The Australian DRG classification: are we ready for structural changes?

DON HINDLE

Don Hindle is Visiting Professor in the School of Health Services Management, Faculty of Medicine, University of New South Wales.

The Australian DRG variant (Australian National DRGs from versions 1 to 3, and Australian Refined DRGs thereafter) compares favourably with other variants in use around the world. This view is supported by a variety of empirical studies (see for example Reid, Palmer & Aisbett 2000) and by experiences of its adoption by several health care systems. An example is its recent selection for use in Germany (Hindle & Lenz, 2000).

The success has mainly originated in the review process that was established in 1992 by the Commonwealth Department of Health and Aged Care (the Department), and in particular as a consequence of the involvement of practising clinicians through the Australian Casemix Clinical Committee. The process has been informed by practical experiences in the use of Australian DRGs in resource allocation, and by several research studies (including one undertaken by the Department which led to the establishment of the combinatory PCCL logic for use of significant secondary diagnoses).

Australia has also been relatively successful in improving the data that are used in DRG assignment. For example, the development of our own diagnosis and procedure classification (ICD-10-AM) was a major achievement that reflects well on both the Commonwealth and State health authorities and the many clinicians that supported the work (Roberts, Innes & Walker 1998).

The processes of both design and use of the DRG classification have been facilitated by the ongoing refinement of a national database containing all hospital discharges, and the associated data definitions and data quality control processes including our excellent coding standards. Australia has also strengthened its data on product costs (Jackson 2001; Oates, Murray & Hindle 1998).

The beneficial effects of the use of DRG data have been widely discussed. There have been demonstrable improvements in resource allocation in the Australian public hospital sector (Duckett 1998), and other countries are following a similar path (Forgione & D'Annunzio 1999). Australian experiences have been taken into account in many countries including Germany, Singapore, Bulgaria, and Malaysia. The value of the DRG classification in support of local comparisons of clinical practice continues to grow. It is common to criticise aspects of imprecision in the classification, but there are seldom any equally useful options.

The Australian DRG variant has been updated every two or three years, and version 5 is currently being finalised. The evidence shows that its classification performance has improved with each version (although only marginally between versions 1 and 2). The interest of users has been a major stimulus to rapid and sensible modifications.

However, the updating process has experienced significant time pressures. Large numbers of suggestions for change have had to be analysed empirically and considered by expert clinical panels, coding and DRG grouper changes have had to be implemented with care, costing studies have had to be undertaken, supporting documentation has had to be prepared, and so on.

Given the high degree of dependency between these elements, there has seldom been much time to rest and think. While the time between DRG versions has averaged two-and-a-half years, the lead times of several of the component tasks have been such that there have been very limited opportunities for the exploration of major modifications.

This was of little concern while the use of DRGs was in its exploratory stages. The consideration of minor changes led to sufficient improvements to satisfy most users.

However, I believe now may be the right time to take a different approach for four main reasons. First, further significant improvements in performance are unlikely to be achievable if we continue to restrict ourselves to marginal changes (in part because the potential for improvement has been largely exhausted over ten years of analysis). Second, the evidence continues to grow of important structural weaknesses. Third, better data systems are available now compared with ten years ago. Finally, there are increased expectations of a growing number of informed users - not only in terms of more precise classification but also with regard to a wider range of analytical uses.

In short, I believe it would be advisable to establish a process that is able to explore changes over a much longer period of time than that available between successive versions. It has been possible to take a longer-term perspective in the past. Indeed, some research has spanned more than a single period between versions. However, there have been no formal and widely accepted processes to encourage and support it.

Ideas for DRG refinement that might be explored

There are many opportunities for improvement that do not directly relate to the structure of the classification itself. Some concern the source data. For example, Alechna, Westbrook, and Roberts (1998) reviewed the coding of burns cases in two Australian public hospitals and found that 89% had errors, of which 10% led to assignment to the wrong DRG.

Problems have been noted with regard to methods of computation of the average costs. For example, Phelan et al (1998) argued that "... inadequate cost weights are a major problem in casemix funding systems" and clinicians needed to be more involved if the weaknesses were to be recognised and alleviated.

The process of appraisal of optional designs has also been criticised. For example, Palmer and Reid (2001) noted that there were many differences in the evaluation criteria used by agencies responsible for the updating of DRG variants. They argued that "... without the evidence provided by properly designed studies, policy-makers and managers may place undue reliance on subjective judgments and the views of the most influential, but not necessarily best informed, healthcare interest groups." It has been frequently argued that more use should be made of direct estimates of cost, rather than length of stay, as the dependent variable when searching for better classifications.

Many authors have concentrated on classification and consequent funding problems for subsets of patients. For example, some useful analyses have been undertaken in Australia with respect to problems of categorisation of Aboriginal patients (Fisher et al 1998; Ruben & Fisher 1998; Russell-Weisz & Hindle 2000). Good arguments have been presented with respect to the problems of discrimination of severity within the principal condition for children (Hanson et al, 1998).

However, these kinds of issues are being addressed through progressive enhancements, and my main concern relates to aspects of DRG structure that have remained largely unchanged since the earliest versions were produced in the late 1970s. I will outline a few of these for the purpose of illustration.

First, there is the continued reliance on principal diagnosis, defined as the main reason for the decision to admit. Most countries apply this definition, for the same reason that Australia changed to it ten years ago: because it was used in the data set from which the first version of DRGs was developed in the United States of America.

Australia had previously defined principal diagnosis as the condition best explaining the stay in hospital - in other words, the one that accounted for most of the costs of the episode. This is the definition currently used in Germany, but there are plans for change: someone has incorrectly formed the view that Germany must use the Australian definition if it wants to use Australian DRGs. In contrast, several other countries, including Canada, took a DRG variant that used 'main reason for admission' but retained their own (and different) definition of principal diagnosis.

My main concern is not, however, that the wrong choice might have been made - there is little evidence to support either definition. Rather, I question the need to select a single definition, whether it is the cause of the admission or of most of the costs of care. I have heard no adequate justification for departing from clinical practice, whereby patients are variously classified in terms of one dominant diagnosis or a constellation of

interacting conditions. The Patient Management Categories classification seems much more sensible in this respect: there is no requirement to indicate the principal diagnosis, and episodes are defined in terms of one or more clusters of related diagnoses.

There are several important possibilities if we are willing to contemplate a different approach. One is that we could explore the definition of classes according to the presence of constellations of diagnoses that are associated with the elderly frail. This idea is becoming increasingly important in geriatrics, and is widely termed the 'geriatric syndrome'.

Take the case of the elderly person admitted to hospital with a fractured leg, which was a consequence of a chronic problem with balance. I do not have strong views about selecting the fracture or the problem of balance as the principal diagnosis. I merely want someone to prove it is necessary to select one, rather than to define the patient in the way that clinicians do as a matter of course.

If nothing else, we should be determined to become better at describing, and consequently measuring and managing chronic conditions. Most of the inpatients of our acute care hospitals are admitted because of an acute phase of a chronic condition, an acute condition caused by the chronic condition, or an acute condition that is exacerbated by a chronic condition that needs attention during the episode. I have become increasingly frustrated over the years at the inability of our DRG data to say anything useful about the way we handle chronic care, especially since it is the dominant problem.

A related point is that there is good reason to categorise episodes in different ways depending on the purpose of the analysis. There are circumstances in which it would be more helpful to group episodes by main reason for admission, and others in which it would make more sense to categorise them by main cause of resource consumption (or a combination of both).

A second elementary problem concerns the failure adequately to take account of multiple procedures during the same episode. I understand the origins of this practice. In the late 1970s, the US federal government was concerned about slowing down the rate of growth in hospital expenditures, and thought (probably incorrectly) that it would have to pay more if account were taken of more than one procedure. This argument has never made sense in Australia, where we have used DRGs to re-allocate resources rather than as a device to contain total costs. After all, we have been more successful in controlling total expenditures in more direct ways - such as by capping the total budget.

Third, there is the matter of the set of variables used to define DRGs. Some new variables may be helpful. Perhaps the most obvious are those related to nursing (such as nursing diagnosis or nursing acuity level). Others include more sophisticated measures of condition, such as the body burden of disease score suggested by Roe et al (1998). Particularly interesting is the possibility of incorporating measures of outcome. It is clearly desirable if feasible to incorporate splits based on estimated value and not only on estimated cost.

The literature on DRGs that is now emerging in Germany mostly addresses risks to hospitals as a consequence of the planned move to per case payment based on DRGs (Lutkes & Kribben 2000; Kienapfel & Hinrichs 2001). However, there are increasing numbers of proposals for new variables, as German clinicians begin to look more carefully at DRGs directly related to their work. A good example is the study by Kugler et al (2000), who investigated the effectiveness of Australian DRGs in classifying stroke patients. They found that classification was deficient relative to the use of an existing method of categorisation that employed more precise measures of disability and impairment than ICD diagnoses alone.

It may be possible to make better use of variables already in the discharge data set, such as transfer to intensive care as suggested by Butt and Shann (1998). An interesting possibility that I have attempted to study of late involves the use of procedure variables to split medical clusters. The DRG logic applies procedure data only to 'surgical' clusters and involves the use of significant procedures - meaning those that were typically performed in an operating room ten or more years ago. However, the procedure data for 'medical' cases are becoming increasingly accurate and relevant. The fact that we have done hardly anything to explore this possibility suggests there are many closed minds with regard to the basic structure of the DRG classification.

Fifth, we should explore ways of defining and using continuous rather than only discrete descriptions of the patient's clinical attributes. This is a complicated idea, but may be illustrated by example. The patient clinical

complexity level score introduced in AR-DRG version 4 is a measure of the total effect of secondary diagnoses, and is a continuous variable (CDHAC 1998). However, it is subsequently used to define a small number of discrete classes (such as condition X with or without significant complications or comorbidities) thus losing information. The use of a small number of discrete classes was justified before computers became ubiquitous, but less so now. The idea of describing individual episodes by continuous variables on multiple dimensions offers significant potential for increased precision of categorisation and payment.

Sixth, we need to do more to manage the relationships between DRGs and classifications used for other care settings. Several years ago, I argued that the goal of casemix classification development in Australia should be the creation of a classification that covered all kinds of hospital episodes, rather than one which suited only acute admitted patients. This was probably a poor idea, but it reflected a serious concern: that single-minded pursuit of DRG improvements might exacerbate the problem of poor continuity of care.

At least, more needs to be done to address the most obvious groups of patients that need care in more than one setting. Again, recent German work is iterating ideas that are not new to Australia, but which have had little practical impact. For example, the study of stroke patients by Kugler et al (2000) led to the view that new classes were needed that "... cover overlapping healthcare sectors" and which would therefore support the integrated management of both acute and post-acute care (including rehabilitation). We have been taking a more sensible approach in particular cases in various parts of Australia, for example, by defining care-years for maintenance dialysis and cystic fibrosis. However, there has been no strategic planning.

Finally, there is the matter of the extent to which DRGs should be defined to take account of desirable methods of care, rather than only on the basis of actual average patterns of care. The Australian Clinical Casemix Committee has proposed marginal adjustments to DRGs from time to time that have reflected their views about desirable trends. State and Territory health authorities have also made local adaptations to the way DRGs are applied to resource allocation, in order to reflect their views about good clinical practice. However, there has been no formal policy on this matter in the context of the Commonwealth's DRG refinement process.

I would like to see a serious effort to incorporate best practice into classification refinement that might (say) begin with clinical practice guidelines and include the design and standard costing of illustrative clinical pathways. Incidentally, this would hardly be a novel idea. The process of development of the Patient Management Categories classification more than 15 years ago included asking expert clinical teams how a particular health care problem should be handled (and not only how it was usually managed).

The most obvious advantage of using actual average costs as the basis for classification is that it requires little more than empirical analysis, and this may be sufficient if the dominant aim is a fair allocation of resources to reflect current clinical practice. However, we have been moving towards the view that a more important objective is to encourage and reward improvements in practice - and this requires us to augment empirical analysis with the views of expert clinicians about future methods of care.

Establishing the right process

There is a wide range of possibly useful research hypotheses that might be worth testing over a longer (or at least more flexible) time scale. It is not necessary to decide what they should be at this stage. Rather, it would be better to see what might emerge from the many interested parties once they were given a different scenario than that which has dominated thus far.

One possibility is that the two- or three-year process of updating would continue, but be augmented by a related developmental process that involves the exploration of more significant changes that may require more than two or three years for completion. In the latter case, research activities would not be restricted by version deadlines. They may lead to changes in the next version, but the intention should be to be driven by a desirable research goal rather than by tight deadlines that are unrelated to the complexity or importance of the analysis.

Thus this stream might comprise a variety of research activities of varying durations that is managed against a rolling horizon of perhaps three to five years. Experience would show how long the horizon should be, and the same may be said of the scheduled duration of particular research projects. In some cases, prolonged research

might be unwise simply because relevant attributes are likely to change. For example, it might make little sense to conduct research in an area of clinical practice where major changes are expected in the near future.

This second stream would be better able to generate and manage collaborative research projects. These might not only include collaborations between particular States, but also between States and the Commonwealth. Equally important, the longer time scale would facilitate the conduct of collaborative research involving countries that are also using Australian DRGs.

Collaboration has often been in short supply within Australia, and this may be a consequence of the time pressures that are an ongoing feature of State and Territory budget-setting processes. It may also be a consequence of a strategic error: the attempt to separate DRG refinement from resource allocation applications. Whatever, the cause, Duckett (1998) is surely right in arguing that "...learning across State boundaries should be encouraged, with knowledge of what is effective and what is ineffective in casemix funding arrangements being used to develop Australian best practice in this area."

Of course, it is not sufficient to have a potentially useful research idea. Other factors must be taken into account including the practicalities of obtaining relevant data. It would be necessary to develop criteria against which research proposals could be appraised - in the same way that applies to any well-run research program.

In summary, there are a few new and many old ideas about improving DRGs by making more significant changes than we have managed thus far. I am not worried about the possibility that the Germans will take our ideas and implement them while we continue to focus on minor details. Indeed, it would be helpful to Australia if this were to happen, since I suspect we would be more likely to follow their practice than to listen to our own experts. However, I expect they will be even busier than us worrying about the immediate details because they have further to go before a reasonable method of health care financing is established.

My concern is simply that we do not miss the opportunity to take the next step towards improved output-based funding. The Americans have largely put aside their 'prospective payment' systems and moved on to managed care. This is not because it is a better idea, but rather because it is different: they do not have the basic health care structures to allow any idea to work well. In contrast, Australia can and should take a more incremental approach and move on to the next level of sophistication of output-based funding that includes better DRGs - and indeed some of the sensible elements of managed care including the definition of appropriate methods of care (through pathways and guidelines) rather than merely the average patterns in a previous period.

References

Reid B, Palmer G & Aisbett C (2000), 'The performance of Australian DRGs', *Australian Health Review*, vol 23 no 2, pp 20-31.

Hindle D and Lenz M (2000), Germany and per case payment using Australian DRGs: a commentary. A*ustralian Health Review*, vol 23 no 4.

Kugler C, Freytag S, Stillger R, Bauer P, Ferbert A 2000, Australian Refined Diagnosis Related Groups: Formal and inherent problems of grouping with the example of stroke care, *Dtsch Med Wochenschr* 2000 Dec 22; 125(51-52), pp1554-9.

Lutkes P, Kribben A 2000, The Diagnosis Related Groups for Germany: clinician, take the lead!, *Dtsch Med Wochenschr* 2000 Dec 22;125(51-52): p1553.

Kienapfel H, Hinrichs F 2001, Case-mix adjusted reimbursement system based on the Australian Refined Diagnosis Related Groups (AR-DRG) for orthopedic and trauma surgery clinics, *Z Orthop Ihre Grenzgeb* 2001 Jan-Feb;139(1): ppM2-10.

Alechna N, Westbrook J & Roberts R 1998, The quality of burns coding: compliance with standards and the effects on clinical data, *Health Inf Manag* 1998 Dec-1999 Feb; 28(4), pp181-5.

Phelan PD, Tate R, Webster F & Marshall RP 1998, DRG cost weights—getting it right, *Med J Aust* 1998 Oct 19;169 Suppl, ppS36-8.

Roberts RF, Innes KC & Walker SM 1998, Introducing ICD-10-AM in Australian hospitals, *Med J Aust* 1998 Oct 19;169 Suppl: pp S32-5.

Fisher DA, Murray JM, Cleary MI, Brewerton RE 1998, The Aboriginal and Torres Strait Islander casemix study, *Med J Aust* 1998 Oct 19;169 Suppl: ppS11-6.

Ruben AR, Fisher DA 1998, The casemix system of hospital funding can further disadvantage Aboriginal children, *Med J Aust* 1998 Oct 19;169 Suppl: ppS6-10.

Russell-Weisz D, Hindle D 2000, High length-of-stay outliers under casemix funding of a remote rural community with a high proportion of aboriginal patients, *Aust Health Review*, vol 23 no 2, pp47-61.

Hanson RM, Phythian MA, Jarvis JB & Stewart C 1998, The true cost of treating children, *Med J Aust* 1998 Oct 19;169 Suppl: ppS39-41.

Palmer G & Reid B 2001, Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues, *Health Serv Manage Res* 2001 May;14(2), pp71-81.

Jackson T 2001, Using computerised patient-level costing data for setting DRG weights: the Victorian (Australia) cost weight studies, *Health Policy* 2001 May;56(2): pp149-63.

Oates B, Murray J, and Hindle D 1998, The second national hospital costing study: background, results, and implications, *Australian Health Review*, vol 21 no 3, pp.

Duckett SJ 1998, Casemix funding for acute hospital inpatient services in Australia, *Med J Aust* 1998 Oct 19;169 Suppl, ppS17-21.

Forgione DA & D'Annunzio CM 1999, The use of DRGs in health care payment systems around the world, *J Health Care Finance* 1999 Winter;26(2), pp66-78.

Roe CJ, Kulinskaya E, Dodich N & Adam WR 1998, Comorbidities and prediction of length of hospital stay, *Aust NZ J Med* 1998 Dec;28(6), pp811-5.

Butt WW & Shann FA 1998, Transferred patients—more complex and more costly?, *Med J Aust*, vol 19, ppS42-3.

CDHAC (1998), *'Australian Refined Diagnosis Related Groups, version 4.1: definitions manual'*, Commonwealth Department of Health and Aged Care, Canberra.