The value of linked data for research into health services use and expenditure: An Australian perspective

Chris Kelman

Medical Advisor, Information and Research Branch, Department of Health and Ageing

Abstract

In this country, significant health service data are collected in the process of managing payments to providers. Despite the investment involved and the extensive coverage of this resource, access to these data for research has not been made a priority – one of the main reasons being that data is collected by different levels of government. This situation makes 'whole of care' analysis impossible unless specific agreements are developed between parties to allow linkage of the separate datasets.

With the growing focus on safety in health care, a gradual cultural change is occurring. This new culture of 'safety and quality' requires accurate and timely information. There is an obvious justification and need for the use of all available health data to support this.

Privacy protection is fundamental, but adequate protection can arguably be provided with current technology while allowing the appropriate research use of these data. This approach would provide a resource for health care monitoring and maximise the use of this resource for community benefit.

Introduction

I am going to talk about how we have used Australian linked data to examine the use of health services and their costs.

I will describe some of the datasets that we have available, discuss the limitations and special barriers that exist in this country, talk about some of the work that has been done with these data and finish with some ideas on where we are going.

In theory, linked administrative data are of course well suited to the measurement of costs and rates of service use – these datasets were designed for accountants rather than researchers. So it is only when we try to take the analysis the next logical step to outcomes analysis, cost-benefit analysis and so on, that we come up against their limitations.

It is fairly straightforward to determine simple measurements of how much is spent by whom but the problems start when we ask questions like: how much does it cost per year to treat emphysema, or how cost-effective is our management of hepatitis B and so on. To use these data for anything so sophisticated is stretching the friendship.

So, my point is that until such time as we have complete and accurate data in the form of an integrated electronic health record, the monitoring of our health care system will require ingenious use of all the data resources we have at our disposal. To this end, the linkage of available datasets is essential to leverage the sparse information available.

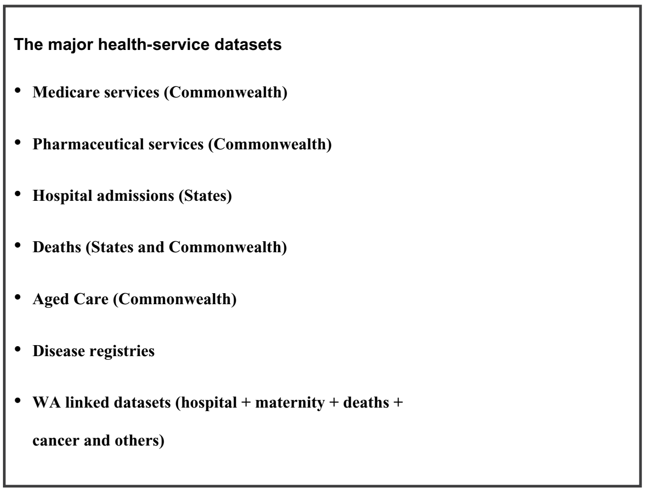
So are we making the most of what we do have? I would say no, not yet. There are a number of unique challenges involved in doing what may seem to be common sense.

One of the main barriers in this country is a consequence of the fragmented system that we have for health funding. The Commonwealth funds doctor services and pharmaceuticals directly through the Medicare Benefits Scheme (MBS) and Pharmaceutical Benefits Scheme (PBS) schemes, whereas hospital services, although partially funded by the Commonwealth, are managed by the states and territories. Both these entities keep good service records, but unfortunately, have for various reasons been reticent in sharing these data. In fact in many cases these data are legally protected from such uses. The result is that until recently, it has been impossible to get a complete view of health care in Australia. It has been very difficult to determine the actual 'holistic' cost of treatment for medical conditions or to assess treatment outcomes. In fact any analysis at the patient level has been difficult.

So, what are we doing about it?

Data resources

Australia has a number of valuable data resources:



The major health-service datasets

- **Medicare services (Commonwealth)**
- **Pharmaceutical services (Commonwealth)**
- **Hospital admissions (States)**
- **Deaths (States and Commonwealth)**
- **Aged Care (Commonwealth)**
- **Disease registries**
- **WA linked datasets (hospital + maternity + deaths + cancer and others)**

Figure 1 Health service datasets

The Health Insurance Commission (HIC) collects records of all services covered under the MBS, and also PBS claims for a subset of transactions. These datasets are archived by the Department, but unfortunately little clinical detail is contained in these claims.

Each of the state health Departments collects comprehensive and useful data about all hospital admissions. If these records are combined with the Commonwealth data, a fairly complete record is created of health services for individuals. However, a serious limitation for research use of these data is the fact that very little clinical detail is stored in the MBS and PBS datasets.

Leading the field, WA has established a major linked health data resource that includes hospital, maternity, cancer and deaths. You will be hearing more about their exciting work during this symposium. Other states have conducted various studies into injury prevention poisoning and cancer using their hospital admissions datasets.

I should add that provision already exists in legislation at both the Commonwealth and also often at the state level to use these data for research and monitoring of health care. These provisions generally take the form of “necessary in the public interest.” In fact quite a lot of work has been done with the data available in Australia, however, in general this has utilised either hospital admissions data or MBS data but not both due to the difficulties in arranging data exchange and linkage between the states and the Commonwealth.

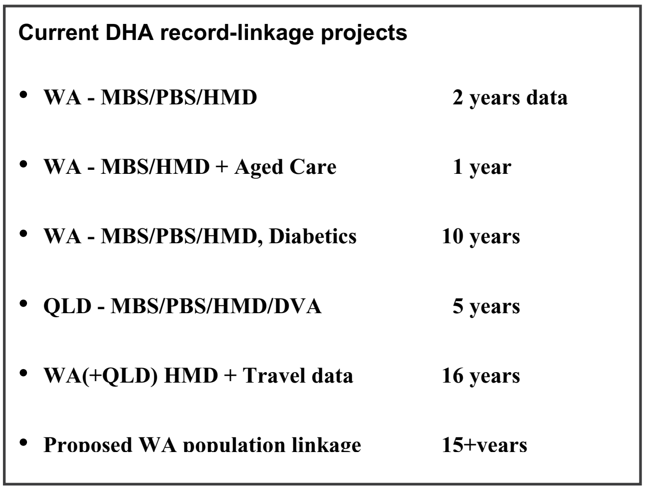
Extensive hospital event data has been collected by all states since 1991 to form the National Hospital Morbidity Data Set (NHMDS) for the purpose of maintaining the Australian National Diagnosis Related Groups (AN-DRG) and calculating weightings for Casemix funding . These data, which include International Classification of Disease (ICD–9, ICD–10) coded information on diagnosis and procedures, are however, de-identified. This process is intended to prevent patient-level analysis being performed. So despite the huge investment in its collection, the great potential for this dataset as a resource for outcomes analysis, and even as a nationwide disease registry is lost. Using these data, we can’t tell the difference between two patients being admitted for a particular problem and one patient being admitted to two different hospitals with the same problem – quite an important distinction if you are interested in treatment outcomes!

Commonwealth data linkage projects

To investigate the potential for ‘whole of care’ research, a record linkage unit was established within the Commonwealth Department of Health in 1997.

Negotiations commenced with WA and data was exchanged in 1998 allowing for the first time, linkage of the hospital and MBS datasets. I must thank Health WA for their input to this project. As this was new territory, privacy protection was considered to be the number one risk, so names or hospital identifiers were removed from the data. Instead, we used probabilistic linkage based on date of birth, gender and Postcode. This produced only a 70 per cent linkage rate. A report of the results of this study was released in 2001. (Copies of this ‘Occasional Paper No. 9” are available on the Department’s website at www.health.gov.au).

The first few years work for this unit have principally been devoted to the planning and negotiation necessary to bring together the necessary data. Even when all parties are eager, the process is remarkably time consuming. The diabetes project, for example, has taken almost four years from initial negotiations to the actual supply of data for a research project. A list of current research projects is presented below.



Current DHA record-linkage projects

• **WA - MBS/PBS/HMD**	**2 years data**
• **WA - MBS/HMD + Aged Care**	**1 year**
• **WA - MBS/PBS/HMD, Diabetics**	**10 years**
• **QLD - MBS/PBS/HMD/DVA**	**5 years**
• **WA(+QLD) HMD + Travel data**	**16 years**
• **Proposed WA population linkage**	**15+years**

Figure 2 DHA linkage projects

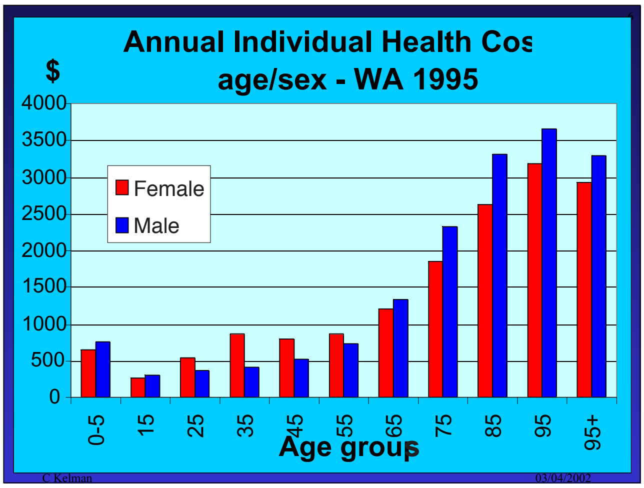


Figure 3 Annual costs by age group.

Starting with these ‘whole of care’ costs, Figure 3 shows the varying costs of treatment by age group and sex. (These data represent around 83 per cent of the WA population).

While these totals include the bulk of service costs, it should be noted that a number of services are not included, in particular outpatients, nursing, residential care and other ancillary services, we use the 'whole of care' term loosely. The pattern is as one would expect with higher costs for females in the reproductive years, and increasing for males towards the end of life.

Note the reduction of costs over the age of 95 – perhaps one could speculate that if average life expectancy continues to increase, total costs may actually decrease in the long term!

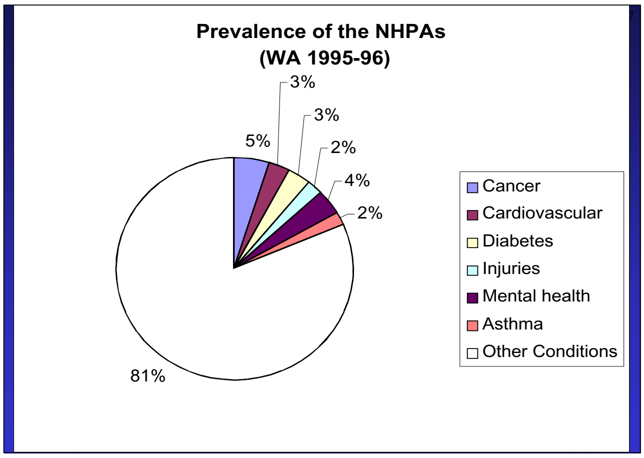


Figure 4 NHPA prevalence

Now looking more closely at the patterns of disease, in particular, the National Health Priority Areas (NHPA). I should note again that this is the first time that this information has been available, these datasets have not been linked previously.

Here the proportion of the (matched) population affected by each of the six NHPA conditions (Cancer, Cardiovascular disease, Diabetes, Injuries, Mental Health and Asthma) is shown in relation to the total population. You can see that this group of patients makes up around 19 per cent of the total. However, when we looked at the costs of these patients we found that the six NHPAs made up around 53 per cent of the total treatment costs, demonstrating that these groupings, though fairly arbitrary, do in fact capture a large proportion of the morbidity. Cancer and cardiovascular disease account for almost a third of total expenditure via medical, hospital and pharmaceutical treatments.

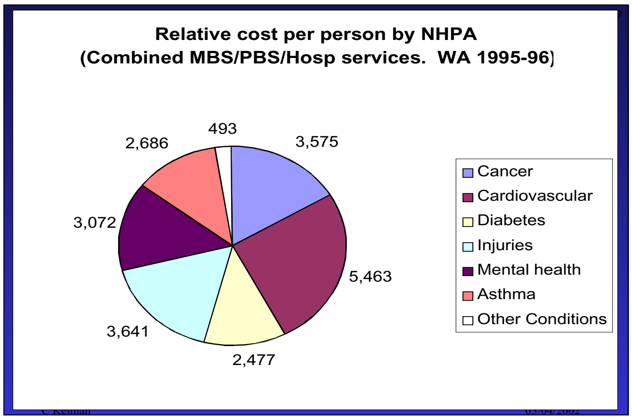


Figure 5 Relative costs of NHPAs

Figure 5 provides a more detailed view, with costs in dollars for 1995–96. Here I have compared the annual costs of treatment for each of the six conditions, one can see particularly the high per-person annual cost for cardiovascular disease ($5,463).

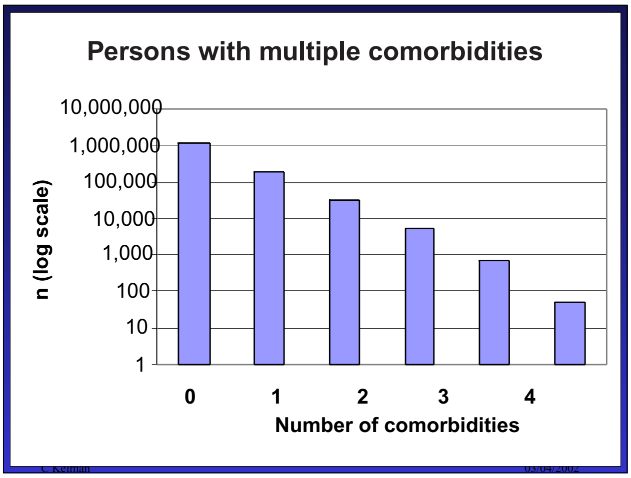


Figure 6 Persons with multiple comorbidities

We also looked at the people who had multiple NHPA conditions, and found an interesting relationship: when plotted on a log scale it is evident that the number of persons in each group decreases by a factor of about 10 for each additional comorbidity. So very roughly, if 120,000 have only one NHPA condition, around 12,000 have two conditions and so on. For those of us who like patterns, this one is surprisingly regular and continues when we look at the additional cost for each additional condition.

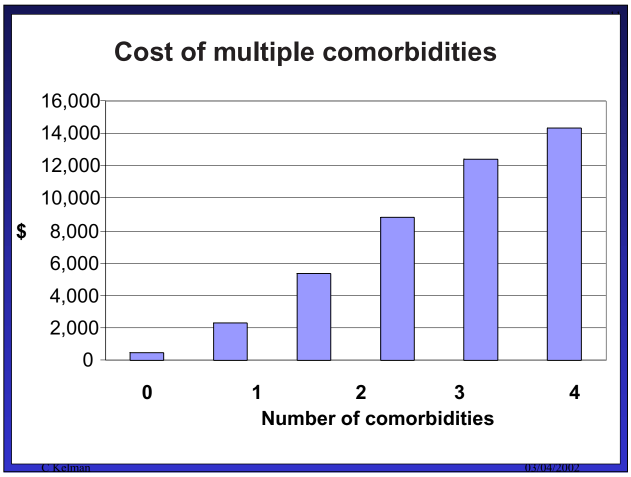


Figure 7 Costs of multiple comorbidities

In Figure 7, looking at the total costs of treatment for the groups, we found a simple linear relationship, in other words, if you add an extra condition, the annual cost of treatment increases by around $3,000.

In another study, the WA data was again used to examine the differences in health service availability in different areas. There was particular interest in comparing service use in rural and remote areas with urban areas. The Accessibility/Remoteness Index of Australia (ARIA) measure was used for these comparisons. The ARIA measure is based on road distances from a finite number of localities to various 'service centres' identified in the 1996 census. (In contrast, the older Rural, Remote and Metropolitan Area (RRMA) classification used crow flying distance). The ARIA is a continuous variable ranging from urban = 0 to remote = 12.

There has been some interest in the concept of 'avoidable admissions' and the possibility that reduced availability of primary care services may lead to 'service substitution' that is an increase in hospital use and also possibly a more serious clinical condition by the time a patient actually presents for care. These issues were able to be investigated as both hospital and General Practitioner (GP) services were included in the linked dataset.

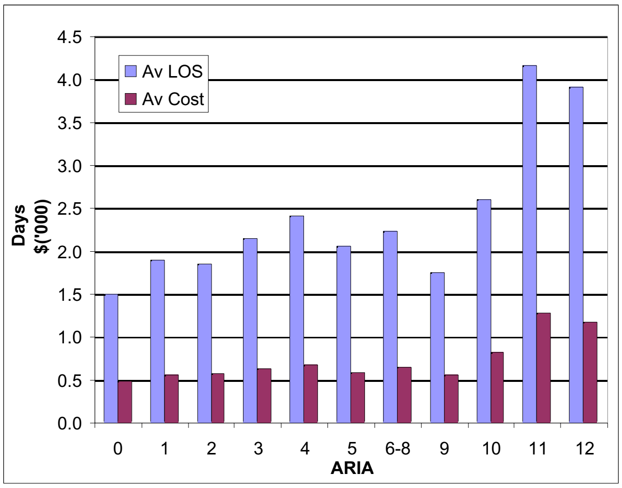


Figure 8 Average Costs and Length of Stay by ARIA

In Figure 8, the relationship between average cost of treatment in hospital and LOS are shown in relation to rurality (using the ARIA scale).

You can see that LOS and average costs increase dramatically in the most remote hospitals, but were pretty constant up to ARIA 9, now ARIA 9 means "very little accessibility of goods, service and opportunities for social interaction." It would seem that anything greater than this must mean 'seriously remote'! So the fact that patients stay longer once being admitted to a hospital, may be influenced more by the effort required to get there than their degree of sickness. Obviously if there is any doubt about a patient's welfare they will be admitted if they live 12 hours drive away and cannot be asked to come back in the morning.

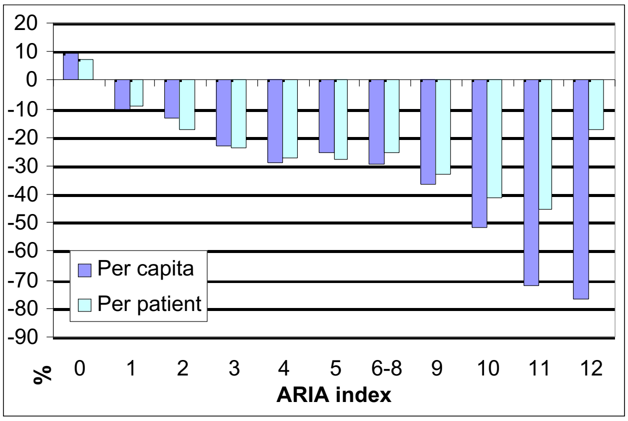


Figure 9 Primary care services by ARIA

Next we had a look to see if the supply of primary care services were related to rurality. We looked at variation in the use of MBS and PBS services compared to the weighted average service use across the state.

It can be seen that the average number of primary care services used decreases with increasing remoteness. However it should be noted that these results need to be adjusted, the per capita results do not tell the full story and it can be seen from the 'per patient' measure that in very remote settings, a large proportion of the population appears not to have used any services at all. When you look at those who did, they appear to make fairly 'normal' use of the services.

So overall, we can say that persons living in remote areas use 20–40 per cent less services, but note that in very remote areas, many seem to find the services to be totally inaccessible and thus do not use them at all.

Also of interest is the possibility that the urban anomaly here is due to supplier-induced demand, though if this is true, the effect is quite small when compared to the state average.

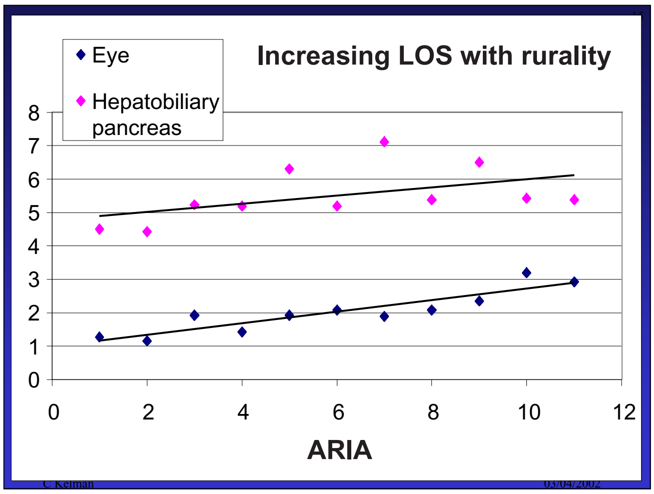


Figure 10 Length of Stay by ARIA

Another proposition was that due to less available primary care services, delayed diagnosis may occur, and thus patients may end up being sicker when they finally present to hospital. An indication of this may be the Length of Stay (LOS) for these admissions (Figure 10).

We examined the LOS for all the Major Diagnostic Categories, I have shown a couple here that showed some significant association with degree of rurality. It is interesting to speculate whether the three fold increase in LOS for 'Diseases and Disorders of the Eye' is related to the type of services provided or whether it is an indication of worsening clinical condition in rural areas. Unfortunately, ICD codes were not available in the hospital data to permit further investigation of this question. However this information is now available to a subsequent study and we will investigate this further.

Aged care

Another area that has evaded examination due to absence of integrated data is the primary care – aged care – hospital interface. We were able to have an initial look at this area again using WA linked data.

Hospital data were linked to MBS and Aged Care residential data using DOB, Gender, and Postcode. These data were linked probabilistically and possibly rather loosely. Nevertheless, a workable linked dataset was produced and some suggestive, if not conclusive, results have appeared.

Last week, the AIHW released a report on dementia showing the condition to be rapidly growing into one of the biggest health problems in this country. We looked at service use by these patients and found some things of interest. It appears to be the case that those with cognitive impairment are less likely to be admitted to hospital from a nursing home than others, and once admitted, less likely to receive services. This may be the result of many factors and needs further investigation.

We also found that those patients who already had a nursing home placement to return to generally had shorter admissions than those who were waiting for placement – this observation would seem to support the move towards provision of step-down care and rehabilitation services.

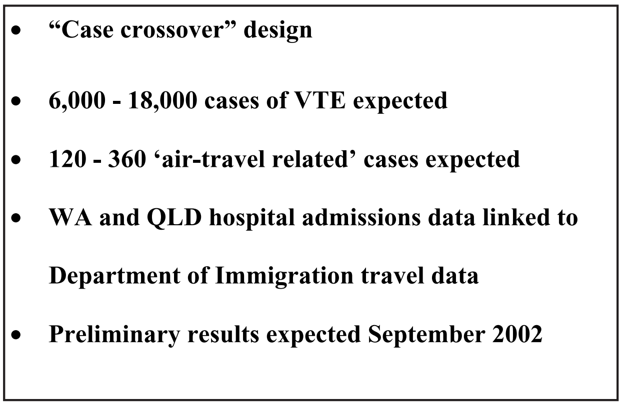
These results are all fairly speculative, however, the point is that interesting areas are revealed using these data and provide a stimulus for more detailed examination. The results also further justify the collection of data and provide a stimulus for improving data quality and access for research use.

The DVT Study

For this study, we have several benefactors all of whom I would like to thank: The Health Departments of WA and QLD and the Department of Immigration, Multicultural and Indigenous Affairs (DIMIA) are providing the data, while the project has been funded by the Department of Transport and Regional Services.

This is a study looking at the strength of the association between prolonged air travel and VTE (DVT and PE), a topic of some international interest in the media over the last 18 months.

Cases (subjects) will be selected from the hospital admissions database, ie those who have ICD codes that identify them as having been treated for DVT or PE over the 15 year period of the study. Records for these persons will then be linked with travel data from the DIMIA, travel histories for all cases will be collated for the month preceding each episode and also, for a previous month to provide the control period for each case. The data will be examined to calculate if there is an increased risk of developing DVT or PE after travel by air. We expect some initial results within six months.



- **"Case crossover" design**
- **6,000 - 18,000 cases of VTE expected**
- **120 - 360 'air-travel related' cases expected**
- **WA and QLD hospital admissions data linked to Department of Immigration travel data**
- **Preliminary results expected September 2002**

Figure 11 The DVT study.

Privacy protection

It is interesting to note that record linkage has been widely adopted for the purpose of fraud and crime detection but that the use of linked health data in the monitoring and management of our health system has been resisted. The importance of privacy protection for health data should not be underestimated, particularly when future health records may include information not only about ourselves, but our progeny, in the form of genetic information.

Nevertheless, it is possible that there is no conflict here after all. As has been mentioned, there are a number of approaches available for protecting individual privacy while using health data for research. With modern information technology we already have the potential to enforce the protection of privacy and confidentiality.

For the purpose of health outcomes research and analysis of treatment costs, the use of patient-level data is necessary. Ideally, a national health identifier would be used to allow accurate linkage of records both within and between datasets. This would facilitate linkage and allow analysis to proceed at the individual level rather than the episode level – essential if we wish to assess the outcomes and cost-effectiveness of services provided.

We must also continue our work on state-Commonwealth collaboration – the sharing of health data is essential if we are interested in what's actually going on in health care – surely a rhetorical question.

Conclusion

To summarise, we have quite significant health data resources in this country. However, these data have been collected with service payments in mind and some are lacking in clinical content. Access to these data for research has not been made a priority and there are a number of reasons why this has continued to be difficult. One of the main problems being that data are collected by different levels of government and this makes 'whole of care' analysis impossible unless agreement is reached to allow linkage.

Over the last few years, progress has been made in the use of linked data in Australia. With the growing focus on quality and safety in health care, a gradual cultural change is occurring. Some may say not before time. Concepts that were foreign only a few years ago like evidence based practice, quality assurance and outcomes monitoring have become part of the language of health care. This new culture requires accurate and timely information. There is an obvious justification for the use of all available health data to achieve these ends.

Barriers resulting from the state-Commonwealth interface can be reduced by focussing on the shared goal of improving quality and safety and reducing the cost of health services. This process is already underway.

While privacy concerns in the use of these datasets are absolutely valid, I propose that there are now ways to fully protect privacy while allowing for the research use of these data and that this approach will provide a resource for health care monitoring and thus maximum community benefit.