Casemix and nursing

DONNA DIERS

Donna Diers is Annie W Goodrich Professor, Yale School of Nursing, United States.

Abstract

The American Nurses’ Association did not embrace the introduction of diagnosis related groups, believing they would not recognise nursing activity nor acuity and would bring about the economic demise of nursing. Australian nurses, by contrast, recognised the window of opportunity that the work towards Australian national diagnosis related groups and funding mechanisms provided to move nursing resources into the political and policy mainstream. This paper reviews the American and Australian nursing experience with casemix, acuity and cost weighting. It uses examples from more recent work to argue for the use of casemix information in new ways, for ‘process improvement’ or ‘evidence-based management’. The paper concludes that the next great leap forward in casemix may require attention to building the information and human infrastructures, so that the valuable clinical–financial information produced by casemix-based information systems can truly inform management and policy.

Introduction

When the prospective payment system based on diagnosis related groups (DRGs) was being debated in the committees of the United States House and Senate, the American Nurses’ Association took a firmly uninformed position that DRGs were evil (Cole 1982). DRGs would not recognise nursing activity nor acuity, and they would be responsible for the economic demise of nursing. There were many, including Professor John Thompson (himself a nurse) who also believed that if length of stay could be truncated as the incentives in the payment system would force, there might not need to be so many nurses. It would amuse John later to observe that one of the unintended consequences of DRGs and the payment system was to cause a nursing shortage in the United States as demand for care exceeded supply and hospitals turned into huge intensive care units (Thompson 1988, pp 25–51; Thompson & Diers 1991, pp 121–83; Diers 1992, pp 139–56).

DRGs entered health service at a time in the United States when nursing was struggling economically, educationally and politically for recognition and respect in its own right, away from the ‘handmaiden to the physician’ stigma. DRGs carried a double whammy:
they were thought to be too ‘medical’ for nursing’s attention; and they were tarnished with the ugly stain of money, long an anathema in the profession.

It would be my impression that the nursing response to DRGs in Australia, led by the redoubtable Acting Professor Debora Picone, then president of the New South Wales Nurses’ Association, was to recognise the window of opportunity that the work towards Australian national (AN)-DRGs and funding mechanisms here provided to move nursing resources into the political and policy mainstream – and not, incidentally, into the research funding arena as well.

The focus of casemix on costs, payment and financial models in both countries has, in my opinion, had the unintended effect of making us lose sight of the original relationship between cost and quality. Embedding DRGs in payment systems as complicated as the ones that have evolved in Australia (Podger 1999, pp 111–13) and in the United States has had the unfortunate consequence of tilting us away from clinical and managerial applications as casemix has found its way into the hands and minds of economists. Indeed, this may be the only paper in the present collection to take casemix application to patient care.

In this paper I will review briefly both the American and Australian nursing experience with casemix, acuity and cost weighting. Then I will use examples from some of our more recent work to build an argument for using casemix information in new ways, for ‘process improvement’ or ‘evidence-based management’. I will conclude that the next great leap forward in casemix may require attention to building the information and human infrastructures, so that the rich vein of clinical–financial information produced by casemix-based information systems can truly inform management and policy.

**Casemix and nursing acuity**

It is somewhat ironic that nursing intensity measurement would be the last element in the DRG system to be developed. No other hospital services has been examined more frequently than … nursing care (Thompson 1984, pp 51–58).

When the DRG-based prospective payment methodology was originally tested in New Jersey in 1980, there was a specific requirement, engineered by the New Jersey State Nurses’ Association, for a study of ways to allocate nursing intensity to DRGs (Joel 1984, pp 42–9; Thompson & Diers 1991). That study was carried out but it was fatally flawed and the method was never implemented (Caterinicchio 1983, pp 36–39; Caterinicchio 1984, pp 61–70; Grimaldi & Micheletti 1982, pp 12–22). The major problem with the study was that it assumed that nursing intensity was independent of DRGs. Nevertheless, the study provided important methodological information which helped shape the United States and Australian research that followed. First, it showed the necessity for a per case nursing weight, a difficult concept for nurses used to thinking about ‘my patient, today’. It also revealed the necessity of separating intensive care nursing from floor or ward care. Most importantly, the study showed the promise of
using existing nursing patient classification systems for allocating nursing resource consumption to DRGs.

Nursing patient classification systems have been around since the 1950s. Today there are very sophisticated computer-based systems of estimating nursing care requirements for patients for the next shift or day. Those estimates can be aggregated for the ward to total required hours which, when divided by eight, indicate the number of full-time equivalent nurses to be rostered for the next shift or day. As the range of nursing care requirements (acuity) in United States hospitals has narrowed, the utility of these methods for staffing has diminished, and our accreditation agency, JCAHO, no longer requires them. But as volume and acuity have both changed in the present payment environment, the usefulness of these measures for new lines of inquiry is being re-established. I shall return to that notion later.

In Australia Evelyn Hovenga, later to be a student of Professor Palmer, has devoted a considerable amount of her professional life to developing and testing a nursing patient classification system called PAIS (Hovenga 1994), which is widely adopted for determining nursing acuity here. That has made it possible to do nursing acuity studies to develop nursing cost weights across institutions, the work Picone has led for some years.

In the United States there is no one acuity system so often used, so the Health Systems Management Group at Yale had to acquire data sets from various sources based in very different nursing patient classification systems (Fetter & Thompson 1987). That was an important turning point in this line of inquiry because it forced us to note that the constant in any nursing resource measurement is simply time. Different systems of defining patient care requirements all eventually standardise the resource (nursing) to hours of care for patient types. Our task was to examine the relationship of the nursing definition of patient types, with their time standards, to DRGs. Since an hour is an hour in New Haven or Sydney or Toronto or Paris, this also opened the possibility of international communication.

To cut a long story short, we obtained data from nine hospitals. Eventually the project produced a powerful allocation statistic as an institution-specific beta weight, which controlled for the anomaly that nursing resources and length of stay are very highly correlated – all patients get some nursing care on every day of stay. The nursing intensity work at Yale was supported by a Health Care Financing Administration contract and our Health Care Financing Administration case officer encouraged us to simplify the model. I was asked to put together a small group of expert nurses to see if expert judgement of nursing care requirements by DRG could produce a measure that would have some validity. The resulting measure – six ‘clusters’ of DRGs which shared similar nursing resource requirements, with estimated average times per patient day – tested remarkably well against more complicated methods (Thompson & Diers 1991) that set the stage for the Australian–United States work.
What we learned from the expert panel approach was that it was possible to produce a small number of DRG groups (six) which were clinically credible, and which had estimated time standards that tested well against actual resources provided. We reasoned further that a measure of nursing resources ought to respect how nursing resources are actually deployed in the operational world. Nurses are rostered in 8, 10 or 12-hour increments, in whole numbers. There is no real world equivalent of .10 of a nurse.

Unfortunately, when our various measures of nursing intensity were tested against the Health Care Financing Administration’s per diem method of weighting nursing, all of our measures would have produced more Medicare expenditure and the Health Care Financing Administration was not amused. In addition, elegant as the methods were, public funding policy in the United States cannot be based on a study of only nine hospitals, using different measures of nursing intensity, and there is no way to exercise political will or administrative fiat to prescribe universal nursing intensity measurement.

But not so in Oz. Picone and her colleagues drew together first 7, and eventually 11, hospitals into a nursing research consortium to develop nursing costs weights for AN-DRGs (Picone, Ferguson & Hathaway 1993). Where actual PAIS data were available, they were used to sort DRGs into ‘bands’ similar to our clusters. The study was funded by the Commonwealth. It produced relative nursing cost weights or ‘relativities’. Analysis of data from the various hospitals began to hint at other uses of this information, to which I will refer later. For example, the nursing resource patterns for the DRG for stroke were remarkably similar across hospitals, while the length of stay was remarkably different. Further, fully 34% of patient days were in the lowest level of nursing care requirements, suggesting that those hospitalisation days might not have been necessary.

This study provided the nursing service weights for AN-DRG version 1. The next study extended the methodology using South Australian data, with a wonderful new twist. Where empirical data were not available, a process of calling nationally upon nurses specialised in the care of various types of patients in a Delphi methodology was used to fill in the nursing weights (Picone et al. 1993). These studies have been extended and expanded to rural hospitals (Picone et al. 1995a) and to version 3 of AN-DRGs (Picone et al. 1995b), with increasing refinements for paediatric data, all with government support.

The Australian nursing intensity studies have gone directly into public policy-making as cost weights in a financial system in which it is not possible to attach dollars directly to nursing resources at the level of patient data. In the United States it is only when hospitals actually bill differentially for different levels of nursing care requirements that direct information about nursing resources can be attached reliably to patients (Sovie et al. 1985). While the clinical methods which began in the United States came Down Under and, as I will discuss next, came back, the application of the methodologies is very different in the two countries.
The interest in measuring nursing intensity revived in the very institution whose data were first used to create DRGs, namely, Yale–New Haven Hospital.

The hospital had never used a nursing patient classification system for anything. Knowing the work of the Sydney Metropolitan Teaching Hospital Nursing Consortium, and being particularly instructed by the method of involving nursing experts, we mounted a project to create nursing intensity weights for American Medicare DRGs. Having learned from the New Jersey experience and our own experts’ intuition, we determined to define different weights for the intensive care and non-intensive care portion of patient stays. By this time, information systems had evolved to the point where we could have actual hospital-specific patient utilisation information by DRG, such as emergency department admission rate, use of intensive care unit, diagnostic imaging utilisation, as well as patient age and secondary diagnoses (Diers, Bozzo & RIMs Nursing Acuity Project Group 1997, pp 124–30, 137).

Our project also produced six clusters of DRGs grouped by expert estimates of relative similarity of nursing care requirements. For those DRGs where an intensive care unit experience occurred, we did separate weights and separate time estimates.

We validated this measurement against hours actually worked and against budgeted nursing care hours by nursing unit. In general, the measurement was quite robust and, where there were variances, they could be explained.

Nursing intensity has already moved into costing models under public policy in Australia. In the United States the interesting application will be institution-specific as ‘flexible budgeting’, a process through which the products of a given cost centre are weighted for the relative resource intensity of their production, and then the budgets are flexed for volume in period analyses. Our DRG-specific nursing weights are the relative value units which will be applied in the flex budget review system. Nurse managers will be provided with budget to actual variance reports quarterly, with the budget and the actual expenses flexed for volume. Where the volume is up, it would be expected (and justified) that the expenses would be as well. Where the volume is down, it would be expected that the dollars would follow. Nursing weighted DRGs will be used to calibrate cases. The general principles are already in place for diagnostic imaging and clinical laboratories (pathology) where relative value units have long been in use. We have extended the methodology into the outpatient setting, creating nursing weights for ICD codes for emergency room cases. Those weights have already been used in the flexible budget analysis exercise for the emergency department.

This may be a peculiarly American application and suited only for a situation subject to violent and rapid volume changes, as is the case in our acute care hospitals. Where there is a queueing or waiting list method and occupancy is nearly always 98% or so, budget to actual variances in staffing may reflect acuity differences, but not volume differences. Our new relative value unit proposal produces ‘nursing weighted patient days’, the product of nursing management, a standard to which nurse managers can be held accountable.
Nursing intensity is where nursing and casemix first collided. In the United States there is a very small body of studies using casemix or DRGs to describe interhospital or intrahospital relationships. Most of the studies were done very early in the life of American DRGs when data systems were not nearly as sophisticated as they are now. And most of the early studies tried to anticipate what everyone thought was coming – a need to justify the costs of nursing in a new payment environment (Riley & Schaefer 1983, pp 40–43; Lagona & Stritzel 1984, pp 15–18; McClain & Selhat 1984, pp 27–34; Mitchell et al. 1984, pp 29–32; McKibbin et al. 1985).

About the same time other potential uses of casemix in nursing began to appear. For example, one of my students looked at the effects of short nursing staffing on patient complications and cost in two general medical units in one hospital (Flood & Diers 1988, pp 34–6, 53). Within high-volume DRGs, the complications were higher in the short-staffed unit, as was length of stay and cost. Margaret Sovie at Strong Memorial Hospital of the University of Rochester used casemix to examine the costs of nursing for specific DRGs between her institution and other hospitals in town (Sovie 1988, pp 131–50). Another of my students and I looked at the notion of ‘nursing intensity outliers’ – patients whose nursing care requirements and length of stay were simultaneously extreme. That study began to identify patient types within DRGs that were worth a further look (Talerico & Diers 1988, pp 27–33). A small study in San Diego found very large variances in nursing time in two specific DRGs – esophagitis and heart failure and shock – suggesting a lack of stability in the processes of care (Fosbinder 1986, pp 18–23). It has even been proposed that ‘severity’, in its vast literature, is really nursing intensity (Sovie & Smith 1986, pp 216–26).

**Nursing practice and casemix**

The largest of the studies in casemix and nursing practice is the one Acting Professor Picone has directed from 1994 to the present (Picone et al. 1995c). With Commonwealth funding to start, and New South Wales Health Department funding to continue, Picone and her colleagues conducted an enormous systematic literature review on the relationship between nursing interventions and patient outcomes. Then they developed consensus practice guidelines, specified outcome standards and tested the interventions in one casemix-specified condition across several teaching hospitals in a randomised controlled clinical trial. This is the first study of evidence-based practice in nursing in Australia, and it will set the curve for whatever comes after. The results await public presentation. Picone also moved this work to a level of intellectual sophistication such that it can begin to bridge the gap that exists in the United States between the nurse academics and those in the clinical field.

On our side of the Pacific pond there is now a trickle of studies, most emanating from Yale, that put nursing and casemix together, in a clinical sense. Czaplinski looked at the effect of nursing specialisation on patient outcomes such as length of stay and mortality. She selected 11 DRGs where the majority of patients would be cared for in nursing-
specialised units in one teaching hospital. She found that, for the majority of casemix-defined patients, those cared for in specialised units (‘on service’) had shorter lengths of stay and lower mortality (Czaplinski & Diers 1998, pp 1626–38). Clarke, a psychiatric liaison nurse, examined the difference between patients in certain DRGs who had or did not have an ICD-coded psychiatric co-morbidity on length of stay. Patients with psychiatric co-morbidity in the same DRG had longer lengths of stay and there were more length of stay outliers in the co-morbidity groups (Clarke 1994).

Diabetes is a ubiquitous diagnosis in hospitals. We mounted a study, at the request of a newly appointed clinical nurse consultant/nurse practitioner, to identify where the patients with diabetes were in the house and what their treatment was. We found that, at any moment, about 14% of patients in our institution have diabetes. But only a small fraction of them are admitted for the diagnosis and treatment of their diabetes, and those who are are the very sick people whose disease is out of control. More often, patients are admitted for other things, including a great deal of cardiac and peripheral vascular surgery, which are the major manifestations (Bozzo, Carlson & Diers 1998). Once patients had been located, the clinical nurse consultant put in place a number of projects to improve their care, with the result that the number of normal finger stick blood glucose readings increased dramatically. The standard of care was improved, particularly in the medical units (Bozzo & Carlson 1998).

Now that our institution has an integrated clinical–financial information system with data at both the cost centre and the patient level, we have developed the in-house capacity to conduct casemix studies on request (Diers et al. 1998c). A couple of small studies illustrate this.

The nurses on the postpartum floor noticed that there had been a change in the epidural anaesthesia ‘cocktail’ as the anaesthetists had decreased the nerve block and increased the opiate analgesia. The nurses were observing women experiencing shivers, nausea and vomiting after delivery, suggesting a withdrawal effect. In our data system we do not record anaesthesia by drug, dose or any other data element. But we do record data about the medications used to treat these reactions. Having identified those medications, we could go into the data system and find the patients who had the antidotes and compare them with patients who did not, in casemix-defined groups using the DRGs for delivered women. About 10% of delivered women actually had the antidotes, not an insignificant number in an obstetrical service with about 5000 deliveries a year. More to the point, the women who had the antidotes, meaning they had the new anaesthesia, stayed in the hospital an average of one day longer – this in a setting where the average length of stay for women delivered vaginally is now about 2.1 days.

Another clinical nurse consultant was concerned with a practice pattern difference among paediatricians in the extent to which they used intravenous Terbutaline for treating children with asthma. This treatment requires the child to be in intensive care. We could locate children with asthma by their ICD codes, and children who had Terbutaline by the fee code for that drug. We found that the overall average length of stay was longer, as well as the length of stay in intensive care, than for children with
asthma not treated with Terbutaline, controlling for status asthmaticus, the illness severity at issue here. The clinical nurse consultant used the information to inform the paediatric quality council discussions towards a more stable standard of practice.

**Nursing management and casemix**

The clinical studies use the capacity of a casemix-based information system’s patient-level data. The management studies use the cost centre-level data capacity.

For example, the nurse manager of our newly merged surgical/neurosurgical intensive care unit came to us bothered by the observation of her nurses that 'the work is so much harder this year'. While this might be a subtle plea for more staff, the nurse manager was more interested in understanding what that feeling was. We disposed early of the notion that patient acuity had changed. It had not. Nor had census. What had changed was the mix of patients and their clinical demands on the nursing staff. In this 20-bed intensive care unit, over a one-year period, there were 189 different DRGs, in all but three of the major diagnostic categories. The neurosurgical nurses were used to caring for patients with craniotomies, endarterectomies and spinal cord injury. Faced with general surgical intensive care patients, often victims of gunshot wounds, surgery gone bad or automobile accidents, they felt that 'the work was harder'. The analysis led the nurse manager away from merely requesting more staff to shoring up the clinical intelligence of the nurses through consultation, inservice education and mentoring (Diers et al. 1998a).

The nurse manager of a 42-bed surgical unit had a reputation in the hospital for being the worst possible nurse manager. Her overtime was out of control, which was the only measurement hospital administration really had. Her unit went through a facility and service redesign which reduced its bed capacity to 27 and concentrated the cases to orthopaedic surgery. We examined the effect of this change before and after it happened (Diers & Potter 1997, pp 27–32).

The number of different DRGs housed in the unit dropped by about half. Because the unit had been too large for the orthopaedic service to fill all the beds, it got everybody’s overflow. It is not reasonable to expect that nurses can be equally good at caring for all possible kinds of patients, to say nothing of negotiating with a large range of physician specialties and medical staff teams. When the unit became specialised to orthopaedic patients again, the length of stay dropped by nearly 1.5 days, which was twice the secular trend in length of stay in the hospital for the applicable period. In all of the high-volume orthopaedic DRGs for which we had before and after data, the length of stay dropped. Overtime decreased from 218% over budget to 31% under budget.

Sometimes very simple casemix analyses can have a powerful operational impact. The nurse manager of the unit in our institution that specialises in caring for persons with AIDS came to us to help her understand what was happening with her casemix, and provide appropriate levels of staff. Local initiatives had produced new discharge options
for AIDS patients, and protease inhibitors reduced length of stay and readmission rates. Her unit was filling up with overflow medical patients, often elderly with one or another of the four competing diseases older persons have – congestive heart failure, diabetes, chronic obstructive lung disease and pneumonia. We tracked her DRGs and she used the information to develop training programs for her staff as well as to bring to the attention of hospital administration the consequences of a lack of bed control policy.

The clinical director for the general medical units came to us because her first quarter financial results showed three of her units seriously over budget in nursing resources. The nurse managers alleged that patient acuity had increased. We examined acuity using our own measurement and it had not increased over the first quarters of three consecutive fiscal years. But what had changed was the mix of cases in the unit. Again, it is not reasonable that nurses can be instantly equally good at caring for patient types that are different from what they are used to. This finding headed the clinical director away from arguing for more nursing staff and towards thinking with her expert clinicians about better ways to manage the casemix shifts.

Finally, in a payment-defined information system, ‘outliers’ are trimmed out as aberrant cases before further analysis. From a clinical point of view, we need to understand these aberrant cases to determine whether there are patterns there that could be discerned to produce a more stable process or production of care. There is practically no literature on outliers, and what exists is on high-cost cases as health services research is not very helpful in operations management. We created a definition of high-cost cases as those cases whose total hospital costs consumed 20% of total inpatient hospital costs in a given fiscal year. In each of the three years we have analysed so far, the actual number of high-cost cases is about 1.6% of discharges. We have found interesting and disturbing patterns in cases between and within DRGs that have questioned the stereotypical definition of the high-cost patient as the uninsured, teenage, ethnic minority, pregnant, cocaine-abusing girl with HIV. Our high-cost patients were concentrated in the high-tech DRGs, were more often transfers-in from other hospitals, were white, male and not from the inner city. The mortality rate for high-cost users was about 20%, in contrast to the overall hospital mortality of under 2% (Diers et al. 1998b). The results of this study have had the effect of shifting the explanation for high-cost cases from the patient to the processes of care.

**Conclusion**

Casemix moved into the financial realm in the United States and Australia leaving clinicians of any stripe in the dust. While *case mix* as two words was the original term used (Thompson, Fetter & Mross 1975, pp 300–12), that phrase now combined into one word is used outside the United States but not within. A search on *casemix* in library information systems will produce nearly exclusively Australian or European citations. And most of the citations will be about costing, cost modelling, case weights, or adjustments to the grouper.
It is indisputable now that casemix-based government or even private payment initiatives, with their efficiency incentives, can actually produce efficiencies. But there will inevitably be a point at which the ratchets are tightened so much that public policy, politics or the health care delivery system itself cannot tolerate any more. While the use of government clout towards payment reform clearly works, eventually the stabilisation of the cost–quality equation will depend upon an understanding of the work of service delivery. That will require the development of information capacity, human infrastructure, and new methodologies for analysis within and between institutions, areas, specialties, hospital types and countries.

Some of that work has already begun as clinical path development and testing, critical indicators, benchmarking and other tools of the casemix method. The micro-level studies reported here are yet very scarce, but they represent another area of inquiry easily supported by integrated clinical–financial information systems.

I suggest here that until clinicians – nurses and physicians – can find within casemix information their own work, they will remain on the periphery of the policy arguments that centre on costs and funding. That will be a pity because the balance or stabilisation that reliably produces affordable high-quality care requires interdisciplinary concentration on the elements of practice as well as the ability to use information systems for tracking and prediction. The raging capitalism in the United States has at least produced an information capacity built from our intricate billing systems that can now feed clinical interests at the atomistic level clinicians require. So long as information systems are aimed only at the economic policy agenda, casemix will remain estranged from those who do its work.

I am encouraged by the sense of a sea change I have felt in the attitudes and interests of clinicians and managers in both the United States and Australia towards casemix information. Where once attendance at casemix workshops I have provided in both countries was sparse and hostile, it is now oversubscribed and eager. There is a growing sense of a need to know among clinicians that makes one believe that it might just be possible to return casemix to its original role in managing for quality. Our experience in the sample of small studies reported here suggests that interdisciplinary work centred in clinical operational management moves clinicians away from the endless risk adjustment and severity arguments and moves managers away from beady-eyed obsession with the bottom line towards mutual engagement in process improvement, measured by the tools of economics. Casemix funding is already here. Casemix management is, perhaps, an idea whose time has finally come.
References


Fetter RB & Thompson JD 1987, *Diagnosis Related Groups and Nursing Resources*, Final progress report to the Health Care Financing Administration (#15-C-98500), Yale University, New Haven, Connecticut.


Lagona TG & Stritzel MM 1984, ‘Nursing care requirements as measured by DRG’, *Journal of Nursing Administration*, vol 14, no 5, pp 15–18.


