Methods and preliminary results for a data linkage project to determine long-term survival after intensive care unit admission

Daniel V Mullany, Carole L Foot, Susan E Smith, Janelle Johnson and Adrian G Barnett

In patients requiring intensive care, the determinants of mortality in the short term (in the ICU and in hospital) are well described by validated scoring systems, including the Acute Physiology and Chronic Health Evaluation (APACHE) score, Mortality Prediction Model, Simplified Acute Physiology Score, Multiple Organ Dysfunction Score, Sequential Organ Failure Assessment and Logistic Organ Dysfunction Score. The probability of in-hospital death is largely determined by the severity of physiological derangement. It is also increasingly affected by discharge practices arising from, for example, the growing capacity to provide mechanical ventilation outside the ICU in skilled nursing facilities, particularly for the elderly. However, events during the acute illness that necessitated ICU admission may also determine prognosis and survival for up to 2 years after the admission.

Although age is consistently identified as an important predictor of long-term survival, there is significant variation in the literature on other factors that determine long-term outcomes in particular patient groups. To identify predictors of short-term and long-term survival, we require detailed information on patients during their ICU stay and linked information on mortality. While respecting patient privacy, it is vital to make maximum use of the available data in a cost-effective manner to assess the adequacy of medical therapy. This is so for patients who require ICU care and particularly those who require prolonged ICU care, with associated substantial costs and uncertain prognosis.

Western Australia has a well-established database linkage system for determining long-term outcomes within the Data Linkage Unit of the Western Australian Department of Health. However, other Australian states do not as yet have this capacity.

Our study aimed to describe methods of determining long-term survival of ICU patients in the absence of an established linkage program and to present preliminary findings.

Methods

Patients and setting
Prince Charles Hospital is a tertiary-referral, university-affiliated public hospital in Brisbane, Queensland, with

ABSTRACT

Aims: To describe a local data linkage project to match hospital data with the Australian Institute of Health and Welfare (AIHW) National Death Index (NDI) to assess long-term outcomes of intensive care unit patients.

Methods: Data were obtained from hospital intensive care and cardiac surgery databases on all patients aged 18 years and over admitted to either of two intensive care units at a tertiary-referral hospital between 1 January 1994 and 31 December 2005. Date of death was obtained from the AIHW NDI by probabilistic software matching, in addition to manual checking through hospital databases and other sources. Survival was calculated from time of ICU admission, with a censoring date of 14 February 2007. Data for patients with multiple hospital admissions requiring intensive care were analysed only from the first admission. Summary and descriptive statistics were used for preliminary data analysis. Kaplan–Meier survival analysis was used to analyse factors determining long-term survival.

Results: During the study period, 21 415 unique patients had 22 552 hospital admissions that included an ICU admission; 19 058 surgical procedures were performed with a total of 20 092 ICU admissions. There were 4936 deaths. Median follow-up was 6.2 years, totalling 134 203 patient years. The casemix was predominantly cardiac surgery (80%), followed by cardiac medical (6%), and other medical (4%). The unadjusted survival at 1, 5 and 10 years was 97%, 84% and 70%, respectively. The 1-year survival ranged from 97% for cardiac surgery to 36% for cardiac arrest. An APACHE II score was available for 16 877 patients. In those discharged alive from hospital, the 1, 5 and 10-year survival varied with discharge location.

Conclusions: ICU-based linkage projects are feasible to determine long-term outcomes of ICU patients.
ICU with six to eight beds and 300–500 admissions per year. Data were obtained for the study on all patients aged 18 years or older who were admitted to either of the ICUs between 1 January 1994 and 31 December 2005.

Datasets were obtained from the Cardiac Surgical Data Management Unit for all cardiac surgical procedures and the ICU Research and Data Management Unit for all general ICU admissions for this period. Data were extracted from two cardiac surgical databases: one containing limited data fields for 1994–2001 inclusive, and the other with comprehensive data on all cardiac surgical procedures) for 2002–2005 inclusive. Comprehensive ICU data were available for non-cardiac surgical admissions for the entire study period.

The cardiac surgical and ICU datasets were merged into a single dataset with 24,938 records. Each patient was allocated a unique patient identifier, and each surgical procedure was allocated a unique procedure identifier. Some patients were admitted to the ICU but did not have a surgical procedure, and some patients had several ICU admissions during their hospital stay. Some patients had several surgical procedures during their hospital stay, with some procedures being performed on the same day.

A dataset was also obtained from the Hospital Based Corporate Information System (HBCIS), including date of death (DOD) and date of last contact, which is compulsory information for submission to the National Death Index (NDI) of the Australian Institute of Health and Welfare (AIHW). All data were cross-checked to account for all admissions. They were thoroughly cleaned and prepared in accordance with NDI guidelines, and a file was generated including all known names, with date of birth (DOB), sex, DOD, date of last contact and State of residence at last contact.

Data linkage and cleaning
We submitted 21,366 records to the NDI for linkage. The NDI uses a proprietary algorithm to give a probability of matching based on submitted data. Some patients were submitted more than once because of uncertainty about the correct name or its spelling. Each proposed matched pair was examined according to its weight (defined as the probability of a match), and in accordance with NDI guidelines. In this process, the probability of the match was assessed using other sources, including the HBCIS, online White Pages (http://www.whitepages.com.au) and online electoral records from the Australian Electoral Commission (AEC; http://www.aec.gov.au).

For patients who had matches for all data except surname, the surnames of next of kin were obtained from the HBCIS. If the surname of the patient’s partner matched the NDI record, then the patient was also regarded as a match. The HBCIS was also used in cases where the DOD did not match with the NDI DOD but was out by only 1 day. In all these cases, the patient had died in hospital with a DOD recorded just after midnight. These were all regarded as a match.

For some patients, the DOD was not a match but was within a few months of the NDI DOD. After consultation with the hospital medical records department, we were advised that, until recently, if actual DOD was not known or given, the recorded DOD on the HBCIS was the date of notification of death. In these instances only patients with a high probability and a match in all other areas were regarded as a match. For patients with a high probability and a match in all other areas but a State of death recorded as New South Wales, we checked the patient’s last documented address, as well as the next of kin, on the HBCIS. In most instances, these were regarded as a match as the patient’s address was listed as NSW. Patients whose DOD was within months of the last date of contact and who did not have a DOD to match were checked on HBCIS to determine whether they had been transferred to another facility. Their inpatient ICD coding was also checked to help determine the likelihood of a match.

In cases where there was a query about the likelihood of a match, the online White Pages was used to check whether the patient was still listed at our last recorded address. If patients had a current White Pages listing, then these were considered unlikely to be a match, especially if their recorded DOD was more than 12 months previously. A double check was also undertaken with the AEC website. If patients had a current White Pages listing, then these address details were checked against the AEC to confirm the patients were alive.

On completion of the clerical review, the files were returned to the NDI for deletion of all rejects. NDI returned a final file with all positive matches. There was some duplication of matches, as a person can have two or more records in the NDI because of ambiguous or doubtful details. In addition, two people might have met the same criteria. Duplicates were reviewed, leading to deletion of three matches from our final review. The details of these matches were linked with our cohort of patients.

All patients with name, DOB and DOD matching were regarded as a match regardless of probability. If DOD was not available to match with the NDI, then all were investigated for probability of a match.

Patients were assumed to be alive on 14 February 2007 if there was no record of death in the NDI.

Security, privacy and ethics review
We emailed data to the AIHW in the form of a zipped password-protected Excel file. The password was emailed separately. A password-protected NDI Data Linkage Report was returned to us, and a clerical review was conducted using the two-line text file. Security at the Cardiac Surgical
and Intensive Care Unit Clinical Information Service is ensured by password access to the respective databases, which are supported on a secure hospital network with safe storage of computers and files. Confidentiality is facilitated by the use of unique identifier with no use of medical record numbers or names for the statistical analysis.

The AIHW has strict policies on access to the NDI, including a requirement for all search applications to be approved by its own ethics committee. This committee has strict guidelines on all aspects of accessing information contained in the death register.

The project involved no patient interventions or changes to clinical management. The new information obtained was on patients who had died and comprised only date and cause of death. No new information was obtained on living patients. Because the research involved evaluation of an existing current health service with the intention of monitoring this service for improvement, it was considered a quality assurance activity under Part 7, section 62H of the Queensland Health Services Act (Part 7 Confidentiality Guidelines para 6.7), which allow this use of patients’ clinical information without specific consent.

Statistical analysis

Preliminary data analysis comprised summary and descriptive statistics. Missing variables were not imputed, but levels of missing data were reported. Survival was determined from the time of the first ICU admission. In specific analyses, the denominator was the number of ICU or hospital admissions.

Survival analysis was used to describe and determine factors that affected time to death. Survival was calculated from the first ICU admission in any hospitalisation. The Kaplan–Meier product limit method (a graphical analysis) was used to describe survival times in pre-specified patient subgroups. The log-rank test was used to examine differences between the groups by survival times. This test assumes that subjects are censored independently of their risk of death, and that average survival probabilities remained the same over the study period.

Data accuracy was assessed by comparing records of in-hospital death from hospital databases with death and time of death recorded from the NDI.

The ICU and cardiac surgery databases were used to provide preoperative, intraoperative and postoperative vari-

### Table 1. Diagnostic classification of patients according to APACHE II diagnostic codes

<table>
<thead>
<tr>
<th>Diagnostic group</th>
<th>First admission</th>
<th>All ICU admissions</th>
<th>APACHE II diagnostic codes and definitions used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac surgery</td>
<td>17 741 (83%)</td>
<td>19 793 (80%)</td>
<td>Coronary artery bypass graft, valve, aortic, thoracic transplant surgery or other cardiac surgery</td>
</tr>
<tr>
<td>Cardiac medical</td>
<td>984 (5%)</td>
<td>1458 (6%)</td>
<td>109–112, 115–117, 302</td>
</tr>
<tr>
<td>Medical</td>
<td>724 (3%)</td>
<td>1005 (4%)</td>
<td>101–108, 123–124, 303–308</td>
</tr>
<tr>
<td>Thoracic surgery</td>
<td>562 (3%)</td>
<td>758 (3%)</td>
<td>209, 303</td>
</tr>
<tr>
<td>Cardiac arrest</td>
<td>352 (2%)</td>
<td>424 (2%)</td>
<td>114</td>
</tr>
<tr>
<td>Sepsis</td>
<td>324 (1.5%)</td>
<td>480 (2%)</td>
<td>113, 204</td>
</tr>
<tr>
<td>Other surgery</td>
<td>320 (1.5%)</td>
<td>530 (2%)</td>
<td>All other surgical codes</td>
</tr>
<tr>
<td>Trauma</td>
<td>159 (0.7%)</td>
<td>188 (0.8%)</td>
<td>118, 119, 207, 208</td>
</tr>
<tr>
<td>Drug overdose</td>
<td>156 (0.7%)</td>
<td>172 (0.7%)</td>
<td>122</td>
</tr>
<tr>
<td>Neurology/central nervous system</td>
<td>83 (0.4%)</td>
<td>111 (0.5%)</td>
<td>120, 121, 217–219, 301</td>
</tr>
<tr>
<td>Missing</td>
<td>10 (0.05%)</td>
<td>13 (0.05%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>21 415</td>
<td>24 936</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Characteristics of the patient sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age years, mean (SD)</td>
<td>62 (13)</td>
</tr>
<tr>
<td>Women</td>
<td>6313 (29%)</td>
</tr>
<tr>
<td>Elective surgery</td>
<td>14990 (70%)</td>
</tr>
<tr>
<td>APACHE II score, median (interquartile range)</td>
<td>13 (11–16)</td>
</tr>
<tr>
<td>Comorbidities (APACHE II definitions)</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>2315 (12%)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>369 (2%)</td>
</tr>
<tr>
<td>Renal</td>
<td>170 (1%)</td>
</tr>
<tr>
<td>Liver</td>
<td>28 (0.2%)</td>
</tr>
<tr>
<td>Immune</td>
<td>287 (1.7%)</td>
</tr>
<tr>
<td>Cancer</td>
<td>90 (0.5%)</td>
</tr>
<tr>
<td>Haematological</td>
<td>14 (0.1%)</td>
</tr>
<tr>
<td>ICU length of stay (h), median</td>
<td>25</td>
</tr>
<tr>
<td>Mechanically ventilated</td>
<td>19 243 (90%)</td>
</tr>
<tr>
<td>Inotropes (&gt; 5 μg/kg/min dopamine)</td>
<td>2440 (12%)</td>
</tr>
<tr>
<td>Any renal replacement therapy</td>
<td>259 (1.2%)</td>
</tr>
<tr>
<td>In-hospital mortality</td>
<td>900 (4.2%)</td>
</tr>
</tbody>
</table>

* Unless otherwise indicated.
Variables were defined according to the APACHE II system unless otherwise specified.

Results

The final dataset comprised 24,936 records. There were 21,415 unique patients with 22,552 hospital admissions during the study period. There were 19,058 surgical procedures performed with a total of 20,092 ICU admissions. There were 4,930 deaths (23% of all patients) recorded in the NDI. Fifteen patients who died in Prince Charles Hospital according to the ICU or cardiac surgical databases out of 900 in-hospital deaths were not in the NDI list of deaths (1.7%). Six of these were confirmed by the HBCIS as dying in hospital with a recorded DOD, giving 4,936 deaths at follow-up. Nine of the 15 cases in which hospital databases and the NDI disagreed on categorisation of death were from 1994–1996, the period when most of the missing data occurred.

Patient age at first admission ranged from 18 to 94 years, with a median of 65 years (interquartile range [IQR], 55–72 years). There were 14,818 (69%) men, 6,313 (29%) women, and 284 (1.3%) with missing data on sex.

The casemix of the sample for first admission and all admissions is shown in Table 1. Most patients (80%) were classified as cardiac surgical, followed by cardiac medical (6%), and other medical (4%). In-hospital mortality based on the first admission was 4.2% (95% CI, 3.9%–4.5%). For the 20,515 patients discharged alive from hospital, there were 134,047 years of follow-up, with a mortality of 3% per year.

Characteristics of the sample are shown in Table 2, and the deaths at follow-up for each annual cohort in Table 3. Table 4 is a life-table analysis of the sample with yearly survival. It shows the number at risk at the beginning of each year, and those who died and were censored in each year.

The Kaplan–Meier survival curves for the entire sample are shown stratified by age in Figure 1, and by sex in Figure 2.
Figure 3. Kaplan–Meier survival estimates for patients discharged alive from hospital, by admission diagnosis at first admission

Figure 3 shows Kaplan–Meier survival curves for patients discharged alive from hospital, stratified by diagnostic category. The cardiac surgery cohort had a low mortality after discharge, as did the drug overdose and trauma cohorts. Figure 4 shows the survival according to discharge destination of those discharged alive from hospital in the period 1997–2005. The long-term outcome of those not discharged home (5.6% of the sample) was considerably poorer, particularly for those transferred to other hospital ICUs.

Discussion

We have described in detail our methods for long-term follow-up of ICU patients using the AIHW NDI. Where possible, we attempted to follow the procedures of previous studies\textsuperscript{25-27} to allow comparisons. Although Charlson comorbidity variables were not available, we were able to use APACHE II comorbidities for all years and APACHE III comorbidities for the years 2000–2005 (not shown). The lack of detailed comorbidity data is a limitation of this study. Linkage to hospital records that include International Classification of Diseases version 10 (ICD-10) codes should be considered in future studies.

The casemix of our sample limits generalisability to more-general ICUs as 80% of the sample were cardiac surgery patients. However, we used diagnostic groupings similar to those in previously reported studies. Cardiac surgery accounts for about 30% of ICU admissions in the Australian and New Zealand Intensive Care Society database, so is an important cohort. The good long-term survival of most cardiac surgery patients will influence any analysis, and should be accounted for in prognostic models.
Repeat admissions are difficult to deal with, as the results for the two admissions are not independent (as assumed by most statistical models). In this study, we simplified the problem by using only information from each patient’s first admission. A more sophisticated model would censor results for patients who return to the ICU, and then re-start their time since discharge from their new visit, while also using their updated information, such as age and APACHE score. However, the bias created by our simpler approach should be relatively small, as the number of repeat visits was low. Nevertheless, we plan to use the more sophisticated design in future analyses.

Assessing the accuracy of our data is problematic. We reported an error of 1.7% based on comparison of in-hospital deaths. In the absence of a unique identifier such as a social security number, algorithm-based probabilistic matching is reported to be 92%–95% accurate.28 This may not be sufficiently accurate for short-term follow-up of a condition with a low mortality rate, such as 90-day survival after elective cardiac bypass grafting, where a 1%–2% mortality error may be important. Studies in the United States report a sensitivity of more than 83% and specificity of 92% with use of social security numbers for matching.32 In our study, we “hand” checked additional records to improve sensitivity and specificity. A total cost analysis was not performed, and costs do not include unpaid research time.

Large series now suggest 10%–30% of ICU patients are discharged from the index hospitalisation to other care, while about 60%–65% are discharged home. In-hospital mortality analysis may record these transfers to other care as survivors or exclude them from analysis. Our study suggests that the transfer group has much lower long-term survival than patients who can be discharged home. Hospital discharge practices may substantially affect the standardised mortality ratio.33

Long-term studies using our methods may better inform decision-making and allow predictive modelling, benchmarking and targeted interventions in high-risk ICU survivors.8,34 Further investigations to determine quality of life are also required, such as might be obtained from the ICON (Intensive Care Outcome Network) Study35 and RECOVER (Outcomes and Needs Assessment in Intensive Care Unit Survivors and Their Caregivers) study,36 to examine the impact of critical care on survivors and their care givers.

Author details
Daniel V Mullany, Director of Intensive Care
Carole L Foot, Intensive Care Specialist
Susan E Smith, Data Manager, Department of Cardiac Surgery
Janelle Johnson, Data Manager
Adrian G Barnett, Senior Statistician
1 Prince Charles Hospital, Brisbane, QLD.
2 Royal North Shore Hospital, Sydney, NSW.

References


