process of care. During this period we also received support from the Bureau of Quality Assurance of the United States Government to develop applications of casemix classification in quality assurance. Our initial efforts, for which I served as principal investigator, included John D Thompson of Yale's Department of Public Health, and several MS/OR graduate students, notably Ronald E Mills and Richard F Averill. We were joined in 1976 by Jean L Freeman and in 1980 by Professor George Palmer.

The origin of DRGs

In 1967 a group of physicians at the local university hospital asked for help with a problem in utilisation review. At the time, two years after the advent of the Medicare program, all hospitals were required to operate utilisation review and quality assurance programs as conditions for receiving Medicare payments. The physicians asked whether or not industrial methods of cost and quality control could be adapted and applied to the hospital setting. Thus began our efforts to measure hospital production as a means of evaluating what takes place in these settings. The ultimate result was the system of casemix classification known as diagnosis related groups (DRGs).

The first problem which we had to address was how to measure productive activity in a hospital. Hospital production is not the same as production in a factory, where there are usually clear standards and criteria pertaining to the utilisation of material, labour and equipment. In manufacturing, design quality, production quality and performance are subject to continual monitoring, measurement, feedback and adaptation of the processes of production. The physicians at the local hospital were asking us, in effect, how to begin a process of measurement and evaluation that would improve the processes and, simultaneously, the efficiency and effectiveness of the utilisation of resources.

Once one has a means of measuring performance, one can develop a system for understanding, predicting and ultimately controlling the processes of production.

In 1969, the year we first seriously attempted to classify patients, we lacked a basic system for interpreting clinical data. A two-year digression was necessary to develop the technology that would allow us to analyse hundreds of thousands of patient records (Mills et al. 1976). This technology enabled us to look at these records statistically and clinically so as to discern relatively homogeneous processes of care.

We initially convened a panel of physicians and asked them to describe care processes in terms of their important elements. The result was many thousands of different types of patients (Fetter 1969). At some level every patient is unique, just as every item produced by a factory is unique. Nevertheless, we needed to identify similarities rather than point out differences. We needed to discover a structure that would enable us to measure and evaluate the activities that take place in a hospital.

Our problem was to identify the ordinary, the usual and the routine, and then, applying the techniques of statistical process control, to filter out and examine the aberrant cases to understand the causes of the aberrations.

The extension of DRGs to reimbursement

Although DRGs were created to serve managers as a tool in running hospitals, the Federal Government, seeing the potential of prospective payment schemes in better understanding and restraining soaring health care costs, supported DRG research, and we directed our efforts simultaneously at cost and reimbursement control (Fetter, Thompson & Mills 1976).

In the late 1960s and early 1970s it was difficult to obtain support for a project aimed at defining hospital products. Most people in the hospital business found this notion, at best, uninteresting; the information we sought was not what they wanted to know. Since hospitals were reimbursed for their costs whatever they might be, there was little incentive to investigate cost and quality trade-offs.

We received funding from the Social Security Administration which, in 1974 and 1975, was investigating methods of containing health care costs. In the office of the Social Security Administration's director of research, we discussed the ability of case-based payments to achieve the mandate of the Medicare law to pay hospitals the cost of producing the required services.

The Social Security Administration sought the simplest regulatory mechanism that could substitute for the absence of an open market (a market based on public information about costs and quality) in health care. The notion of a regulatory method to ensure a minimum level of performance stems from the lack of a market that establishes value among providers. Health consumers not only do not have the information upon which to assess value and quality but, in most cases, do not pay directly for the services they receive. This lack of information and absence of incentive to minimise costs distort the market forces that operate effectively in other industries. DRGs could establish a rate of payment that would discourage hospitals from producing at a higher cost than an open market would tolerate.

However, to ensure that a cap on revenue would not result in low-quality output, we would need to establish some kind of peer review mechanism to assure a minimum level of quality. This would have the effect of identifying providers who attempted to produce lower quality services than would be acceptable if information were available. Most providers would actually operate to produce services analogous to those that would be found in a marketplace limiting the scope and cost of operations. This is the idea behind DRG-based prospective payment systems.

Defining the products of a hospital

The major function of a hospital is to provide the diagnostic and therapeutic services required by physicians in the clinical management of their patients. In doing so, the hospital also provides certain hotel and social services. Unlike many other kinds of enterprises, a hospital actually consists of two separate, separable production functions (Figure 1) (Fetter & Freeman 1986).

The first function is to convert raw materials (labour, supplies, equipment) into standard outputs (meals, clean linen, laboratory procedures, medications). But these outputs do not constitute the real business of the hospital, namely, caring for patients. They are really intermediate products. In the traditional organisational structure of a hospital, the line organisation consists of layers of administration that ultimately manage the various service departments. The implication is that somehow by managing these departments you will end up managing the institution.

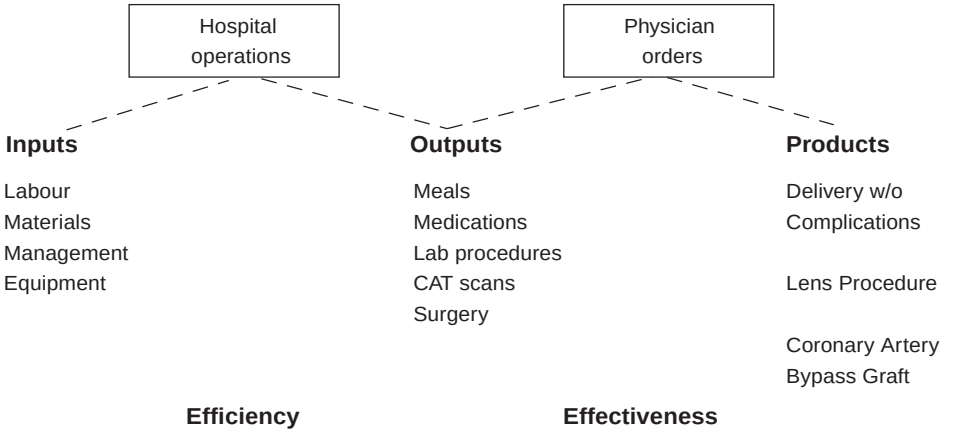


Figure 1: The products of a hospital

The second function and the main business of the institution is to accept, one at a time, human beings who have a problem, a disease or a disorder, and to evaluate and treat, through physicians and other professionals, the problem and the patient. Under the direction of these professionals, the institution provides a set of goods and services deemed appropriate to the diagnosis and treatment of the illness. It is this bundle of things that we define as the product of the hospital.

The labour, supplies and equipment used in a hospital are similar to the bill of materials for a chair in the manufacturing environment – arms, legs, rails, cushions, nails, pegs, screws, and so on. Although there are obvious differences between a health services institution and a chair factory, Levitt (1972, p 42) has pointed out the importance of the analogy:

> *So many things go wrong because companies fail to adequately define what they sell. Companies in so called service industries generally think of themselves as offering services rather than manufacturing products; hence they fail to think and act as comprehensively as do manufacturing companies concerned with the efficient, low cost production of customer-satisfying products.*

The main problem in managing a hospital is to separate issues of *efficiency* in the production of intermediate products from issues of *effectiveness* in the utilisation of these intermediate products. To manage an institution competently, the issues and measures of differential efficiency and differential effectiveness must be clearly separated. It is quite important to produce each laboratory test efficiently in the sense of utilising a standard set of inputs for each output. If, however, the test is not used effectively or is ordered inappropriately, it is a waste of resources no matter how efficiently it is produced. Most of the observable differences in hospitals' performances are a function of differential effectiveness in utilising services; these differences have little to do with the relative efficiency with which the hospitals produce intermediate products.

To focus simply on the intermediate products is to miss the point of the enterprise: to treat patients who have illnesses. The effective utilisation of a hospital's resources is primarily a function of its ability to treat specific kinds of illnesses.

To evaluate, compare and provide feedback regarding a hospital's performance, one must identify the specific products they provide. If one can identify classes of patients with the same clinical attributes and who are treated by similar processes of care, then one has established the framework needed to aggregate patients into case types.

Major obstacles

To make the method work, one must identify the cost of delivering services and let managers know where they stand in respect to the estimated cost established. An episode of illness must be defined in a manner that permits a reasonably accurate prediction of the goods and services to be delivered. In attempting to do so, we encountered three major obstacles:

1. Not all diseases are equally well understood. Physicians find some diseases easy to define while they find it is difficult even to reach agreement on a label for others.
2. The treatments provided for the same diseases often differ, which makes accounting for the treatment chosen very difficult. Physicians understand some diseases very well and have established consensus, at least locally, as to how they ought to be treated. The treatment of many diseases, however, varies among physicians. Significant barriers exist to defining norms in the practice of medicine for many types of illness.
3. Coding illnesses presents difficulties, that is, deciding what descriptive labels to use. If we look at the codes of the International Classification of Diseases-Ninth Revision-Clinical Modification (ICD-9-CM) (United States Department of Health and Human Services 1980), there are areas that, from the point of view of the process of care, are significantly over-identified (for example, cataracts are described in almost 40 different ways). In other areas (for example, cerebrovascular accident) one label is used to describe patients receiving a broad spectrum of treatment regimens.

The DRG classification system was developed in spite of these obstacles. Variability is bound to occur in any system. Variability in the utilisation of goods and services in treating a given illness is not itself the problem. The problem is to establish a mechanism for predicting the variability. If the variability can be predicted, the process can be managed. The problem of analysing resource consumption in a hospital therefore centres on discovering the illnesses for which stability in the utilisation of resources, at least at the institutional level, can be determined.

Guidelines for grouping diagnoses

We had to discover stable patterns of resource utilisation using standard, commonly collected review data. We therefore chose to use the Uniform Hospital Discharge Data Set (UHDDS), a standardised set of demographic, diagnostic and procedural data that hospitals are required by law to produce for each hospitalised patient in the United States (National Center for Health Statistics 1980). Practical utility was a second requirement of the classification: a given case type had to occur with enough frequency to merit formation of a separate group. In addition to requiring statistical stability of resource utilisation within a class, we recognised the need to ensure that patients in a given class comprised a clinically coherent group. Otherwise physicians would reject the classification.

In sum, we used the following four characteristics as a guide in developing the DRGs (Fetter et al. 1980):

- class definitions based on information routinely collected by hospitals (the UHDDS)
- a manageable number of classes
- similar patterns of resource intensity within a given class
- similar types of patients in a given class from a clinical perspective.

DRGs have been revised in response to changes in disease-coding and procedure-coding schemes, to new conceptual models of health service utilisation, and to feedback from the health care community regarding both clinical interpretability and statistical evaluations of resource use.

Developing the structure of the DRG classification: An organ systems approach

The DRG classification scheme uses an organ systems approach insofar as this is possible (Averill 1991). This approach divided the classification process into two major steps. The first step considered the diagnosis. We organised ICD-9-CM codes representing diseases or disorders that could serve as a patient's principal diagnosis (that is, illnesses which would bring patients to acute care hospitals) by medical specialties or organ system involvement. This resulted in a set of 23 mutually exclusive and exhaustive categories that were called major diagnostic categories (MDCs) (Table 1). The 23 MDCs included about 10 000 individual ICD-9-CM diagnosis codes.

Our second step in the process of classification was to identify from the surgical codes of ICD-9-CM, codes for procedures that require acute care hospital surgical facilities. A patient undergoing such a procedure was classified into one of 22 'surgical MDCs' based on the organ system inferred from the principal diagnosis (one MDC, 20, does not include any patients who undergo surgical procedures). Any record of patient care in an acute care hospital, by virtue of its diagnoses and procedural codes, can be placed into one of these 45 classes.

Table 1: Major diagnostic categories

1.	Diseases and disorders of the nervous system
2.	Diseases and disorders of the eye
3.	Diseases and disorders of the ear, nose and throat
4.	Diseases and disorders of the respiratory system
5.	Diseases and disorders of the circulatory system
6.	Diseases and disorders of the digestive system
7.	Diseases and disorders of the hepatobiliary system and pancreas
8.	Diseases and disorders of the musculoskeletal system and connective tissue
9.	Diseases and disorders of the skin, subcutaneous tissue and breast
10.	Endocrine, nutritional and metabolic diseases and disorders
11.	Diseases and disorders of the kidney and the urinary tract
12.	Diseases and disorders of the male reproductive system
13.	Diseases and disorders of the female reproductive system
14.	Pregnancy, childbirth and the pueperium
15	Newborns and other neonates with conditions originating in the perinatal period
16.	Diseases and disorders of blood and blood forming organs and immunological disorders
17.	Myeloproliferative diseases and disorders, and poorly differentiated neoplasms
18.	Infectious and parasitic diseases (systemic or unspecified sites)
19.	Mental diseases and disorders
20.	Alcohol/drug use and alcohol/drug induced organic mental disorders
21.	Injuries, poisonings and toxic effect of drugs
22.	Burns
23.	Factors influencing health status and other contacts with health services

The third step we took was to examine the process of care by considering all variables available in the hospital discharge abstract. This involved taking all the records of patients in a given MDC who had surgery, for example, and submitting them first to a statistical clustering algorithm (Mills et al. 1976) to establish common patterns of resource consumption, and second to a group of clinicians who were asked to attempt to determine the fundamental clinical organising principles underlying the results.

Our final step was to investigate diagnoses representing problems that would influence the treatment process but were not the fundamental problem that brought the patient to the hospital. This exercise was done with 1979 data, which were less than complete with regard to the coding of additional diagnoses. Many of our hypotheses about what ought to make a difference could not be supported in the data. However, we were able

to isolate some clusters of secondary conditions that were important in identifying differential treatment. For the most part, however, we had to rely upon generic identification of secondary problems. When all else failed, we used age. Age is not the best indicator of state of health since some old people are very healthy and some young people are unhealthy, but at the time we were forced to use it. Because the coding of patient health has improved and now includes co-morbidities and complications, Medicare has recently dropped age as a variable in DRG classification. Recently, this fourth level has been improved through detailed analysis of new data. The result is a differentiation of secondary diagnoses into classes which, in interaction with the principal illness or procedure, require different sets of resources.

The result of our efforts is a set of product definitions for each MDC (Figure 2). It is important to look at the diagnosis and procedure codes which describe each of these classes because the names alone are not always descriptive.

The first level describes organ systems, the second level distinguishes between surgical and medical procedures, the third level describes a hierarchy of procedures and a hierarchy of medical problems, and the fourth level outlines other indicators that differentiate processes of care.

The evolution of DRGs since 1983

Figure 3 traces the evolution of DRGs from the original grouper used by the Health Care Financing Administration (HCFA) for the Medicare hospital payment system, HCFA version 1, which was developed at Yale in 1980–81 (Fetter 1984), and which is described above.

As Figure 3 shows, until 1988 annual updates to the DRGs were relatively minor, mainly addressing changes in medical practice and coding. In 1988 the 'CC exclusion' lists were adopted. They modified the complications and co-morbidities (CC) generic list for specific DRGs so as to deal with the problem of secondary diagnoses, which are manifestations of some given principal diagnosis and, in that instance, not as serious as when unrelated to the principal diagnosis. At the same time, HCFA funded a project at Yale to attempt to refine the use of CCs based on the more complete data which hospitals had produced since the introduction of the prospective payment system (Fetter 1991). Also, New York State had determined that HCFA DRGs were no longer representative of the entire patient population, especially as concerns newborns and children.

Thus in 1989 the New York DRG grouper was introduced, adding the following features:

- HIV infection MDC
- multiple trauma MDC
- newborn DRGs based on birth weight

- new DRGs for
  - cystic fibrosis
  - lead poisoning
  - paediatric patients (18)
  - high risk obstetrical patients (3)
- tracheostomy DRGs.

This grouper is now known as the All Patient (AP) DRG and is used for payment purposes in several states in addition to New York. Additional features have been added since 1989, such that version 10.0 for use in 1993 contains 617 DRGs (3M/Health Information Systems 1992b).

The Yale refinement project used the HCFA grouper version 4, but completely restructured the fourth level by eliminating all partitions beyond the medical groups based on principal diagnosis and the surgical groups based on principal procedure. The CC list was organised into 124 categories based on the same logic as that used for the medical group definitions, and then for each medical group clustered these CC categories into three classes – those with no or a minor effect, those with a moderate effect, and those with a major influence on cost of care. A minor effect was defined as

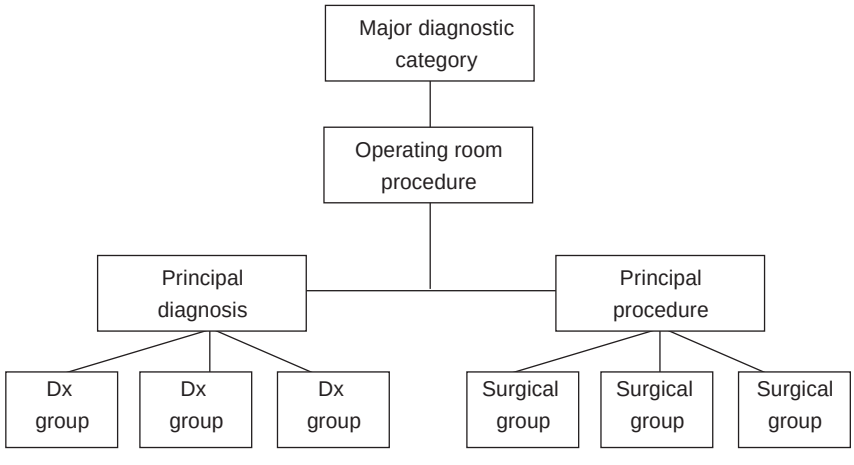


Figure 2: Product definitions for each major diagnostic category

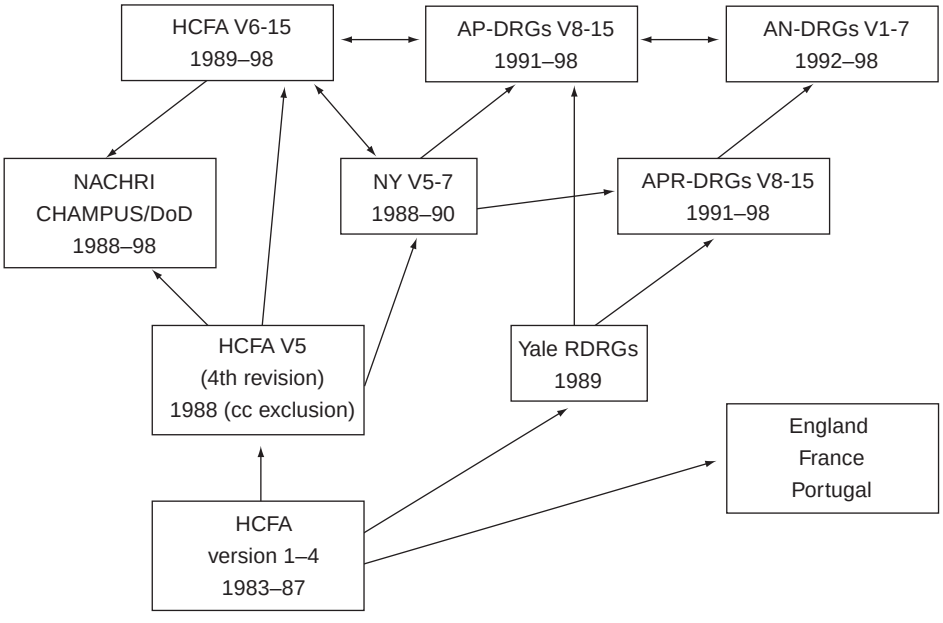


Figure 3: Evolution of DRGs

involving less than one extra day of care or $300 in cost; moderate as more than a day but less than a week or $2000 in cost; and major as in excess of that limit. On the surgical side, the CC categories were organised into four classes for each surgical group – the three defined above, plus a fourth category for extreme or catastrophic effect.

The result was a structure much simpler than the original or AP-DRG logic, as shown in Figure 4. Improvement in explanation of cost of care was obtained for every MDC on both the surgical and medical side, with the overall explanation almost 50% of variance in the 1986 Medicare hospital claims file.

Based on the AP-DRG improvements and the Yale refined DRGs, 3M/Health Information Systems developed in 1991 the All Patient Refined DRGs (APR-DRGs) (3M/Health Information Systems 1992c). As a first step, the AP-DRG model was consolidated to eliminate all splits based on CCs, death and age, plus any groups split based on complicated/uncomplicated principal diagnosis. Then *all* diagnoses in the ICD-9-CM code book were assigned to one of four complexity subclasses as follows:

- minor
- moderate
- major
- extreme.

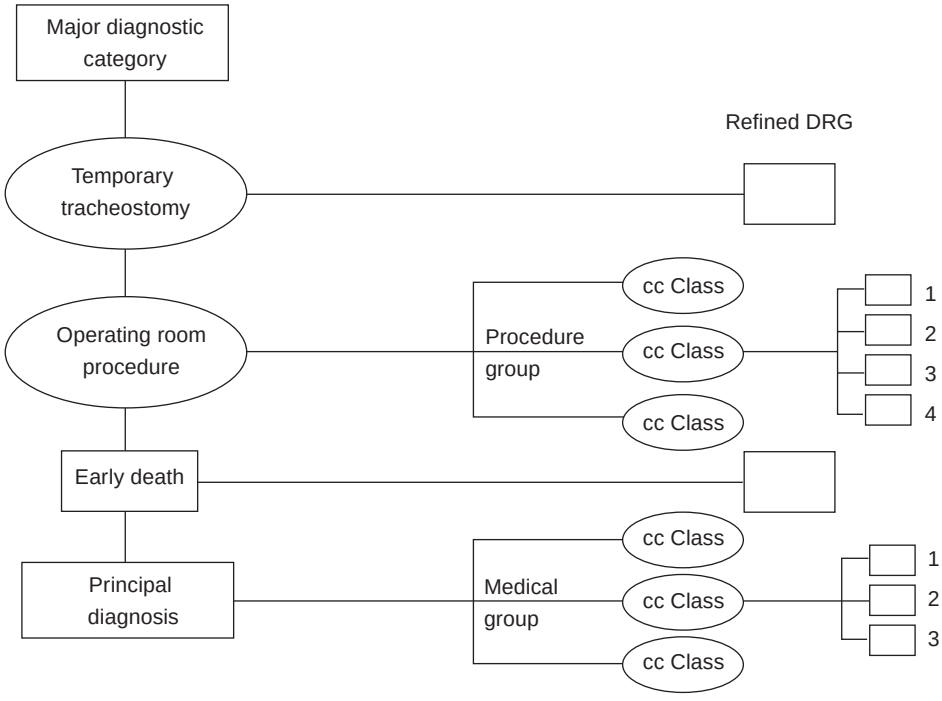


Figure 4: Structure of refined DRG model

These are known as 'default subclasses' and are used initially to partition each of the 337 consolidated AP-DRGs defined above. This means that for the first time since the introduction of DRGs, the CC list has been completely redefined. As a consequence, for example, 1448 diagnoses not previously considered a CC (subclass 1) became a CC (subclass 2, 3, 4). Also, 481 diagnoses moved from CC (subclass 2, 3, 4) to non-CC (subclass 1) status.

Then, to improve the subclass assignment process, a number of modifications were made for some consolidated DRGs based on:

- consolidated AP-DRG
- age
- non-operating room procedures
- principal diagnosis
- additional secondary diagnoses.

Thus the final grouping logic for subclass assignment based on secondary diagnoses is as shown in Figure 5. This results in a model with 1380 groups, but one which is simpler in structure than any other, easier to understand, more appealing to clinicians, and which produces a 75% improvement in explanation of resource consumption over HCFA version 6.

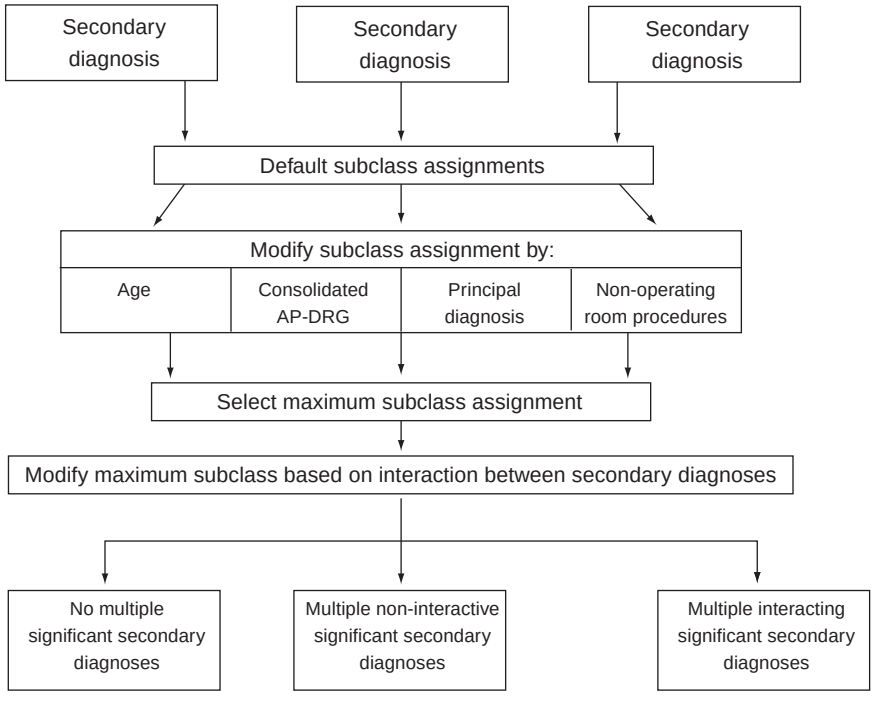


Figure 5: APR-DRG subclass assignment

Table 2 shows the results for consolidated AP-DRG 1, Craniotomy Except for Trauma, in 1989 Maryland data. In HCFA version 6, the complexity subclasses 2, 3 and 4 would be aggregated into one class. The APR-DRGs are being used by some states for payment purposes but will, I believe, find their greatest use initially in developing comparative reports at the hospital level, and in comparing hospital performance.

Table 2: Consolidated AP-DRG 1, Craniotomy Except for Trauma, 1989 Maryland data

Class	Count	Average length of stay	Average charges
4	602	34.36	35504
3	139	17.76	17143
2	417	13.58	12308
1	222	10.22	10207
Total	**1330**	**16.72**	**18684**

The severity problem and other casemix classification systems

The initial version of DRGs was criticised by many for addressing the 'severity' issue in too crude a manner. It was felt that the system as then defined did not differentiate sufficiently among patients with different degrees of illness in the same class. We believe that the issue is better defined as that of 'casemix complexity' and its impact on cost of care. This may be described in a number of ways, as follows.

Severity of illness refers to the relative levels of loss of function and mortality that may be experienced by patients with a particular disease.

Prognosis refers to the probable outcome of an illness, including the likelihood of improvement or deterioration in the severity of the illness, the likelihood of recurrence, and the probable life span.

Treatment difficulty refers to the patient management problems that a particular illness presents to the health care provider. Such management problems are associated with illnesses without a clear pattern of symptoms, illnesses requiring sophisticated and technically difficult procedures, and illnesses requiring close monitoring and supervision.

Need for intervention relates to the severity of illness that lack of immediate or continuing care would produce.

Resource intensity refers to the relative volume and types of diagnostic, therapeutic and bed services used in the management of a particular illness.

DRGs consider resource intensity as the major effect to be measured by casemix. Their purpose is to relate a hospital's casemix to the resource demands and associated costs experienced by the hospital.

A recent study by MacKenzie (1991) at Queen's University in Canada and one by Miller at The University of Florida have compared eight 'severity indexing systems' to assess their predictive validity and reliability. The Voluntary Hospital Association of America (1992) has recently published a summary of these findings, which concludes that 'APR-DRGs is (sic) as good as or better than any system available on the market as a predictor of cost and resource consumption'. Also, they concluded that 'APR-DRGs performed as well as generic clinical data systems and better than diagnosis-dependent systems in predicting mortality'. The systems evaluated were all developed as alternatives to DRGs and included:

- Disease Staging
- Patient Management Categories
- Acuity Index Method
- APACHE II
- MedisGroups
- Computerized Severity Index

- Body Systems Count
- APR-DRGs.

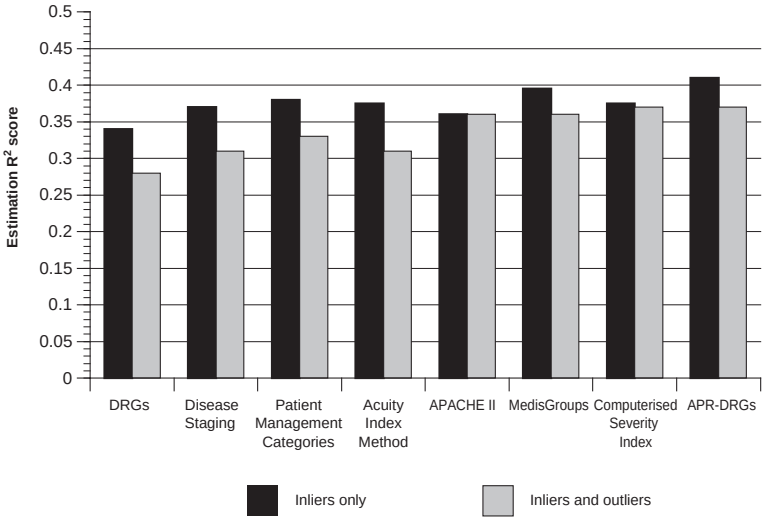
Disease Staging attempts to predict morbidity and mortality within the context of degrees of illness among specific diseases. It focuses on approximately 400 diseases that are common in United States short-term, acute care hospitals. Most of these diseases are separated into four main stages using a rubric borrowed from the staging concept employed in oncology. The stages reflect increasing disease severity as it relates to the risk of death. Stage 1 defines patients with no complications or problems of minimal severity; stage 2 identifies problems limited to an organ or organ system with significantly increased risk for complications; stage 3 patients are those with multiple site involvement, generalised systemic involvement, and poor prognosis; stage 4 includes only patients who die. Assignment is based on UHDDS data and is retrospective.

Patient Management Categories (PMCs) consider severity of illness to be a disease-specific concept. A computer algorithm uses standard discharge abstract data to assign a patient to one or more PMCs depending on the clinical relationships of the diagnostic codes used to describe the illness without regard to order. The PMCs were defined by panels of physicians who attempted to describe illnesses and their management based solely on clinical criteria and judgement. Multiple PMCs may be assigned since the principal diagnosis is not used as the major assignment variable, and multiple diagnoses may identify more than one patient management path.

The Acuity Index Method attempts to explain variation in length of stay within each DRG. Given the DRG assignment, the Acuity Index Method algorithm subdivides each DRG into subgroups based on the relationships between pathological processes identified as secondary diagnoses/procedures and pathological processes identified as the principal diagnosis/procedure. The assignment is retrospective and uses only UHDDS data.

APACHE II, the Acute Physiology and Chronic Health Evaluation II system, attempts to predict risk for in-hospital death for critically ill patients. The method assigns a score from 0 to 71 to a case, using information on age, chronic health status and 12 physiological measurements: temperature, mean arterial pressure, heart rate, respiratory rate, arterial oxygenation, arterial pH, serum potassium, serum sodium, serum creatinine, hematocrit, leukocyte count, and Glasgow Coma Score. The data must be collected from primary sources, including the medical record. The score is determined from the most abnormal values for each of the 12 measurements within the first 24 hours of a patient's stay in intensive care.

MedisGroups produces severity scores between 0 and 4 that are independent of patient diagnosis and are intended to predict the risk of imminent organ failure. Either two or three reviews are conducted for each patient stay, one at admission, one at discharge and another in between if the length of stay is more than six days. The admission review records both acute conditions and also more chronic conditions that may affect clinical status. The mid-stay review identifies complications that arise acutely during the hospital



Notes: Prediction of death (30–60 days post discharge) – All systems performed well. APACHE, MedisGroups, Computerised Severity Index and APR-DRGs led the group and were not statistically distinguishable from one another in measuring prediction of death.

Thus, APR-DRGs performed as well as generic clinical data systems and better than diagnosis-dependent systems in predicting mortality.

Figure 6: The ability of various systems to account for variations in resource consumption

stay. Up to 260 possible key findings are identified, representing clinical abnormalities drawn from patient history, physical examination, laboratory, pathology, radiology and other findings from the medical record. As the system is independent of diagnosis, the abstractor must look for any key clinical findings. Between 5 and 10 such findings are identified in the average review.

The *Computerised Severity Index* assigns severity scores between 0 and 4 based on the extent of the primary diagnosis and its interaction with other disease processes. This score is assigned to each diagnosis on the record and then computes an overall score based on interactions among the diagnoses. In computing this overall score, the Computerised Severity Index tracks more than 600 clinical characteristics encompassing all patient conditions. These clinical indicators must be abstracted from the hospital medical record. Included are patient history, symptoms, physical examinations, vital signs, laboratory tests, radiology studies, and the like. The software asks 32 questions about the average patient.

The *Body Systems Count* method assigns severity based on the number of body systems requiring treatment during a hospitalisation. Records are assigned retrospectively to one of three levels – high, moderate and low. The measure is determined by assigning each diagnosis on the record to an MDC as determined by the DRG grouper. The number of unique MDCs assigned represents the body systems count.

Figure 6 shows the results of the evaluation of these systems which was conducted at Queen's University (MacKenzie 1991). Based on these results as well as its own evaluation, the Voluntary Hospital Association of America (1992) has stated that the use of APR-DRGs 'will strengthen the ability of hospitals and physicians to use CFIS (Clinical/Financial Information Systems) to critically contrast resource appropriation and efficiency of practice styles'. In addition, they concluded that APR-DRGs show the greatest potential for reimbursement equity.

Research and development work on casemix classification continues today, much of it under the auspices of 3M/Health Information Systems. In addition, casemix classification of ambulatory and geriatric patients has seen continuing efforts. It is to be hoped that ultimately a casemix system which is capable of representing episodes of care across multiple settings can be developed. This is the future of casemix classification.

References

Averill RF 1991, 'Development of the DRG patient classification system', Chapter 2 *in* RB Fetter (ed) *DRGs: Their Design and Development,* Health Administration Press, Ann Arbor, Michigan.

Codman EA 1914, 'The product of a hospital', *Surgery, Gynecology and Obstetrics*, vol 18.

Codman EA 1934, *The Shoulder,* T Todd Company, Boston.

Ericson WA 1964, 'A note on partitioning for maximum between sum of squares', Appendix C *in* A Sonquist & JN Morgan, *The Detection of Interaction Effect*, Institute for Social Research, University of Michigan, Ann Arbor, Michigan.

Feder J, Hadley J & Zuckerman S 1987, 'How did Medicare's prospective payment system affect hospitals?' *New England Journal of Medicine*, vol 317, pp 867–73.

Feldstein MS 1965, 'Hospital cost variation and case-mix differences', *Medical Care,* vol 3, pp 95–103.

Fetter RB 1969, 'Financing health care: Discussion', *in* GK Chacko (ed) *The Recognition of Systems in Health Services*, Operations Research Society of America, Health Applications Section, pp 209–21.

Fetter RB 1981, *Development, Testing and Evaluation of a Case Payment Reimbursement System,* Final report, Social Security Administration, Health Care Financing Administration Contract No 600-75-0180, Health Systems Management Group, School of Organization and Management, Yale University (January).

Fetter RB 1984, *The New ICD-9-CM Diagnosis Related Groups Classification Scheme,* Health Care Financing Administration Grants and Contract Report, HCFA No 01367, Washington DC.

Fetter RB (ed) 1991, *DRGs: Their Design and Development*, Health Administration Press, Ann Arbor, Michigan.

Fetter RB & Freeman JL 1986, 'Diagnosis related groups: Product line management within hospitals', *Academy of Management Review,* vol 11, no 1, pp 41–54.

Fetter RB & Freeman JL 1989, *DRG Refinement with Diagnostic Specific Comorbidities and Complications: A Synthesis of Current Approaches to Patient Classification,* Final Report of Cooperative Agreements Nos 15-C-98930/1-01 and 17-C-98930/1-0251 between HCFA and Yale University.

Fetter RB, Shin Y, Freeman JL, Averill RF & Thompson JD 1980, 'Case mix definition by diagnosis related groups', *Medical Care,* vol 18, no 2 (supplement), pp 1–53.

Fetter RB, Thompson JD & Mills RE 1976, 'A system for cost and reimbursement control in hospitals', *Yale Journal of Biology and Medicine*, vol 49, pp 123–36.

Guterman S & Dobson A 1986, 'Impact of the Medicare prospective payment system for hospitals', *Health Care Financing Review,* vol 7, no 3, pp 97–114.

Guterman S, Eggers PW, Riley G, Greene TF & Terrell SA 1988, 'The first three years of Medicare's prospective payment: An overview', *Health Care Financing Review,* vol 9, no 3, pp 67–77.

Lave JR 1989, 'The effect of the Medicare prospective payment system', *Annual Review of Public Health,* vol 10, pp 141–61.

Levitt T 1972, 'Product line approach to services', *Harvard Business Review,* vol 50, no 5, September–October, pp 41–52.

Mills RE, Fetter RB, Riedel DC & Averill R 1976, 'AUTOGRP, An interactive computer system for the analysis of health care data', *Medical Care,* vol 14, no 7, pp 603–15.

National Center for Health Statistics 1980, *The Uniform Hospital Discharge Data Set,* Report of the National Committee on Vital and Health Statistics, Public Health Service Publication No. 80-1157, US Government Printing Office, Washington DC (April).

Rodrigues J-M 1989, 'L'Europe des DRG', *Sozial-und Praventivmedizin,* vol 34, pp 152–5.

Russell LB 1989, *Medicare's New Hospital Payment System, Is It Working?* The Brookings Institution, Washington DC.

Scheffler R & Gibbs J 1986, *Blue Cross Utilization and Payment Rates Before and After Prospective Payment,* Blue Cross and Blue Shield Association, Chicago, Illinois, December.

Schieber GJ & Poullier JP 1988, 'International health spending and utilization trends', *Health Affairs,* vol 7, no 4, pp 105–12.

Singer J, Karon S, Pendleton S & Bachman S 1986, *The Effects of Medicare PPS on Medicaid Programs,* Brandeis Health Policy Consortium, Waltham, Massachusetts, December.

3M/Health Information Systems 1992a, *Australian National DRGs, Version 1.0, Definitions Manual,* 3M/HIS, Wallingford, Connecticut.

3M/Health Information Systems 1992b *All Patient DRGs, Version 10.0, Definitions Manual,* 3M/HIS, Wallingford, Connecticut.

3M/Health Information Systems 1992c, *All Patient Refined DRGs, Version 7.0, Definitions Manual,* 3M/HIS, Wallingford, Connecticut.

Voluntary Hospital Association of America Inc. 1992, *Severity Indexing Systems,* Irving, Texas.

United States Department of Health and Human Services 1980, *International Classification of Diseases-Ninth Revision-Clinical Modification,* Public Health Service Publication 80-1260, US Government Printing Office, Washington DC.

Discussion

A discussion of the paper, 'Casemix classification systems', presented by Robert Fetter.

Jim Pearse

Jim Pearse is Director, Structural and Funding Policy, New South Wales Health.

It is fitting that Professor Fetter's paper begins with a reference to that eminent Victorian, Florence Nightingale. You may recall that in his short and amusing book, *Eminent Victorians,* Lytton Strachey selected Miss Nightingale as one of the four representatives of the Victorian age. When someone comes to write of the influential figures of the late twentieth century, Professor Fetter may well rate a mention. I find it hard to recall any other figure who has had such a profound effect on how health systems are organised and funded across the world over the last few decades. It is with a great deal of humility that I have accepted the invitation to discuss his paper.

Professor Fetter's paper provides us with an overview of the origins and development of the casemix concept, in particular diagnosis related groups (DRGs) classification. Presented in their bare bones, the concepts are simple:

- Casemix schemes quantify the final products of hospital services.
- Diagnosis related groups represent satisfactory classifications of final products and have evolved to address many of the initial problems.
- By being clear about the products of our services, we can be clear about the issues that require attention and the nature of the management task. By quantifying the final products of hospital services, we are able to improve:
  - the clinical management task
  - the hospital management task
  - comparison of hospital performance
  - methods of funding hospitals
  - methods of planning for hospital services.

To be accepted, good ideas need evangelists. I'm not quite sure I'd call George Palmer's lecturing style evangelical, but it has to be accepted that he has been one of the most effective promoters of the notion that casemix has a vital role to play in Australian health policy.

In my discussion I would like to address where Australia is positioned with casemix classification and some of the strategic directions that need to be adopted to move further ahead. My message is a relatively simple one: the casemix agenda remains a vitally important micro foundation to many of the system reforms that have been debated over the last decade.

The Commonwealth Government should count the National Casemix Development Program as one of the genuine successes of policy innovation. The program has laid the groundwork for many innovations.

As a result of the program, all Australian States and Territories, and the private sector, have available a common classification of acute hospital products. The extent and implication of this achievement is often vastly underestimated. Many other public sector services do not have a standardised method of accounting for their products, yet I believe it is an accurate reflection that the hospital and health services embody a level of complexity and variety that is rarely found in education, community services, corrections and justice. The fact that the classification is common across States is also important. All States now use the Australian national DRG system in their funding and performance comparison, including the so-called non-casemix State of New South Wales.

The program has sponsored excellent work on extending casemix principles outside the acute care setting into mental health, rehabilitation, palliative care, geriatric care and other non-acute care.

But there still remain gaps and challenges in this enterprise. The challenges can be represented across two dimensions: gaining more depth in our understanding of casemix and its potential in hospital care; and broadening the casemix enterprise to better understand the health system as a whole.

By depth, I mean further enhancing the application of DRGs in the acute care setting. There are two levels at which these issues need to be further tackled. As Professor Fetter points out, DRGs allow us to separate the issues of the efficient production of intermediate products from the effective management of the episode of care. But a technical impediment to this separation is the lack of robust and commonly agreed ways of classifying and costing intermediate products. The development of classifications and relative value units for intermediate products such as pathology, pharmacy and allied health may not at first sight appear an item for the national agenda, but it is essential if we are to understand the contribution of efficient production and effective ordering, and promote a more meaningful benchmarking framework.

Other ways in which national initiatives deepen our understanding of the hospital production process include the following.

- Developing an effective and efficient system for promoting the use of clinical pathways. In this respect we need to balance the imperative of local clinical ownership with the need to avoid duplicated effort.

- Investing in developing a framework where casemix information is used to investigate and understand medical practice variations, as a matter of routine.
- Getting a grip on the severity issue. The development of the Australian refined (AR)DRGs version 4 built on the lessons of the all patient refined DRG model Professor Fetter describes. However, I suspect we will still be challenged by the severity issue when AR-DRGs are introduced. An aspect of the issue is to develop a more sophisticated understanding of how clinical services are networked between hospitals, and the role referral hospitals play in that network.
- Promoting the linkage of routine hospital morbidity data with the various national disease treatment registers that exist across Australia. This linkage is one of the keys to unravelling the clinical quality question. The linkage is technically possible tomorrow. It is not a question of an additional data collection effort. The challenge is to carefully negotiate with the various professional bodies.

By broadening the casemix enterprise, I mean that casemix principles need to be pushed out to cover a wider set of health services. We need a comprehensive set of classifications of the products of health care services. The logic of this agenda is identical to the original logic of the DRG enterprise. If we don't understand our products in a systematic and comparable way, we cannot really hope to effectively manage our services. If we are genuine about the desire to shift the balance of care within the health system, to shape and tailor services to meet individual needs, then we need to understand the products we are attempting to manipulate.

Of course extension of casemix, particularly in the non-inpatient settings, poses a number of difficult questions about the nature of the product. We can identify at least three levels at which these products can be defined:

- the individual service event
- the episode of care
- the episode of illness.

In my view all three levels of product classification are required to effectively manage health care and develop more appropriate funding arrangements. The imperative should be to start building the foundations, first focusing on classifications at the service event and the episode of care. The various forms of budget-holding and coordinated care need these building blocks. Nationally we face a choice: to let the various individual groups attempt to grapple with these complex and sometimes daunting tasks; or to tackle these matters on a national basis.

Amongst my priorities for the national agenda I would include the following.

- Finalising a national outpatient classification. Once a national classification is agreed most States will implement the classification.
- Working to explore the practical issues to be engaged in implementing episode of care collections in ambulatory and community health services.

- Extending the episode of care to primary care, in particular general practice care.
- Setting up a national research program to develop a set of measures of functional status and dependence that can be applied across rehabilitation, geriatric assessment, palliative care, home and community care for most of these services, and most community care services. The set of measures should be hierarchically related and allow expansion to deal with the particular requirements of certain service streams. This research program would be quite expensive, and can't expect pay-offs for three to five years.

A key lesson to draw from the Australian experience to date is that concerted national efforts can be mounted, and these often represent the most efficient way to progress matters. A challenge is to properly engage all the relevant parties in the process, including the States and Territories and the professional bodies. Professor Fetter's paper reminds us of how far the casemix agenda has brought health policy over the last two decades. It is time for a re-affirmation of the importance of the National Casemix Development Program in Australian health policy development.

National leadership on coordination focused on this micro-level foundation has enormous potential to benefit and promote the national reform agenda.

Some day, some fool will attempt to write a book called 'Eminent Australians in Health Policy'. I'm sure Jeff Kennett would suggest the book be retitled 'Eminent Victorians'. You might have to agree, but for a few exceptions. George Palmer has to be one of those exceptions. Like an isolated and besieged General Gordon in the Sudan, he has had to fight the good fight for injecting a little science into policy. In many ways the National Casemix Development Program is one of the outcomes of his vision and persistence. Fortunately for us, unlike General Gordon, Professor Palmer has lived to see his efforts bear fruit. I wish him well for his retirement.