Data linkage to estimate resource and service utilisation for palliative care clients

A. McAlindon

Vectra Corporation, Mile End, SA 5031

A. Abernethy

Department of Palliative and Supportive Services, Flinders University of South Australia, Bedford Park, SA 5042.

Divisions of Hematology and Medical Oncology, Duke University Medical Center, Durham, North Carolina, 27710, USA

Abstract

Background

In Australia, palliative care is increasingly being offered to clients with terminal cancer and other end-stage illnesses. Up-to-date and accurate costs associated with the provision of these services are required for the successful negotiation of resources from state and Commonwealth governments and for local and system-wide service planning.

Setting

Southern Adelaide Palliative Services (SAPS) is a comprehensive palliative care programme serving the southern region of Adelaide, South Australia, including specialist nursing and medical support, social work, inpatient acute hospital and hospital care, outpatient visits, home care, nursing home consultations, a bereavement program, volunteers, complementary care, and day care. SAPS interfaces with public and private hospitals, district nursing and domiciliary care.

Objective

To determine the optimal record linkage methods for linking together a range of databases in order to estimate the resource and service utilisation for palliative care clients in a regional palliative care service.

Methods

Access to a range of databases from local service providers and state and Commonwealth governments was negotiated for the 931 SAPS clients who died in 1999. These databases contained service utilisation and claims data for public and private hospitals (inpatient and non-patient), Medicare, Pharmaceutical Benefits Scheme, Royal District Nursing Service (RDNS), Southern Domiciliary Care (SDC) and aged care services. A Master Client Index (MCI) for the SAPS clients was constructed from these databases and from the South Australian Enterprise-wide Patient Master Index (EMPI) using a combination of deterministic and probabilistic record linkage methods. Using an iterative process, the MCI was progressively refined and used to extract the final utilisation data.

Results

The creation of a MCI from a number of data sources and subsequent iterative updating of demographic elements maximised the number of potential linkages. Supplementary client lists such as EMPI and the South Australian Cancer Registry were also rich data sources that identified 72% of the client Unit Record Numbers (URNs) detected. The use of an iterative linkage process detected an additional 10% of all client URNs.

Conclusion

A combination of deterministic and probabilistic record linkage methods are required to accurately estimate the resource and service utilisation for palliative care clients. The construction of a MCI allows patient identifiers from a range of databases to be spliced together to further increase the accuracy of the linkage process.

Introduction

Palliative care is increasingly being offered to clients with terminal cancer and other end-stage illnesses leading to a growing demand for palliative care services within the Australian healthcare system. Up-to-date and accurate costs associated with the provision of these services are required for the successful negotiation of resources from the state and Commonwealth governments and for local and system-wide service planning.

Palliative care clients are not uniquely identified within the available local, state and Commonwealth healthcare databases. In order to develop a picture of the costs of a palliative care service, a list of clients who use the particular service must be married with the de-identified health resource utilisation data. Record linkage provides a method for associating each client with all of their episodes of care recorded in the databases and extracting the costs. The data elements used for the linkage processes are often affected by data quality and data capture issues, thereby decreasing the reliability of the cost estimate.

Previous studies have indicated that these issues can be partly overcome. Combining data from a number of sources has

been demonstrated to increase the accuracy of the linkage process.[1,2,3] In addition, further improvements can be made using an iterative process to improve the accuracy of individual data elements.[1]

Based on this evidence we set out to build a Master Client Index (MCI) for a palliative care cohort that contains the combined patient demographics from a number of databases, is iteratively updated using a combination of probabilistic and deterministic linkage techniques, and will lead to the reliable extraction of the final service utilisation and cost data for the cohort.

Methods

Objective

To determine the optimal record linkage methods for linking together a range of databases in order to estimate the resource and service utilisation of palliative care clients in a regional palliative care service.

Setting

Southern Adelaide Palliative Services (SAPS) is a comprehensive palliative care programme serving the southern region of Adelaide, South Australia. Services include specialist nursing and medical support, social work, inpatient acute hospice and hospital care, outpatient visits, home care, nursing home consultations, a bereavement program, volunteers, complementary care, and day care. SAPS interfaces with private and public hospitals, district nursing, and domicilary care. Caring for a regional population of 350,000 and covering approximately one fourth of the population of the state of South Australia, SAPS has over 1000 new referrals, 15% of which are non-cancer patients.

Linkage fields

A list of all 931 SAPS clients who died in 1999 was extracted from the electronic SAPS database. Fields included Medicare Card Number, Veteran Card Number, gender, date of birth, date of death, cancer flag, hospice flag, hospital code and hospital patient unit record number (URN) (Table 1).

Probabilistic	Deterministic
Given Names (partial weights for nick name or SOUNDEX agreements)	Health Unit Code
Family Name (partial weights for agreements on NYSIIS code)	Patient Unit Record Number
Gender	
Date of Birth (partial weights for partial agreements)	
Date of Death (+/- one day)	
Address	
Postcode	
Medicare Card Number (where available)	
Veteran Card Number (where available; partial weights for incorrect prefix)	

Table 1 Linkage Fields by Linkage Method

The data quality and completeness for most fields was very high, with the exception of the Medicare Card Numbers and Veteran Card Numbers.

The Medicare Card Number is a ten-digit number used by the Health Insurance Commission for the processing of doctor and specialist payments within Australia. It is useful for isolating linkages to a given family unit but must be combined with other demographic identifiers to identify an individual within a family. Further, a number of individuals may have more than one number due to changed family circumstances (e.g. marriage, divorce, leaving home).

While the use of the Medicare Card Number is not strictly approved for linkage purposes, it is widely recorded within the hospital setting to verify eligibility for Medicare services. Although the proportion of records with a Medicare Card Number recorded varies considerable from hospital to hospital, the data quality is quite good due to the use of the 10th digit as a check digit. While SAPS does record the number, most community-based services do not record the number and as a result it cannot be used for the linkage of community-based datasets.

The Veteran Card Number is a nine-character identifier provided to Australian War Veterans for the provision of treatment services. While this identifier is unique, it is only available for a small proportion of all hospital patients and there is often data quality problems associated with the entry of the alphanumeric prefix. In this study a partial weight was assigned where the Card Number agreed but the prefix was partly incorrect.

Linkage method

While the linkage processes and service utilisation extracts were undertaken for all of the specified datasets, only the steps for the linkage of hospital admissions data are presented in this paper.

Access to a range of databases from local service providers and state and Commonwealth governments was negotiated. These databases contained service utilisation and claims data from the South Australian Department of Human Services, local public and private hospitals, Medicare Benefits Scheme, Pharmaceutical Benefits Scheme, Royal District Nursing Service (RDNS), Southern Domiciliary Care (SDC) and aged care services.

A combination of deterministic and probabilistic record linkage techniques was used to link the SAPS clients with the de-identified client data with claims and service utilisation data from the negotiated databases. Supplementary client lists from the metropolitan-wide Enterprise Master Patient Index (EMPI) and from the South Australian Cancer Registry were used to increase the accuracy of the linkage process.

We first constructed a Master Client Index (MCI) containing client Unit Record Numbers (URNs) for SAPS clients. The MCI contained the demographic elements and hospital URNs for every hospital and health unit that SAPS clients had attended in the previous five years. The index was initially populated using probabilistic linkage techniques and was iteratively updated using a combination of deterministic and probabilistic linkage techniques (Figure 1). When no further improve-

ments could be made to the MCI a deterministic linkage using client URNs was made to extract the final service utilisation data (Figure 2).

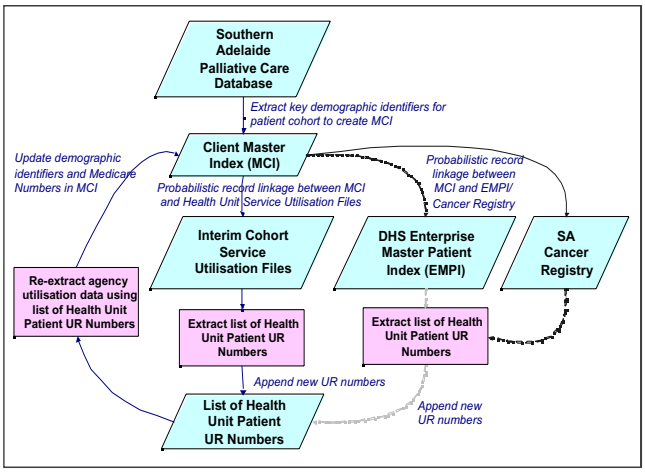

Figure 1 Method Overview – Building and Iteratively Updating the MCI

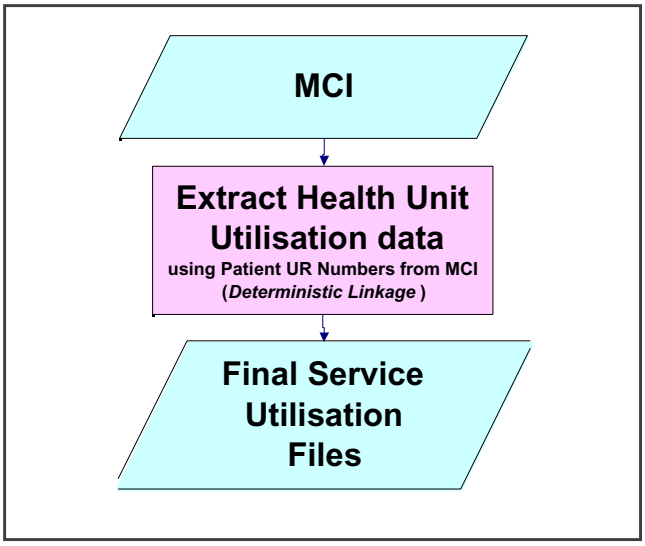

Figure 2 Final Extraction Process

The creation of the MCI allowed the completeness of the demographic elements to be progressively improved by combining these fields from a number of data sources and from the same data source over time. The accuracy of these elements is vital to the accuracy of the probabilistic linkage process.

The following example indicates the potential benefits of using an iterative process to maximise the number of record linkages (Figure 3). A dummy SAPS client was created for this purpose.

Figure 3 – Example of Iterative Process

Step 1 – Identify SAPS client

Client ID/UR Number	Date of Birth	Gender	Medicare Number
SAPS Client Number 600203	3/10/63	M	**5812356789**

Step 2 – Extract all hospital records with the same Medicare number, date of birth and gender:

Hospital URN= **9845**	3/10/63	M	**5812356789**

Result – 1 linkage.

Step 3 – Extract all hospital records with the same URN (9845):

Hospital URN= 9845	3/10/63	M	***9876543210***
Hospital URN= 9845	3/10/63	M	5812356789

Result – 2 additional linkages.

Step 4 – Extract all hospital records with the new Medicare Card Number (9876543210) identified in the previous step:

Hospital URN= 8874	3/10/63	M	***9876543210***

Result – 1 more additional linkage; total of 4 linkages.

Results

The linkage results were measured by counting the number of unique patient URNs present in the MCI after each major linkage step (Table 2).

Linkage Step	Number of Unique Hospital URNs (cumulative)	Number of Public Hospitals Attended (cumulative)	Number of Private Hospitals Attended (cumulative)
Initial MCI (after initial population with SAPS cohort data)	582	6	2
Append known URNs for SAPS clients from a major Adelaide private hospital to MCI (from manual search by hospital).	681	6	2
Link SAPS clients to list of URNs contained in the South Australian EMPI (using probabilistic record linkage on patient demographics for 4 major public hospitals) & append linked URNs to MCI.	1,999	6	2
Link SAPS to URNs in the South Australian Cancer Registry (using probabilistic record linkage on patient demographics) & append linked URNs to MCI.	2,892	25	24
Iterative processing of patient identifiers in MCI to produce the final list.	3,221	39	35

Table 2 Linkage Metrics

The results demonstrated that the supplementary patient lists such as EMPI and the Cancer Registry were rich data sources that identified 72% of the URNs detected. Furthermore, the use of an iterative linkage process detected an additional 10% of all URNs. If the supplementary patient lists had not been used then the iterative approach would have been even more effective and necessary.

Conclusions

A combination of deterministic and probabilistic record linkage methods are required to accurately estimate the resource and service utilisation for palliative care clients. The construction of a MCI from a number of data sources and subsequent iterative refinements is pivotal to this process and considerably increases the number of linked records. The use of multiple data sources also increases the power of the linkage process, particularly where the patient demographic identifiers are incomplete or missing.

In the upcoming Palliative Care Trial commencing in April 2002 we plan to expand our data sources to include additional datasets for private allied health services, ambulance use and patient "out of pocket" expenses. Further, we will maximise the power of the linkage process by including supplementary demographic information from private health insurers and the Enterprise Mater Patient Index from four new hospitals.

References

Clark DE and Hahn DR. Hospital trauma registries linked with population-based data. J Trauma-Injury Infection & Critical Care, 1999;47(3):448–54.

McClish DK, Penberthy L, Whittemore, et. al. Ability of Medicare Claims Data and Cancer Registries to Identify Cancer Cases and Treatment. Am J Epidemiology, 1997;145:227–33.

Russell J and Conroy C. Representativeness of Death Identified through the Injury-at-Work Item on the Death Certificate: Implications for Surveillance. Am J Public Health, 1991;81:1613–1618.