

**OECD AGEING RELATED DISEASE STUDY**

**TECHNICAL REPORT**

**USING HOSPITAL ADMINISTRATIVE DATABASES FOR A DISEASE-BASED APPROACH  
TO STUDYING HEALTH CARE SYSTEMS**

*Pierre Moise*

*August 2001*

This work has been completed as part of the OECD project on Ageing Related Diseases, which has received the support of the US National Institute on Aging (IAG Y1-AG-8363-01). The authors thanks all the experts in countries who have provided information on their national information systems. Remaining errors are his sole responsibility.

## SUMMARY

This paper provides a preliminary overview of the main hospital administrative data sets potentially available in several OECD countries (Australia, Belgium, Canada, Denmark, Finland, France, Great Britain, Italy, Sweden and the United States). The information is provided within the context of the OECD project studying how cross-national differences in the treatment of ageing related diseases affect health outcomes, for which these administrative data bases may offer the core tools for analysis.

Hospital administrative data bases have evolved from their origins as monitoring tools used by health policymakers and planners for tracking hospital activity. Today they are also used by health care researchers as invaluable tools for conducting clinical, policy and economics-related research. Despite the wealth of studies using these data, there have been few that have pooled together information about these data bases and analysed them from a health care research perspective. The aim of this paper is to provide an inventory of such sources across countries, which had not been available to policy analysts and researchers in the past.

The analysis herein focuses on the usefulness of using hospital administrative data for comparing treatment costs and outcomes and they could be used for a variety of diseases with well identified acute episodes of care. This is the case for example for heart attacks, stroke, hip fractures or breast cancer to name a few. Information on the core data elements, such as way of entry, primary diagnosis, main procedure performed and reason for discharge, are available in most countries where these datasets exist, although the range of information they provide may vary. This paper provides an in-depth analysis of the range of information such databases may provide. The analysis considers the identifiability of diseases, the description of patient characteristics and possibilities for investigating health outcomes and costs. The paper also briefly explores characteristics of available information on ambulatory and out-patient care and the links with inpatient data.

Depending on the disease, the depth of analysis achievable with inpatient data will be limited by the degree to which the disease in question is treated on an inpatient basis. Several limiting factors of inpatient data such as the ability to link to other data sets, a lack of detailed cost information and the degree to which they are relevant to studying specific diseases will be investigated. These limitations necessitate the need for complementary sources of data to bridge these gaps. Therefore, this paper will also examine the usefulness of information sources created from a disease-specific perspective and often within an epidemiological context. These sources of information will complement the inpatient data, the degree of complementarity depending upon the characteristics of the disease in question. The paper presents the characteristics of each disease being studied in the ageing-related diseases project with respect to the in-patient data and explores how the disease-specific sources can help address them.

## TABLE OF CONTENTS

SUMMARY .....	2
INTRODUCTION.....	4
1. General overview and features of the datasets .....	5
1.1. An overview.....	5
1.2. General features .....	5
2. Patient characteristics, identification and the impact of privacy rules .....	7
3. Identification of diseases and comorbidities .....	7
4. Identification of health interventions .....	8
4.1 Acute interventions in inpatient setting .....	8
4.2 Other characteristics of admissions in an inpatient setting .....	9
4.3 Characteristics of the hospital where inpatient care was provided .....	9
4.4. Care provided in an outpatient setting .....	9
4.5. Specific data on pharmaceutical consumption .....	9
5. Relevance of the datasets for health services research .....	10
5.1. Relevance for studying disease outcomes .....	10
5.2. The needs of disease specific studies, general datasets and more targeted sources.....	10
5.3. From general to more disease specific datasets: from general datasources to more specific disease targeted datasets.....	11
CONCLUSION .....	12
REFERENCES.....	13

## INTRODUCTION

1. Hospital administrative data bases were originally designed for tracking hospital discharge summaries. As government have felt an increasing need to assess the volume of hospital production and link it with financing procedures, these data bases evolved to become important tools for monitoring hospital activity. Researchers took notice of the wealth of data available in one place and the consequent potential for conducting research. With advances in information technology they have been able to exploit this potential for clinical, policy and economics-related health research. Despite the wealth of studies it is difficult to find a single source that identifies these data and their potential for conducting international health care research.

Among the variety of sources used by OECD countries to monitor their health care systems, administrative data sources developed since the beginning of the 1980s for monitoring hospital systems are of special importance for health services research. Member countries have started to develop discharge abstract information collection systems following the experience of Diagnosis-Related-Groups (DRG) in the US (Wiley 1992). In other countries, such as in Scandinavia or the United Kingdom, even if systems have not been developed formally into DRG type data sets, OECD countries have large national registries or large national administrative databases, collecting diagnoses and treatments for the whole population.

2. The purpose of this paper is to provide a handy reference guide for available sources of micro-data. This has been done within the Ageing Related Diseases Project (ARD) at the OECD but should be of general relevance for researchers interested in the field. This paper will focus on analysing the specific characteristics of each of the data sets, using the information which was available to the author. These data sets differ in both scope and technical details. It may be particularly important to assess these differences with regard to the practical usefulness of the data sets for the diseases under consideration. The countries identified in this study represent a varying array of both complexity and comprehensiveness in their data, providing an excellent opportunity to illustrate the various possibilities of the type of information available. Many national data sources may only be referenced in national publications that are relatively inaccessible, which could explain why in certain cases they have not been referenced here. This paper can be considered as a step to more comprehensive efforts describing available data sources for most OECD countries.

3. The paper is structured as follows. First, a general overview of the available sources will be provided, together with a description of the main features of the datasets. Second the information on patients will be discussed in light of privacy rules. Third, the information on diseases will be presented, which could transform these datasets into useful epidemiological tools. Fourth, the paper provides information on health care interventions, including both inpatient and in some cases outpatient settings. Fifth, the paper discusses the usefulness of these sources for the purpose of applied health services research, including outcomes, costs and disease specific characteristics.

## 1. General overview and features of the datasets

### 1.1. An overview

4. The General Administrative Data Sources table in Annex 1 provides a summary of the main health care administrative data sources for several OECD countries. The table is the first to data to present such a synthetic overview of these datasources. Most of the information displayed in Table 1 pertain to inpatient databases, including descriptions of case-mix (DRG) systems, coding of health conditions and procedures and capacity to link to other sources of data. Table 1 also provides descriptions of pharmaceutical, outpatient and ambulatory care data that are collected by these countries.

**TABLE 1 General Administrative Data Sources**

5. Hospital inpatient data are generally fairly comprehensive and are available in great detail and for an increasing number of countries. They now can compare favourably with other specific data sources such as epidemiological studies and registries. Inpatient data may provide significant information for acute conditions such as heart attacks and hip fractures. In general, individuals suffering an occurrence of either of these events would be sent almost immediately to a hospital for treatment, thereby reducing selection bias in identifying these patients. For other conditions, diagnosed individuals are not necessarily treated as inpatients, such as is the case for breast cancer, or individuals may not immediately present themselves at a hospital for treatment which can occur for individuals suffering from the less acute phases of ischemic heart disease and stroke. In this case, these datasources have to be supplemented by other outpatient and ambulatory care databases, epidemiological studies and registries.

### 1.2. General features

6. In general, in-patient data sets were originally created as public monitoring tools of the hospital system. Once in place, health services researchers began to realise the potential of these largely administrative databases to provide robust analyses of factors that affect health. They are referenced in Table 1 by their main title. However, in many countries, hospitals may send different levels of information of their discharge abstract data to various public agencies. For example, patient specific identification, or information on costs might be lost when the data is transmitted to the central collection agency. Therefore, for a number of countries (e.g. Australia, Canada), specific state/province level data sets have been mentioned since they are likely to provide more detailed information, involving linkage possibilities not always available at the central/federal level. In most cases, the principal data set will refer to the inpatient data available. However, for some countries such as the United States and their Medicare database, the principal data set encompasses outpatient and other sources of information in addition to inpatient data.

7. The availability of these data sets over time depends on the range of implementation on the part of central public data collection monitoring efforts. Not only are these datasets useful to see whether or not treatments vary across countries, but also across time. The ability to analyse the data over a significant period of time is affected by the evolution of these data sets, especially with regards to the types of coding systems used. To mitigate the impact of evolutionary changes to these data sets, any substantive revisions to the data which have not been identified, especially for the latest year available, should be noted.

8. With respect to DRG type systems, they were originally developed in the US. However, other countries may have developed their own systems for collecting information on resource use and hospital outputs, even if these did not formally assume the DRG format (e.g. Scandinavia, United Kingdom). The

availability of data sets spanning a significant period of time is crucial to ascertain trends using micro data. However, this may not be possible for countries where these systems have been developed more recently, such as France or Italy, necessitating the need for alternative strategies.

9. Originally, most countries developed independent systems. However, the impact of the US DRG model has been felt strongly across OECD countries, and many OECD countries have adapted their systems along the U.S. Medicare DRG line, with some adjustment for national characteristics. Despite this apparent convergence toward a universal type of system, coding still differs from country to country. DRGs are baskets of diagnostic/treatment procedures of health services grouped by statistical analysis in order to facilitate the public payment process of health care financing. Contrary to what might have been thought when describing DRGs as a pure prospective tool depending solely on diagnosis, they in fact involve a significant amount of coding related to treatments (McClellan 1996). As a result, countries have had to develop their own scales and classifications in order to use DRGs<sup>1</sup>. Countries also differ in how they use DRGs: in some countries it directly specifies the payment such as in the US; in other countries the impact is more indirect and relates to the adjustment of public budgets to activity levels. In order to properly identify diagnostic and treatment procedures from inpatient data in countries that rely extensively on DRGs to code their procedures, further investigation will be needed to ‘tease out’ the relevant diagnostic and treatment procedures.

10. For most OECD countries, the data stem from databases that are used to administer public financing of health care. In some countries, the government may only provide coverage for stays in public hospitals. In other countries, both public and private hospital stays may be covered. Across OECD countries, variations in public hospital insurance schemes will mean inpatient discharge abstracts may not cover all hospital stays. Therefore, the analysis using these datasets should be cautious and take into account potential measurement bias, or make sure that all types of activities are covered since private hospitals may also play an important role in a wide range of interventions included in the analysis (e.g. cataract surgery, angioplasty). The column coverage provides relevant information to describe the relative coverage of each data set.

11. One of the main advantages of large administrative data sets is the possibility of obtaining representative samples on a broad basis. Thus, even if there are measurement errors created through mis-coding or imperfect data collection, these errors tend to be isolated and are relatively minor in number compared to the large size of the data sets, effectively rendering these errors insignificant. For the majority of countries, the total population would be represented. However, the population covered by some of these administrative data sets may be limited. For example, Medicare in the U.S. covers only the 65 and over population, in other countries coverage may be limited to public hospitals only. In addition, coverage of the data sets is not only limited by its representative population, but in some instances to institutional factors involved in the data collection process. For example, in Canada, several provinces have chosen not to submit in-hospital data for the national level data set identified in Table 1. Thus, as local variations and variations of practice between public and private facilities might be significant, results may need to be compared to global activity or economic aggregates to ascertain the overall reliability of the results obtained.

---

<sup>1</sup> See for example PMSI 1994 for France, or Australian similar efforts.

## 2. Patient characteristics, identification and the impact of privacy rules

12. In most data sets, age and gender are coded in the data. Place of residence may be included with a varying degree of detail. In countries like the U.S., the geographic coding has been precise enough to offer additional dimensions for analysis. For example, using these codes, the initial distance from the patient's residence to the hospital can be estimated and used to adjust for unobserved heterogeneity in empirical econometric estimates (McClellan Newhouse 1998). Geographic coding has also been used to obtain a proxy of socio-economic status, thereby allowing an assessment of the redistributive impact of Medicare (McClellan Skinner 1999). In other countries, such as Canada, postal codes make it possible to obtain information on small areas of residence of no more than a few hundred housing units. Detailed geographical codes would also be available in the U.K. For other countries like France, privacy rules mean detailed residency codes are suppressed when gathering the data at the central level. Therefore, aggregate data would not permit very detailed analyses using these codes. In certain countries, data may include other socio-economic characteristics, such as the availability of supplemental insurance (e.g. Australia, U.S.).

13. The linkage of discharge abstracts through time and with other databases using patient identifiers plays a key role in analyses of health outcomes and also for assessing socio-economic impacts. In contrast to the requirements for health services research, patient identification remains the most sensitive issue with administrative data sets in the public policy debate. There is a strong link between patient privacy and rights policies and public confidence in data gathering systems. However, the practical levels of requirements, and their potential consequences for performing socio-economic research have differed greatly. The interest in obtaining linked information, both over time and over care settings, has been acknowledged in countries such as the United States, Australia and Canada, and these countries have made significant strides in linking this information. In Scandinavia, a strong public tradition with central registries has also made it possible to gather linked information. By contrast, European countries such as France and Italy have strongly limited the possibility of keeping patient identifiers. France is currently considering the possibility of having patient identifiers to track rehospitalisations over time, without making it possible to link hospital discharge data with any other central social administrative files. In the United Kingdom the Oxford Record Linkage study has linked inpatient data to birth and death registries on a local level.

14. As an international organisation heavily involved in data collection, the OECD is sensitive to the issue of data confidentiality. The Organisation has embodied its commitment to this issue with a set of principles set forth in "Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data." <sup>2</sup>. An internal OECD scoping study has recently been released<sup>3</sup> to examine the issue of confidentiality regarding health research data and to assess the potential for further work in this area.

## 3. Identification of diseases and comorbidities

15. In the micro data sets, the coding systems for diseases are more or less convergent across countries. For most countries this will be ICD-9, but some countries may have already switched to ICD-10. Furthermore, the farther back in time the data span, the more likely it will be that earlier versions of

---

<sup>2</sup> OECD, *Recommendations concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* (September 23, 1980); available on the Internet at [www.oecd.org/dsti/sti/it/secur/prod/priv-en.htm](http://www.oecd.org/dsti/sti/it/secur/prod/priv-en.htm).

<sup>3</sup> See Directorate for Science Technology and Industry, (DSTI/STP/BIO(99)3).

ICD were used. If more than one version of ICD has been used during the history of a particular data set, then which version and when it was first implemented should be identified.

16. Individual discharge abstract data often makes the distinction between the principal diagnosis and secondary diagnoses. It is important to understand how principal and secondary diagnoses have been allocated, taking into account which factors are used in deciding which condition is allocated as the first diagnosis. The number of secondary diagnoses is also an important element to identify. Among other things, this contributes to identifying the influence of confounding conditions that may adversely affect health outcomes. Secondary diagnoses will be used to identify comorbidities, especially for diseases such as diabetes, which itself is a comorbidity for many of the diseases under consideration. However, it will not be necessary to know all secondary conditions, especially for some cases such as France where it is possible to have up to 50 secondary diagnoses coded.

#### **4. Identification of health interventions**

##### ***4.1 Acute interventions in inpatient setting***

17. Information coded in such large administrative data sets may be of uneven quality. The core set of the information is related to the needs of hospital administration and financing. Therefore most of the main interventions are likely to be well coded, particularly when they play a role in establishing a better DRG grading. This in turn will affect the type of conditions and health information which are likely to be studied using those data sets. Therefore, in all likelihood only the main interventions (e.g. bypass, angioplasty, hip replacement, etc.) can be studied. For other important health interventions such as thrombolytics, stents, and some types of chemotherapy, may not be reported in detail, unless there is a specific payment attached to all items, as is the case in Belgium for pharmaceuticals.

18. In addition, the classifications used for procedures are more heterogeneous than for diagnoses. Many countries rely upon specific national systems. There is a need for further harmonising of definitions. For those countries that do not use ICD-9 CM, it is important to determine if they have in place an algorithm that links their domestic codes to ICD-9 CM codes, otherwise the relevant codes for each of the diseases have to be identified and linked to the appropriate ICD-9 CM codes.

19. Discharge abstract data may include a varying number of procedures, from 3 to 50 depending on the country. However, even if a high number of procedures are included, coding for more minor interventions such as echocardiograms (EKGs) in the case of heart disease, is likely to be unreliable, thus reinforcing the need to concentrate on the main interventions. For cataract surgery, which may be performed in alternate settings, a careful evaluation will likely be required to see whether these interventions could be tracked using inpatient data sets.

20. Another problem in identifying minor interventions is the lack of linkage to outpatient data, where many of these interventions may take place. This makes it all the more likely that inpatient data will be most useful only for acute health interventions of highly defined conditions (e.g. stroke, heart interventions). Apart from procedure codes, other relevant information may be included in the data set. These may include use of specific hospital services such as intensive care unit, day surgery and rehabilitation use, which might prove to be highly relevant for some diseases.



#### ***4.2 Other characteristics of admissions in an inpatient setting***

21. Most hospital discharge data include information on patients' length of stay, for the most part because in the past this played a key role in hospital reimbursement. This may be particularly useful to illustrate different patterns of treatment across countries, with "high tech intensive countries" such as the United States providing more intensive treatments requiring longer stays.

22. Another important source of information is the way of entry (source of admission) included in the data set. This will help to identify those patients who have been admitted from another acute care facility where it is possible they may have received treatments relevant to the disease in question. It is particularly important to assess the impact of transfers, and to see whether patients first admitted to a facility equipped for heavy interventions are likely to receive a different treatment from patients first admitted to second level facilities. Moreover, the way of exit (discharge status) will help identify where patients have been sent after hospitalisation. This information can help in identifying in-hospital deaths, transfers to other hospitals and perhaps even the use of rehabilitation services (for example after a stroke intervention).

#### ***4.3 Characteristics of the hospital where inpatient care was provided***

23. Many data sets include a variable to identify the hospital to which the patient was admitted. One potential benefit of having this information will be the possibility of linking to hospital surveys. The hospital surveys are available in most countries and provide a wealth of relevant information for analysing hospitals' structures and efficiency. This in turn has important implications for health economic analysis, linking treatment patterns to hospital status, thereby providing an opportunity to analyse economic incentives. This could also be used to assess the impact of the planning process.

#### ***4.4. Care provided in an outpatient setting***

24. In countries where inpatient data can be linked with outpatient data, this information would be used to increase the chances of identifying treatments. Information for outpatient beds, for which diagnoses and procedures are likely to be coded, will be more easily accessible than information for outpatient clinics, where it is less likely this information will be coded. This can vary between countries. Further investigation is needed to assess the level of detail of such information and the possibility of its practical use.

#### ***4.5. Specific data on pharmaceutical consumption***

25. Apart from acute health care interventions, drugs play an important role in the treatment of these diseases. This is true for both inpatient and outpatient settings. Therefore, it is necessary to know whether the inpatient database is linked to any pharmaceutical databases, and if so then what type of information on drug consumption is recorded, and what drug identification is used.

## **5. Relevance of the datasets for health services research**

### ***5.1. Relevance for studying disease outcomes***

26. Apart from the peculiarities of each national data set, one crucial question is their usefulness for studying disease outcomes in a health status perspective, the key element being the longitudinal characteristics of these data sets. Thus, in order to conduct a rigorous longitudinal study using these data, it is important to be able to link temporally through the use of unique patient identifiers, which may not be possible for some countries as was discussed earlier.

27. The first health outcome to be considered for vital conditions such as heart attacks, stroke and hip fracture is mortality, including the length of time between the intervention and death. By default, for those countries that lack the proper linkages in their data sets, inpatient mortality may be a factor of interest, provided enough data standardisation could be performed to provide reliable information. In hospital planning and evaluation, in-hospital mortality is often a very sensitive measure: control for case mix and type of admission are key elements in assessing the reliability of in-hospital mortality as a measure for hospital-quality benchmarking. At the international level however, the general mix of patients may not vary as much from country to country, so it might be quite useful to shed some light on such aspects. However, hospital discharge patterns may vary, due to different lengths of stay at the international level, which requires caution in the analysis.

28. For those countries able to link information over time, it is more appropriate to gather information at specific points in time after the initial admission for the acute diagnosis (7 days, 30 days, 90 days, or a year). For diseases that span a longer period of time, such as breast cancer, specific data sets might be needed to assess survival rates. In a couple of countries, the possibility of linkage with death registries associated with administrative data sets and gathered over a significant number of years makes it theoretically possible to track the impact over a span of a couple of years.

29. The presence of unique patient identifiers is crucial in exploring the possibility of linking micro data with other data sets. The first likely candidates for linkage may be census data and data from socio-economic surveys. However, it may also be useful to link the data with information systems for ambulatory care in some countries or death registries in others. Even with countries that have inpatient data with linkable patient identifiers, confidentiality restrictions may restrict the level of detail that may be examined. However, this restriction can be overcome if the initial analysis of the micro data is done within the institutions that house the databases.

### ***5.2. The needs of disease specific studies, general datasets and more targeted sources***

30. In a certain number of countries, these administrative data provide information to assess costs and/or prices of interventions. This is true in most countries using DRG information to finance hospitals. Table 1 is limited in the amount of detail that can be imparted regarding how this cost information is captured within the data sets. The cost information may describe the cost of each type of admission, or at a more desegregated level, the main types of procedures performed. Cost information for determining resource use based on fee-for-service information, while not as reliable as DRG cost information in calculating the total cost of a patient stay, can provide an approximation to the cost of patient stays in hospital. Questions remain as to whether the costing could be extended to the outpatient setting and also include interventions such as the use of pharmaceuticals.

### ***5.3. From general to more disease specific datasets: from general datasources to more specific disease targeted datasets***

31. The range of information available from the general administrative data sources presented here may not be available for all countries. Moreover, their use might be problematic for the study of some diseases due to specific longitudinal requirements or to difficulties in proper identification of the disease. Disease specific studies may be supplemented by registers, surveys and studies. As previously discussed, inpatient data will be most useful for the highly defined acute conditions where the first intervention will likely occur in hospital. For some diseases, such as breast cancer, practice patterns may vary considerably depending on a case by case basis, so inpatient data sets may not capture the management of the disease if surgical intervention is not involved. For chronic conditions such as diabetes that are usually managed on an outpatient basis, inpatient data are even less capable of capturing treatment patterns. For dementia, where patients need to be followed over very long periods of time, there might be a need for long-term longitudinal cohort studies.

32. In general, there are three types of information data sets available for the disease specific data sources:

- registers;
- surveys; and
- ad hoc studies.

33. The data sources have usually been created from a disease specific perspective, quite often within an epidemiological context. While they may be rich sources of information on incidence, prevalence and outcomes, it is unlikely they will reveal much information on expenditures related to treatments. How much these sources of data supplement the inpatient data depends on how each disease is treated. For diabetes and breast cancer, these alternative data sources will be far more important than for the more highly defined conditions, such as AMI and stroke.

34. In order to track mortality and survival, inpatient data will be able to capture individuals that die during their stay in hospital; once these patients are released, disease specific sources are needed to capture death as a consequence of the particular condition, especially if no linkage exists between death registers and inpatient data derived from administrative sources. Information on morbidity should be captured by the disease specific data sets although the type of information will vary depending on the disease. It is difficult to ascertain at this point how easily it will be to capture quality of life information. For some conditions where this is an extremely important issue, such as cataracts, some of this information may be captured. For stroke as well, the facility to which the patient is discharged may provide some of this information.

## CONCLUSION

35. This represents an inventory of hospital administrative databases that has examined their usefulness in comparing treatment costs and outcomes for a disease-based approach to studying health care systems. The aim of this exercise was to provide a summary analysis of the appropriateness of using a micro-data based approach to comparing health care systems by focusing on one of the more commonly available sources of data that would be used in a micro-based approach, hospital discharge databases.

36. The analysis has shown that reliable, well constructed hospital administrative databases are well suited to examining treatments that are delivered in an acute care setting, and at least for in-hospital care, much can be said about easily defined outcome measures such as in-hospital mortality. Beyond the hospital, only jurisdictions able to identify patients through unique identifiers can more properly assess the true impact of hospital-based treatment by extending the analysis to include longer-term survival and other measures of outcome such as readmissions.

37. The use of hospital administrative databases is a necessary first step in comparing health care systems from a disease-specific point of view. This paper has provided the reader with a sense of the limits of these data and how they could be used properly. However, any study that focuses on a specific set of diseases, no matter how well defined or prevalent the condition is in the acute care setting, will need to rely on disease-specific sources of data in order to assess the true effectiveness of health care interventions. The value of this paper could be enhanced with a similar type of analysis on some disease specific sources of data and how they could complement hospital administrative databases.

## REFERENCES

- Antczak-Bouckoms A, Burdick E, Klawansky S, et al. (1991) Using Medical Registries and Data Sets for Technology Assessment, *International Journal of Technology Assessment in Health Care*, 7:2, 123-128.
- Barnes C, Krinsky T (1999) Classifications Systems, Casemix and Data Quality: Implications for International Management and Research Applications, *Casemix Quarterly*, 1, 2, 7-15.
- Blais R (1991) Using Administrative Data Bases for Technology Assessment in Health Care: Results of an International Survey, *International Journal of Technology Assessment in Health Care*, 7:2, 203-208.
- British Columbia Ministry Of Health Linked Data Project (1996, updated 1998) Access Policy for Research Uses of Linked Health Data, Ministry of Health, Centre for Health Services and Policy Research, April.
- Burdick E, Falotico-Taylor J, Young JM (1991) Technology Assessment in the Coronary Artery Surgery Study, *International Journal of Technology Assessment in Health Care*, 7:2, 171-181.
- Chamberlayne R, Green B, Barer ML, et al. (1994) Creating a Population-based Linked Health Database: A New Resource for Health Services Research, *Revue Canadienne de Santé Publique*, 89, 4, 270-273.
- Coffey R M (1999) Casemix Information in the United States: Fifteen Years of Management and Clinical Experience, *Casemix Quarterly*, 1, 1, 7-16.
- Cooper GS, Yuan Z, Stange KC, et al. (1999) The Sensitivity of Medicare Claims Data for Case Ascertainment of Six Common Cancers, *Medical Care*, 37, 5, 436-444.
- Cutler DM (1999) Data Needs for Studying the Health Care System, preliminary, Harvard University and NBER.
- D'Agostino RB, Dillon M, Odell P (1991) Technology Assessment in the Framingham Heart Study, *International Journal of Technology Assessment in Health Care*, 7:2, 156-170.
- Duckett SJ (1998) Casemix Funding for Acute Hospital Inpatient Services in Australia, *Medical Journal of Australia*, 169, S17-S21.
- Duckett SJ (1999) Casemix Funding for Acute Hospital Inpatient Services in Australia, *The Medical Journal of Australia*, 169: S17-S21,  
[www.mja.com.au/public/issues/oct19/casemixduckett/duckett.htm](http://www.mja.com.au/public/issues/oct19/casemixduckett/duckett.htm)
- Fetter RB, Shin Y (1980) Case Mix Definition by Diagnosis Related Groups *Medical Care* – Philadelphia, 18, 2, Supplement.

- Hanson RM (1999) Casemix: Evolution, not Revolution, eMJA, Casemix Supplement.
- Jackson T, Watts J, Lane L (1999) Data Comparability in Patient Level Clinical Costing Systems, Casemix Quarterly, 1, 1, 36-45.
- Kazanjian A (1998) Understanding Women's Health through Data Development and Data Linkage: Implications for Research and Policy, Journal of the Canadian Medical Association, 159, 4, 342-345.
- Klawansky S, Burdick E, Adams M, et al. (1991) Use of the Seer Cancer Registry for Technology Assessment, International Journal of Technology Assessment in Health Care, 7:2, 134-142.
- Klawansky S, Antczak-Bouckoms A, Barr J, et al. (1991) Using Medical Registries and Data Sets for Technology Assessment: An Overview of Seven Case Studies, International Journal of Technology Assessment in Health Care, 7:2, 194-199.
- Lowrance WW (1999) "Scoping Study" for an OECD Project on Data Protection in Transborder Flows of Health Research Data, DSTI/STP/BIO(99)3.
- Malenka DJ, McLerran D, Roos N, et al. (1994) Using Administrative Data to Describe Casemix: A Comparison with the Medical Record, Journal of Clinical Epidemiology, 47, 9, 1027-1032.
- Mcclellan M., (1996), Are the Returns to Technological Change in Health Care Declining? Proceedings of the National Academy of Science U.S.A., 93(23), Nov. 12 1996, 12701-8.
- Mcclellan M., J. Skinner (1999), Medicare Reform: Who Pays and Who Benefits? Health Affairs, Vol. 18, No. 1, Jan.-Feb., 48-62.
- Mountney L (1998) The Increasing Role of Casemix in Understanding the NHS in England, [www.casemix.org/0\\_number/art1.htm](http://www.casemix.org/0_number/art1.htm)
- Newhouse J.P., M. Mcclellan, (1998), Econometrics in outcomes research: the use of instrumental variables. Annual Revue of Public Health, Vol. 19, 17-34.
- The Nova Scotia-Saskatchewan Cardiovascular Epidemiology Group (1989) Estimation of the Incidence of Acute Myocardial Infarction Using Record Linkage: A Feasibility Study in Nova Scotia and Saskatchewan, Canadian Journal of Public Health, 80, 412-417.
- OECD (1980) Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, [www.oecd.org//dsti/sti/it/secur/prod/PRIV-EN.HTM](http://www.oecd.org//dsti/sti/it/secur/prod/PRIV-EN.HTM)
- Orueta JF, Lopez-de-Munain J, Báez K, et al. (1999) Application of the Ambulatory Care Groups in the Primary Care of a European National Health Care System, Medical Care, 37, 3, 238-248.
- Phelan PD, Tate R, Webster F, et al. (1998) DRG Cost Weights—Getting It Right, Medical Journal of Australia, 169, S36-S38.
- Pilla J (1994) Development of AN-DRGs: Meeting the Concerns of Clinicians, The Medical Journal of Australia, 161, September, Supplement.

- Pinfold PS, Goel V, Sawka C (2000) Quality of Hospital Discharge and Physician Data for Type of Breast Cancer Surgery, *Medical Care*, 38, 1, 99-107.
- Roos LL, Sharp SM, Cohen MM (1991) Comparing Clinical Information with Claims Data: Some Similarities and Differences, *Journal of Clinical Epidemiology*, 44, 9, 881-888.
- Roos LL, Nicol JP (1999) A Research Registry: Uses, Development, and Accuracy, *Journal of Clinical Epidemiology*, 52, 1, 39-47.
- The Swedish Heart Surgery Register (1996/97) 23-24, 47-48, 61-64, 75-77, 105-108.
- Warren JL, Feuer E, Potosky AL, et al. (1999) Use of Medicare Hospital and Physician Data to Assess Breast Cancer Incidence, *Medical Care*, 37, 5, 445-456.
- Wiley M., (1992), Hospital Financing Reform and Case-mix Measurement: An International Review. *Health Care Financing Review*, Vol. 13, No. 4, Summer 1992, 119-133.
- Williams JI, Young W (1996) Inventory of Studies on The Accuracy of Canadian Health Administrative Databases, Institute for Clinical Evaluative Sciences, Technical Report.
- Wray NP, Hollingsworth JC, Petersen NJ, et al. (1997) Case-Mix Adjustment Using Administrative Databases: A Paradigm to Guide Future Research, *Soc. Sci. Med.*, 40, 12, 1707-1715.





**ANNEX 1 TABLE 1 General Administrative Data Sources**

<b>Inpatient Data Information</b>									
<b>Country</b>	<b>Principal data set and Time Period</b>	<b>DRG (Case Mix) System</b>	<b>Coverage</b>	<b>Representativeness</b>	<b>Linkage capacity (follow patients over time)</b>	<b>Age</b>	<b>Sex</b>	<b>Place of residence</b>	<b>Other socio economic characteristics, availability of supplemental insurance</b>
AUS -  National	National Hospital Morbidity Database (NHMD)  1989 - 1996/97	AN-DRG (Adapted from US classifications to reflect Australian clinical standards & practice); first released 1992	All patients discharged from both public and private acute-care hospitals	All individuals residing in Australia	Possible Personal ID number is collected but is not available beyond collection authority level (record linkage has been tried in some states - see below)	Yes	Yes	State and Local Area; Rural, Remote and Metropolitan Areas (RRMAs)	Accommodation status (eligible public patient, eligible private patient ineligible patient); insurance status (private insurance), compensable status (entitlement to claim compensation)
AUS -  W. Aus.	Western Australia (Hospital Morbidity Data System) HMDS 1970 -	AN-DRG (version 3.1) introduced 1996/97	All inpatient episodes for defined admitted patients to public, private and freestanding day hospitals in Western Australia.	All individuals residing in Western Australia	Yes The database has been linked to the Medicare Claims Database	Yes	Yes	Postcode	Indigenous status, marital status, employment status, insurance status (private insurance), payment classification
BEL	Résumé Clinique Minimum & Résumé Financier Minimum (RCM-RFM)  1988 - 1997	All Patients Diagnosis Related Group (APDRG - version 10)	Includes all patients discharged from 15 - 75 hospitals per year (teaching and general, public and private)	All hospital stays	Yes, but can only follow patient for multiple stays with same hospital. IHE - mortality data;	Yes	Yes	Postal Code	Mutuality code, private insurance status
CAN -  Ontario & National	Discharge Abstract Database; CIHI & Statistics Canada;  1979/80 - 1998/99	Case Mix Group (CMG) CIHI methodology and P1x - a refinement to CMG which takes into account variation in LOS and comorbidities	All acute-care hospitalisations in Ontario & Canada	All individuals residing in Ontario & Canada	National Possible (Ind. prov. health card no. is collected; use of this info. would require consent from each province) Ont. - ICES receives patient health card no with the DAD. Has been linked with OHIP phys. claims d-base Ont. Myocar. Inf. D-base; Ont. Reg. Pers. D-base - mortality data.	Yes	Yes	Postal code	Possible (can be linked with census data using postal code)

<b>Inpatient Data Information</b>									
<b>Country</b>	<b>Principal data set and Time Period</b>	<b>DRG (Case Mix) System</b>	<b>Coverage</b>	<b>Representativeness</b>	<b>Linkage capacity (follow patients over time)</b>	<b>Age</b>	<b>Sex</b>	<b>Place of residence</b>	<b>Other socio economic characteristics, availability of supplemental insurance</b>
CAN - Quebec	MED-ECHO 1985 - 1995	No DRG grouping	All acute-care hospitalisations for the province	All individuals residing in Quebec	Yes. Can be linked to RAMQ phys. claims database; also with RAMQ vital statistics data	Yes	Yes	Postal code	Possible (can be linked with census data using postal code)
CAN - Manitoba	Manitoba Health Research Data Repository - (MCHPE) 1985/86 - 1998/99	No DRG grouping	All hospital inpatient discharges in Manitoba (includes outpatient services)	All individuals residing in Manitoba	Yes Inpatient data linked with: Office of Vital Statistics; Manitoba Cancer Treatment and Research Foundation	Yes	Yes	Postal code, municipality and other aggregations, e.g. hospital area	Can provide ecological measures of socio-economic status
DEN	National Patient Register 1978 – 1998	NordDRG (Nordic countries version of DRG) 1987 - 1998	Includes all inpatients discharged from Danish public hospitals.	All individuals residing in Denmark (and persons visiting Denmark)	Yes, based on patient identifier it is possible to link with registers on: operations and cancer Can be linked with Cause of Death Register (COD)	Yes	Yes	County, municipality	Yes, by linking to other population registers. Not possible to view whether patient has supplementary insurance.

**TABLE 1 General Administrative Data Sources**

<u>Inpatient Data Information</u>											
Country	Principal data set	Diagnosis code	Diagnoses	Procedure code	Procedure	Length of stay	Costs/Fees	Way of entry	Way of exit (discharge)	Hospital identification	Other information
AUS - National	NHMD	Australian version of ICD 9 CM	Primary diagnosis + 31 additional diagnoses	Australian version of ICD 9 CM	Principal procedure + 31 possible additional procedures	Yes	Estimated average cost for each AN-DRG	Yes (transfer)	Yes (transfer to another hospital or nursing hom, death)	Yes	Overnight patients, type of care (rehab, palliative, non-acute, other); waiting list status
AUS - W. Aus.	HMDS	ICD-9; ICD-10 AM (as of July 1, 1998)	Primary diagnosis + co-diagnosis + unlimited number of additional diagnoses	ICD-9-CM? ICD-10-AM (as of July 1, 1998)	Principal procedure + all "significant" procedures	Yes	Estimated average cost for each AN-DRG	Yes (other institution, private practice, etc.)	Yes (transfer to another hospital or nursing hom, death)	Yes	Days in ICU, clinician performing principal procedure
BEL	RCM-RFM	ICD 9	Primary diagnosis + XX additional diagnoses	ICD 9 CM	Primary procedure+unlimited number of additional procedures	Yes	Fee billing information (divided into 70 categories)	Yes	Yes	Yes	Number of days in ICU, type of department
CAN - Ontario & National	Discharge Abstract Database: CIHI & Statistics Canada	ICD 9	Primary diagnosis + 15 possible additional diagnoses	Canadian Classification of Procedures (CCP) and ICD 9 CM	Principal procedure + 9 possible additional procedures	Yes	No (possible to estimate using information collected on resource use)	Yes (Other institution, nursing home, death, home care, etc.)	Yes (Other institution, nursing home, death, home care, etc.)	Yes	Spec. of physician performing procedure; date of main procedure; type of admission (emergency, urgent, elective)
CAN - Quebec	MED-ECHO	ICD 9	Primary diagnosis + XX possible additional diagnoses ??	ICD 9 CM	Principal procedure + XX possible secondary procedures ??	Yes	Cost estimates are possible	Yes	Yes	Yes ??	??
CAN - Manitoba	Manitoba Health Research Data Repository	ICD 9 CM	Most responsible diagnosis + 15 possible additional diagnoses	ICD 9 CM	Up to 12 procedures	Yes	Yes	Yes	Yes	Yes	
DEN	National Patient Register	ICD 8 up to 1993; ICD 10 since 1994	Primary diagnosis + an indefinite number of secondary diagnoses	SKS - code; NOMESCO	Primary procedure+ an indefinite number of secondary procedures (only surgical interventions are coded)	Yes	No	Yes	Yes	Yes (hospital and department where patient was treated)	No

**TABLE 1 General Administrative Data Sources**

Country	<u>Inpatient Data Information</u>			<u>Outpatient and Ambulatory Care Information</u>		<u>Pharmaceutical Data Information</u>	
	In-hospital mortality	Rehospitalisations or other outcomes information	Possibility of linkage with other data sets:	Type of information and services captured	Costs/fees	Type of info., services captured, drug consumption and codes	Costs
AUS - National	National Hospital Morbidity Database			Medicare Claims Database		Drug Utilization Sub-Committee Database	
	Yes	No rehospitalisation; Personal identification number is unique for each hospital stay and does not follow patient after discharge		Private medical services provided out-of-hospital + medical services provided in public and private hospitals for private patients (no distinguishing between services provided in-hospital and those out of hospital)	Fee charged	Reason for dispensation not available; does not include prescriptions in public hospitals ; defined daily dosage; ATC code	Cost can be estimated but only at an aggregate level, ie by drug subsidised by the PBS/RPBS
				National Hospital Morbidity Database		National Hospital Morbidity Database	
				Outpatient services (incl. radiology, medical and diagnostic)	No data	No data	No data
AUS - W. Aus.	Hospital Morbidity Data System			Medicare Claims Database		Drug Utilization Sub-Committee Database	
	Yes	?? (client identifier is used for all admissions to a particular hospital); should be feasible, linked to Medicare Claims dbase		See above;	See above	See above	See above
				Hospital Morbidity Data System		Hospital Morbidity Data System	
				Same day patients are included	No data	No data	No data
BEL	Résumé Clinique Minimum et Résumé Financier Minimum			CIES-INAMI		Résumé Clinique Minimum et Résumé Financier Minimum	
	Yes	Rehospitalisation is available if patient is rehospitalised in same hospital.		All ambulatory care services covered by mutuelles. Information is available within each mutualité. A national level single source of information will not be available until 2000 (see previous).	Yes (billed charges)	Anatomical Therapeutic Chemical code (ATC)	Total costs by ATC are available. Includes amounts billed fee-for-service.

	<u>Inpatient Data Information</u>			<u>Outpatient and Ambulatory Care Information</u>		<u>Pharmaceutical Data Information</u>	
Country	In-hospital mortality	Rehospitalisations or other outcomes information	Possibility of linkage with other data sets:	Type of information and services captured	Costs/fees	Type of info., services captured, drug consumption and codes	Costs
CAN - National	Discharge Abstract Database			National Physician Data Base		Discharge Abstract Database	
	Yes	Rehospitalisations		All fee-for-service claims to physicians - includes inpatient, outpatient and ambulatory care services (patient demographic information but no patient identifier, no diagnoses)	Physicians' billings for services rendered	No data	No data
CAN - Ontario	Discharge Abstract Database - ICES-OHIP data			Ontario Health Insurance Plan Claims for physicians' services data		Ontario Drug Benefit Program Claims Database	
	Yes	Rehospitalisations		Ontario - link via patient identifier available	Physicians' billings for services rendered	Claims paid for drugs listed on formulary for Ont. residents covered over age 65	Claims paid for drugs listed on formulary
CAN - Quebec	MED-ECHO			Régie de l'assurance maladie du Québec (RAMQ) Medicare database		Régime d'assurance-médicaments data??	
	Yes	Rehospitalisations		All diagnostic or therapeutic procedures performed in Quebec where physicians were reimbursed by the RAMQ	Physicians' billings for services rendered	linkage with special provincial drug scheme claim data	??
CAN - Manitoba	Manitoba Health Research Data Repository						
	Yes	Yes		All fee-for-service claims paid to physicians	Physicians' billings for services rendered	Physicians' reimbursements for services rendered ?	
DEN	National Patient Register						
	Yes	Yes, however it is not possible to imply a relationship with the outcome of a previous admission		No information	Primary procedure + an indefinite number of secondary procedures (no diagnosis)	No	No

**ANNEX 1 TABLE 1 General Administrative Data Sources**

<b>Inpatient Data Information</b>									
<b>Country</b>	<b>Principal data set and Time Period</b>	<b>DRG System</b>	<b>Coverage</b>	<b>Representativeness</b>	<b>Longitudinal</b>	<b>Age</b>	<b>Sex</b>	<b>Place of residence</b>	<b>Other socio economic characteristics, availability of supplemental insurance</b>
FIN	Hospital Discharge Register 1987 - 1997	NordDRG (Nordic countries version of DRG) - can be used from 1997 onwards	Includes all patients discharged from public hospitals in Sweden, private facilities not included	All individuals residing in Sweden (and persons visiting Sweden)	Yes	Yes	Yes	County, municipality, parish	Yes, by linking to other population registers
FRA	Données PMSI 1994 - 1997	Groupe Homogène de Malades (Adapted from US classification using weights reflecting French clinical standards)	Public database (includes most public hospitals); cost data base (includes 40 public hospitals); private database	Public and private hospitals (all individuals residing in France); Cost data base (around 5% of admissions)	No	Yes	Yes	In the national database, regional code (département)	No
GBR	Hospital Episode Statistics 19?? - 1998	Healthcare Resource Groups	Population admitted to public hospitals	Public hospitals	No	Yes	Yes	Yes	No
	Oxford Record Linkage Study 1968 - 1986	No DRG system	Population in the Oxford area		Yes	Yes	Yes	Yes	To be investigated, linked with Census
ITA	Schede di Dimissione Ospedaliera (SDO) 1993 - ?; MOD ISTAT prior to 1993	Raggruppamenti Omogenei di Diagnosi (based on U.S. DRG version 10)	Includes all patients discharged from Italian hospitals	All individuals residing in Italy		Yes	Yes	Region	Marital status
SWE	Hospital Discharge Register 1987 - 1997	NordDRG (Nordic countries version of DRG) - can be used from 1997 onwards	Includes all patients discharged from public hospitals in Sweden, private facilities not included	All individuals residing in Sweden (and persons visiting Sweden)	Yes	Yes	Yes	County, municipality, parish	Yes, by linking to other population registers
USA	5 or 100 % of Medicare claims data (Research Identifiable File; Beneficiary Encrypted File) 1984 - 1996	US DRG grouping	All acute-care hospitalisations (Part A)	99 % of population aged 65 and over; includes population under 65 eligible for program under disability or End Stage Renal Disease	Yes	Yes	Yes	Zip Code (RIF); County (BEF), can be aggregated to MSA (Metropolitan Statistical Area)	Other SES implicit with MSA; supplemental insurance

**TABLE 1 General Administrative Data Sources**

<b>Inpatient Data Information</b>											
Country	Principal data set	Diagnosis code	Diagnoses	Procedure code	Procedure	Length of stay	Costs/Fees	Way of entry	Way of exit	Hospital identification	Other information
FIN	Finnish hospital Discharge Register	ICD 9	Primary diagnosis + XX secondary diagnoses	NOMESCO	Primary procedure+ XX possible secondary procedures	Yes	Yes, prices and fees for procedures bought by municipalities	Yes	Yes	Yes	No
FRA	Données PMSI	ICD 9	Primary diagnosis + 50 possible secondary diagnoses	ICD 9 CM	Primary procedure + 50 possible secondary procedures	Yes	Yes, for the cost base, disaggregated	Yes	Yes	Yes, with SAE hospital survey	ICU use
GBR	Hospital Episode Statistics	ICD9	Yes	Yes	To be investigated	Yes	No	Yes	Yes		
	Oxford Record Linkage Study	ICD9	Yes	Yes	To be investigated	Yes	No	Yes	Yes	Yes	
ITA	Schede di Dimissione Ospedaliera (SDO)	ICD 9	Primary diagnosis + 3 possible secondary diagnoses	ICD 9 CM	Primary surgical procedure + 3 possible secondary surgical procedures	Yes	Yes (Total cost of stay)	Yes	Yes	Yes (hospital code and U.S.L. code - local health unit)	
SWE	Hospital Discharge Register	ICD-9 (1987 - 1996; ICD-10 (1997 - )	Primary diagnosis + 5 secondary diagnoses; Primary diagnosis + 7 secondary diagnoses (depends on how dagger-asterisk codes are used)	Surgery codes only, Swedish classification; Surgery codes only, NOMESCO (Nordic classification)	8 procedure codes, none specified as primary; 12 procedure codes, none specified as primary	Yes	No	Yes (home, another department)	Yes (home, another department)	Yes	Type of department
USA	5 or 100 % of Medicare claims data (Research Identifiable File; Beneficiary Encrypted File)	ICD 9	Primary diagnosis + 9 possible secondary diagnoses	ICD 9 CM	Reliable claims reports for all "major" procedures (10 possible codes)	Yes	Yes, with extensive possible disaggregation	Yes (home, other hospital)	Yes (home, nursing home, death, other hospital)	Yes, link with American Hospital Association Survey	ICD & CCU use; information following acute admissions

**TABLE 1 General Administrative Data Sources**

Country	Principal data set	Outpatient Information			Pharmaceutical Data Information		Health Outcomes Information		Possibility of Linkage With:	
		Procedure information	Type of information	Costs/Fees	Drug consumption and codes	Costs	Mortality	Rehospitalisations or other outcomes information	Other data sets	Other outcome info.
FIN	Finnish hospital Discharge Register	No information	No	No	No	No	Mortality (1 and 7 day mortality), in hospital mortality over time	Yes	No	
FRA	Données PMSI	No, administrative coding in claims data with no linkage possible	No	No	No	No	Inpatient mortality; no link with death registry	No	No, except local specific data	No
GBR	Hospital Episode Statistics	No information			No		No	No	No	No
	Oxford Record Linkage Study						Yes, linkage	Yes	Yes	
ITA	Schede di Dimissione Ospedaliera (SDO)	Includes information on whether or not a patient was admitted to a "day-hospital" and the number of days in a day-hospital	No	No	No	No	In-hospital mortality (includes information on results of autopsy)		No other linkage possible	No
SWE	Hospital Discharge Register	Project has been initiated to develop registers for out-patient care.	No	No	No	No	Can be linked with Cause of Death Register (COD)	Yes	Unique personal identifier allows linkage to other registers (MI, Cancer, Stroke, Hip, Fractures, Diabetes)	
USA	5 or 100 % of Medicare claims data (Research Identifiable File; Beneficiary Encrypted File)	Ambulatory care (Part B); similar to inpatient data set + more reliable coding of minor treatments	Care delivered covered by Medicare, including physician and laboratory services	Total cost (or public reimbursement)	Generally no linkage possible. Specific drugs dispensed are rarely coded; linked data on inpatient prescription drug use for 1994 - 95 only		Mortality (7 days, 1 month, 3 months, 1 year, possibility of link with death registry)	Follow up of readmissions (plus comorbidities)	Pharmaceutical	



