The performance of Australian DRGs

Beth Reid, George Palmer, and Chris Aisbett

Beth Reid is Professor and Head, School of Health Information Management, Faculty of Health Sciences, The University of Sydney. George Palmer is Professor and the Director of the Centre for Hospital Management and Information Systems Research at the University of New South Wales. Mr Aisbett is the Director of Laeta Pty Ltd Randwick.

Abstract

This paper reports the results of an evaluation study of the Australian National Diagnosis Related Groups (AN-DRGs). The evaluation was based on statistical rather than clinical criteria with the principal goal being to provide information for the future development of the classification system. As well as comparing versions 1.0 to 3.0 of AN-DRGs, the project included a comparison of these systems with the most recent versions of the DRG systems from the United States. Taking all the evaluation criteria together, Version 3.0 of AN-DRGs performed best of all the systems except for the All Patient Refined (APR)-DRGs with its much larger number of groups. However, the differences between all the classifications were slight. Data of higher quality are needed if further refinements of the AN-DRGs are to produce substantial improvements in performance.

Introduction

Casemix systems are classifications of patient treatment episodes designed to create classes which are relatively homogeneous in respect of the resources used, and which contain patients with similar clinical characteristics. The best known example of a casemix system is the Diagnosis Related Group (DRG) system which is now used extensively for hospital funding and management purposes in the United States (US), Europe, Australia and elsewhere. DRGs are a system for classifying acute hospital inpatients. Australian National (AN)-DRGs, based on US DRGs, were developed in Australia as part of the Casemix Development Program and the first version was released by the Commonwealth in July 1992. The decision to localise a US DRG version meant that the AN system benefited from extensive US research and experience. Furthermore, the localisation process increased the clinical acceptability of the system and facilitated the inclusion of local terminology and coding conventions.

The process of development of each version of the DRGs has been described elsewhere (Commonwealth Department of Human Services and Health 1995; Pilla 1994). The clinically driven recommendations compiled by the Australian Casemix Clinical Committee (ACCC) were evaluated by the Commonwealth Department of Health and Family Services by reference to a number of statistical criteria. The ACCC, in conjunction with the Commonwealth Department of Health and Family Services, has undertaken an extensive clinical review of AN-DRGs version 3.0 as part of the process of producing further Australian versions of DRGs

(Australian Casemix Clinical Committee 1996). The latter report contained details of the statistical tests and criteria applied in evaluating the proposals. One of the problems in the short time frame involved in developing new versions of AN-DRGs is that the testing of the clinical recommendations for changes is mainly conducted on a piecemeal basis. The statistical review reported here provides a total systems perspective as well as analysing the Major Diagnostic Category (MDC) level of performance where appropriate.

Data and methods

The data for the study, from every State and Territory in Australia for the 1993/94 financial year, were provided already grouped using AN-DRGs version 3.0 by the Commonwealth Department of Health and Family Services. These data were further edited by us to ensure that only acute hospital stay records were included, leaving a total of 4,276,752 records. Several data items, such as discharge status and birth weight, required manipulation and coding to the format required by the US versions.

Versions 1.0 to 3.0 of AN-DRGs were compared, along with three US versions. The modifications in the All Patient (AP)-DRGs addressed the needs of paediatric patients better than the previous versions used by the Health Care Financing Administration (HCFA) for the Medicare Prospective Payment System. The All Patient Refined (APR)-DRGs were developed to combine the best features of the AP version developed in New York with the Refined DRGs developed at Yale University. Comprehensive details of the six DRG systems and the methods used in the project are set out in the Final Report (Palmer *et al.*1997). The versions and numbers of DRGs in each of the systems compared are set out in Table 1.

Table 1: Number of DRGs in each of the DRG systems used in the study

AN-DRG	AN-DRG	AN-DRG	AP-DRG	APR-DRG	HCFA
Version 1.0	Version 2.0	Version 3.0	Version 12	Version 12	Version 13
527	530	667	641	1530	492

For the purposes of this paper, further comparisons were made using data from New South Wales (NSW) for 1996/97 grouped using AN-DRGs version 3.1 and version 4.0, now named Australian Refined (AR)-DRGs. Version 3.0 and 3.1 were virtually the same except for up-dating some of the disease and procedure codes used.

It is important to note that although the first revision of AN-DRGs was based more closely on the AP-DRGs than on the HCFA DRGs, subsequent revisions introduced many changes, so that there are now substantial differences between AN-DRGs and AP-DRGs. The extensive use of resource intensity classes in APR-DRGs based on secondary diagnoses is one important difference between that and the other systems, but the DRGs without age and complication or comorbidity splits (adjacent DRGs) in the APR system are virtually the same as those in AP-DRGs. A major structural difference is that APR-DRGs eliminated almost all age splits, except for neonates, whilst AN-DRGs version 3.0 introduced a large number of age splits. The extensive use of age may be regarded, in part, as a method of compensating for inadequacies in the Australian data in respect of the reporting of secondary diagnoses.

The main performance criteria used to compare the DRG systems are set out below.

The predictive validity of a classification system is a function of the amount of variation in the dependent variable – length of stay in this case – explained by the classification of items. It is measured by the coefficient of multiple determination R^2. This statistic gives a measure of the reduction in variation of the dependent variable produced by the classification (Bland, 1995). The coefficient of variation (CV) provides a separate measure of within-group variation for individual DRGs. A CV equal to or greater than 100 for trimmed data indicates poor within-group homogeneity.

The proportion of outlier cases is also a useful measure of system effectiveness where the same trimming algorithm is used for all systems. A less effective DRG version will define more cases as outliers, whereas a more effective one will assign these outlier cases to DRGs as inliers rather than removing them by trimming.

A DRG version with a large number of groups will have more classes with low volumes of cases unless only high volume groups are sub-divided during the revision process. The main problem with low volume DRGs is related to establishing reasonably accurate cost weights. The small numbers lead to instability from year to year for reasons that are inherently a reflection of sampling variation. Even where the overall sample size is large, the sampling errors in the cost estimates may be very high. On the other hand, the greater accuracy of splitting into more groups with a consequent fairer funding allocation between hospitals might reduce the impact of the difficulty of establishing the cost weights. These problems are accentuated when AN-DRGs are applied at the individual hospital level. In a previous evaluation of AN-DRGs version 1.0, attention was drawn to several low volume groups that were not justified, on the grounds of extreme high cost (Reid *et al*.1992). Some of these groups disappeared because of changes made in version 2.0. We used the same criterion as previously, namely fewer than 50 cases per million, to define the low volume classes.

In addition, two supplementary analyses were undertaken. The first addressed the performance of the system in achieving the objective of providing a fairer basis for the funding of hospitals. Principal teaching/referral hospitals have always claimed that their patients, within DRGs, were sicker and consumed additional resources, and some evidence from the US supported this claim (Health Care Financing Administration 1994). We used a simple funding model based on the cost weights for each AN-DRG version, and an outlier payments policy, to simulate the aggregate share of funding notionally allocated to the major referral hospitals and compare it with that allocated to other types of public hospitals. An increase in the proportion going to the former group would provide a measure of the improved performance of the version in question. The magnitude of the changes between versions would also provide an indication of the scope for further refinements of AN-DRGs, including increases in the numbers of groups, to improve funding models.

The aim of the second supplementary analysis was to examine the validity of the surgical hierarchy in version 3.0. A US publication defined the surgical hierarchy as an ordering of surgical classes according to their resource intensity; its use ensures that cases with multiple procedures are assigned to the DRG associated with the most resource intensive surgical class (Department of Health and Human Services 1996). The surgical class consists of adjacent DRGs, that is one or more DRGs without splitting on the presence or absence of age, complication/comorbidity, or other procedures. It was possible that the separate changes in the

various versions of the AN-DRGs had produced a hierarchy which no longer reflected well the costs involved. The national cost weights for AN-DRGs version 3.0 were used to verify the position in the surgical hierarchy within an MDC of each surgical class.

The purpose of defining outlier observations in statistical analysis is to remove them from the data in order to provide a more satisfactory basis for examining the characteristics of the population under review. For example, outliers may be removed to obtain more robust and meaningful estimates of the central tendency and variability of the distribution of the variables which are the subject of the analysis. Both the arithmetic mean and the standard deviation are affected greatly by even a relatively small number of extremely large values. Similarly, the coefficient of multiple determination is distorted considerably by the presence of outliers. All these statistics play a key role in the evaluation process. The interquartile range method was used to define an upper trim point beyond which all cases were defined as outliers. This method has been used widely to define outliers although several other methods have been advocated and investigated (Bender & McGuire 1995; Coombes Eckstein & Gomes 1995; Palmer & Aisbett 1996). The data were trimmed for high outliers only. This was done for several reasons, but chiefly because we found that the removal of low outliers had little impact on the measures of performance used in the study.

The results reported below for Australian data were calculated using trimmed data (which included same day cases) except for the low volume criterion which was based on untrimmed data.

Results

The results for each of the main performance criteria are set out in Table 2. There was a slight improvement in the R^2 values for each successive version of the AN-DRGs. The results for the US DRGs are as expected; the HCFA DRGs have the lowest values and the APR-DRGs the highest, with AP-DRGs results lying in between the other two. AN-DRGs version 3.0 performed slightly better than AP-DRGs (and HCFA) but not as well as APR-DRGs.

Table 2: Results for all DRG systems for the performance criteria, Australian data 1993/94

Criterion	AN-DRG Version 1.0	AN-DRG Version 2.0	AN-DRG `Version 3.0	AP-DRG Version 12	APR-DRG Version 12	HCFA Version 13
R^2	0.498	0.500	0.505	0.501	0.525	0.477
Number of DRGs where CV>= 100	34	34	41	33	74	18
% cases trimmed	6.1	6.0	5.5	5.7	5.7	6.9
Number of DRGs with <50 cases per million	47	37	35	62	725	15

In all the versions there was considerable variation between MDCs in the R^2 values, ranging in version 3.0 from 0.077 for MDC 20 to 0.603 for MDC 6. The R^2 for each MDC for version 3.0 are set out in Table 3, and Figure1 shows graphically the wide variation in the values at the MDC level. A definite improvement between versions 1.0 and 3.0, which was also statistically

significant, corresponded to an increase in R^2 of at least 10 per cent. The nine MDCs that met this criterion are shown in Figure 2. In a few instances, the R^2 value decreased between these two versions, namely for MDCs 1, 15, 16 and 19.

Table 3: MDC R^2 Values for AN-DRG Version 3.0, Australian Data 1993/94, trimmed

MDC	MDC Title	R^2	No. of Separations
Pre-MDC	MDC not applicable	0.209	6592
01	Diseases & disorders of the nervous system	0.382	176204
02	Diseases & disorders of the eye	0.324	105373
03	Diseases & disorders of the ear, nose, mouth & throat	0.420	242906
04	Diseases & disorders of the respiratory system	0.337	222449
05	Diseases & disorders of the circulatory system	0.456	307364
06	Diseases & disorders of the digestive system	0.603	508114
07	Diseases & disorders of the hepatobiliary system & pancreas	0.361	75082
08	Diseases & disorders of the musculoskeletal system & connective tissue	0.521	377546
09	Diseases & disorders of the skin, subcutaneous tissue & breast	0.531	181305
10	Endocrine, nutritional & metabolic diseases & disorders	0.347	39388
11	Diseases & disorders of the kidney & urinary tract	0.534	355115
12	Diseases & disorders of the male reproductive system	0.569	81408
13	Diseases & disorders of the female reproductive system	0.567	213458
14	Pregnancy, childbirth & the puerperium	0.537	403128
15	Newborns & other neonates	0.558	158061
16	Diseases & disorders of the blood, blood forming organs, immunology disorders	0.373	45172
17	Myeloproliferative diseases & disorders, poorly differentiated neoplasm	0.493	138028
18	Infectious & parasitic diseases, systemic or unspecified sites	0.400	50429
19	Mental diseases & disorders	0.113	90240
20	Alcohol/drug use & alcohol/drug induced organic mental disorders	0.077	22111
21	Injuries, poisonings & toxic effects of drugs	0.519	91540
22	Burns	0.464	5661
23	Factors influencing health status & other contacts with health services	0.463	120266
Error Classes	Error & edit classes	0.335	24757
All Classes	all cases	0.505	4041697

Figure 1: R^2 values for each MDC, for AN-DRGs version 3.0 Australian data 1993–94, trimmed

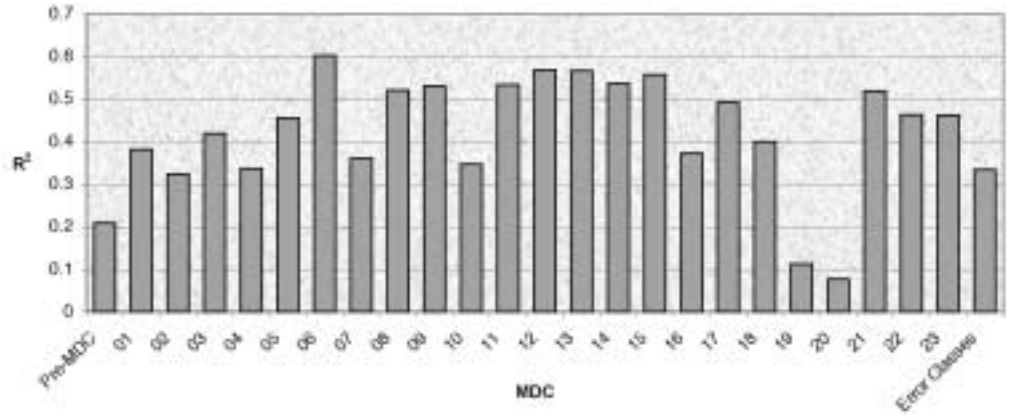
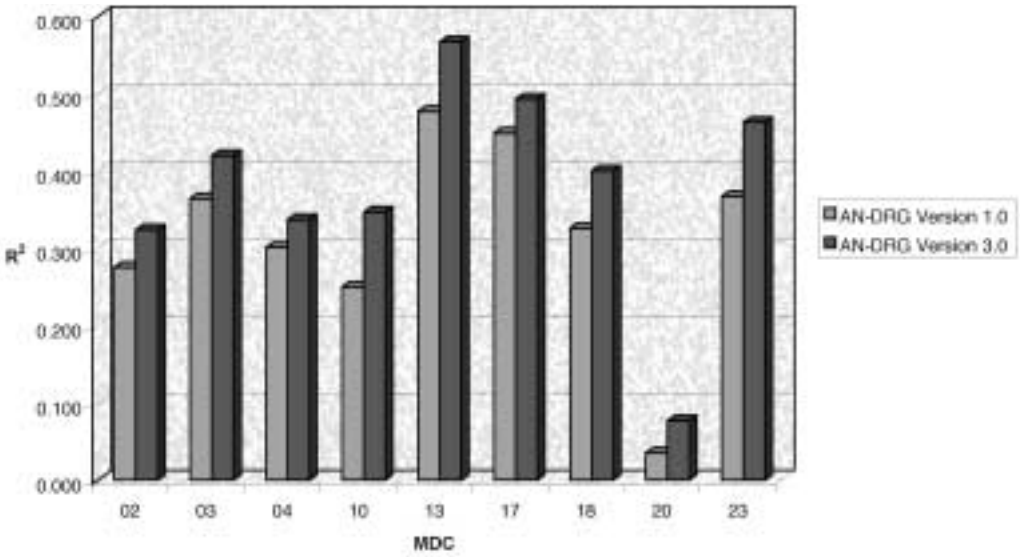


Figure 2: R^2 values for MDCs where there was a definite improvement between AN-DRGs versions 1.0 and 3.0, Australian data 1993–94, trimmed



A criterion of a CV greater than or equal to 100 for trimmed data was used to identify the more heterogeneous DRGs. For AN-DRGs there was a small increase in the number of heterogeneous DRGs from 34 in version 1.0 and 2.0 to 41 in version 3.0 (Table 2). The HCFA version had the smallest numbers of more heterogeneous DRGs and APR-DRGs the largest. The latter is a reflection of the large increase in the total number of groups and the low volumes of cases in the higher severity classes.

Table 2 also shows that version 3.0 performed best on the measure of the proportion of outliers trimmed (with 5.5 per cent of outliers removed) followed by the AP and APR versions.

Using a criterion of fewer than 50 cases per million for untrimmed data, the numbers of low volume DRGs decreased with each version of AN-DRGs. Among the US versions the HCFA DRGs had the fewest number of low volume DRGs (Table 2). As expected, the APR-DRGs paid a high price for its extensive use of complexity classes with 725 DRGs (47 per cent) having fewer than 50 cases per million. Low volume and low cost weight DRGs are less justifiable than low volume DRGs where the costs are high, so we created a criterion of fewer than 50 cases per million and a cost weight less than three. Adopting this criterion, the number of low volume/low cost weight DRGs for the APR system decreased to 322, that is 21 per cent. Excluding those APR-DRGs where the average length of stay exceeded 12 days decreased this further to 206, or 13.5 per cent. While this was an impressive improvement, it still represents a considerable challenge to estimate cost weights with any precision.

When each Australian version of DRGs was applied to data for four groups of NSW public hospitals, in conjunction with a simple funding model, AN-DRGs version 3.0 produced a slightly better result for principal referral hospitals than the other versions. However, the change in the distribution of funds between the different types of hospitals was very small (Table 4). This result was consistent with the very modest improvement in predictive validity of version 3.0 as compared with the other versions.

Table 4: Proportion of funds allocated to different public hospital types, for three Australian DRG versions, NSW data 1993/94 untrimmed

Peer Group	Proportion of funds AN-DRG version1.0 %	Proportion of funds AN-DRG version 2.0 %	Proportion of funds AN-DRG version 3.0 %	Total cases
Principal Referral	34.05	34.10	34.21	346 284
Major Referral	18.20	18.20	18.24	224 360
Other hospitals with >50 beds	34.43	34.38	34.28	446 378
Other hospitals with <50 beds	13.32	13.32	13.27	160 899

It is not possible in this paper to reproduce fully the results of comparing the resource hierarchy derived from the national cost estimates by AN-DRG with the hierarchy used in version 3.0 (Palmer *et al*, 1997). There was often good agreement between the ordering of the average cost estimates and the surgical hierarchy. However, there were some instances where the hierarchy was questionable, namely in the pre-MDC group, and in MDCs 1, 2, 3, 5, 6, 8, 10, and 11. Some of these problems were addressed in AR-DRGs version 4.0, but these efforts were hampered by the lack of cost data (Commonwealth Department of Health and Family Services 1998). Further evaluation of the surgical hierarchy should be undertaken once these cost data are available.

AR-DRGs became available after the evaluation project had been completed. We calculated an R^2 of 0.560 for version 4.0 using 1996/97 data from NSW compared with 0.503 for version 3.0 using NSW 1993/94 data. Mindful that some of this improvement may be due to more recent and therefore improved data, we measured the R^2 for version 3.1 using the 1996/97 data. Indeed, the resulting R^2 value of 0.533 indicated that better, more recent, data do account for some of the

improvement (Table 5). The MDC level analysis for AR-DRGs version 4.0 showed that the greatest increase in R^2 occurred in MDC 19 where the R^2 was 0.440 as compared with 0.113 for version 3.0.

Table 5: R^2 values for NSW Public Hospital Data using AN-DRGs Version 3.0, 3.1 and AR-DRGs 4.0

Year	R^2	DRG version
93/94	0.503	AN-DRGs Version 3.0
96/97	0.533	AN-DRGs Version 3.1
96/97	0.560	AR-DRGs Version 4.0

To assess the impact of the quality of the data on the performance of the DRGs we applied these systems to data from the State of Maryland in the US. The R^2 values were substantially better for the Maryland data no matter which system was used. For example, the R^2 for AN-DRGs version 3.0 increased from 0.505 to 0.540. The main difference between the Australian and Maryland data was that the latter were much richer in diagnoses and procedures. We calculated the casemix adjusted average number of diagnosis and procedure codes per case for the Australian and Maryland data excluding the day cases which were largely absent in the Maryland data. For Australian data the average for diagnoses was 2.44 compared with 3.46 for Maryland, and for procedures the average was 0.95 for Australia and 1.65 for Maryland.

Discussion

Across all the main performance criteria, version 3.0 performed best of the AN-DRGs. It had a slightly improved R^2, the smallest number of heterogeneous DRGs, proportion of cases trimmed and low volume DRGs. Version 3.0 also performed well when compared with the US versions. As expected due to its much larger number of classes, the APR system yielded a better R^2 than AN-DRGs version 3.0, but the APR-DRG system paid a price for its size through more heterogeneous and low volume DRGs. For these latter criteria the AP version performed better than the APR version but at a cost of a smaller R^2.

The wide variation in R^2 values at the MDC level noted previously for US DRGs, was also a feature of the Australian versions (Reid Palmer & Aisbett 1991). Furthermore, there were often marked differences between the surgical and medical sections of an MDC, a poorer R^2 performance for the medical DRGs being the common pattern. For recent revisions, the review process has concentrated on clinical recommendations, made through the ACCC, that were subsequently evaluated for technical performance by the Commonwealth, and supplemented by additional research on specific issues (Commonwealth Department of Health and Family Services 1998). The results presented here suggested that a more systematic approach to improving the R^2s for the poorly performing MDCs (particularly MDCs 2, 4, 10, 16, 19 and 20), and medical sets of DRGs has considerable merit.

A salient example was the gain in R^2 yielded by the improved performance of MDC 19 in AR-DRGs. The creation of a same day DRG for mental health treatment without electro-convulsive therapy was most effective in reducing the variation. This was because all the same day cases for this MDC are allocated to the one DRG and all these cases have exactly the same length of stay. The significantly improved R^2 for the MDC reflected this. Worthwhile gains in R^2 are achievable by improvements to poorly performing MDCs and medical groups of DRGs. However, the improved performance of MDC 19 is somewhat artificial because placing all the same day cases in the same DRG ignores the diversity of the clinical conditions treated in psychiatry on a same day basis. It is important that any recommendations for improvement be based on the application of the clinical coherence criterion rather than on same day attendance alone. Similarly, a review of the individual AR-DRGs that perform relatively poorly based on the homogeneity criterion would also be desirable.

One important difference between AN-DRG version 3.0 and the AP-DRGs was the more extensive use of age as a classificatory variable in the Australian system. This has been justified by the relative lack of secondary diagnoses in the Australian data compared with US data. However, the slight difference in R^2 between version 3.0 and the AP system indicated that little was gained where age was used as an alternative to secondary diagnoses for allocation of cases to more complex classes. Furthermore, this acts as a deterrent to improving the quality of coding. Thus we recommended that the use of age be phased out except for the oldest and youngest age groups. A modified version of this strategy was adopted in version 4.0 where the number of adjacent DRGs split on age and a complication/comorbidity decreased from 60 to 32.

It is important to work simultaneously on improving the quality of the hospital discharge data. The strong impact of better quality data on predictive validity indicated that further attempts to refine the Australian DRGs by the more extensive use of the secondary diagnosis codes will require significant improvements in the comprehensiveness of the diagnosis reporting and coding. Details of the data quality aspect of the project are reported elsewhere (Palmer *et al.* 1997).

Despite shortcomings in the quality of the Australian data, the results for the APR-DRGs indicated that there was scope for improving the DRGs performance using this approach. Averill (1995) reported a similar difference in R^2 for the APR-DRGs compared with AP-DRGs using US data. The difference between APR-DRGs and AP-DRGs was a good deal better than was found several years ago for the Refined system compared with the AP system (Reid Palmer & Aisbett, 1991). One of the reasons for this was that the Refined DRGs were based on the HCFA DRGs while the APR-DRGs were based on the AP system, as was the first version of AN-DRGs.

The good performance of the APR-DRGs makes this version very attractive despite its large number of classes. As mentioned above, the main problem with a large number of classes is deriving valid and reliable cost weights. However, this problem could be overcome by using regression analysis to determine cost increments between the individual severity classes and pooling data over more than one year. One of the attractive features of the APR-DRGs is the availability of mortality risk classes for application to hospital outcomes measurement. There has been little application of casemix data for outcomes assessment in Australia and yet this is an area where the casemix adjustment of data is sorely needed.

Conclusion

The modest increases in the R^2 values, and the results obtained from the hospital funding simulations, indicated that the changes from version 1.0 to 3.0 of AN-DRGs did not produce much improvement in overall performance as measured by these criteria. Furthermore, the extensive use of age as a classificatory variable in AN-DRGs version 3.0 did not produce significant gains overall as compared with the somewhat greater use of secondary diagnosis-based classes which characterise AP-DRGs.

The main driver for change in AN-DRGs was the ACCC and the main reason for these changes was to improve the clinical coherence of the DRGs, especially for interpretation by Australian clinicians. We would not wish to argue, on the basis of the R^2 results, that the creation of AN-DRGs has not been worthwhile. Apart from the gains in clinical coherence and the credibility of the AN-DRGs, the process of development has exposed a wide range of clinicians to casemix concepts. The acceptability of the tool by medical and other clinicians has been enhanced considerably as a consequence.

How is it that the extensive clinical and statistical input into the development of the AN-DRG versions and the correction of what seem to have been a large number of anomalies in the placement of disease and procedure codes in the US versions have not been reflected more strongly in the R^2 results for the first three Australian versions? There are several possible explanations for this apparent contradiction.

First, while extensive changes may have been justified according to the clinical coherence criterion, many of these changes affected relatively small numbers of cases. Second, the cost of the new groups formed may have warranted their creation, especially where the activities are performed in a small number of hospitals. The use of length of stay as the dependent variable in this and previous evaluation studies may not have done justice to improved performance which would only be revealed if data on costs for individual patients had been available. Third, improved performance may have been masked by data quality problems.

While the results for AR-DRGs version 4.0 are limited to more recent and improved data, there was a marked improvement compared with AN-DRGs. However, this was not as spectacular as the improvement in the results for the Maryland data as compared with the Australian data. An important element in improving the performance of DRG versions is the quality of diagnosis reporting and coding. However, unlike previous modifications of AN-DRGs, the improvement with AR-DRGs is more significant than found with versions 1.0 to 3.0. This appears to be partly a result of the modification of MDC 19 through consolidating the day cases, but other changes such as modifications to the complication/comorbidity classes have no doubt had some impact. A more comprehensive evaluation of AR-DRGs would be required to identify the main sources of improvement.

In light of the considerable resources consumed by the existing DRG refinement process in Australia, it would be wise to use the APR-DRGs as a benchmark for future work. Further research should also include the application of the APR version for mortality and other types of risk adjustment of data. This unique feature of the APR version has the potential for improving considerably the information value of casemix data as one measure of patient care.

Acknowledgements

This project was supported by a grant from the Commonwealth Department of Health and Aged Care. Jane Alanthwaite and Sheree Crick assisted with preparing the manuscript and tables.

References

Australian Casemix Clinical Committee 1996, *Report on the Clinical Review of the Third Version of the AN-DRG Classification*, Commonwealth of Australia, Canberra.

Averill RF 1995, The evolution and current use of casemix measurement using DRGs, *in* J Hofdijk (ed) *Proceedings Patient Classification Systems/Europe, 11th working Conference*, PCS/E, Oslo, 30–43.

Bender JA & McGuire TE 1995, A focussed look at the L3H3 exception policy, *in* J. Hofdijk (ed) *Proceedings Patient Classification Systems/Europe, 11th working Conference*, PCS/E, Oslo 266–277.

Bland M. 1995, *An Introduction to Medical Statistics*, 2nd edition, Oxford University Press, Oxford.

Commonwealth Department of Human Services and Health 1995, *Australian National – Diagnosis Related Groups: Report on the Development of the Third Version of the AN-DRG Classification*, Commonwealth Department of Human Services and Health, Canberra.

Commonwealth Department of Health and Family Services 1998, *Development of the Australian Refined Diagnosis Related Groups (AR-DRG) Classification Version 4 Volume 1*, Commonwealth Department of Health and Family Services, Canberra.

Coombes J Eckstein G & Gomes R 1995, *Determination of AN-DRG trim points for NSW acute hospitals*, Health Services Research Group, The University of Newcastle, Newcastle.

Department of Health and Human Services, Health Care Financing Administration 1996, Medicare Program: Changes to the Hospital Inpatient Payment Systems and Fiscal Year 1997 Rates; Final Rule, *Federal Register*, vol 61, no 170, pp 46166–239.

Health Care Financing Administration 1994, *Severity Adjustment of DRGs*, HCFA, Washington, D.C.

Palmer G & Aisbett C 1996 Defining and paying for outliers: an evidence-based clarification of conceptual issues, *in* J. Hofdijk (ed) *Proceedings Patient Classification Systems/Europe 12th working Conference*, PCSE, Sydney, 12–21.

Palmer G Reid B Aisbett C Fields S Kearns D & Fetter R 1997, *Evaluating the performance of the Australian National Diagnosis Related Groups: Report to the Commonwealth Department of Health and Family Services*, The Centre for Hospital Management and Information Systems Research, The University of NSW, Kensington.

Pilla J 1994, Development of AN-DRGs: Meeting the concerns of clinicians, *Medical Journal of Australia*, 161, S9–S11.

Reid B Aisbett C Ng L-M. Lohmann J & Palmer G 1992, *Evaluation of the performance of the Australian National DRG Grouper*, Centre for Hospital Management and Information Systems Research, University of NSW, Kensington.

Reid BA Palmer GR & Aisbett C 1991, Choosing a DRG grouper for Australia: issues and options, *Australian Health Review*, vol 14, no 3, 285–300.