Australian Government
Australian Institute of Health and Welfare

Better information and statistics
for better health and wellbeing

Measuring and reporting mortality in hospital patients

David Ben-Tovim[1], Richard Woodman[1], James E Harrison[2], Sophie Pointer[2], Paul Hakendorf[1], Geoffrey Henley[2]

1. Clinical Epidemiology Unit, Flinders Medical Centre and Flinders University

2. AIHW National Injury Surveillance Unit, Flinders University

March 2009

Australian Institute of Health and Welfare

Canberra

Cat. no. HSE 69

The Australian Institute of Health and Welfare is Australia's national health and welfare statistics and information agency. The Institute's mission is *better information and statistics for better health and wellbeing*.

A complete list of the Institute's publications is available from the Institute's website <www.aihw.gov.au>.

ISBN 978 1 74024 904 1

Suggested citation

AIHW: Ben-Tovim D, Woodman R, Harrison JE, Pointer S, Hakendorf P & Henley G 2009. Measuring and reporting mortality in hospital patients. Cat. no. HSE 69. Canberra: AIHW.

Any enquiries about or comments on this publication should be directed to:

Professor David Ben-Tovim

Clinical Epidemiology Unit

Flinders Medical Centre

Bedford Park SA 5042

Phone: (08) 8204 3093

Email: david.ben-tovim@health.sa.gov.au

Published by the Australian Institute of Health and Welfare

Printed by Bluestar Print ACT

Please note that there is the potential for minor revisions of data in this report. Please check the online version at <www.aihw.gov.au> for any amendments.

Contents

Acknowledgments

The Australian Commission on Safety and Quality in Health Care (the ASCQHC) has funded a major project being undertaken by the Australian Institute of Health (AIHW) to develop national indicators of safety and quality in health care. This is a report of a support project undertaken in conjunction with the major project.

The authors wish to acknowledge the following people for their contribution to this report: Kylie Thomas and Dr Robert Elzinga (Flinders Medical Centre); Vicki Bennett, Jenny Hargreaves, Belinda Emms and David Bulbeck (AIHW). Thanks also go to NIAG committee members for their feedback on an early draft.

Abbreviations

AAA	abdominal aortic aneurysm
ABS	Australian Bureau of Statistics
AIHW	Australian Institute of Health and Welfare
AMI	acute myocardial infarction
APR-DRG	All-patient Refined Diagnostic Related Groups
CABG	coronary artery bypass grafting
CIHI	Canadian Institute for Health Information
ERM	elaborated risk-adjusted mortality (model)
GP	General Practitioner
HCFA	USA Health Care Financing Authority
HMO	Health Maintenance Organisation
HSMR	Hospital Standardised Mortality Ratio
ICC	Intraclass Correlation Coefficient
ICD-10	International Classification of Diseases 10th Revision
ICD-10-AM	International Classification of Diseases Edition Australian Modification
ICU	intensive care unit
LR	likelihood ratio
NHCDC	National Hospital Cost Data Collection
NHMD	National Hospital Morbidity Database
NHS	National Health Service
NISU	National Injury Surveillance Unit
PMC	Patient Management Categories
RACM	risk-adjusted Canadian referred mortality (model)
ROC	receiver-operated curve
SEIFA	socioeconomic indexes for an area
UK	United Kingdom
USA	United States of America
VA	Veterans Affairs
WCH	Women's and Children's Hospital

Summary

Measuring and reporting mortality in hospital patients aims to develop national indicators of in-hospital mortality and is one of several projects conducted for the National Indicators Project commissioned by the Australian Commission on Safety and Quality in Health Care. The project has two parts: a literature review focusing on methods for analysing and reporting in-hospital mortality, and a modelling project aimed at establishing national indicators of in-hospital mortality that can be implemented now, and in the future.

Literature review

Papers on in-hospital mortality have been appearing in the scholarly literature since the middle of the 19th century. A large and growing body of modern literature describes the methods used to measure in-hospital mortality to allow comparison of mortality levels between different hospitals. Valid comparisons require methods that adjust for the differing risks of patient mortality that arise from hospitals having patients with different mixes of illnesses.

There is an emerging international consensus on which measure to use (the risk-adjusted Hospital Standardised Mortality Ratio, HSMR), on patient characteristics (such as age and diagnosis) to be included in risk-adjustment models, on modelling methods, and on types of cases to exclude (e.g. palliative care cases). Routinely collected data from good quality systems appear to provide an adequate basis for measuring in-hospital mortality, though discussion continues about data quality. Risk-adjusted in-hospital mortality rates—calculated using routinely collected data—are now reported regularly and publicly in several countries or jurisdictions within countries (United Kingdom, The Netherlands, Canada, and Queensland, Australia).

Three main methods are used for presenting comparative in-hospital mortality data: tables, caterpillar plots and funnel plots. For individual hospitals, the methods generally feature the ratio between the actual or observed mortality rates and the expected rates calculated from the models. Because there is some random variation in mortality rates, and expected rates fall within a range, the confidence intervals for the expected rates are usually also presented.

Longitudinal analysis of in-hospital mortality is an emerging and powerful new theme in the literature.

Measuring in-hospital mortality in Australia

The routinely collected data from the Australian National Hospital Morbidity Database were analysed. We applied a method used in Canada, England and the Netherlands, and referred to in this report as the risk-adjusted Canadian referred mortality (RACM) model. Logistic regression modelling of in-hospital mortality was used to calculate expected mortality: adjusting risk according to principal diagnosis, age, sex, comorbidity, length of stay, emergency or elective admission status and whether transferred from another hospital. The expected mortality estimate for each hospital was then combined with observed deaths to calculate risk-adjusted HSMRs.

The model was tested to determine how well it predicted or explained the actual variation in mortality rates.

HSMR analysis was conducted on three groups of cases, which exemplify types of general-purpose indicators of in–hospital mortality:

- high-risk cases (20% of cases, 80% of in-hospital deaths)
- lower risk cases (all other in-scope cases; that is, the other 80% of cases including 20% of in-hospital deaths)
- all cases and all in-hospital deaths.

Data for 1 year were analysed initially. Longitudinal analysis was then done using 3 years of data. This was a two step process. The first step was to calculate risk-adjusted HSMRs in a similar way to the 1–year analysis. The second step was two-stage multi-level logistic regression.

The hospital peer group classification developed by the AIHW was used to group hospitals for comparisons.

Results

Overall, the results demonstrated that, using the Australian data, the RACM model predicted or explained the variation in mortality rates to a similar extent as models reported in the international literature. Some differences in the strength of the model were apparent when applied to the three mortality groups (80%, 20% and 100%): with better prediction of mortality rates for the 20% and 100% groups.

Single-year analysis (2005–06)

The single-year analysis resulted in the production of HSMRs and confidence intervals for public hospitals in peer groups. They are presented using HSMR ranked tables, funnel plots and caterpillar plots. Funnel plots illustrated that some hospitals had HSMRs that were relatively high or low compared with peer hospitals.

Longitudinal analysis (2004–05 to 2006–07)

The longitudinal analysis showed that most variation in HSMRs was between different hospitals, with much less variation between repeated measurements for the same hospital. The lack, on the whole, of large variation between measures of HSMR for the same hospital suggests that values largely reflect the phenomenon of interest (mortality rates), and are not dominated by 'noise' in the data. This is less true for peer groups of small hospitals.

The results presented for the longitudinal analysis demonstrate a modest decline in overall risk-adjusted mortality during the 3–year period. This is similar to the findings of a recent Dutch study using the same method. Although replication of analysis and refinement of the method used should be undertaken before too much weight is place on this finding, the possibility remains that it is a true decline. If so, perhaps an increased emphasis on hospital safety in recent years is beginning to have a demonstrable effect on in-hospital mortality.

Conclusions

This project shows that indicators of in-hospital mortality can now be produced using the Australian National Hospital Morbidity Database. The present study produced indicators based on the three mortality groups specified above, reported by hospital in public hospital

peer groups. Our findings suggest that the available data are generally sufficient for this purpose.

How should the HSMRs be used? Variations in hospital mortality should be viewed as screening tests rather than being diagnostic of poor safety or quality. High or rising HSMRs signal that a problem might exist and that further investigation is required. Low or falling HSMRs might signal good performance, from which lessons could be learned.

Further work

The report also describes a refinement to the RACM model—referred to as the elaborated risk-adjusted mortality (ERM) model—which consistently predicted mortality rates better than the RACM model. This model could be further investigated for its potential to generate indicators of in-hospital mortality.

Further work is also warranted on several matters where data limitations prevented us from undertaking desirable aspects of analysis.

Data matching to include deaths up to 30 days after discharge is technically feasible in Australia, as is internal linkage of the data on multiple episodes of care for individual patients. Both of these forms of linkage are routine parts of data linkage activities in some states. They could be used for a project to test these refinements of the data underlying HSMR, in preparation for later use of linked data for national indicators.

The lack of a hospital identifier for many private hospitals prevented analysis of this sector—a limitation that needs to be overcome in future analyses.

Emerging data developments (national coding of conditions 'present on admission' to distinguish pre-existing comorbidities from complications of care) and analytical innovations (e.g. use of Bayesian regression, especially for data from small hospitals) are also likely to improve results.

1 Introduction

1.1 Context of the report

This project, *Measuring and reporting mortality in hospital patients,* is one of several conducted as part of the National Indicators Project commissioned by the Australian Commission on Safety and Quality in Health Care (the Commission). The purpose of the National Indicators Project is to systematically identify and develop information that can be used to monitor Australia's performance in safety and quality in health care, for intra jurisdictional, inter-jurisdictional and international benchmarking and reporting purposes.

Measuring and reporting mortality in hospital patients aims to develop National indicator(s) of in-hospital mortality. The specified outcome for the project is:

- The development of indicators of in-hospital mortality taking into consideration different types of measurement and/or presentation (e.g. disaggregation) that may be required at the national level, at jurisdictional level and at facility or unit level.

In reaching this outcome, this project provides a detailed review of available literature: evaluating current developments in the measurement and presentation of in-hospital mortality indicators, as well as recommending in-hospital mortality indicators that can be produced using current administrative data sets.

One of the strengths of mortality measures is that the fact of death is unequivocal and generally accurately reported. The task is to identify how measures of hospital mortality can be further developed to generate an indicator, or indicators, of hospital safety and quality more generally.

Measuring and reporting mortality in hospital patients has two separate, but interdependent, components: a literature review focusing on methodologies for analysing and reporting in-hospital mortality, including methods and models for risk adjustment; and a modelling project aimed at establishing what national indicators of hospital mortality can be developed now, and in the future.

The modelling project provides a variety of worked examples of methods for analysing and presenting mortality rates using national data sets.

1.2 Structure of the report

The report contains six chapters of which this introduction is the first. Chapter 2 details the review of the literature and Chapter 3 introduces the modelling project itself. Chapter 4 provides a description of the method employed, while Chapter 5 presents the results of the modelling. The report concludes with a discussion of the findings in Chapter 6. The report also contains Appendixes, including one on data issues.

2 Review of the literature

2.1 General introduction: in-hospital mortality

Although users of hospital care might consider variations in mortality rates to be of significance in their own right, the increased interest in them in recent years relates primarily to their role as indicators of broader issues in relation to the safety and quality of care provided within hospitals.

2.2 Introduction to the literature review

The narrative review that follows contains a broad introduction, and a description of the search process. Then there is an analysis of what is known about the extent of variations in hospital mortality rates, and of sources of variation; this incorporates a discussion of risk adjustment. There is a section devoted to the analysis of the relationship between variations in hospital mortality and other measures of safety and quality. Hospital mortality as an indicator is then assessed against a series of general and technical issues in relation to criteria for indicator development (Scobie et al. 2006).

2.2.1 Developments from 1860 to present

The issues around hospital mortality rates were clearly articulated in the middle of the nineteenth century (Spiegelhalter 1999). Between 1861 and 1865, the *Journal of the Statistical Society of London* published a series of articles describing hospital mortality rates, probably at the urging of Florence Nightingale. Nightingale advocated the publishing of uniform hospital statistics because these would 'enable us to ascertain the relative mortality of different hospitals, as well as of different diseases and injuries...'(Nightingale 1863).

Nightingale was very interested in the issue of quality within hospitals. She hoped that such statistics would ensure that 'As regards their sanitary condition, hospitals might be compared with hospitals and wards with wards' (Nightingale 1860). The kinds of dilemmas that the publication of such statistics would raise were also clearly understood by Nightingale, including the importance of risk adjustment for age, sex and complications (Nightingale 1863). These issues were well canvassed in the comments of Guy (1867) in relation to variations in the mortality of London hospitals.

Guy stated that 'it would be no less invidious than unjust to attribute the differing death–rates of our hospitals, in an appreciable degree, to any difference in the professional skill and ability of their professional staff, chosen, as it is, from among those members of the profession [including himself] who have already given proofs of sound training, ability and skill in practice' (Guy 1867).

Although Guy (1867) attributed the variations in hospital mortality to casemix ('…the mortality of hospitals is mainly due to causes which determine the nature and severity of the cases admitted within their walls…'), the mortality rates he quoted were not in fact adjusted for such variations, so the basis of his assertion is unclear.

Interest in variations in hospital mortality remained sporadic until the end of the 1980s, despite the unexpected findings of substantial inter-hospital variation in post surgical mortality in the National Halothane study in the USA conducted in the 1960s (Moses & Mosteller 1968). A review published in 1989 (Fink et al. 1989) could only find three articles (from 22 identified after a search) that contained any kind of adjustment for severity of illness, as well as demographic and health status issues.

At that point, the forced release by the USA Health Care Financing Administration (HCFA) of mortality rates of Medicare patients for all Medicare provider hospitals led to a major surge in interest in the analysis of hospital mortality. It appears that there was concern that the introduction by HCFA of a fixed-fee prospective payments system for Medicare patients—based on diagnostic related groups—might lead to a decrease in the quality of care provided (e.g. Stern & Epstein 1985; Iglehart 1986). Calculations of hospital mortality were a monitoring activity related to the introduction of the prospective payments system.

The public release of the HCFA information sparked considerable professional and community interest. Although subsequent studies confirmed that there were indeed variations in hospital mortality rates (e.g. Dubois et al. 1987; Chassin et al. 1989 Bradbury et al. 1991; Manheim et al. 1992; Thomas et al. 1993), a debate ensued as to the extent to which hospital level variations in mortality measures were sufficiently reflective of variations in the quality of hospital care to be broadcast to a non-professional audience, or to influence funding or purchasing decisions by insurance groups or other funders (Green et al. 1991; Hofer & Hayward 1996).

The intensity of the questioning was such that in 1993 HCFA ceased producing mortality measures. But interest in mortality measures did not decline, and as concerns have been examined and health-care providers have become more used to the release of mortality data, the frequency with which comparative risk-adjusted mortality measures have been made available to institutions and the public at large has increased year by year, and country by country.

2.3 Search method

We searched PubMed (last search updated June 2008) with a focus on studies where mortality was the primary outcome.

We searched with a variety of strategies using the following search terms: hospital mortality, review quality + risk-adjusted mortality, review risk-adjusted mortality, risk-adjusted mortality methods, risk-adjusted mortality rates, risk-adjustment methods, hospital mortality classification, history mortality measurement, quality risk-adjusted mortality rates, quality + risk-adjusted mortality rates, hospital standardised mortality ratios.

We focused on studies that compared whole of hospital mortality rates and related the results to any evidence of quality and or safety. Although we did find many studies looking at only a single condition—such as acute myocardial infarction (AMI), coronary artery bypass grafting (CABG), and pneumonia—or only one hospital, those were not our primary interest. We aimed our review at studies that compared at least two hospitals. Studies that looked at mortality through the lens of organisational/structural variables, nurse–patient or physician–patient ratios, and public versus private funding were not our prime focus.

In our search we paid particular attention to national mortality rate reporting that has recently been undertaken in the United Kingdom, United States, Canada and Holland.

For all studies, the authors decided final inclusion/exclusion by discussion and consensus.

2.4 Considerations in the development of mortality as an indicator

2.4.1 Random and systematic variation

Before any attempt is made at interpreting or using hospital mortality data, a basic issue needs to be understood and responded to.

Hospital mortality is a special case of a more general issue related to the analysis of variations in the outcomes of any intervention (Thomas & Hofer 1998) when the factors involved are both systematic and random. As with any other outcome in biomedicine, variations in mortality outcomes following hospitalisation can be thought of as having at least two components:

- systematic variations in factors that may influence [mortality] outcomes; those variations being assumed to relate to the quality and effectiveness of the interventions that affect the outcome in question
- random variations.

The random variations may have a variety of origins. There are random variations in the moment-by-moment effectiveness of biomedical interventions, even when they are optimally applied. There are random variations in the interaction between optimally applied interventions and the immediate states of the people to whom those interventions are applied, and random variations during attempts to implement evidence-based interventions (the systematic consequences of the longer term characteristics, or traits, of those people are best thought of as confounders of systematic variation and are considered below).

In biomedical research, the uncertainties due to random variations are optimally dealt with by a process of randomisation. When patients are randomly allocated to the settings or interventions of interest, the presence of a systematic effect is confirmed by assessing the magnitude of differences in outcomes between sites or interventions, taking overall variability into account. The fundamental analytical question is whether the observed differences are so large that they are unlikely to have occurred by chance alone.

Hospital mortality measures are measures of outcomes in the usual care provided by hospitals. There is no possibility of random allocation of patients to different sites. The question of whether observed differences are so large that they are unlikely to have occurred by chance can only be assessed by comparing the outcomes for a patient or group of patients treated in any one hospital against a hypothetical outcome that might have occurred if the patient(s) had undergone treatment elsewhere.

The most straightforward way to do that would be to assess the average outcome across the population being assessed and use that to calculate the expected outcome (and confidence limits around the value) for the number of patients treated at any one hospital. The observed (actual) and expected values for the numbers of patients treated would be compared and a decision made as to whether any hospitals stand out as being 'extreme' in terms of

differences between observed and expected outcomes. However, a simple comparison on that basis is likely to be both inaccurate and misleading.

Patients are non-randomly allocated (and self referred) across institutions. The use of crude averages ignores patient-level differences between institutions that might systematically influence outcomes. These confounding factors, combined, may be described as variations due to the clinical, demographic and casemix differences between patients present at the point of arrival in hospital (V_C). In which case, total variation in in-hospital mortality (V) comprises:

- systematic variations in factors influencing mortality outcomes; those variations being assumed to relate to the quality and effectiveness of the interventions that affect the outcome in question (V_Q)
- variations due to the clinical, demographic and casemix differences between patients present at the point of arrival in hospital (V_C)
- random variations (V_R).

In most studies of hospital mortality, efforts are made to discount V_C before assessing the magnitude of any inter-hospital differences (Thomas & Hofer 1998). The measurement of V_C for this purpose is usually described as risk adjustment because pre-existing patient-level factors influence or confound any other institutional-level factors that might influence the risk of dying in hospital. There is also the possibility that there are some confounding factors related to the characteristics of the functional catchment areas of hospitals that are not captured in existing individual-level measures, and that need to be accounted for by inserting measures of social disadvantage into analyses (Jarman et al 1999). Whilst there is disagreement as to whether such influences should or should not be adjusted for, the question of the influence of catchment population measures on in-hospital mortality in the Australian context is assessed empirically in this project (Section 5.9).

Much of the criticism of the release of the HCFA mortality studies of the 1980s (Rosen & Green 1987; Berwick & Wald 1990; Green et al. 1991) related to the fact that the risk adjustment was confined to the impact of each patient's principal diagnosis and four secondary diagnoses, and demographic factors of age, sex, race, and whether the patient had been transferred from another hospital. Critics argued that this was too simplistic to adequately adjust for patient-level variations between institutions (Green et al. 1991).

2.4.2 Mortality at what point: in-hospital, 30 days after discharge, or longer?

Another common complaint in the literature following the release of the HCFA data was that many of the effects of hospital care do not become evident until after patients leave hospital. Also, if studies of variations in mortality rates were to be confined to deaths during hospital stays, hospitals might be tempted to discharge poor prognosis patients to minimise in-hospital mortality (Omoigui et al. 1996)

By linking hospital data with relevant information from death registers, a number of investigators have assessed the relationship between mortality during hospital stay and mortality 30 days after discharge (Jencks et al. 1988; Chassin et al. 1989, Rosenthal et al. 2000) or longer (Fleming et al. 1991; Garnick et al. 1995). Inclusion of deaths in the thirty-day period after discharge appears to be sufficient. After an exhaustive study, Garnick et al.

(1995:693) concluded that 'mortality occurring after 30 days has little to do with hospital-specific effects…'

As may be expected, mortality up to 30 days after discharge is tends to be similar to in-hospital mortality (e.g. Rosenthal et al. 2000), but this is not necessarily so, and variations have the potential to be informative. Assessing mortality up to 30 days after leaving hospital provides the opportunity to assess effects of variations in discharge policy (Jencks et al. 1988) and of immediate post-discharge care.

Whilst it may thus be preferable to assess mortality in a manner that includes deaths up to 30 days after discharge, it is not always feasible to do so, and the gain in precision by taking account of mortality after discharge has to be traded against the greater complexity involved in linking hospital administrative information with other registry data (Krakauer et al. 1992). However, developments in population-level data linkage capabilities, such as the Western Australian Data Linkage System and the work of the Centre for Health Record Linkage in NSW, are reducing this barrier and will offer further opportunities in the future.

2.5 Model development

2.5.1 What variables to include in risk adjustment

Demography and diagnosis

The risk-adjustment hypothesis is that observed rates of in-hospital mortality will be systematically influenced by the characteristics of patients on arrival at the hospital.

It seems reasonable to assume that the risk of death during a hospital stay is likely to be influenced by factors such as age, sex, primary clinical diagnosis and secondary or complicating diagnoses present at admission. Information on these types of factors is commonly collected within administrative data sets—that is, within information about individual patients collected by hospitals for internal and external administrative reasons and mandatory reporting requirements. Hospital-level administrative data sets in Australia and elsewhere also commonly contain information about arrival and discharge dates, home address, source of referral, whether the admission was as an emergency or planned, and the nature of discharge. Information about ethnicity may or may not be available, along with other jurisdiction-specific information.

Severity

Administrative data sets do not usually contain much information about the severity of the principal diagnosis, though this varies between diagnoses. For example, Australian data coded according to the International Classification of Diseases Australian Modification of the ICD (ICD-10-AM) do not usually provide information on the severity of an uncomplicated case of community-acquired pneumonia over and above the diagnosis itself. The same classification does, however, distinguish depressive episodes as mild, moderate, severe and severe with psychotic symptoms, and liver lacerations as minor, moderate and major.

Severity is neither a simple nor a uniform characteristic, nor easily or uniformly assessed. For instance, the severity of heart disease may be inferred from physiological or medical imaging data reports, whereas the severity of schizophrenia is best determined by clinical judgment.

Institutional characteristics

Many administrative data sets that report patient-level data also characterise the reporting institutions in some way. The basic requirement is for a field in patient-level records that records the treating hospital[1]. This is particularly important when a data set contains outcomes from both large principal referral hospitals, and small institutions. The case loads of small hospitals are often quite different from those of tertiary institutions. Patients in smaller institutions can appear to be at lower risk than patients in larger institutions, even after risk adjustment. However, it is not appropriate to assume that the smaller hospital could achieve similar types of outcomes if they were confronted with the kinds of patients that tertiary institutions have to deal with. A low-risk hospital is only low risk for the kinds of cases it is familiar with (Shahian & Normand 2008). So, institution type is a relevant issue when making comparisons. Risk adjustment itself is, however, best undertaken at the level of the patient, not the institution (e.g. Hadorn et al. 1993).

2.5.2 Logistic regression and risk adjustment

The 'mechanics' of risk adjustment—once potential risk modifying factors have been identified—are well established. Taking hospital mortality as the dependent variable, the influence on outcome of various independent variables (or contributors of mortality) is assessed by means of logistic regression: the appropriate analytic strategy for binary (survive/dead) outcomes. Logistic regression allows development of a linear equation for the log (odds) of a positive outcome. The log (odds) increases by the magnitude of the coefficient for each unit increase in the independent variable. For example the log (odds) of a positive outcome for male versus female increases by the coefficient for sex, if male is coded 1 and female is coded 0.

The exponentiated coefficients can then be interpreted as the change in the odds of a positive outcome for a unit increase in the associated independent variable (i.e. covariate).

The coefficients from logistic regression can also be applied to create a predicted probability of an outcome of interest (i.e. death) for each individual in the data set. The probabilities for each particular pattern of covariate values effectively create a set of reference weights that relate to the population of hospitals as a whole, enabling standardisation of each individual hospital to a reference hospital population. The aim is to profile how the results for a particular hospital compare with what would be expected if that hospital functioned in a way that was typical for the whole population of hospitals studied.

2.5.3 Logistic regression, indirect standardisation and HSMR

Each patient in any one hospital will survive or die. The sum of all the deaths divided by the total number of hospital separations is the crude in-hospital mortality rate for that hospital. By calculating the probability that any one patient in a population of patients will die (or survive) using the logistic regression coefficients and covariate values relevant to that patient, it becomes possible to compute the standardised mortality rate for that institution; that is, a mortality rate that is adjusted for its casemix.

1 Some private hospitals are not identified as separate establishments in the Australian hospitals data available for this project (see Appendix 5 Data issues).

Indirect standardisation of hospital mortality rates is the term given to the comparison of the observed mortality rates against the expected rates as generated from the study of all the patients within the hospital populations studied. Those expected rates become the denominator of the ratio of observed to expected outcomes (O/E). A ratio value less than 1 is favourable and a ratio of greater than 1 unfavourable. When the ratio is multiplied by 100 the convention is to describe that value as the Hospital Standardised Mortality Ratio (HSMR) (Jarman et al. 1999).

Although the computational method used in risk adjustment for the calculation of hospital level HSMRs is now fairly settled, the range of contributing variables that might be included in the regression equation is almost without limit. In practice, there is an emerging consensus on which variables to include in studies analysing the majority of deaths occurring within hospitals (as distinct from studies dealing only with deaths of specific types).

2.5.4 Narrowly focused or broad studies

Which patients should be included in the study of mortality rates? Should the study be narrow focused or more broadly based?

Over the years, studies have examined mortality rates in single conditions, small groups of diagnoses with high predicted short-term mortality, patients from diagnostic groups in which the majority of deaths within hospital occur, or all patients treated with a limited number of exclusions. Despite the substantial potential differences involved, there has been little discussion of the rationale behind any one choice, though there are some practical issues to be considered.

Studies of hospital mortality easily accumulate very large numbers of subjects; for example, the national data set for all separations from Australian hospitals in the financial year 2005–06 contains some 6 million individual records. Data sets from countries with higher populations, such as the UK or USA, will be proportionately larger.

The surge in interest in hospital mortality began in the late 1980s. Although it is not explicitly discussed in the literature, very large data sets were not as easy to handle at that time as they are now. The greater expense then of acquiring access to sufficient computing power would have been a consideration in favour of opting to confine analyses to subsets of the whole population of patients treated in hospitals.

A limited number of clinical conditions accounts for the majority of deaths within hospital. When analyses examine mortality rates within the diagnosis groups that account for 80% of all hospital deaths, clinical diagnoses—albeit somewhat simplified or consolidated—can be included directly within risk-adjustment methods (e.g. Kahn et al. 1990). Once studies encompass all deaths within a population of hospital patients, then some means of aggregating diagnoses into larger groups is required because the numbers of individual diagnoses are just too large for all those diagnoses to be individually included in risk-adjustment computations.

In all studies, provision is made to exclude those patients for whom death in hospital is integral to the service provided. Strategies have been developed to deal with palliative-care-type hospital separations (CIHI 2007). In Australia, palliative care is designated within administrative data sets as a care type that can only be provided in a designated Palliative Care service. It is straightforward to exclude such patients. In settings where that is not possible, other arrangements are required to deal with potential palliative-

care issues, such as excluding patients with a primary diagnosis of cancer (e.g. Lakhani et al. 2005).

Restricting the analysis of mortality to a small number of conditions may be relevant if there is a strong interest in linking mortality outcomes with specific process measures. Otherwise, a broader sample of in-hospital deaths is likely to provide a more representative population for analysis. The case for confining a more broad-based analysis to the higher risk diagnoses that account for 80% of deaths—instead of all in-hospital deaths—has not been formally argued, and relates more to convenience and the capacity to include primary diagnoses as they stand within the risk-adjustment process, than to other issues of substance. The analyses further include high-risk diagnoses, low-risk diagnoses, and all causes of mortality.

2.5.5 Summary measures of model performance

The underlying rationale for logistic regression is that the risk of an event in relation to risk factors falls along a logistic curve. The s-shaped logistic curve is shown below, where 0 on the y-axis is alive, and 1 the outcome dead, and the values between are the probabilities of the outcome.

Logistic regression analyses are mathematical models that attempt to fit the data to the logistic curve. Commonly asked question of such models are 'How good is it? What is its predictive validity—how well does the model account for the actual variation in patient-level risks (Shwartz & Ash 2003)?'

There is some controversy in the technical literature about what, if any, are the best summary measure to use to answer such queries. There are two issues to be considered: null model and goodness of fit.

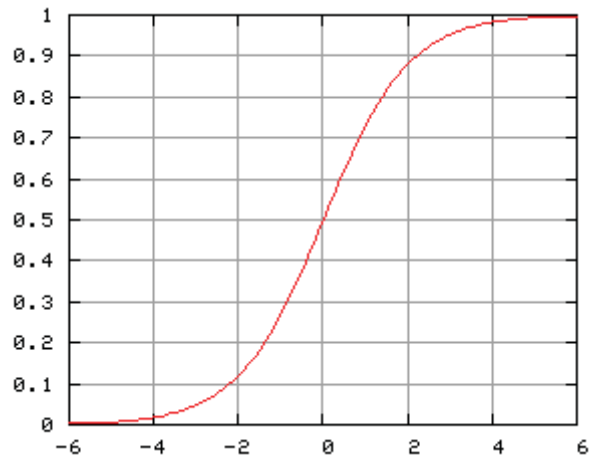


Figure 1: Example of an s-shaped logistic curve

2.5.6 Null model

Firstly, do the models created improve upon the 'null model'? Say we are interested in examining the mortality at St Elsewhere—one of a population of hospitals for which in-hospital mortality is being studied. If there is no opportunity to risk adjust by reference to additional variables, the only way to define the expected numbers of deaths in St Elsewhere is to take the average death rate for all hospitals and apply that rate to the total number of patients treated in St Elsewhere, deriving a predictive 'null' model using that information alone.

If patient-level confounders are important, adding them to the model will improve predictive power over a model with no other adjustment variables. Whether any improvement is statistically significant may be tested by means of a likelihood ratio (LR) test. LR tests examine the predicted probabilities of living among those who lived, and the predicted probabilities of dying amongst those that died. Better models have higher LRs (i.e. more of the living were predicted to have lived, and more of the dying were predicted to have died).

2.5.7 Goodness of fit

Goodness of fit is a somewhat different question. The issue is not 'does it fit better than the null model?' What is being asked is 'how well does the model fit?' It may be better than chance, but how strong is the relationship?

The challenges posed by such questions are best appreciated by comparing logistic regression models with the more straightforward measures generated for linear relationships. There, the relationships between the dependent and independent variables can be considered as potentially falling along a straight line. When increases in the independent variables are perfectly mirrored in increases in the dependent variables, an equation linking the two groups of variables will predict 100% of the variability in the values of the dependent measure. If there is no link at all, then the equation will predict 0% of the variability. By calculating the R^2 statistic, the percentage of variability explained by the equation can be calculated (i.e. how closely do the points in the scattergram linking independent and dependent relationships fit to a straight line?).

R^2 (or pseudo R^2, a related measure) can be calculated in logistic regression, but the results cannot be interpreted in the same way as in a linear regression. The issue of interpretation goes back to the fact that a logistic regression is an attempt to predict the degree to which a group of variables (such as age, sex, and admission status) predict a binary (alive/dead) outcome, not a graded one. Conceptually, the analytical question asked is 'does a risk-adjusted equation produce a result that, when applied to a population, sharply separate the population who are alive at discharge from those that die in hospital, with limited overlap between the two groups?' The problems with interpreting R^2 as a measure of 'model fit' for logistic regressions were summed up (Schwartz & Ash 2003) in a discussion of the publication of CABG data in New York (Chassin et al. 1996).

> 'In logistic-regression models in which the overall mortality rate ranges from 2 to 4 per cent, however, R2 is almost always less than 0.2. This limitation arises from the nature of logistic regression, in which the dependent variable must have one of only two values (in this case survival or death). When the differences between actual and predicted mortality rates is calculated for each person (as part of the calculation of R2) no matter how accurate the prediction is, the difference between the predicted value and the

observed value for the mortality will be large, because the observed mortality must be either 0 or 1, and the prediction is a proportion between 0 and 1.' (Chassin et al. 1996: 396–7).

Using changes in R^2 to assess the impact of adding or subtracting variables within a logistic regression model remains valid, however, because this is using it in a variable-by-variable comparison, rather than in an attempt to provide a single statistic against which to assess model fit.

2.5.8 The c-statistic

A better measure of discrimination is the c-statistic, which also equals the area under a receiver-operator curve (ROC). The c-statistic has a number of definitions, but one is as follows.

> 'Within a population, take all the possible pairs in which one patient dies and the other survives. Assign a probability of death for each patient in each pair. The c-statistic equals the proportion of cases in which the predicted probability of death is higher for the patient who died than the patient who lived. When the probability is tied, the assigned value is one half—that is, there is a 50:50 chance of being right or wrong. So when models have no ability to discriminate—that is, to truly assign a probability of death while minimising false positives—the c-statistic is 0.5. Although there are no absolute hard and fast rules, models generating a c-statistic value below 0.7 are considered to be poorly discriminatory, models with a c-statistic 0.7–0.8 are more adequate, and above 0.8 a good discrimination' (Aylin et al. 2007).

As will be shown below, many risk-adjustment models for mortality have c-statistics in the range 0.8 and above.

2.5.9 Risk adjustment across the range of predicted probabilities

Many studies of hospital mortality will involve patients across a wide range of risk. One method for assessing the robustness of risk adjusters across the whole range is the Hosmer–Lemeshow method (Hosmer & Lemeshow 2000).

Patients are divided up into deciles of predicted risk and the observed and expected values of mortality (derived from applying the coefficients of the logistic regression to the populations) calculated for each decile. The distribution of the deviations within each decile follows the chi^2 distribution, and the model is accepted if the observed deviations or differences are *less* than would be expected by chance. Despite the elegance of this method, the Hosmer–Lemeshow test, like all chi-square tests, is sensitive to sample size, and may not be suitable for studies with large samples (Schwartz & Ash 2003; Aylin et al. 2007). The direct comparisons between observed and expected values at deciles of risk may be of considerable interest (Aylin et al. 2007), and may provide insights into the impact of risk adjustment without further analysis.

2.5.10 Calibration

An entirely different issue is that of calibration. Because the risk-adjustment process begins with the calculation of an expected or average outcome, the overall observed and expected outcomes will be identical, because the expected is the average of the observed.

When a risk-adjustment equation is calculated in one population and then applied to a quite different one, the calculated expected number of deaths will not necessarily be the same as the observed. The question arises as to whether the expected results should be calibrated, or adjusted in some way, so that the overall expected and observed values resemble each other. A number of calibration methods have been suggested in the literature (see DeLong et al. 1997) but, although this is a theoretically important issue, and would need to be considered carefully if there were any attempt at a
cross-national comparison of HSMRs, it has only received limited empirical study to date.

So, in summary, there are a variety of measures that can be used to assess the robustness of a risk-adjustment process for binary outcomes, but none give a simple answer to the question 'how good is the fit?'

2.6 Inter-hospital variation and risk-adjustment models

2.6.1 Hospitals differ

After interest in variations in hospital mortality picked up following the publication of the HCFA data, the fact of highly statistically significant variations in in-hospital mortality rates have been confirmed in every country where they have been studied (e.g. Chassin et al. 1989; Kahn et al. 1990; Jarman et al. 1999; CIHI 2007; Heijink et al. 2008), in public and private hospitals alike (Devereux et al. 2002).

2.6.2 Risk adjustment—administrative data sets

Table 1 provides a listing for the R^2 and c-statistic values for a variety of reports of risk-adjustment models, and the values for the areas under the ROC where provided.

Numerous reviews of the outcomes of risk adjustment using administrative and other data sets have been published over the years (e.g. Hadorn et al. 1993; Iezzoni 1997a; Thomas & Hofer 1999; Powell et al. 2003; Daley et al. 2003), and it is now possible to draw some overall conclusions.

Administrative data sets contain a restricted amount of information at the patient level. Demographic information, mode of admission (emergency or elective, transfer from other care facility or direct) and duration of admission, care type, mode of discharge, principal and secondary diagnoses, surgical procedures, and institutional identifiers are almost always available. The Australian administrative data sets separate types of care into acute, rehabilitation and palliative care. Information about previous admissions and linkage across hospital and community services are less common.
(The spreading availability of data linkage facilities in Australia is overcoming this limitation.)

In the UK, there have been particular difficulties relating to the use of multiple consultant-completed episodes within a single admission that have had to be overcome (Jarman et al. 1999), but that is not a widespread problem outside the UK.

The most important changes over the years have related to the increase in the number of primary and secondary diagnoses that are contained within administrative data sets, with restricted numbers (e.g. in the HCFA) now commonly replaced by more exhaustive enumerations in many countries. For example, the current Australian National Morbidity Collection allows for the reporting of one primary and
49 secondary diagnoses, and up to 50 procedure codes.

A more subtle issue relates to the notion of what constitutes the principal diagnosis for a patient. In most systems that derive from the Medicare-derived USA prospective payment systems, the convention is that the primary diagnosis is the diagnosis that, after study, was the primary condition leading to hospital admission. But in the large USA Department of Veteran Affairs system, it is the condition primarily responsible for the length of the hospitalisation (Daley et al. 1997, Iezzoni 2003b). In Australia, principal diagnosis is defined within the National Health Data Dictionary as 'The diagnosis established after study to be chiefly responsible for occasioning an episode of admitted patient care, an episode of residential care or an attendance at the health-care establishment' (AIHW 2006). The specification of the principal diagnosis may have an important bearing on the risk rating of each patient.

Although concerns have frequently been raised over the accuracy of coding of diagnoses (e.g. Iezzoni 1997a, Scott & Ward 2006) those concerns have tended to become less prominent in recent years, as countries have become familiar with the work of professional coders, and as work on coding standards and coding practice has become increasingly refined.

Within the National Hospital Minimum Dataset, only a small percentage of cases are recoded due to an error (AIHW 2007), with most errors being in the direction of 'up-coding' in the direction of increased complexity, which would tend to reduce any measure of hospital mortality because observed mortality would tend to be less than expected mortality in those cases.

Studies of the outcome of risk adjustment via administrative data systems have been reviewed on a number of occasions (e.g. Hadorn et al. 1993; Iezzoni 1997b; Thomas & Hofer 1999; Powell et al. 2003; Daley et al. 2003), and a number of different methods for combining the information within administrative data sets have emerged, including a number of proprietary methods developed in the USA (e.g. the APR-DRG system, Disease Staging). However the R^2 model statistics reported in Table 1 have not varied from the 0.2 to 0.3 levels reported by Hadorn et al. in 1993, and the c-statistic levels continue to typically range from 0.7 to 0.8 or slightly above.

In the next section, the addition of clinical elements to risk adjustment is discussed. Because it stretches across both administrative and clinical risk-adjustment methods, a discussion of the integration of comorbidities in risk adjustment is undertaken further on.

2.6.3 Risk adjustment—the addition of clinical factors

Clinicians make judgments based on the clinical characteristics of their patients, so it would seem axiomatic to those practitioners that outcome predictions that include clinical information would be an improvement over those that do not. It is not surprising, therefore, that considerable effort has gone into the search for clinical elements to test in risk-

adjustment models. The rationale for those attempts was put elegantly by Hadorn et al. (1993: 1–2), 'Statistical prediction models rely on the same clinical and demographic factors (e.g. age, blood pressure) used by clinicians to arrive at prognostic judgments. Unlike clinicians, however, models assign explicit weights to these factors based on their observed statistical association with the outcome of interest (e.g. mortality) in some sample of patients. As a result, prediction models render precise (if not always accurate) predictions of outcome or diagnosis.'

The simplest of these strategies has been to model physiological data (e.g. blood pressure in the first 48 hours of stay) or laboratory test values (for potassium, haematocrit, and so on), and include them as confounders within models to risk adjust mortality data. Then there are strategies that generate condition-specific measures combining laboratory and clinical elements, using guidance from clinical panels or other sources of clinical advice to choose from among candidate variables, extracted from case notes by trained reviewers, to test in risk adjustment.

Finally, there are proprietary services (e.g. MedisGroups) whose trained personnel (commonly nurses) review case records and extract and tabulate many different features of interest that can be tested in risk-adjustment studies. Iezzoni (1997a, 1997b) describes the origin of one of the most widely used of these methods, the MedisGroups listing of key clinical findings, in the observations made by two physicians from Saint Vincent's Hospital in Worcester, Massachusetts, after participating in the morning reporting process of medical residents. These observations eventually became the initial list of what are now hundreds of key clinical findings.

Table 1 provides a selection of the model parameters from risk-adjustment models using a variety of clinical risk parameters. Although many of them do improve on the R^2 for the risk-adjustment methods based on administrative data, the gain is often modest.

Given the variety of administrative and clinical risk-adjustment methods that have emerged, the series of studies of Iezzoni and colleagues conducted during the mid 1990s are particularly important (the outcomes are tabulated in Table 1, and overall outcomes summarised in Iezzoni (1997a, 1997b). These researchers compared a wide variety of risk-adjustment methods using a single data set as the test or trial data source. They directly compared a wide variety of risk-adjustment methods for AMI, coronary by-pass artery grafting, pneumonia or stroke, and compared five of the methods on all four diagnostic groups.

The risk-adjustment methods studied included Disease Staging, All-Patient Refined Diagnostic Related Groups (APR-DRGs) and Patient Management Categories (PMC): all three being proprietary risk-adjustment methods that made use of discharge abstracts (i.e. administrative data sets). MedisGroups and the APACHE 111 system represented risk-adjustment methods that made use of physiological and or clinical data.

The results were clear. Although risk adjustment is necessary for valid comparison of hospitals or groups of hospitals, no particular method stood out as preferable. Whilst the methods tend to agree on high and low mortality outliers, no one method provided markedly more specific and consistent discrimination than the others.

Table 1: Risk-adjustment-model outcomes

Year	First author	Condition(s)/severity adjustment	R^2	C	ROC
1985	Knaus	ICU—Apache 1	0.31		0.851
		ICU—Apache II	0.319		0.863
1990	Keeler	stroke—Apache II	0.30		
		pneumonia—Apache II	0.26		
		myocardial infarction—Apache II	0.22		
		heart failure—Apache II	0.12		
1991	Knaus	ICU—Apache III on initial day	0.41		0.90
1992	Krakauer	multiple—demographic model		0.64	
		multiple—claims model		0.84	
		multiple—clinical model		0.90	
1994	Hannan	CABG		0.79	
1995	Green	CABG	0.073		
1995	Romano	AMI—model A		0.766	
		AMI—model B		0.844	
		Lumbar diskectomy—model A		0.722	
		Lumbar diskectomy—model B		0.73	
		Cervical diskectomy—model A		0.702	
		Cervical diskectomy—model B		0.744	
1996a	Iezzoni	pneumonia—medisgroups or	0.13	0.81	
		pneumonia—medisgroups exp	0.19	0.85	
		pneumonia—physiology 1	0.10	0.78	
		pneumonia—physiology 2	0.15	0.82	
		pneumonia—body systems count	0.05	0.71	
		pneumonia—comorbidities index	0.06	0.74	
		pneumonia—disease staging	0.13	0.80	
		pneumonia—PMC severity score	0.11	0.79	
		pneumonia—AIM	0.05	0.73	
		pneumonia—APR DRGs	0.10	0.78	
		pneumonia—PMC RIS	0.1	0.78	
		pneumonia—R DRGs	0.28	0.83	
		pneumonia—age sex interact only	0.03	0.67	
		pneumonia—age sex interact, DRG	0.04	0.71	
1996b	Iezzoni	AMI—medisgroups or	0.17	0.80	
		AMI—medisgroups exp	0.23	0.83	
		AMI—physiology 1	0.18	0.82	
		AMI—physiology 2	0.23	0.83	
		AMI—disease staging	0.27	0.86	
		AMI—PMC severity score	0.18	0.82	
		AMI—comorbidity index	0.06	0.70	
		AMI—APR DRGs	0.20	0.84	
		AMI—R DRGs	0.15	0.80	
		AMI—age sex interacted	0.05	0.69	

(continued)

Table 1 (continued): Risk-adjustment-model outcomes

Year	Author	Condition(s)/severity adjustment	R^2	C	ROC
1997a	Iezzoni	AMI—medisgroups	0.227	0.83	
		AMI—Physiology score	0.229	0.83	
		AMI—disease staging	0.27	0.86	
1997b	Iezzoni	AMI—PMC severity score	0.176	0.82	
		AMI—APR DRGs	0.198	0.84	
		CABG—Medisgroups	0.036	0.73	
		CABG—Physiology score	0.028	0.72	
		CABG—Disease staging	0.069	0.77	
		CABG—PMC severity score	0.079	0.8	
		CABG—APR DRGs	0.066	0.83	
		Pneumonia—Medisgroups	0.19	0.85	
		Pneumonia—Physiology score	0.149	0.81	
		Pneumonia—disease staging	0.132	0.8	
		Pneumonia—PMC severity score	0.115	0.79	
		Pneumonia—APR DRGs	0.101	0.78	
		Stroke—Medisgroups	0.265	0.87	
		Stroke—Physiology score	0.242	0.84	
		Stroke—Disease staging	0.112	0.74	
		Stroke—PMC severity score	0.101	0.73	
		Stroke—APR DRGs	0.105	0.77	
1997	Silber	Adult surgical—Medisgroups full model		0.92	
		Adult surgical—without severity score		0.83	
		Adult surgical —without everity/emergency		0.74	
1997	Pine	AMI, cerebro, CHF, pneumonia—admin			0.75–0.87
		AMI, cerebro, CHF, pneumonia—clinical			0.86–0.87
1997	Khuri	Non-cardiac surgery—10 variables		0.87	
		Non-cardiac surgery—44 variables		0.89	
1998	Polanczyk	CHF		0.83	
1999	Ansari	Prostatectomy	0.24	0.89	
2001	Austin	AMI			0.775
2003	Tekkis	Gastrooesphageal cancer			0.78
2003	Reed	CAB—Parsonnet/recalibrate		0.752–0.805	
		CAB—Canadian/recalibrate		0.693–0.755	
		CAB—Cleveland/recalibrate		0.748–0.769	
		CAB—New York/recalibrate		0.735–0.768	
		CAB—Northern New England/recalibrate		0.772–0.803	
		CAB—New Jersey/recalibrate		0.787–0.839	
2005	Geraci	CABG		0.698	
2005	Gordon	Non-cardiac surgery		0.65–0.83	
2007	Aylin	isolated CABG, AAA, colorectal			0.66–0.803

2.6.4 Over-fitting

The Iezzoni study touched on an important issue in relation to risk adjustment based on clinical parameters. Risk adjustment involves assessing the extent to which patient-level parameters—present at the point of admission—predict an outcome at a future point. The closer a risk-adjustment model is tailored to a particular condition, or to a particular clinical setting, the less likely it is be as precise when applied to other conditions or other settings. There is no intrinsic reason why a risk-adjustment method that is tailored to predict the outcome of one condition, such as myocardial infarction, should predict the outcome of another condition, such as pneumonia, because the physiology, pathology and the range of potentially beneficial interventions are quite different.

Statistically, the risk of adjusting too closely to a particular casemix, is called over-fitting. It is assessed by means of cross-validation measures, but the problem of over-fitting represents a natural ceiling for the development of clinical risk-adjustment methods for studies of mortality across a wide range of patients. Risk-adjustment methods that have been developed on specific patient groups, or within specific clinical settings, will lose precision when applied across a broader range of patients and settings. This reinforces the utility of risk-adjustment methods that make use of the more general information in administrative data sets.

One simple test for over-fitting is to divide a data set into a test set and a confirmatory set. When the model developed with the test set is fitted to the confirmatory set, if the precision deteriorates markedly with the confirmatory set, over-fitting is likely to have occurred.

2.6.5 Further comparisons between risk adjustment from administrative and clinical databases

In an important recent study, Aylin et al. (2007) compared the discriminatory capacity of risk-adjustment models for in-hospital mortality derived from an administrative data set with models based on clinical databases compiled by professionals.

The clinical databases were compiled by the Society of Cardiothoracic Surgeons, the Vascular Surgical Society of Great Britain, and the Association of Coloproctology of Great Britain and Ireland. The conditions whose mortality was recorded were isolated CABG, repair of abdominal aortic aneurysm (AAA), and colorectal excision. The administrative data set was the UK hospital episodes statistics, with the completed consultant episodes that comprised each admission merged together.

The authors calculated the c-statistic for both a simple model derived from the administrative data (just the year of procedure, age and sex), and more complete models with primary and secondary diagnoses, method of admission, Charlson index for secondary diagnoses (see below), and socioeconomic deprivation. The models derived from the administrative data sets were compared in relation to discriminatory power against the published results of risk-adjusted models using the clinical data in the database, as generated by the holders of the databases.

The results clearly demonstrated that the models based on administrative data were as successful in discriminating cases as those derived from the clinical databases. For the repairs of AAA and colorectal excision for cancer, the models based on the administrative data showed better discrimination, and for isolated CSBG, the c-statistic was only different by 0.02.

Models derived from administrative data systems have also proved to be adequately discriminatory in a study of post surgical outcomes in the Department of Veteran Affairs surgical clinical improvement program (Geraci et al. 2005, Gordon et al. 2005).

2.6.6 The Charlson Index

Although biomedical knowledge and evidence-based practice are often derived from studies of isolated clinical disorders, patients themselves will commonly suffer from a mixture of conditions. This is increasingly important as the population ages. So risk-adjustment methods need to reflect that complexity. The dilemma is that there are so many potential individual and combined clinical comorbid confounders, that some method of data reduction or simplification becomes necessary if comorbid complexity is to be included in risk adjustment for mortality or morbidity.

In 1987, Mary Charlson and colleagues (Charlson et al. 1987) published a paper describing an index—since widely known as the Charlson Index—in which groups of clinical conditions were assigned numerical weights whose additions combined to generate an interval score that predicted increasing likelihood of death over a 1 year or longer period. The original paper described a score with values from 1 to 16.

There is now a very extensive literature relating to the use of the Charlson index as a measure for predicting mortality in many settings, and it soon became apparent that it was a useful method for grouping comorbidities in hospital mortality studies (Iezzoni et al. 1996a; Polanczyk et al. 2002; Romano & Mutter 2004; CIHI 2007; Heijink et al. 2008; Aylin et al. 2007).

Although the original version was in the ICD-9 diagnostic system, it has been converted to the ICD-10, (Sundararajan et al. 2007) with no loss of precision.

Computerised systems exist for grouping secondary diagnoses in administrative data systems, such as the Australian National Hospital Morbidity Collection, into their respective Charlson group. In addition, the widespread use of the Charlson groups for the development of risk-adjustment models for hospital mortality studies makes it clear that the groups within the Charlson index are the de-facto standard method for grouping complicating conditions both for studies of specific conditions, or broad-based measures.

Although the Charlson index groups conditions into groups of increasing 'severity', and aggregates those groups into an interval score that can range from 1 to 16, most studies of hospital mortality have truncated the score. In an unpublished study of hospital mortality in South Australia in 2002 (Ben-Tovim 2002), the score was truncated at 5. In the Canadian study described above (CIHI 2007), it was capped at 2, and so on. An alternative to the identification of regression coefficients related to the score assigned to the comorbidity is to aggregate the comorbidites into their Charlson group, then insert the groups into the logistic regression, and generate a group-specific coefficient (Polanczyk et al. 2002). That was also the method used in the unpublished South Australian study (Ben-Tovim 2002). When used in that way, the coefficients cannot be applied to a different population of patients without testing for over-fitting. The Charlson index in its various guises continues to be developed as a valuable tool in risk adjustment.

2.6.7 Summary of risk adjustment and hospital mortality

When studies of comparative hospital mortality are presented to clinicians, one of two stereotypical reactions often occurs. If the hospital or service involved scores 'well', then satisfaction is taken with the outcome. If the hospital or service scores 'poorly', then doubt is likely to be expressed about the data and method used, focusing on whether the method has adequately accounted for the 'difficulty' of the institution's casemix. The discussion of risk adjustment in this section has been provided with this in mind.

Some conclusions can be drawn from the sections above. Firstly, real and substantial differences can be found between hospitals in relation to in-hospital mortality. The differences are not affected greatly by whether measurement is restricted to deaths during hospital stays, though it is better to include deaths soon after discharge.

Attempts to create a level playing field for inter-institutional comparisons have their problems. There are limits to the precision of existing risk-adjustment models. Models can be developed that have acceptable discriminatory power overall, but are poor predictors of individual outcomes. This is not simply a technical problem. As practising clinicians will acknowledge, their accuracy in predicting survival or death during any one hospital stay for an individual patient who is not clearly terminally ill is limited, even in the case of the most severe illness. Survival 'against the odds' is a driving force for much clinical effort, and there are countless patients and their families who have enjoyed extra years of life as a result of those efforts. The limits of statistical methods are the limits of our understanding of the nature of illness itself.

It is also clear by now that the early concerns about the limitations of administrative data systems are unfounded. Contemporary administrative data systems—professionally extracted and coded, with a wide variety of primary and secondary diagnoses—are an acceptable source for further study of the causes of variation in hospital mortality, and there is little difference in terms of discriminatory power between models derived from them and models derived from clinical databases
(e.g. Smith 1994). This is reassuring, because the cost and complexity of extracting clinical, or even simple laboratory, information on a large scale from existing record systems on a national scale in countries such as Australia are prohibitive. This is true even in countries such as the USA, where, as Birkmeyer et al. (2006: 417) put it in 2006:

> 'Although it is not clear whether our results would have differed if we had access to detailed clinical information for better risk adjustment, this question may be moot from a practical perspective. With the exception of cardiac surgery, clinical data for determining risk-adjusted mortality rates with other procedures are currently not on the horizon.'

Finally, however much we wish it, advanced statistical modelling will not reveal factors that are otherwise obscure. When a clinician complains that a risk-adjustment process is inadequate, or does not correspond with clinical experience, the challenge is to find a way to enable the clinician to articulate his or her concern in such a way that it is open to measurement. Until that happens, the only reasonable assumption from the work to date is that severity of illness—at least as measured by clinical databases or laboratory results—does not account for all of the differences in death rates between hospitals.

2.7 Inter hospital variation and random variation

In Section 2.4.1, it was argued that V— variation in hospital mortality rates—would be made up from three components: V_Q = systematic variations in factors that influence mortality outcomes, V_C = variations due to the clinical, demographic and casemix differences between patients present at the point of arrival in hospital, and V_R = random variations.

With methods for the computation of V_C established, the issue of random variation now has to be tackled.

If mortality is to be used as an indicator of safety and quality, then, like all indicators, it has to be reliable and valid. In psychometric practice, reliability is examined before validity. A reliable indicator may not be valid, but an unreliable indicator cannot be valid as its values cannot be interpreted.

The reliability and validity of indices of in-hospital mortality depend on the quality of measurement of relevant characteristics of hospital cases (e.g. number of diagnoses, vital status at the end of an episode of care).

2.7.1 Measurement

Although concerns have at times been expressed as to the accuracy of coding of diagnostic information within administrative data sets (Scott & Ward 2006), the extent of such disagreements in Australia at least are modest, and certainly appear to be no greater than found in the daily interactions between colleagues within the same team or discipline. Apart from diagnoses, the data elements in administrative data sets have generally been chosen because they are robust, straightforward to collect and enumerate and, in the case of the Australian National Hospital Morbidity Data collection, come with very explicit rules for their definition and tabulation. Coding audits constitute the test of inter-rater reliability relevant to assessing the utility of risk-adjusted measures of hospital mortality. Those audits commonly lead to no more than a small percentage of cases being re-coded: implying an acceptable level of inter-rater reliability (AIHW 2007).

It must be noted that although the fact of death will be accurately recorded, it is likely that there can be differences in relation to the proximate cause of death, as reported at death certification (Scott & Ward 2006). Fortunately, hospital mortality measures do not make use of the aetiological factors reported in death certificates, so that is not an issue of relevance.

2.7.2 Random and systematic variation

From Nightingale onwards, variations in hospital mortality rates have been taken to indicate variations in the safety and quality of the care provided. If hospital mortality rates are subject to large amounts of random variation, then they are outside the control of the staff in the hospital. Labelling a hospital as unsafe, when its results at any one time could vary between those considered safe and those considered unsafe solely due to chance, would be unreasonable for the staff and cause undue concern among current and potential patients. The reliability of the measure is clearly of great importance.

Random variation is present in the observation of all phenomena, though that is minimal in relation to the fact of death. The issue here is not the fact of death; it is variations in the observed death rate. Because patients are not randomly assigned to hospitals, the test that is applied to any one hospital is: 'does the observed mortality rate differ significantly from the rate that would have been expected if the patients had been treated in the 'average' hospital in the population of hospitals studied?'. Because inferences about hospitals are based on the size of the differences between the observed and expected mortality rates, the 'test-retest' question in relation to hospital mortality is whether the magnitude of differences remains similar when a hospital is studied again at a later time (assuming that the hospital's casemix did not change materially).

This question has been assessed in a number of ways: some more directly relevant than others.

A small group of studies in the 1990s (reviewed in Thomas & Hofer 1998) were conducted with the stated aim of examining the role randomness played in explaining hospital death rates (Zalkind & Eastaugh 1997: Thomas & Hofer 1999). Those studies all used broadly similar strategies, though they varied in scale and method. They all took as their starting point the assumption that variations in hospital mortality were a consequence of poor care, with poor care being identified via adherence to process measures. Then pre-existing external sources of information were used to specify the mortality implications of poor care, and these external parameters were then used to test the extent to which mortality outcomes in specific data sets could be attributed to poor quality. Simulation techniques were used to test the strength of relationships between mortality and poor quality, with Monte Carlo simulation being used to create multiple runs of the simulation equations under conditions of variation of the specified model parameters.

Because of the reliance of process measures as the measures used to infer poor quality, the studies were all on restricted ranges of diagnoses. The Zalkind (1997) study was entirely hypothetical, whereas Thomas and Hofer (1999) examined patients with CABG, AMI, stroke, pneumonia or congestive heart failure.

A careful examination of the analyses makes it clear that all those studies were in fact assessing the sensitivity and specificity of hospital mortality rates as indicators of hospital quality, with adherence to processes being the 'gold standard' against which hospital mortality was being assessed. Considered in that context, hospital mortality had low specificity in that there was a considerable risk that a hospital with a varying mortality rate might be flagged as low quality even though its quality, as measured by process adherence, was acceptable or high. Monte Carlo simulation showed that the poor performance in terms of this criterion was mainly the consequence of random variation.

Such studies are of interest, but they are of secondary importance here. They have cross-sectional designs rather than longitudinal, and so cannot measure variation over time in the absolute and relative performance of hospitals. As will be discussed later, the realisation of the extent and seriousness of adverse events during hospital care and the scale of their mortality outcomes, as crystallised in reports such as 'To err is human' (Kohn et al. 2000), and the 'Quality in Australian Health Care' study (Wilson
et al. 1995), have altered the landscape in hospital mortality studies, and challenged pre-existing assumptions about how quality is identified and assessed.

Of more direct relevance are studies that look directly at test-retest or repeated measures issues. Marshall et al. (1998) described the development of a time series monitor of outcomes for patients undergoing CABG procedures in Veterans Affairs (VA) hospitals in the USA. Their concern was that monitoring the performance of
VA hospitals by cross-sectional comparisons of performance would miss issues such as hospitals whose rates remained static although there was a general trend towards improvement, or hospitals whose results improved or deteriorated in a substantial way over time, despite the absolute mortality rates not being deviant enough to attract attention on cross-sectional study.

Implicit in such a design is the assumption that mortality rates are sufficiently predictable and stable over time that variations from the usual patterns will stand out. The study examined 11 six-month periods. The risk-adjusted mortality rates for patients undergoing CABG in 30 out of 43 hospitals were stable over the whole period, four hospitals had significantly high ratios over the whole period, and one significantly low. There was some movement in the rates for the remaining eight hospitals.

Birkmeyer and colleagues in the USA have been investigating the relationship between volumes of procedures performed, and subsequent mortality, for some time. As part of their work (Birkmeyer et al. 2006), they examined the value of historical mortality rates and procedure volume as predictors of subsequent performance on four high-risk surgical procedures (CABG, elective aortic aneurysm repair, oesophageal cancer resection, and pancreatic cancer resection).

They accessed the Medicare and Medicaid records for all patients undergoing these procedures over the period 1994 to 1997. Risk adjustment was undertaken for each procedure using the information in the Medicare data file: namely age, sex, race, admission status, socioeconomic status (defined as mean Social Security Income for the postcode of residence), and comorbidities aggregated into Charlson Index scores.

Morbidity rates in each hospital were then transformed using the t-statistic. The
t-expected mortality is the difference between the observed and expected mortality, divided by the standard error of the expected mortality. This allows an adjustment for the variance due to small sample size, and tends to dampen the extreme mortality rates observed in hospitals with small case loads, moving them towards the mean. They then divided hospitals into quintiles of mortality for the period in question. Assignments to a quintile for the period 1994 to 1997, along with procedure volumes, were used to predict mortality during the subsequent two year period 1998–1999. Predictions were per procedure, and historical mortality predicted subsequent mortality for CABG, AAA and pancreatic resections, but not oesophagectomy. Historical mortality predicted 54% of subsequent mortality in CABG (compared with hospital volume, which only predicted 9%). It predicted 35% of mortality in AAA repair, and 41% in pancreatic resection. The same analysis was undertaken for the periods 1996–1999, and 2000–2001, with similar results.

Although not a conventional test-retest study, risk adjustment in these studies renders the populations similar to each other in relation to patient-level variability over time. The location of a hospital in a mortality quintile predicts its future location for the same procedure, implying that the measure—relative risk-adjusted hospital mortality—remains stable over time. If the differences between hospitals were solely a consequence of random variation, this would not be the case, and historical mortality could not predict subsequent mortality.

Finally, a study published recently has examined this directly (Heijink et al. 2008). Heijink and colleagues examined hospital mortality in all hospitals in Holland over the period 2003 to 2005. Risk adjustment was by means of age, sex, primary diagnosis, length of stay and admission status. The analysis was confined to those primary diagnoses causing 80% of hospital deaths. The HSMR was calculated for each hospital on the Dutch National Medical Registration. Nine of the 101 registered hospitals were excluded because of insufficient registration of separation data.

The aim of the study was to assess variation within hospitals over time and between hospitals in relation to a variety of organisational and environmental factors. Only the results in relation to variation within hospitals over time will be discussed at this point.

A two-level multi-level model was constructed to look at time trends. The results showed that there was a significant overall decrease in HSMR over the period in question, and that most of the variation in HSMR was caused by variation between hospitals rather than variation within hospitals over time.

Smith (Smith, 1994)—drawing on an earlier study (Smith et al. 1991) of Medicare data of over 41,000 patients in 81 hospitals and other studies—used complex statistical reasoning to partition the variance in hospital mortality into the three components described earlier (V_C = 50%, V_R = 15%, V_Q = 35%). Although no subsequent analyses have confirmed his partitioning, it is possible to draw some overall conclusions on the fact and partitioning of variability of hospital mortality.

First, it is clear that hospitals vary substantially in their mortality rates. Second, risk adjustment using the data elements in administrative data sets provides an acceptable level of discrimination in relation to hospital-level outcomes (though not, of course, prediction for individual cases). Third, after risk adjustment, the residual variations between hospitals have a substantial systematic element and the extent of random variation is not so great as to invalidate the use of hospital mortality as an indicator.

Whatever the factors that cause hospitals to differ, they tend to persist over time. Thus, in a recent publication from the Canadian Institute for Health Information (CIHI 2007), the HSMR outcomes from a large number of named Canadian Hospitals were computed and tabulated for the three 1–year periods from 2004–2005 to 2006–2007. The HSMRs and the confidence limits for the in-hospital mortality of each hospital were reported.

Using the simple expedient of saying that an HSMR that was above 100, accompanied by confidence limits that did not cross the 100, indicated a high-mortality hospital, and a hospital whose HSMR was below 100 and whose limits did not cross 100 indicated a low-mortality hospital (and all others were intermediate): 12 hospitals were low for each of 3 years, 36 were intermediate in each of 3 years, and 10 were high for the 3 years. Fourteen hospitals shifted between intermediate and low in one or more years, and twelve between intermediate and high. Only one hospital moved between all three levels in the 3–year period: and it went from being a high-mortality hospital to low, then intermediate.

Once it is agreed that variation is a fact, that it tends to persist after risk adjustment, and that in the absence of intervention it tends to remain stable over time, it becomes meaningful to examine to what the variation may be attributed.

2.8 The relationship between variations in hospital mortality and other measures

A modern general hospital is among the most complex of all human enterprises. Thousands of staff from a myriad of professional backgrounds—deploying varied and complex technologies, faced with patients whose combinations of principal and secondary diagnoses and other care needs are effectively infinite in number—make decisions whose implications are uncertain yet which can materially influence the very survival of the patients under their care. Is it not surprising then, that the literature on what it is that influences variations in hospital mortality rates is at times confusing and hard to follow. Some things, however, are fairly clear.

2.8.1 Structural characteristics

The structural characteristics of hospitals, including their staffing, their facilities, and possibly their role as teaching hospitals, are important but inconsistent predictors of in-hospital mortality (Silber et al. 1995).

For instance, Krakauer et al. (1992), in a broad-based study of mortality of Medicare patients treated in 84 hospitals across the USA, found that hospitals with a higher proportion of registered nurses or board-certified physicians, or with a greater level of access to high-technology equipment, had lower risk-adjusted mortality rates.

New York City has municipal public acute-care hospitals, and a large number of voluntary (private) hospitals. Shapiro et al. (1994) studied mortality rates for AMI, pneumonia, stroke, head trauma and hip repair in both municipal and voluntary hospitals. After risk adjustment using a wide range of secondary diagnoses, they found that there was increased mortality in the municipal hospitals for stroke and head trauma.

In an early study from New South Wales, Corben et al. (1994) looked at the variation in risk-adjusted mortality rates between different kinds of hospitals in New South Wales. The analysis showed that there were differences in mortality outcomes between hospital types (e.g. Principal Referral, District Hospitals) but the differences were not tested statistically.

Birkmeyer is a consultant to the Leapfrog Group in the USA, which promotes evidence-based purchasing amongst funders and purchasers of health care. In a series of large scale studies (e.g. Birkmeyer et al. 2002), Birkmeyer and colleagues have explored the relationship between volumes of surgical cases treated in hospitals, and hospital mortality, They demonstrated that, within the USA hospitals studied, there is a relationship between high volumes of certain high-risk cases treated and lower levels of hospital mortality.

In the study described previously, Jarman et al. (1999) looked at mortality rates for hospitals throughout England, and found that the best predictors of variations in hospital mortality were the numbers of hospital doctors per 100 beds and the numbers of general practitioners (GPs) per 100,000 population of the population served by hospitals.

In the recent large scale study of Dutch hospitals, Heijink et al. (2008) studied the relationship between variations in HSMR and a wide variety of structural characteristics of hospitals throughout Holland. The study used a sophisticated two-level multi-level random effects model to assess within hospital variation over time (previously discussed) and the influence of structural and input factors on inter-hospital variation. In the final analysis, factors such as socioeconomic status of the

patients treated, numbers of nurses and doctors per bed and bed occupancy rate did not have an independent influence on mortality, though numbers of GPs per 10,000 occupants, and hospital type (teaching or non-teaching) did.

It is difficult to interpret the significance of the influence of community-based medical care in the Jarman and Heijink studies. Although it may be inferred that a relative lack of GPs might lead to patients who arrive at hospitals in a more severely ill state, that relationship has not been demonstrated empirically.

2.8.2 Performance, safety, quality

Performance

Since the publication of the HCFA studies in the 1980s, there has been a continuing interest in the search for measures of hospital performance. This has been fuelled by two major concerns. First, as health care has become increasingly expensive— particularly in the USA, but elsewhere also—but without clear evidence of the benefits of increased expenditure, efforts have been made to evaluate the performance of hospitals, to improve them; second, to provide guidance both to patients and to insurers or other purchasing groups, such as HMOs.

Safety

The *Compact Oxford English Dictionary* defines safe as 'protected from danger or risk; not causing or leading to harm or injury, and; (of a place) affording security or protection'. So, hospitals with relatively higher mortality rates are less safe overall than hospitals with lower mortality rates. That is self evident; when, after risk adjustment and allowing for random variation, mortality rates differ between hospitals, those hospitals with higher mortality rates afford their patients less security or protection than those with lower rates.

But this only applies at the hospital population level; and it is an increase in relative risk. It is quite inappropriate to deduce a conclusion for an individual on the basis of aggregate or population data: this is known as the ecological fallacy. Its force may be gathered from trying to deduce Sir Donald Bradman's batting average from the average for the Australian teams that he played in. The characteristics of a group may not be shared equally by all its members and, in population terms, the risks of the population at a whole are not equally shared by all its members. A patient with a particular risk profile may still be better off in a high-mortality hospital (Heijink et al. 2008), if that hospital is used to dealing with his or her condition. Furthermore, there is no a priori reason to assume that a hospital with a low-risk profile and a low-risk casemix would continue with that profile if it was faced with a higher risk case load.

Quality

What characterises quality in health care is not easy to pin down, and the relationship between in-hospital mortality and hospital quality measures is not clear. A dictionary definition of quality is that it is an essential or distinguishing characteristic. In common usage, the term tends to imply positive characteristics. What then, are the essential or distinguishing characteristics of high-quality health care?

Campbell et al. (2000) made a useful distinction between generic and disaggregated definitions of quality. A number of generic definitions of quality from fields outside health care base their definition on the viewpoint that a quality product or service is one that meets the requirements of those who use it. Thus a quality product or service is one that is fit for purpose or fit for use. Montgomery (2001), arguing from a statistical quality control viewpoint, defined quality as being inversely proportional to unwanted or harmful variability.

Within health care, the Institute of Medicine defined quality as the 'degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge' (Lohr 1991). Whatever the appeal of the generic definitions put forward by such bodies, they are hard to operationalise, and although disaggregating quality into a set of component parts emphases its complexity and multidimensional nature, the components are generally easier to measure.

A characteristic and much quoted multidimensional model is that of Maxwell (1984) who defined quality in relation to access to services, relevance to the needs of the whole community, effectiveness, equity, social acceptability, and efficiency and economy. That kind of multidimensional view is best understood in relation to a health service as a whole, rather than to an individual encounter within that service.

Donabedian has been the most influential voice in relation to quality at the level of the individual encounter. As he says in his landmark article 'Evaluating the quality of medical care' (Donabedian 1966: 163), he 'remained, by and large, in the familiar territory of care provided by physicians and has avoided incursions into other types of health care.'

Donabedian proposed that the quality of medical care be assessed in relation to three components—structure, process and outcome—of which structure has been dealt with above. Donabedian (1966: 186) recognised that outcomes validate other measures ('the validity of all other phenomena as indicators of quality depends, ultimately, on the relationship between these phenomena and the achievement of health and satisfaction') but introduced into common parlance the notion of measures of process as indicators of quality.

Process quality relates to an assessment of the interactions between clinicians and patients, and may be considered to have two elements (Schuster et al. 1998): technical process quality and care in relation to professional standards. Technical medical quality was simply described by Donabedian (1966) as 'whether what is known to be 'good' medical care has been applied'. By that he meant the skilful application of clinical care in the broad sense. It is clear that holistic assessments of that kind can only be made by judges who are themselves skilled: examining a range of information collected during an encounter. Such a strategy has come to be termed an implicit evaluation of care.

Broad-based evaluations can be distinguished from process quality as measured by process indicators. There, an assessment is made of the extent to which a specific process of care has been performed, defined either by reference to the scientific literature, or an expert panel, and deemed to represent appropriate care for a particular condition or set of circumstances. Most feasible process measures are usually indicators for a very specific element of the care process rather than comprehensive measures of how care is actually delivered (Rubin et al. 2001)—the hope being that the part is indicative of the larger whole.

It is the link between measured process and hospital mortality outcome that is most problematic. The underlying dilemma is clear.

Process measures provide direct feedback to professionals about measurable changes of practice; for example, 'the percentage of eligible cases of patients with AMI who leave hospital with evidence-based treatments that will reduce the risk of recurrence' is a measure that provides information that can be acted on.

But the link between specific process steps and overall hospital mortality is less clear because many of the factors that might affect mortality are outside the direct control of the practitioner. As Donabedian (1966: 181) puts it 'Care can be good in many of its parts and be disastrously inadequate in the aggregate due to a vital error in one component'. Nevertheless, for the patient who is the subject of treatment, the process steps in his or her care are of little direct interest—what interests the patient is the outcome and, most interesting of all, the question of survival.

So is survival the gold standard of quality, and are measures that do not correlate with mortality poor measures of quality? Or is adherence to process standards the essence of quality, and measures that do not relate to variations in process adherence inappropriate measures of quality? Although this is clearly a matter of viewpoint, it is not simply a matter of semantics.

Take the following contrasting views. The Hospital Quality Alliance is a national public reporting program in the USA—initiated by the US Department of Health and Human Services—collecting data on a set of process measures for heart attack, heart failure, pneumonia and surgical site infection prevention. As Jha et al. (2007: 1105) point out, the indicators were developed with a broad consensus from experts, and from the research literature, but their performance 'in the real world in identifying hospitals with better outcomes, such as lower risk-adjusted mortality across a number of clinical conditions is unknown'. Only if this relationship is confirmed can the measures be useful for quality improvement programs

However, in a review of a series of studies of the relationship between the Health Quality Alliance-supported process measures and hospital mortality, Shih and Schoenbaum (2007) found only a modest relationship between the measures and short-term mortality. As they say, equivocal results of this kind lead to criticism that such measures have only a limited value as tools for informing consumers about quality of care, or guiding payers seeking value in pay-for-performance programs (Horn 2006). Werner and Bradlow (2006) were similarly concerned that their findings of only a modest relationship between performance on process measures and risk-adjusted mortality rates—in a large scale study of Health Quality Alliance—supported process measures and mortality outcomes—would be inferred as meaning that the ability of performance measures to detect clinically meaningful differences across hospitals would be questionable.

A contrary view is exemplified in a recent comprehensive review of the relationship between quality of care and risk-adjusted mortality by Pitches et al. (2007: 1). The reviewers begin by partitioning mortality into patient casemix factors, random variation and a residual unexplained mortality (described as systematic variation above). The authors state that this unexplained component may 'implicate quality of care' and lead naturally to the ranking [in league tables] of hospitals with an implied correlation with quality of care. They go on to explicitly equate quality of care with adherence to existing evidence-based standards of clinical care, and seek to determine if hospitals with higher risk-adjusted mortality rates, provide poorer quality of care so defined. So, in this view, adherence to evidence-based standards of clinical care is the gold standard of quality against which mortality is assessed. This position has been

re-stated particularly clearly by Shojania & Forester (2008: 153) who state that 'for the hospital standardised mortality ratio to represent a valid performance measure, it must correlate with accepted measures of quality'.

With these basic issues in mind, it is possible to begin to look for underlying patterns in the extensive literature that has accumulated in this area. This is a partial review only: more comprehensive analyses can be found in Iezzoni (1997a), Thomas and Hofer (1998), and Pitches et al. (2007).

Firstly, there are those studies that have gone from outcome back to process: that is, risk-adjusted mortality rates have been calculated, then processes within contrasting groups of hospitals have been examined. In an early study, Knaus et al. (1986) ranked intensive care units (ICUs) on mortality outcomes using the APACHE 11, and then undertook management audits of the units. The hospital with the lowest mortality ratio had a number of structural characteristics thought to be associated with good ICU care (e.g. 24 hour cover by a unit physician) and these were in contrast to the worst performing unit, where poor communication between the unit physician and the nursing staff was also noted. The small numbers and the very subjective aspects of the management audit make it hard to draw conclusions from this study.

The issue of small patient numbers is also found in a much quoted study by DuBois et al. (1987). In a rather complex design, they first created a crude risk adjustment that made no attempt to take comorbidities into account, and used that to rank hospitals in a provider-owned chain. They then studied six of the high-mortality outliers, and six of the low mortality outliers. Case records for a total of 378 patients with AMI, stroke or pneumonia were studied. A structured review against explicit criteria (generated by a panel of experts) was conducted by one of the researchers who was a physician. That physician also dictated case summaries for the 182 patients who died during their hospitalisations.

The extracted data was used for two purposes: firstly, a severity based analysis was conducted, allowing for more sensitive risk adjustment for each primary diagnosis. The performance against the explicit criteria was also reviewed and found not to vary between the high and low performing hospitals. The case summaries were then reviewed by external assessors, who looked at the overall care provided and rated the deaths as preventable or not preventable. After risk adjustment, the high-mortality hospitals were rated as having a greater proportion of preventable deaths for pneumonia and stroke, but not AMI.

The study is described in some detail because it reveals the complexity of the methods required to undertake an implicit review. Also noteworthy was that there was only modest inter-rater reliability between the assessors in relation to the outcome of implicit review.

Similar methods were then used in studies by Best and Cowper (1994), Goldman and Thomas (1994) and Gibbs et al. (2001). In each case, the potential preventability of deaths of patients who had died in hospitals with high (Goldman and Thomas 1994) or high and low mortality rates, (Gibbs et al. 2001) was assessed by independent assessors. Although in both cases higher overall mortality was associated with deaths that were deemed more likely to have been preventable, the associations were generally modest.

Park et al. (1990) in a RAND Corporation study, used HCFA data to identify high outlier hospitals, and compared a representative sample of over 2000 patients with either congestive cardiac failure or AMI. Quality of care was examined by a detailed case note review, in which quality of care was assessed in relation to an explicit set of processes, though the processes assessed were quite broad, and included physician and nurse examination, diagnostic tests and use of therapeutic and intensive services. Although, at the patient level,

higher severity and poorer quality of care were associated with higher mortality, no hospital-level effect could be detected
(a demonstration of the ecological fallacy). Interestingly, simulation was used to assess the extent to which variations in hospital mortality could be attributed to random variation. Although that proportion was substantial, the non-random variation was statistically significant and clinically important.

A quite different approach that made it possible to overcome the problems of small sample sizes, but traded size for credibility, made use of the fact that files of patients for whom USA hospitals claimed re-imbursements were independently assessed by peer review organisations in each state (Hartz et al. 1993). The Peer Review Organisations review about one in four records. Nurse reviewers look for a specified set of quality-of-care performance problems (quite diverse and widely drawn) and, once a problem has been identified, a physician review confirms the problem or not.

Although there were modest, but statistically significant, correlations between problem rates as specified by the Peer Review Organisations and risk-adjusted hospital mortality rates, at the state level, there were major differences between the Peer Review Organisations in each state. Hence, the findings of Hartz et al. (1993) and Thomas et al. (1993), which also used Peer Review Organisation assessments, are hard to interpret.

Finally, there are a number of other studies (reviewed in detail by Pitches et al. (2007)) that use explicit review to assess compliance with process measures for one or more specific conditions in patients treated in groups of hospitals, and assess the association of process measure compliance with risk-adjusted hospital mortality for those conditions. The outcomes of these studies are in line with the outcomes of the Health Quality Alliance process measures.

The literature reviewed in this section demonstrates that the relationship between process measures and mortality outcomes is inconsistent. Further work in this area should continue to be monitored.

2.8.3 Studies aimed at changing hospital mortality rates

The recognition in recent years of the pervasive nature of adverse events during hospital care, and their mortality and morbidity implications, has begun to change the context of discussions about hospital mortality.

As the previous section demonstrates, for many years the concentration was on hospital mortality as an indicator of quality, when quality was associated with the performance of clinical practitioners in relation to what might broadly be termed evidence-based care. Do practitioners do what is thought to be necessary, or at least practise in conformity with the best evidence for what ought to be done? In that context, a good quality hospital is one that provides the right care. But as Donabedian (1966: 182) points out, the relation between structure, process and outcomes is not simple:

> 'In healthcare, each event is an end to the one that comes before it and a necessary condition to the one that follows. This indicates that the means–ends relationship between each adjacent pair requires validation in any chain of hypothetical or real events. This is … a laborious process. More commonly… the intervening links are ignored. The result is that causal inferences become attenuated in proportion to the distance separating the two events on the chain.'

There are very many steps between a specific process measure (giving aspirin on arrival in hospital for patients with AMI) and overall in-hospital mortality. And studies are now emerging that describe hospitals' efforts to reduce overall mortality rates directly, rather than looking solely at specific process steps.

In 2000, the Walsall Hospital NHS trust had a HSMR of 130: the highest of all acute hospitals in England. In response, seven clinical governance groups were formed to implement changes across the whole range of clinical disease areas, together with a wide variety of management areas including bed management, information services, discharge liaison, integrated care pathway development, and many others. By the end of 2004, the HSMR had dropped to 92.8 (Jarman et al. 2005).

It could be argued that what was accomplished here was no more than statistical regression to the mean, or a more causal effect resulting from public scrutiny causing a poorly performing hospital to get back into line (akin to the 'Hawthorne effect'). Any change, no matter what its impetus, would have had the same outcome.

The case study of the Bradford Teaching Hospitals Trust (Wright et al. 2006) is particularly interesting in this light. The Trust is a large (1200 bed) acute service which, in 2002, was a low mortality Trust in terms of HSMR. Nevertheless, in 2002 the Trust chose to focus on hospital mortality, with a commitment to eliminate all unnecessary hospital deaths. The program of work that followed was very diverse. Following a review of a consecutive series of hospital deaths, a high prevalence of sub-optimal clinical observations, medication errors and hospital infections was noted amongst the patients that died. A wide variety of corrective actions were initiated in all relevant areas. Also, a monthly monitoring program for hospital deaths using a statistical control chart for hospital mortality was begun. (Statistical control charts are discussed further on in this report.)

In Bradford, the effect of the mortality reduction program was to significantly reduce the hospital HSMR from 94.6 at the start of the program to 77.5 three years later. The Bradford Trust began its work after enrolling in an Institute of Healthcare program: Improvement Partnership for Hospitals. Gilligan and Walters (2008) described the experiences of the East Lancashire Hospitals Trust (Royal Blackburn Hospital) following enrolment in that same program. Their focus became improving the flow of patients through the hospital by a combination of activities including changes to medication charts and physician rounds, redistribution of bed stock and the introduction of a critical-care outreach service, plus intensive monitoring of outcomes using control charts. The Trust was never a high-mortality outlier—though its HSMR was above the national average—but over the period of the study, the HSMR declined substantially.

The large-scale 100,000 Lives Campaign initiated by the Institute for Healthcare Improvement is aimed directly at hospital mortality: reducing mortality by a series of broad based improvement strategies, rather than through the medium of adherence to narrow-focused process measures for specific conditions. The strategy is not without its critics (Auerbach et al. 2007), but is defended indirectly in Berwick (2008).

2.8.4 Summary

The performance of hospitals will continue to be scrutinised, and measures will continue to be devised to open up the historically rather hidden world of hospital outcomes to external inspection. Mortality is one such measure, and although its status as a measure of safety

seems secure, its role in specifying hospital quality is subject to the difficulties and ambiguities inherent in the concept of quality.

Feasible and reliably measurable process measures tend to be of very specific elements of care, and there are likely to be very many unmeasured steps between such process elements and the survival or death of a group of patients—as the mortality reduction programs described above infer. Furthermore, there is a logical fallacy at the heart of referring back from mortality to quality when (and if) quality is defined in relation to performance levels on a set of specific process measures. If quality is synonymous with p, and p→m (mortality), that does not necessarily mean that p←m, because m is not identical to p. All cherries are red, and all cherries are fruit, but this does not mean that all red fruit are cherries.

Another way of looking at this is the fallacy of composition. The fallacy of composition is committed when a conclusion is drawn about a whole, based on the features of its constituent parts, when there is no justification for drawing this inference. For example, every player on the team is a superstar, so the team is a great team. This is not necessarily so, because the superstars will not necessarily play together well, and so could form a very weak team. Teamwork is a quality of interaction and not a matter of simple addition. Similarly, in a hospital, individual staff may perform specific process measures with great accuracy, but modern health care depends as much on teamwork as individual competence, and the health-care team as a whole may perform poorly (Lemieux-Charles & McGuire 2006), and so increase mortality risk, despite the team being made up from conscientious and caring practitioners.

2.9 Presentation of information about in-hospital mortality

2.9.1 Methods of presentation

Variations in hospital mortality rates are analysed and disseminated in an effort to influence the recipients to look further at health-care practice. Thus the mode of presentation of the results of analyses of mortality rates is of considerable interest.

Goldstein and Spiegelhalter (1996) have provided a review of issues in this area. Drawing on examples from education and health care, they make the important point that comparisons must take context into account. Risk takes account of patient characteristics at entry into a hospital, in the same way that comparisons of schools performances should take account of the status of the children on arrival at a school. Goldstein and Spiegelhalter (1996) further argue that the need to contextualise does not stop at the institutional level, but needs to be considered at state and national levels.

So, accepting that simple comparisons of mortality rates that are not risk adjusted will almost always be confusing, there are only a small number of practical alternatives for presenting the information.

As previously described, risk adjustment of hospital mortality always involves a comparison between observed and expected mortality rates for a set of institutions or services. Although a number of different ways of generating indices from this comparison have emerged over the years (starting, in Australia, with an interesting early paper by Duckett & Kristofferson

1978), the HSMR has emerged as the standard in this area, and so is the index of hospital mortality discussed further.

Institutional HSMRs can simply be listed. But any single HSMR needs to be accompanied by a measure of the uncertainty of the value. The conventional method of doing that in scientific biomedical practice is to calculate the confidence intervals around each HSMR, usually using 95% confidence limits (e.g. CIHI 2007). The 95% confidence limits represent the range within which a particular parameter will be found 95% of the time on repeat testing of a population, so the width of the confidence limit gives an indication of the uncertainty or precision of the parameter. Wide confidence limits commonly occur when sample sizes are small.

The Canadian National Study of HSMRs, referred to above, simply listed each participating institution, together with its HSMR and the confidence limits. It made no explicit inter-institutional comparisons: leaving that to the reader.

League tables

League tables in which hospitals are ranked in relation to their particular HSMRs, are an explicit method of providing inter-institutional comparisons. The Dr Foster group in the UK has provided several non-peer-reviewed reports in which hospital are ranked according to their HSMRs.

Typically, most hospitals in a country have HSMR values that are quite close to one another, especially after adjustment for casemix. League tables tend to encourage unwarranted emphasis on small and unimportant differences in the rates, because they can translate into large differences in the ranking of hospitals with similar rates.

League tables are improved by the addition of confidence intervals. But no matter how much effort is put into explaining uncertainty and variation, it is hard not to assume that being 24th in a table of institutions ranked from 1 to 100 really means that the institution in question is superior to the 25th institution, and much superior to the 35th, even if all of those institutions have overlapping confidence intervals and cannot be said to differ significantly. So, whatever their attraction, from a statistical and epidemiological viewpoint the presentation of HSMRs in a simple league table format is hard to support.

Caterpillar plots

Another method is to present HSMRs (Goldstein and Spiegelhalter 1996) in the form colloquially known as 'caterpillar plots', in which the HSMRs and their confidence intervals are represented as a graphical plot, with individual institutions ranked by HSMR along the x-axis and the HSMR values shown on the y-axis.

The two caterpillar plots below (Figure 2 and 3) are included here to illustrate this type of presentation. Each of the plots summarises data for one peer group of hospitals. The analysis underlying these plots has applied the same adjustment model to both peer groups (rather than analysing each group separately). Hence, it is computationally valid to compare the HSMR values in each of these plots. The values for hospitals in the A1[2] group are spread fairly equally above and below 100 (Figure 2). In contrast, the values for B1 hospitals are mainly below 100 (Figure 3). However, the interpretation of this difference between peer groups is complicated by their different casemixes. Adjustment for casemix based on data available in administrative data allows for part, but not all, of the difference. An apparently

[2] See Table 2 for information on the types of hospitals included in the peer groups.

low-risk group of hospitals will only be low risk for their casemix, not the casemix of larger hospitals.

The extent to which low-risk populations, as well as low-risk hospitals, provide an important opportunity for analysis is discussed in Coory and Scott (2007).

The examples of caterpillar plots presented in this section are typical of those in the literature. There may be potential to improve the performance of this type of plot as a graphical method to convey information about in-hospital mortality. We present and discuss some variations in Appendix 4.

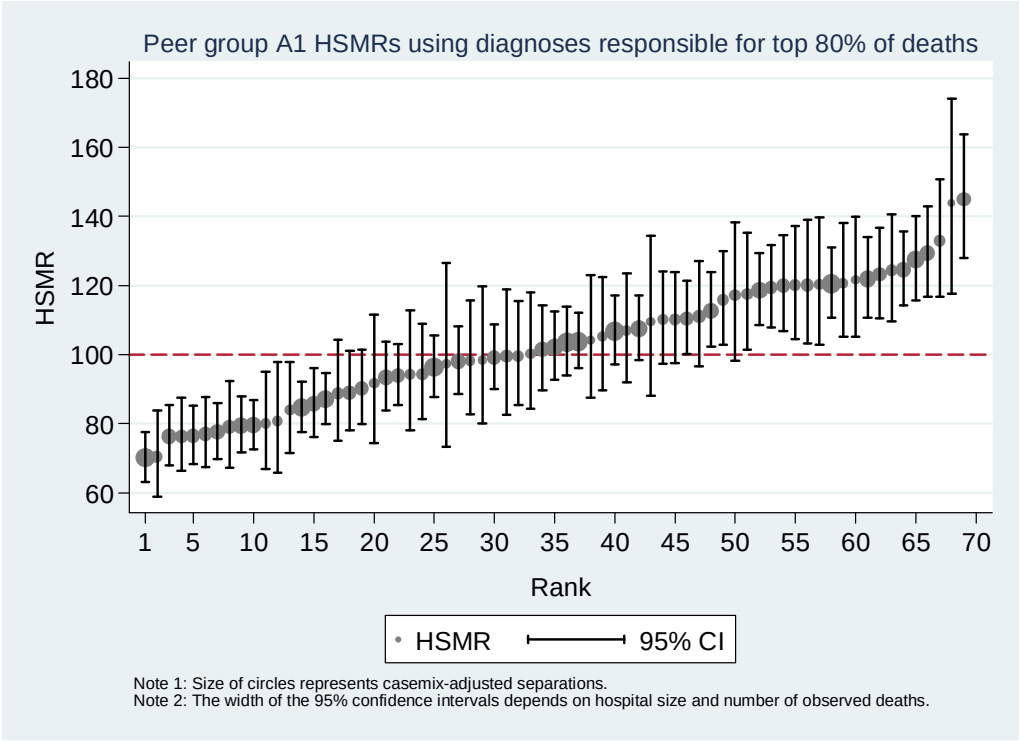


Figure 2: Caterpillar plot of variation in point estimates in HSMR for peer group A1

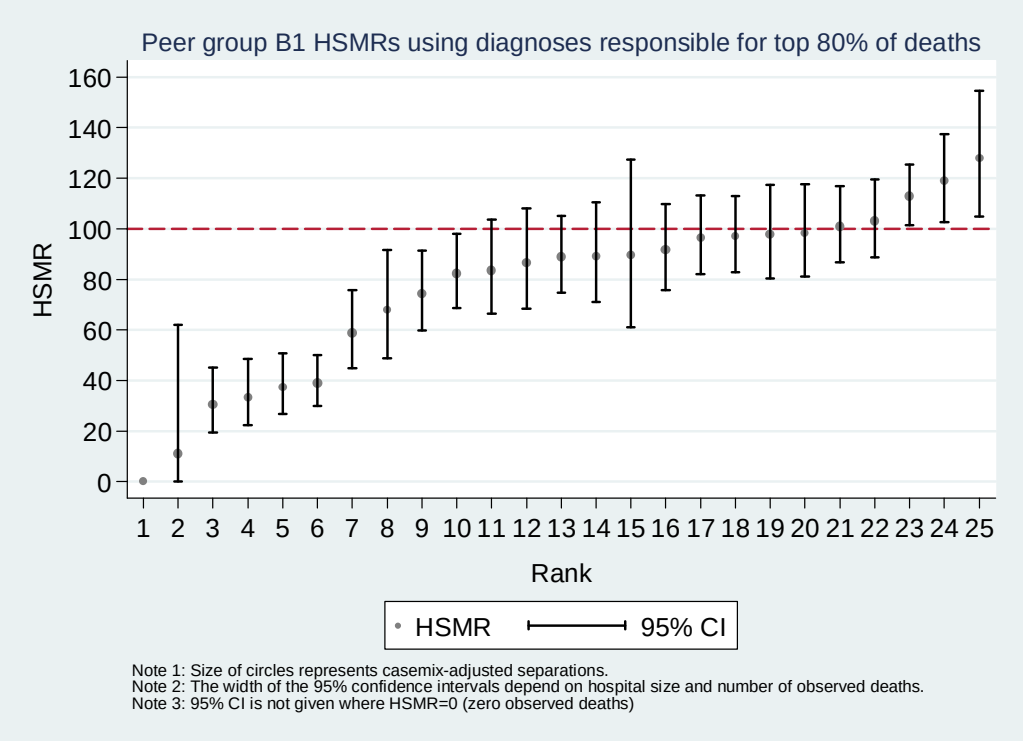


Figure 3: Caterpillar plot of variation in point estimates in HSMR for peer group B1

The obvious question is 'when is a difference between institutions important?' When the lower confidence limit of the estimate for any an institution is above the population average of 100, or the upper confidence limit is below 100, then that institution differs statistically from the population average. When a HSMR is so deviant that the institution not only fulfils the above criterion but is say 15% above or below the average, some analysts would declare the institution to be an outlier. Some would set even more rigorous criteria against which to assess outlier status and some would not set an outlier standard at all, but would just identify institutions at extremes. There is no absolute standard here.

It is also the case that when the confidence intervals of two institutions do not overlap, they are deemed to differ to a statistically significant extent from each other, and that is helpful when undertaking inter-institutional comparisons for sub-samples of institutions that appear at very low (or very high) risk overall—at least in relation to HSMR.

The results of the analysis of the Australian data sets are presented later in the form of a series of caterpillar plots, and their utility can be gauged from those presentations.

Funnel plots

A relatively recent innovation in the area of the analysis and presentation of HSMRs and other hospital performance indicators is the use of funnel plots, which were extensively developed by the Medical Research Council Biostatistics Unit in Cambridge in the UK (Spiegelhalter 2002: 2005) and are now coming to be seen as potentially preferable to caterpillar plots (e.g. Mohammed & Deeks 2008).

In the context of institutional comparisons, a funnel plot is an extension of a Shewart chart, or a statistical control chart. It is a method for detecting when a particular institutional outcome on a parameter, such as the HSMR, is so extreme as to constitute a potential case of 'special-cause' variation. This means that the variation is so great that it is outside the bounds of the underlying, or common, cause variability that is present in the usual outcomes of the parameter in question. When a control chart 'signals' special-cause variation, an investigation into potential causes should follow.

The method of computation of funnel plots is quite complex, although the results are presented in an easy to assimilate graphic. A relatively straightforward description is provided by the Dr Foster group in a recent non-peer reviewed (Dr Foster Intelligence 2007).

> 'Funnel plots (or control charts) are a graphical method used to assess variation in data and are used to compare different trusts over a single time period. These plots (HSMR funnel plots) show the position of each trust's HSMR. Control limits form a 'funnel' around the benchmark and reflect the expected variation in the data.
> [The wide base of the funnel demonstrates that as the number of separations involved fall, the size of the expected variation increases because the measure is less precise].

Each chart has five lines:

- a centre line, drawn at the mean (the national average RR=100)
- an upper warning line (upper 95% control limit)
- an upper control limit (drawn three standard deviations above the centre line-upper 99.8% control limit)
- a lower warning line (lower 95% control limit)
- a lower control limit (drawn three standard deviations below the centre line–lower 99.8% control limit).

Data points falling within the control limits are consistent with random or chance variation and are said to display "common-cause" variation. For data points falling outside the control limits, chance is an unlikely explanation, and hence they are said to display "special-cause" variation.'

Further discussion of methods of presentation is delayed until after the results of the Australian study are provided.

2.9.2 Public or private dissemination of mortality outcomes

There has been lengthy discussion over the years as to the legitimacy of public reporting of mortality data, in comparison to dissemination solely to the institutions themselves. The issue will not be discussed at length here for several reasons.

First, public dissemination of performance indicators is an area that has been comprehensive reviewed on a number of occasions, and there is little to add to recent reviews (e.g. Fung et al. 2008; Hibbard et al. 2005).

Second, the overall results are fairly clear. Public reporting has, at best, a modest impact on the public at large, but it has a more definite impact on providers of care: tending to increase improvement activities of a variety of kinds. It is not without its hazards however (Hibbard et al. 2005).

Third, public reporting of mortality, as well as many other outcomes, is already so widespread as to be the norm in the USA, the United Kingdom and Canada, and in the UK will become increasingly so if the reforms recently advocated by Lord Darzi are enacted. In Australia, the Queensland Measured Quality reports, first produced in 2004 (Queensland Health 2004) provide very detailed mortality and other information about the hospitals in Queensland, and the reports have been elaborated in various ways since then.

Finally, there will always be a necessary tension between the desire of establishments to maintain a good reputation and a public right to know. Media reports based on publicly available information have not always presented a completely accurate, or necessarily fair, representation of institutional or even personal outcomes. Public reporting does, however, guard against a tendency to withdraw support from analyses that may be seen as a source of embarrassment or distress—even if they are accurate—but it also places an obligation on the reporter to stringently guard against bias and misrepresentation.

2.9.3 Future developments of note

As well as providing a review of existing work, we have also been asked to comment briefly on any noteworthy trends in data gathering or analyses. We would say that the two most promising developments that will be implemented in the near future are the decision to require national coding of 'present on admission' indicators for all secondary conditions in the Australian National Hospital Morbidity set, and the wider application of data linkage. Some methodological developments also hold promise.

Present on admission indicators

One of the challenges for risk adjustment of performance indicators is a health-care version of the moral hazard problem. Coders are required to code complicating or comorbid conditions, irrespective of whether they were present at admission or occurred after admission. Some of those secondary conditions may have been the result of problems that occurred as a result of sub-optimal care. To risk adjust for them is to provide an allowance for poor-quality care rather than to reveal it by comparison of outcomes. For example, patient X had a presenting problem of severity Y, and was at low risk of death; having had a series of falls and a surgical site infection, he is now ranked as a high-risk patient, and his death is partially discounted for that reason.

One way to capture this in hospital data is to attempt to record which conditions were present on admission. 'Present on admission' codes require a coder to judge whether a secondary condition was, or was not, present on admission, and are mandatory for Australian public hospital-coded separation data from the beginning of financial year 2008–09. Present on admission coding has been practised for some years in California, and a recent study by Glance et al. (2008) demonstrates that present on admission coding is likely to considerably enhance the precision of mortality measures. Present on admission coding (known as C-codes in Victoria) has been in place in Victoria for some time, and a study by Ehsani et al. (2006) has shown that it is similarly informative there.

Data linkage

A second useful development is the increasing availability of data linkage. Two forms of linkage are relevant. The first is linkage within hospital morbidity data. Some people—particularly those with serious and persisting conditions—are likely to experience more than

one episode of in-hospital care within the period covered by a study of in-hospital morbidity. Hospital inpatient administrative data files have generally been organised in a way that includes a record for each of these 'separations', but does not provide a good basis for grouping together the set of records referring to a particular person or reason for admission. Without this form of linkage, it is not possible to be sure whether a person whose episode of hospital care ended with transfer to another hospital, or with a 'statistical type change', died during the next episode of inpatient care. Even a person who separates with discharge home might have been re-admitted soon after, with the possibility of fatal outcome of that episode.

The second role of data linkage relevant to this type of work is linkage between hospital records and death registers (or the National Death Index). This is necessary to enable studies that include deaths soon after discharge.

Health data linkage systems also have potential to be used to assess individual health status over time. Such information might be found to improve risk adjustment.

Developments that enable such linkage are well-established in some parts of Australia (notably Western Australia and New South Wales) and are being put in place elsewhere (e.g. South Australia), but there is not yet a routine capability to enable the necessary linkage at national level.

Analytical methods

From a methodological point of view, the issue of the development of Bayesian regression models for use in large scale mortality studies (e.g. Austin 2008) is of interest, but will require further study. The approach has potential for analysis of smaller hospitals. Bayesian techniques have a number of adherents in the field of performance measurement, but the techniques can be complex and are not without controversy, and will require quite detailed assessment and testing before their strengths and weaknesses can be assessed in this context (Paul Aylin personal communication, 2008). Nevertheless, this approach is sufficiently promising to warrant exploratory use and further development.

Further developments in statistical process control methods for immediate monitoring of mortality and other performance measures are also clearly an area of great interest however (Duckett et al. 2007)

2.10 Conclusions

In 2006, Scobie et al. (2006)— drawing on the work of the National Performance Committee (NHPC 2004)—provided a useful set of criteria against which to assess the potential utility of a candidate health performance indicator. Those criteria can be used to assess variations in hospital mortality as a candidate indicator of hospital performance.

Scobie et al. (2006) stated that an indicator should:

1. **Be worth measuring—it should represent an important and salient aspect of the performance of the health system**. It is hard to argue against variations in hospital mortality on those grounds.
2. **Be measurable for diverse populations—the measure should be valid and reliable for general populations and the diverse populations in Australia.** Variations in hospital mortality rates are relevant to all populations studied, and are reliably reported.

3. **Be understood by people who need to act.** The fact of variations in mortality is readily and immediately understood. The remedial actions are less clear.
4. **Galvanise action—The indicators are of a nature that action can be taken at the national, state, local or community level by diverse groups of individuals.** Once the fact of variations in mortality are acknowledged, then actions take on some urgency, though, again, this is at an early stage and the necessary roles of the various levels in the health system are not yet clear.
5. **Be relevant to policy and practice.** Although the policies and practices that will directly focus on mortality reduction are yet to be generally agreed, the speed with which institutions have taken up the creation of medical emergency teams as a mortality reduction measure indicates that remedial actions can be developed and implemented on a wide scale.
6. **Reflect results of actions when measured over time.** The studies described earlier demonstrate that.
7. **Be feasible to collect over time.** This is clearly possible.

Variations in hospital mortality appear to fulfil all the necessary criteria to qualify as a performance measure. The more pressing question is 'how they should be used?'

The uncertainty surrounding the relationship between variations in hospital mortality and other measures of hospital structure and process mean that, in our view, variations in hospital mortality should be viewed as screening tools, rather than being assumed to be definitively diagnostic of poor quality. A screening tool is a signalling device. It is intended to signal that a problem may exist and that further detailed investigation is required.

With a screening tool, some lack of precision is accepted, because being too cautious in sounding a warning risks ignoring a problem in its early stages, when it may be more open to change. So, because of the uncertainty in the interpretation of mortality rates, it is inappropriate to use variations in hospital mortality to assert with confidence that a high-mortality hospital provides poor-quality care. That is a premature rush to judgment. High relative mortality should be seen as a prompt to further detailed investigation. The issues were well summed up by Donabedian (1966: 196). 'A final comment concerns the frame of mind with which studies of quality are approached. The social imperatives that give rise to assessments of quality have already been referred to. Often associated with these is the zeal of the social reformer. Greater neutrality and detachment are needed in studies of quality. More often one needs to ask "What goes on here?" rather than "What is wrong; and how can it be made better?" '

3 Measuring in-hospital mortality in Australia

3.1 Current in-hospital mortality reporting in Australia

In recent years, numerous studies have been published describing mortality rates calculated for Australian deaths in hospitals for a variety of conditions and using a number of different methods. Much of this work is carried out by academics and physicians—sometimes in collaboration with government health departments. These types of activities are generally reported in the public domain within peer-reviewed journals (National and International) or as government-badged published reports. A number of examples of this type of work have been cited in the current report.

In addition to what appears in the public domain, an unknown amount of work on in-hospital mortality is effectively hidden and commonly referred to as 'grey literature'. Grey literature refers to materials that are either unpublished, or published but not in the peer reviewed literature. Such material is typically produced by governments, business or industry, and can include government reports, technical reports, white papers, or position papers.

A classic example of grey literature with high relevance to the current topic is the emergence of 'quality reports' that were produced in 2004 for every Queensland public hospital, but not available to the public. These reports came to light during the Bundaberg Hospital inquiry (Van Der Weyden 2005). Currently, the Queensland Government is regularly publishing in the public domain a number of indicators, including condition specific in-hospital mortality indicators (e.g. *Moving ahead*, *Queensland Public Hospitals Performance Report 2006–07* (Queensland Health 2007)).

To the best of our knowledge, no other jurisdiction in Australia publicly reports in-hospital mortality data in this way.

We are aware that each state and territory has developed advisory bodies that examine a variety of elements of patient safety within hospitals. Many, if not all, cover the reporting of adverse events (Wilson & Van Der Weyden, 2005), but only some of this work reaches the public domain.

The Commission may be aware that recently the Health Round Table (a privately owned, not-for-profit organisation), which is voluntarily provided with regular extracts from hospital morbidity collections by a number of Australian hospitals, has undertaken a series of analyses of in-hospital mortality using that data. It is currently feeding that data back to those hospitals. However, the methods used by that organisation are by their nature proprietary, and the outcomes not subject to any further scrutiny. The work of the Health Round Table is likely to promote further interest in this issue, both from the hospitals who subscribe to the Health Round Table, and others.

Each state and territory contributes hospital separations data to the AIHW for collation. The data elements provided by the jurisdictions are governed by National minimum data

requirements. Beyond these minimum requirements, the data items collected by each jurisdiction can range in number and complexity. For example, Victoria has been the only jurisdiction collecting data on comorbidities present on admission for several years (colloquially known as C-codes). The availability of additional data items to individual jurisdictions means that the types of variables that can be used in risk-adjustment models will vary according to each state and territory. Additionally, the types of models used to calculate in-hospital mortality may also vary.

The present review of the literature has revealed one commonly-used method for calculating in-hospital mortality. Variations in the inclusion of factors to be used for risk-adjustment have been described and the results of these variations discussed. It is likely that any recommendation for a single method of calculating in-hospital mortality will create discussion among the jurisdictions regarding whether the recommended method is as appropriate or sophisticated as the variety of methods employed to date by individual researchers, or by health departments or individual hospitals.

It is important that any method that is singled out as the basis for creating a National indicator of in-hospital mortality is replicable by individual researchers, jurisdictional health departments or individual hospitals. By basing an indicator on currently available national minimum standards governed administrative data, such as the National Hospital Morbidity Database (NHMD), the ability of the various stakeholders to validate and replicate in-hospital mortality rates is assured.

3.2 Mortality rates in Australian hospitals

The second part of this report is an analysis of Australian hospital mortality data so as to demonstrate is suitability as the basis for measuring in-hospital mortality and to show what National indicators of hospital mortality can be developed now, and in the future.

The analyses that have been undertaken have been mindful of certain considerations.

1. Although the analyses has been conducted entirely on Australian data, we recognise the importance of (a) allowing for comparisons between components of the Australian health system as a whole with other health systems, and (b) for enabling analyses to be 'rolled down' to the hospital or health-service level and 'rolled up' to state or other jurisdictions.
2. These analyses are provided by way of demonstration. The aim has been to provide the Commission with worked examples of analytic methods, scope of analyses and methods of presentation, so that the Commission can make an informed choice not only on whether to use mortality rates for reporting purposes, but, if so, how those rates might be presented.

3.3 The analytic strategy

The analytic strategy adopted was based on the outcomes of the literature review. The review made it clear that there is an emerging international consensus on best practice for national studies of hospital mortality. Those studies have the following characteristics.

- Observed in-hospital mortality rates are determined from existing nationally mandated administrative hospital morbidity data sets.

- The information contained within those data sets is used to risk adjust those rates.
- When possible, mortality is studied up to 30 days after discharge from hospital, but when linkage with births and deaths registers is not feasible, deaths during hospital stay are an acceptable end point.
- Risk adjustment is by way of logistic regression and indirect standardisation, which is used to calculate expected mortality rates.
- Those expected rates become the denominators of the ratio of observed to expected outcomes (O/E). A ratio value less than 1 is favourable and a ratio of greater than 1 unfavourable.
- When the ratio is multiplied by 100 the convention is to describe result as the HSMR (Jarman et al. 1999).
- HSMRs are presented in a variety of ways: as tables; as caterpillar plots; and, more recently, by way of funnel plots.
- Multi-level modelling has begun to be used to look at intra-hospital and inter-hospital variations in HSMR over time.

Our analyses were framed by a number of specific methodological concerns related to the implementation of the general approach described above. The most pressing were what variables to test and choose for risk-adjustment purposes? What proportion of total deaths in hospital to choose for analyses? What kind of model development process should be undertaken? How should institutional differences be taken into account? How should HSMRs be presented?

Our approach to each of these issues is outlined before a detailed presentation of the methods and results of the analyses. Our overarching strategy was as follows: having the literature review in hand and international practice identified, we came to the view that the recently released study of mortality in Canadian hospitals (CIHI 2007) was particularly relevant for our purposes.

The Canadian method is consistent with those used by the Dr Foster group in the UK and in the Dutch study (Heijink et al. 2008). Although the Canadian hospital system differs in many ways from that of the Australian system, it does not suffer the fragmentation found in the USA, and the lack of a national data system other than the Medicare patient care group. Unlike the USA, Canada has moved to using a clinical modification of the ICD-10, as has Australia. There is a nationally consistent approach to gathering morbidity data in the Canadian study that benefits from familiarity with coding in the context of diagnosis-related grouping and does not suffer from the problems with linking consultant completed episodes that can make comorbidity risk adjustment problematic using data from the British National Health Service (NHS).

The documentation of methods provided by the Canadian Institute for Health Information (CIHI) is noteworthy for its openness and comprehensiveness. The CIHI had previously been contacted by the AIHW and expressed a willingness to provide further information about their method if required, but the quality of their documentation has meant that, as yet, it has not been necessary to take them up on their offer.

Taking all that into account, we decided to base our initial model building on the example provided by the Canadian study, at least in relation to the choice of variables and confounders for testing within a regression model, and for the regression model building process. In that way, the Commission will have access to an example of HSMR creation that is broadly comparable with that used in Canadian, Dutch and

UK studies—all of which have themselves been strongly influenced by the methods developed by the Dr Foster group in the UK.

Later, we describe an exercise in which we built an Australian risk-adjustment model using a somewhat more refined process than the modelling exercise described in the Canadian study documentation. The pros and cons of using a more analytically sophisticated model, which differs from the model that is currently used most widely internationally, are discussed further in the conclusions.

3.3.1 Cross sectional and longitudinal analyses

The main body of the work is a cross-sectional analysis of one year of national data covering the period 1 July 2005 to 30 June 2006. This is based on logistic regression modelling. As noted above, we used an analytic approach closely modelled on current Canadian practice, which is similar to methods used in UK and Dutch work.

As discussed in the literature review, longitudinal studies are emerging as a valuable way to assess data, as well as for investigating the presence of trends in mortality outcomes. We used a two-stage method similar to Heijink et al. (2008), in which logistic regression modelling (as above) is followed by multi-level modelling.

3.3.2 Observed mortality

Observed mortality was confined to deaths in hospitals. Had data been available for this project which included deaths during the 30 days post-discharge, then they would have been used as well. While the availability of such data would have been preferable, our assessment of the literature led us to conclude that it was safe to proceed with an analysis of in-hospital mortality alone (see Section 2.4.2).

3.3.3 Choice of variables for risk adjustment of expected mortality rates

The variables tested for the purposes of risk adjustment were all derived from the hospital data set that were provided by the AIHW. The study of the Canadian data, and the existing literature, made it clear that the variables to be included needed to cover principal and secondary diagnoses, demographic information, modes of admission and length of hospital stay. A small number of exclusion criteria were applied, including admissions for palliative care, neonates (there are considerable difficulties in Australia with issues around coding of qualified and unqualified new-born babies, which makes the identification of denominators problematic) and patients who discharged themselves against medical advice and so did not complete the hospital care judged necessary by their treating doctors. We also tested the value of adding a measure of social deprivation (see *Appendix 5 Data issues* for information on how socioeconomic indicators for areas (SEIFAs) were derived) related to the usual place of residence of patients, but did not include it in our final model (see below). We found that the addition of a SEIFA measure did not materially add to the discriminatory power of our risk-adjustment model, and it makes international comparisons problematic, because social deprivation is measured quite differently in different countries.

3.3.4 What proportion of total deaths in hospital should be chosen for analysis?

The majority of international studies of hospital mortality have been confined to a subset of all primary hospital diagnoses. The underlying rationales for choosing high-risk groups of one kind or another have been discussed in the literature review. However, the recent trend in the international literature has been to confine analyses to the diagnoses assigned most often to cases ending with death in hospital and which account for 80% of in-hospital deaths in the population of interest. We followed that practice. We found that 68 three-character Principal Diagnosis ICD-10 codes accounted for 80% of in-hospital deaths in the 2005–06 NMDS data set (see *Appendix 1*). About one-fifth of records had one of these 68 Principal Diagnosis codes.

We also analysed the complement of the first group—that is, cases with any Principal Diagnosis code except the 68 that were most frequently present in records of in-hospital deaths. By definition, this second group includes 20% of in-hospital deaths. It comprises about four-fifths of all records. Because of the large number of principal diagnoses involved, this required a somewhat different approach to risk-adjustment modelling, which is described in Section 4.5.2. The method presented there is a novel contribution to the analysis of in-hospital mortality, and may be of interest to others.

Thirdly, we analysed the whole set of data meeting the study criteria. This is the sum of the first and second sets. By definition, it includes all records and all deaths.

3.3.5 What kind of model development process should be undertaken?

Most of our analyses were undertaken after a risk-adjustment-model building process, in which we followed the strategy adopted for the Canadian study (CIHI 2007). We refer to this as the risk-adjusted Canadian referred mortality model (RACM) and model parameters are presented for that model.

We also undertook another model-building process that includes a more sophisticated approach to variable preparation and inter-action analysis in the logistic regression model. We refer to that model as the elaborated risk-adjusted mortality model (ERM). The ERM model is described fully and some comparisons with the RACM model are included in Sections 4.8 and 5.8. Although the ERM model has performance advantages, we have opted to use the RACM model for the main part of the report, because it is relatively well-established in the literature. The Commission may wish to consider how the ERM model might be used. A comprehensive analysis using the ERM model has not been undertaken, though it would be straightforward to complete such an analysis if required.

3.3.6 How should institutional differences be taken into account?

The importance of making inter-hospital comparisons only within groups of like hospitals was emphasised in the literature review. There is a well-established peer grouping process for Australian public hospitals, supported by the Commonwealth, and based on a hospital peer group classification developed by the AIHW. Although originally peer grouping was simply by reference to volumes of activity, the process has been somewhat refined, and is described as follows (DoHA 2007: 1)

'Although not specifically designed for purposes other than the cost per casemix-adjusted separation analysis, the peer group classification is recognised as a useful way to consistently categorise hospitals for other purposes, including presentation of other data.'

The AIHW national peer group classifications are determined using several criteria:

- size of hospital determined by number of acute casemix weighted separations and actual separations
- demographic characteristics of major patient groups; e.g. women and children, Aboriginal and Torres Strait Islander status
- teaching and research status
- proportions of acute, rehabilitation, palliative care and non-acute patients treated.

The inclusion criteria and code numbers for these peer groups are shown in Table 2.

Although the model building exercise made use of all available Australian hospital separations, hospital level results for public hospitals are displayed within peer groups, as described above.

Equivalent grouping was not available for private hospitals. Indeed, many private hospitals are not separately identified in the NHMD. Hence, analysis of private hospitals was not undertaken in this project.

3.3.7 How should HSMRs be presented?

The HSMRs produced by our analysis are presented as tables, caterpillar plots and funnel plots.

3.3.8 Confidentiality

Due to the sensitive nature of the work undertaken for this project, we have taken two steps in order to secure the confidentiality of individual institutions:

1. Establishment identifiers have been replaced with study-assigned identification codes.
2. The HSMR values present in the results section for the single-year analysis have been adjusted using a recalibration process in order to mask their true value. The recalibration relates to the process of identifying deaths in all patients receiving palliative care. The effect is to produce values that serve the purpose of demonstrating the operation and performance of the methods, and the distributions of HSMRs, but provides institution-specific HSMR values that differ somewhat from the values that would be obtained when applying the methods without recalibration.

These two steps have the effect of masking individual institutions and preventing other parties from attempting to apply the model to their own institutional data in order to try to compare themselves with other institutional HSMRs present in the report or presented elsewhere. The recalibration was applied only to the production of the HSMR values for the single-year analysis. All other analyses (e.g. discriminatory and explanatory power, goodness of fit) were carried out on unmasked data. Details of the recalibration can be made available on request.

We recognise that individual HSMRs will be of high interest to institutions and other interested parties. However, at this stage of the process, it is important that the focus remains on the method and means of presentation rather than the actual HSMR values. As mentioned previously, a number of institutions who are members of the Health Round Table, are currently in possession of hospital mortality data for their own and other hospitals. The recalibration of HSMR values here means that comparison with results provided by bodies such as the Health Round Table will not allow confident identification of particular institutions.

4 Methods

4.1 Data

National hospital separations data were provided by the AIHW from the NHMD.
A separation is defined as: 'A formal or statistical process by which an episode of care for an admitted patient ceases' (AIHW 2005).

This report uses data for hospital separations that occurred in Australia during the 3 years from 1 July 2004 to 30 June 2007.

Data for the second year in this 3–year period were used for the single-year analysis.

Hospital separations in 2004–05 and 2005–06 were coded according to the 4th edition of ICD-10-AM (NCCH 2004). Those in 2006–06 were coded according to the 5th edition.

4.1.2 Peer groups

The hospital peer groups used in this report are according to the AIHW Peer Group Report Round 10 (2005–2006) AR-DRGv5.0. The groups and their designations are shown in Table 2.

We have not presented results for all peer groups. To do so would require a large number of tables and figures, not all of which are necessary for the purposes of this project.

Table 2: AIHW Peer Groups

AIHW Peer Group	Designation	Definition
Principal referral and specialist women's and children's	A1	Major city hospitals with >20,000 acute casemix-adjusted separations and Regional hospitals with >16,000 acute casemix-adjusted separations per annum
	A2	Specialised acute WCHs with >10,000 acute casemix-adjusted separations per annum
Un-peered and other hospitals	A9	Prison medical services, special circumstance hospitals, Major city hospitals with <2000 acute casemix-adjusted separations, hospitals with <200 separations, etc.
Large hospitals	B1	Major city acute hospitals treating more than 10,000 acute casemix-adjusted separations per annum
	B2	Regional acute hospitals treating >8,000 acute casemix-adjusted separations per annum, and remote hospitals with >5,000 casemix-adjusted separations
Medium hospitals	C1	Medium acute hospitals in Regional and Major city areas treating between 5,000 and 10,000 acute casemix-adjusted separations per annum
	C2	Medium acute hospitals in Regional and Major city areas treating between 2,000 and 5,000 acute casemix-adjusted separations per annum, and acute hospitals treating <2,000 casemix-adjusted separations per annum but with >2,000 separations per annum
Small acute hospitals	D1	Small Regional acute hospitals (mainly small country town hospitals), acute hospitals treating <2,000 separations per annum, and with <40% non-acute and outlier patient days of total patient days
	D2	Small non-acute hospitals, treating <2,000 separations per annum, and with >40% non-acute and outlier patient days of total patient days (D2) plus Multipurpose service (E2) – Small subacute and non-acute hospitals (G)
	D3	Small remote hospitals (<5,000 acute casemix adjusted separations but not 'Multipurpose services' and not 'Small non-acute'. Most are <2,000 separations)

Note: Excludes psychiatric hospitals. Definitions from Australian Hospital Statistics 2005–06 (AIHW 2007).

4.2 Single-year analysis: 2005–06

Rates of in-hospital mortality amongst Australian hospitals in 2005–06 were compared using indirect standardisation. Separate analyses were performed for three sets of cases:

1. Diagnoses accounting for 80% of mortality: cases with one of the principal diagnosis codes listed in Appendix 1. These 68 codes are associated with the largest numbers of cases ending with in-hospital death in the subset of the 2005–06 NHMD file meeting our study criteria. This group accounts for 80% of deaths and about 20% of cases.
2. Diagnoses accounting for the remaining 20% of mortality. This group includes cases with any Principal Diagnosis code that is not in Appendix 1.
3. All diagnoses. This is the sum of the previous two groups, and includes 100% of in-scope cases and 100% of in-scope deaths.

4.3 Calculation of HSMRs

HSMRs were calculated for each hospital using the ratio of the observed to expected number of deaths:

$$\text{HSMR} = \frac{\text{Actual number of in-hospital deaths amongst selected diagnosis groups}}{\text{Expected number of in-hospital deaths amongst selected diagnosis groups}} \times 100$$

Logistic regression was used to calculate the expected number of deaths in each hospital. The 'standard' population was the combined population of all hospitals included in each analysis. Each HSMR was therefore a ratio of the observed hospital mortality rate to the rate for all hospitals combined based on patients with the same characteristics. The logistic regression model used for all principal analyses, referred to here as the RACM model, followed the approach of the CIHI (2007), which has also been used in the UK and Holland.

Hospital-specific expected numbers of deaths were calculated by summing the probabilities of death obtained from coefficients for the logistic regression model. The independent variables included in the RACM model are listed and described in Section 4.5.2.

As stated in Section 3.3.8, the HSMRs were recalibrated before presentation.

4.4 Graphical methods of presentation

Graphs were generated using Stata Statistical Software, Release 10.

4.4.1 Caterpillar plots

Caterpillar plots are simply graphical presentations of HSMRs from each institution within a population of interest. Each institution's HSMR is graphed from three points: the HSMR value and the upper and lower 95% confidence limits. The population average of 100 is provided by way of reference. The plots were generated with the relevant Stata commands.

4.4.2 Funnel plots

Funnel plots allow many points to be plotted simultaneously, with information about whether each point is significantly above or below the expected, or average, value. Funnel plots were developed as a method of displaying data for statistical process control (SPC). They avoid the ranking approach that some methods of displaying performance data use. The Association of Public Health Observatories of the UK (APHO) has adopted and adapted this method for a report on indicators of public health in the regions of England, and the APHO method for generating funnel plots was accessed at http://www.apho.org.uk/default.aspx (June 1 2008) and used here.

Creating a funnel plot requires the superposition of two charts:

1. A scatter plot of the expected number of deaths against the HSMR.
2. A scatter plot with interpolated lines of the SMR; that is, 100% and Poisson control limits around that measure. The '*invgammap*' function in STATA was used to determine the Poisson control limits.

The construction of the control limits depends on distribution of the underlying performance measure. The width of the limits is somewhat arbitrarily equivalent to 95% and 99.8% confidence intervals around the target value (roughly equivalent to two and three standard deviations). The three standard deviations measure is often taken as the boundary between 'common-cause' and 'special-cause' variation in control chart method.

4.5 Case selection

4.5.1 Inclusion/exclusion criteria

Inclusion criteria

- admission for acute care (episode type=1)
- age at admission from 0 to 120 years
- gender recorded as male or female (i.e. not 'missing')
- length of stay up to 365 consecutive days
- admission category: either elective or emergency
- Principal Diagnosis at discharge. The proportion of in-hospital deaths was calculated for the set of cases with each three-character ICD-10-AM code. Diagnosis codes were ranked in descending order of this proportion. Three sets of records were selected, each being used for part of the analysis:
  - (i) high risk: the set of records with three-character ICD-10-AM codes that ranked highest in terms of the diagnosis-specific number of deaths and which, together, account for 80% of all in-hospital deaths. (See Appendix 1.)
  - (ii) lower risk: the set with all other three-character ICD-10-AM codes
  - (iii) all records satisfying the selection criteria. This is the sum of (i) and (ii).

Exclusion criteria

- patients discharged against medical advice (defined using AIHW data element 'mode of separation' = Left against medical advice/discharge at own risk)
- palliative care patients (recalibrated for single year HSMR production)
- neonates (infant age >0 and <= 28 days)

Outcome

- death in hospital was defined as Mode of Separation = died

4.5.2 The logistic regression model

The independent variables used for the RACM model used in all primary analyses were:

- age (in years at time of admission)
- sex (based on sex recorded at discharge)
- length of stay group (as six separate categories; i.e. 1 day, 2 days, 3–9 days, 10–15 days, 16–21 days and 22–365 days) (same day admission/separation cases were included in the '1 day' category)
- admission category (emergency or elective)
- diagnosis group (based on the first three digits of the principal diagnosis coded according to ICD-10-AM 4th Edition) The groups were specified using NHMD data for 2005–06.

  a) High-risk group, accounting for 80% of in-hospital mortality

  The first three characters of the principal diagnosis code (ICD-10-AM) were used to detect the conditions that (i) have the highest number of cases ending with the death of the patient in hospital and (ii) in aggregate, account for 80% of all deaths in hospital. In total, 68 three-character ICD-10-AM codes are in this set (see Table A1.1, Appendix 1).

  b) Lower risk group, accounting for the other 20% of in-hospital mortality

  This set includes cases with all principal diagnosis codes except the 68 in the 'high-risk' group. For the low risk of mortality analysis, rather than creating a separate coefficient for each of the 1,518 three-character principal diagnosis codes present in the group of cases that accounted for the remaining 20% of in-hospital mortality, a 10–category risk variable was created based on the crude risk of in-hospital mortality for each three-character diagnosis category. The risk categories are deciles of the log of the crude risk. Risk deciles were determined by calculating the absolute risk for each diagnosis group (i.e. taking all deaths in each diagnosis group and dividing by the total number of separations in the same diagnosis group). The log of the absolute risks was divided into 10 equal groups. Diagnoses that did not account for any deaths were included in the lowest decile of risk.

  c) 100% of in-hospital mortality

  The same risk approach as that defined in (b) was used. The data include 68+1,518=1,586 three-character ICD-10-AM diagnoses groups.

- comorbidity group (either 0, 1 or 2 and based on the Charlson Index score (Quan et al. 2005)). Comorbidity status was derived from the additional diagnosis codes in the NHMD, which were used to generate a Charlson Index score for each patient based on Quan's method (Quan et al. 2005). The Charlson Index was converted to a score of 0, 1 or 2. Patients whose Charlson Index value is 2 or higher were assigned a score of 2.
- inward transfer status (admission mode=1 indicating whether a patient was transferred from an acute institution)

Women's and Children's hospitals (WCHs)

The effect of including WCHs in the analyses was assessed by comparing the HSMRs based on diagnoses for the leading 80% of in-hospital deaths with and without WCHs included in the logistic regression model.

4.6 Model checking

For each of the three analyses that used the RACM approach, models were assessed for goodness of fit, discriminatory power and explanatory power.

4.6.1 Goodness of fit

Goodness of fit was assessed using the Hosmer–Lemeshow (2000) goodness-of-fit statistic for 10 groups based on deciles of risk (StataCorp 2007).

4.6.2 Model discrimination

An assessment of the discriminatory power of each model was based on the c-statistic (area under the ROC).

4.6.3 Explanatory power

The pseudo-R^2 statistic is reported to assess the degree to which included variables decreased the unexplained variance in the data.

4.7 Calculation of 95% confidence intervals

a) for HSMR point estimates

95% confidence intervals for the tables and caterpillar plots were calculated using Byar's approximation:

Lower confidence limit = $O/E*(1–1/(9*O) – 1.96 / (3*\text{sqrt}(O)))^3 * 100$

Upper confidence limit = $(O + 1)/E*(1– (1/(9*(O+1))) + 1.96 / (3*\text{sqrt}(O+1)))^3 * 100$

where O = observed number of deaths and E = Expected number of deaths.

b) for funnel plots

The confidence intervals displayed in funnel plots were calculated by assuming a Poisson (inverse gamma) distribution for the expected number of deaths. Intervals roughly equal to two and three standard deviations for the HSMR funnel plots were plotted.

4.8 Further development of the risk model

For the high-risk (80%) group, we compared HSMRs calculated using the RACM model with HSMRs calculated using our own ERM model (see Section 3.3.5). Although that model was based on the same variables as the RACM model, the effect of transforming independent variables such as age to more closely approximate the distributional characteristics of the logit curve was tested empirically. All significant two-way interactions in the regression model were tested separately in a main effects model using the LR test. All significant interactions were then included in a final model and removed one at a time: the effect of each removal being assessed using the LR test.

Once the final model was chosen, we split the data set into a development and validation data set (50% of the data for each) and assessed model fit on the validation set using the coefficients obtained from the developmental data set. This allows the assessment of whether or not the chosen model over-fitted the given data and would not perhaps fit so well on other data sets; for example, different years.

In addition, we also assessed the effect of the addition of the SEIFA index of disadvantage (as either a continuous or categorical variable) on the pseudo-R^2 statistic. (The SEIFA index is described in *Appendix 5 Data issues*.)

4.9 Longitudinal analysis

In Section 2.7.2 we described recent growth in longitudinal analysis of in-hospital mortality. This approach allows analysis of trends—a matter of considerable interest. It also provides a way to assess the extent to which information derived from a data source is informative about characteristics of hospitals, rather than reflecting random variation. The latter requires treatment of values of in-hospital mortality at several time-points as repeated measures.

A prominent recent example of this approach is the analysis of Dutch data reported by Heijink (2008). We opted to use the same approach, because this would allow comparison.

The method is described further in Section 5.10.1.

4.10 Statistical software

Data preparation and cleaning were carried out using SPSS Version 14.0 for Windows.

All other analysis, and preparation of caterpillar plots and funnel plots, was done using Stata Statistical Software, Release 10.

5 Results

Results of the one-year analysis are presented separately for analysis of the 80%, 20% and 100% in-hospital mortality data sets, and stratified by peer group (although data from all hospitals was combined for each of the three analyses). Peer groups were identified using the National Hospital Cost Data Collection (NHCDC) Round 10 (2005–2006) Peer Group Report (DoHA 2007). Cost-weighted separations were calculated by applying the AIHW 2005–06 DRG cost weights to each separation and summing these cost weights to calculate the number of cost-weighted separations. A selection of descriptive statistics for the total sample is presented in Table 3.

Effect of Women's and Children's hospitals

Women's and Children's hospitals have very different mortality profiles from other centres, and it makes little sense to compare these specialised centres with anything other than similarly specialised centres. However, there are many more general hospitals that include obstetrics, gynaecology and paediatrics in their casemix. The effect of including WCHs in the principal analyses was assessed by comparing the HSMRs based on diagnoses for the leading 80% of in-hospital deaths with and without WCHs included in the logistic regression model. Because HSMRs were virtually identical with both approaches, the data from WCHs were included in all analyses. The WCHs were also analysed as a specific peer group (A2), though the results of the single-year analysis are not presented in this report.

Table 3: Selected descriptive statistics for the total sample of 2005–06 hospital separations

	N	Per cent
Gender		
Male	3,438,248	47.02
Female	3,873,645	52.98
Persons[(a)]	7,311,983	
Mode of separation		
Discharged at own risk	35,707	0.49
Died in hospital	71,122	0.97
Type of episode of care		
Acute care	7,016,160	95.95
Rehabilitation care	151,527	2.07
Palliative care	25,741	0.35
Other	109,685	1.50
Not stated	8,870	0.12
Health-care sector		
Public hospital	4,450,509	60.87
Private hospital	2,298,437	31.43
Public psychiatric hospital	15,567	0.21
Private free standing day hospital	547,470	7.49
Diagnosis groups		
High risk (80% of total in-hospital deaths)	1,109,758	
Low risk (20% of total in-hospital deaths)	4,924,758	
All diagnoses (100%of total in-hospital death)	6,034,516	

(a) Total does not sum due to a small number of cases with unknown gender

5.1 Inclusions and exclusions

Of the 7,311,983 records in the original 2005–06 data set, 1,277,467 were excluded, as follows: 900,832 due to admission category being neither elective or emergency; 295,823 admitted for a reason other than acute care; 36,553 due to being a palliative care patient (note that the recalibration process described in Section 3.3.8 was confined to the numbers of palliative care patients selected for analysis); 32,856 due to patients being discharged against medical advice; 11,164 due to being a neonate (infants age between 0 and 28 days); 189 due to length of stay being greater than 365 days; 40 due to gender not being recorded as either male or female; and 10 due to having a recorded age that was not in the range 0 to 120 years.

5.1.1 High-risk group (80% of in-hospital mortality)

Of the 6,034,516 records retained after the above exclusions, 4,931,241 records were omitted because the principal diagnosis was not one of the 68 diagnoses in the 'high-risk' group, associated with 80% of deaths in hospital (Appendix 1).
The remaining 1,103,275 records were included in the analysis (see Table 4).
Of the 923 hospitals in the original 2005–06 data set, 817 had admitted patients meeting these inclusion criteria in 2005–06.

Table 4: Selective descriptive statistics for the high-risk case group (80% of in-hospital mortality in 2005–06)

	N	Per cent
Gender		
Male	588,106	53.31
Female	515,169	46.69
Mode of separation		
Discharged at own risk	0	0.00
Died in hospital	36,046	3.27
Health-care sector		
Public hospital	744,481	67.48
Private hospital	309,064	28.01
Public psychiatric hospital	9	0.00
Private free standing day hospital	49,721	4.51

5.1.2 Lower risk group (20% of in-hospital mortality)

We also analysed in-hospital mortality for the in-scope records not included in the 'high-risk' group. Table 5 describes this group.

Table 5: Selective descriptive statistics for the lower risk case group (20% of in-hospital mortality in 2005–06)

	N	Per cent
Gender		
Male	2,324,908	46.97
Female	2,624,987	53.03
Mode of separation		
Discharged at own risk	0	0.00
Died in hospital	9,128	0.18
Health-care sector		
Public hospital	2,841,781	57.41
Private hospital	1,669,056	33.72
Public psychiatric hospital	13,113	0.26
Private free standing day hospital	425,952	8.61

5.1.3 Total in-hospital mortality

All in-scope records were included in this part of the analysis. Table 6 presents descriptive statistics.

Table 6: Selective descriptive statistics for the case group including 100% of in-hospital mortality in 2005–06

	N	Per cent
Gender		
Male	2,913,014	48.12
Female	3,140,156	51.88
Mode of separation		
Discharged at own risk	0	0.00
Died in hospital	45,174	0.75
Health-care sector		
Public hospital	3,586,262	59.25
Private hospital	1,978,120	32.68
Public psychiatric hospital	13,122	0.22
Private free standing day hospital	475,673	7.86

5.2 Model building and the effect of covariates on odds of in-hospital mortality

The odds ratios for the effect of each of the included covariates on in-hospital mortality for 80%, 20% and 100% mortality groups were extracted and are presented as point estimates, together with standard errors and 95% confidence intervals, in Tables 7–9. Readers are reminded that these results were obtained without recalibrating the palliative-care variable.

The odds ratios can be interpreted as the effect of the presence of each modelled characteristic on the likelihood that an episode in hospital will end with in-hospital death, after allowing for all of the other variables in the model. For example, considering the high-risk group (Table 7), elective admissions were associated with a little over one-quarter (0.281 times) the likelihood of in-hospital death compared with emergency admissions (used as the reference group). Similarly, the presence of two or more Charlson comorbidity categories was associated with odds of fatal outcome that were more than 6 times higher (6.048 times) than if no Charlson comorbidity was present.

Table 7: Odds ratios for the effect of each of the included covariates on 80% in-hospital mortality

	Odds ratio	95% CI	p-value
Age (years)	1.045	(1.044–1.046)	<0.001
Sex (Male=1, Female=2)	1.007	(0.984–1.031)	0.556
Length of stay			
1 day	1	–	–
2 days	1.035	(0.991–1.082)	<0.122
3–9 days	0.633	(0.613–0.652)	<0.000
10–15 days	0.66	(0.634–0.687)	<0.000
16–21 days	0.831	(0.789–0.874)	<0.000
22–365 days	1.106	(1.058–1.157)	<0.000
Urgency admission			
(Emergency=1, Elective=2)			
1	1	–	–
2	0.281	(0.271–0.291)	<0.001
Canadian Charlson category			
0	1	–	–
1	2.756	(2.637–2.880)	<0.001
2	6.048	(5.780–6.330)	<0.001
Transferred patient	1.578	(1.519–1.639)	<0.001
Logistic regression	Number of obs = 1103275		
	LR chi2(78) = 7748.16		
	Prob > chi2 = 0.0000		
Log likelihood = –120028.66	Pseudo R^2 = 0.2440		

Table 8: Odds ratios for the effect of each of the included covariates on 20% in-hospital mortality

	Odds ratio	95% CI	p-value
Age (years)	1.031	(1.030–1.032)	<0.000
Sex (Male=1, Female=2)	0.929	(0.890–0.970)	<0.001
Length of stay			
1 day	1	–	–
2 days	1.493	(1.365–1.632)	<0.000
3–9 days	1.467	(1.378–1.562)	<0.000
10–15 days	1.994	(1.845–2.155)	<0.000
16–21 days	2.943	(2.689–3.221)	<0.000
22–365 days	3.808	(3.528–4.111)	<0.000
Urgency admission			
(Emergency=1, Elective=2)			
1	1	–	–
2	0.322	(0.305–0.340)	<0.000
Canadian Charlson category			
0	1	–	–
1	2.696	(2.2.542–2.860)	<0.000
2	7.155	(6.742–7.593)	<0.000
Transferred patient	1.819	(1.705–1.939)	<0.000

Logistic regression	Number of obs	=	4949902
	LR chi2(20)	=	45312.60
	Prob > chi2	=	0.0000
Log likelihood = –43931.126	Pseudo R^2	=	0.3402

Table 9: Odds ratios for the effect of each of the included covariates on 100% in-hospital mortality

	Odds ratio	95% CI	p-value
Age (years)	1.036	(1.035–1.037)	<0.000
Sex (Male=1, Female=2)	0.955	(0.936–0.974)	<0.000
Length of stay			
1 day	1	–	–
2 days	1.02	(0.982–1.060)	<0.299
3–9 days	0.686	(0.668–0.705)	<0.000
10–15 days	0.783	(0.756–0.811)	<0.000
16–21 days	1.054	(1.009–1.101)	<0.017
22–365 days	1.466	(1.413–1.522)	<0.000
Urgency admission			
(Emergency=1, Elective=2)			
1	1	–	–
2	0.301	(0.293–0.309)	<0.000
Canadian Charlson category			
0	1	–	–
1	2.165	(2.095–2.236)	<0.000
2	4.571	(4.422–4.726)	<0.000
Transferred patient	1.77	(1.715–1.827)	<0.000

Logistic regression	Number of obs	=	6053177
	LR chi2(20)	=	189758.61
	Prob > chi2	=	0.0000
Log likelihood = –171379.69	Pseudo R^2	=	0.3563

5.3 Discriminatory and explanatory power

Tables 10 to 12 display the c-statistic, pseudo R^2, and the change in pseudo-R^2 for subsets of the independent variables included in the RACM model for the three groups.

The generally high values of the c-statistic largely reflect the large size of the data set analysed. The R^2 values are larger with the fuller models, indicating a reduction in unexplained variance with the addition of the covariates shown.

Although these models are not exactly comparable with any of the results from the literature that are summarised in Table 1, it is worth noting that the values presented in Table 10 of the measures of discrimination and explanatory power for the full models are certainly not low in relation to the ranges of values in Table 1.

Table 10: c-statistic, pseudo R^2, and the change in pseudo R^2 for subsets of the independent variables included in the RACM model for 80% in-hospital mortality

Included variables	c-statistic	Pseudo R^2	∆ Pseudo R^2
Age	0.7058	0.0581	
Age, sex	0.7068	0.0586	0.0005
Age, sex, LOS group,	0.7289	0.0727	0.0141
Age, sex, LOS group, urgency	0.767	0.1017	0.029
Age, sex, LOS group, urgency, pdiag_aihw3	0.8583	0.2186	0.1169
Age, sex, LOS group, urgency, pdiag_aihw3, cancharlson	0.8751	0.2424	0.0238
Age, sex, LOS group, urgency, pdiag_aihw3, cancharlson, transfer	0.8764	0.244	0.0016

Model Un-stratified, 80% mortality *N* = 1,103,275

Table 11: c-statistic, pseudo R^2, and the change in pseudo R^2 for subsets of the independent variables included in the RACM model for 20% in-hospital mortality

Included variables	c-statistic	Pseudo R^2	∆ Pseudo R^2
Age	0.79	0.0795	
Age, sex	0.7911	0.0799	0.0004
Age, sex, LOS group,	0.8767	0.187	0.1071
Age, sex, LOS group, urgency	0.9147	0.2205	0.0335
Age, sex, LOS group, urgency, riskcat	0.9554	0.3045	0.084
Age, sex, LOS group, urgency, riskcat, cancharlson	0.9625	0.338	0.0335
Age, sex, LOS group, urgency, riskcat, cancharlson, transfer	0.9632	0.3402	0.0022

Model Un-stratified,20% mortality *N* = 4,949,902

Table 12: c-statistic, pseudo R^2, and the change in pseudo R^2 for subsets of the independent variables included in the RACM model for 100% in-hospital mortality

Included variables	c-statistic	Pseudo R^2	∆ Pseudo R^2
Age	0.8073	0.1114	
Age, sex	0.8084	0.112	0.0006
Age, sex, LOS group,	0.8603	0.1693	0.0573
Age, sex, LOS group, urgency	0.8997	0.2154	0.0461
Age, sex, LOS group, urgency, riskcat	0.9491	0.3357	0.1203
Age, sex, LOS group, urgency, riskcat, cancharlson	0.9548	0.3542	0.0185
Age, sex, LOS group, urgency, riskcat, cancharlson, transfer	0.9555	0.3563	0.0021

Model Un-stratified,100% mortality *N* = 6,053,177

5.4 Goodness of fit

Tables 13 to 15 display Hosmer–Lemeshow deciles of risk and the observed and expected numbers of cases (and non-cases) of in-hospital mortality for the high-risk case group (80% of deaths), analysed using the RACM model, and the lower risk and the all-deaths groups. The tables are collapsed on deciles of estimated probabilities of death. Figures 4 to 6, accompanying the tables, show the percentages of in-hospital mortality for each decile of risk for both the observed data and the data predicted by the logistic regression model for the mortality outcomes. The predicted values for the high-risk group were derived from the RACM model, using principal diagnoses at the three character ICD-10-AM level (Appendix 1). The predicted values for the other two groups were derived using principal diagnoses assigned to deciles of risk, as described above (Section 4.5.2).

The Hosmer–Lemeshow test did not demonstrate good fit for any of the RACM models. However, as has been discussed previously, the Hosmer–Lemeshow goodness of fit method is sensitive to the very large sample sizes used here. Moreover, the RACM model does not include data transformations or allow for possible interactions between covariates—issues which were tackled when developing the ERM model. The tables and graphical plots of deciles of observed and expected risks show that the RACM model fit is closer for the deciles of higher risk than for the lower deciles, where the model seems to somewhat 'over-call' expected mortality (see tables 13 to 15).

The goodness of fit for the ERM model is discussed in Section 5.7.1.

Table 13: Hosmer–Lemeshow deciles of risk and the observed and expected numbers of cases (and non-cases) of in-hospital mortality for the high-risk group of deaths (using the RACM model)

Decile of risk group	Prob	Obs 1	Exp1	Obs 0	Exp 0	Total
1	0.001	30	61.2	110,306	110,274.8	110,336
2	0.002	69	152	110,455	110,372	110,524
3	0.003	159	282.3	110,180	110,056.7	110,339
4	0.006	271	484.1	109,845	109,631.9	110,116
5	0.009	554	786.4	109,789	109,556.6	110,343
6	0.015	1165	1304.3	109,164	109,024.7	110,329
7	0.026	2412	2230.8	107,894	108,075.2	110,306
8	0.046	4111	3837.4	106,227	106,500.6	110,338
9	0.089	7655	7013	102,704	103,346	110,359
10	0.980	19620	19894.4	90,665	90,390.6	110,285

Note: Obs1 and Exp1 = expected cases; Obs 0 and Exp0 = expected non-cases, Hosmer–Lemeshow $Chi^2(8) = 396.37$, $p > 0.000$

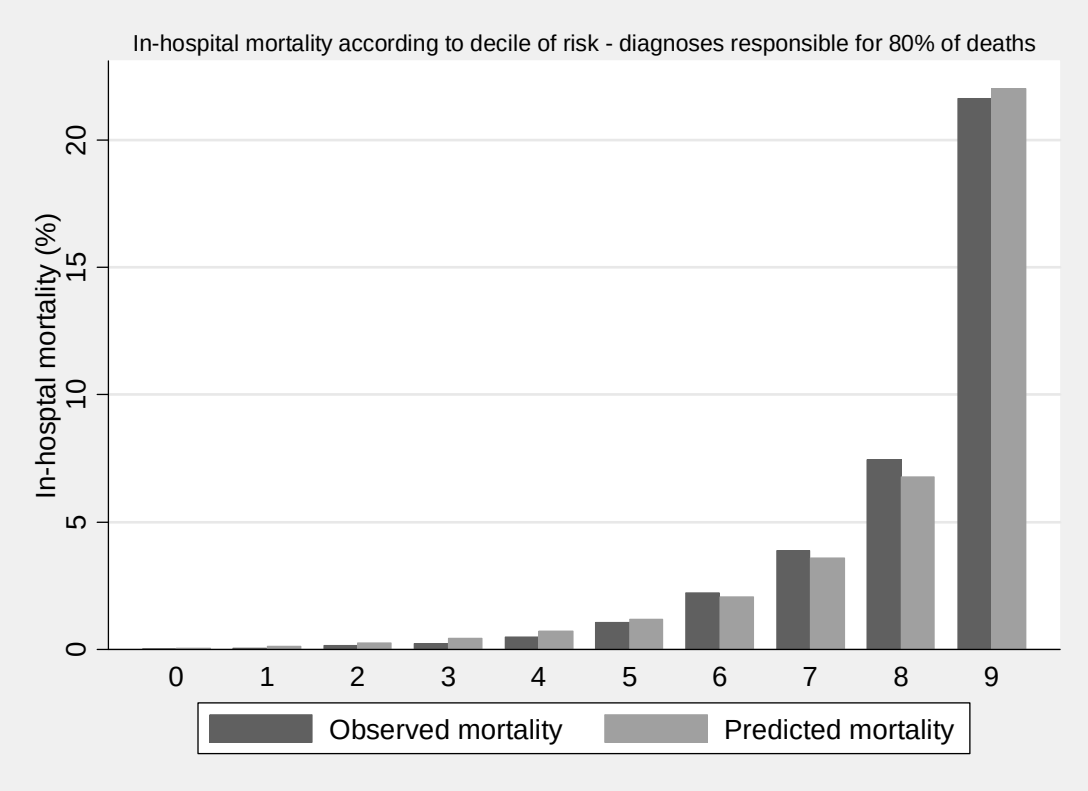


Figure 4: Percentages of in-hospital mortality for each decile of risk for both the observed data and the data predicted by the logistic regression model for the high-risk group of cases accounting for 80% of in-hospital deaths

Table 14: Hosmer–Lemeshow deciles of risk and the observed and expected numbers of cases (and non-cases) of in-hospital mortality for the lower risk group of deaths (using the RACM model)

Decile of risk group	Prob	Obs 1	Exp1	Obs 0	Exp 0	Total
1	1	0	6	5.3	500,313	500,313.8
2	2	0	7	9.9	489,655	489,652.1
3	3	0	8	16.3	495,031	495,022.7
4	4	0.000	14	26.2	496,719	496,706.8
5	5	0.000	20	43.8	497,137	497,113.3
6	6	0.000	23	71.5	491,491	491,442.5
7	7	0.000	53	123.7	495,029	494,958.3
8	8	0.001	126	235.9	494,567	494,457.1
9	9	0.002	547	616.4	494,318	494,248.6
10	10	0.720	8324	7979.1	486,514	486,858.9

Note: Obs1 and Exp1 = expected cases; Obs 0 and Exp0 = expected non-cases, Hosmer–Lemeshow $Chi^2(8) = 171.29$, $p > 0.000$

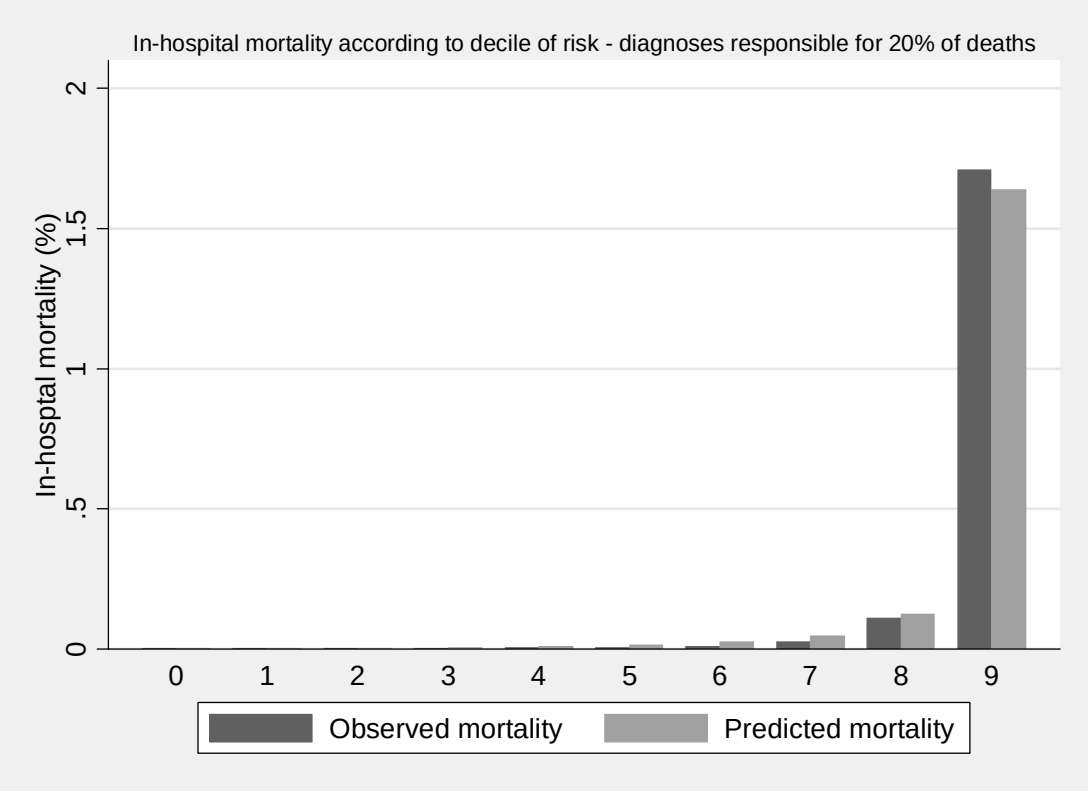


Figure 5: Percentages of in-hospital mortality for each decile of risk for both the observed data and the data predicted by the logistic regression model for the lower risk group including the remaining 20% of in-hospital deaths

Table 15: Hosmer–Lemeshow deciles of risk and the observed and expected numbers of cases (and non-cases) of in-hospital mortality for the group including all in-hospital deaths (using the RACM model)

Decile of risk group	Prob	Obs 1	Exp1	Obs 0	Exp 0	Total
1	0	8	7.3	605,391	605,391.7	605,399
2	0	12	16.4	606,090	606,085.6	606,102
3	0.000	21	30.5	604,450	604,440.5	604,471
4	0.000	26	55.4	612,694	612,664.6	612,720
5	0.000	49	99.9	600,665	600,614.1	600,714
6	0.001	100	203.2	602,952	602,848.8	603,052
7	0.001	259	469.1	604,648	604,437.9	604,907
8	0.004	924	1270.6	604,253	603,906.4	605,177
9	0.014	4021	4483.8	601,364	600,901.1	605,385
10	0.617	39754	38537.7	565,496	566,712.3	605,250

Note: Obs1 and Exp1 = expected cases; Obs 0 and Exp0 = expected non-cases, Hosmer–Lemeshow $Chi^2(8) = 376.26$, $p > 0.000$

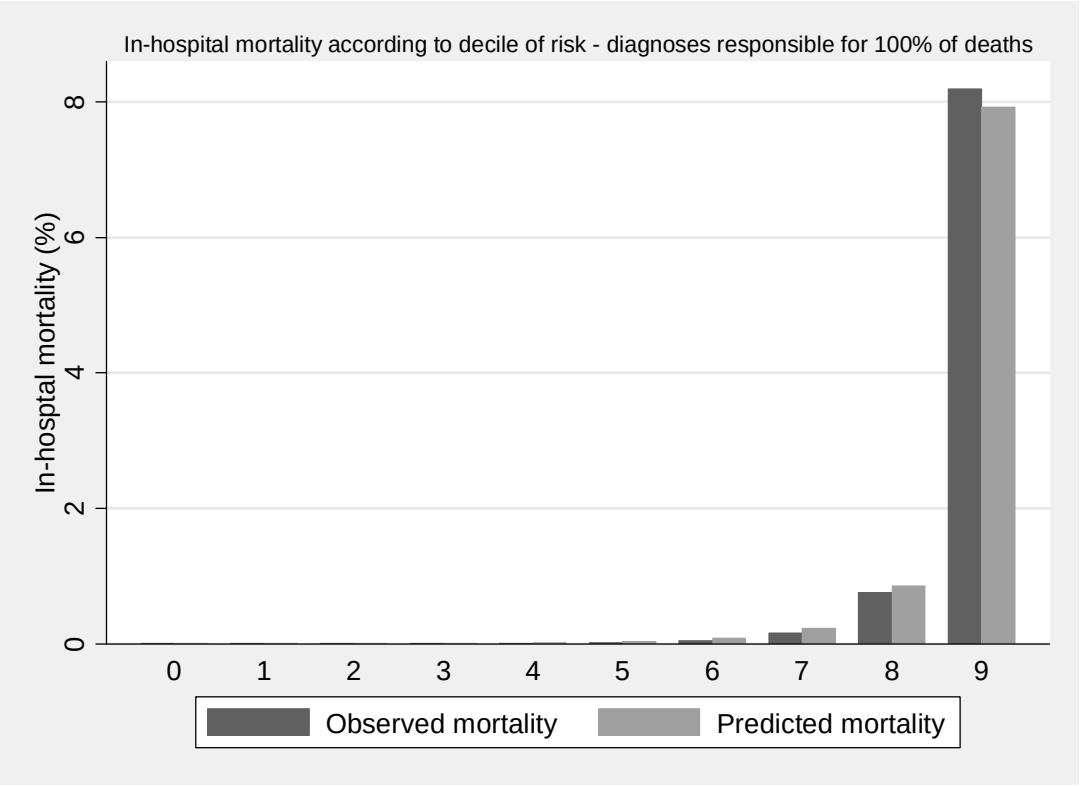


Figure 6: Percentages of in-hospital mortality for each decile of risk for both the observed data and the data predicted by the logistic regression model for the group including all in-hospital deaths

5.5 Individual HSMRs and their 95% confidence intervals

One of the three modes of presentation of HSMRs described in Section 2.9.1 is 'league tables'. This section presents some results of our analysis in this format. Because of the large number of hospitals analysed, we have selected one peer group, A1, to illustrate the approach (equivalent tables of recalibrated risk-adjusted HSMRs for peer groups B1, C2 and D1 are in Appendix 2).

Table 16 shows, for peer group A1, the observed and expected numbers of deaths, the HSMRs (after recalibration) and 95% confidence intervals, and the peer group rankings for the case groups including 80%, 20% and 100% of in-hospital deaths. Readers are reminded that these demonstration values have been recalibrated in order to protect the confidentiality of individual institutions.

Results are arranged in ascending order of risk-adjusted HSMR for the high-risk group of cases (which includes 80% of in-hospital deaths).

Table 16: Observed and expected number of deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group A1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
A1001	44965.51	226	321.25	54	100.89	280	379.70	70.35 (61.5–80.1)	53.52 (40.2–69.8)	73.74 (65.4–82.9)	1	1	3
A1002	46593.44	161	224.57	33	49.24	194	267.46	71.69 (61.0–83.7)	67.02 (46.1–94.1)	72.53 (62.7–83.5)	2	3	2
A1003	56580.62	321	444.35	84	104.59	405	544.52	72.24 (64.6–80.6)	80.31 (64.1–99.4)	74.38 (67.3–82.0)	3	12	4
A1004	20299.45	133	180.47	23	36.68	156	229.17	73.70 (61.7–87.3)	62.70 (39.7–94.1)	68.07 (57.8–79.6)	4	2	1
A1005	74787.9	559	701.25	136	154.17	695	845.75	79.72 (73.2–86.6)	88.21 (74.0–104.3)	82.18 (76.2–88.5)	5	15	6
A1006	66842.18	576	717.05	154	177.22	730	885.59	80.33 (73.9–87.2)	86.90 (73.7–101.8)	82.43 (76.6–88.6)	6	14	7
A1007	80017.26	617	739.08	136	187.79	753	927.60	83.48 (77.0–90.3)	72.42 (60.8–85.7)	81.18 (75.5–87.2)	7	5	5
A1008	20061.79	99	117.68	29	28.48	128	144.61	84.12 (68.4–102.4)	101.82 (68.2–146.2)	88.51 (73.8–105.2)	8	33	11
A1009	30667.47	171	199.64	48	44.00	219	242.17	85.65 (73.3–99.5)	109.09 (80.4–144.6)	90.43 (78.9–103.2)	9	42	14
A1010	44398.09	435	507.72	84	89.23	519	565.82	85.68 (77.8–94.1)	94.14 (75.1–116.6)	91.73 (84.0–100.0)	10	21	20
A1011	81502.59	553	639.58	164	147.47	717	747.03	86.46 (79.4–94.0)	111.21 (94.8–129.6)	95.98 (89.1–103.3)	11	44	24
A1012	19691.53	142	162.31	37	37.73	179	195.93	87.49 (73.7–103.1)	98.07 (69.0–135.2)	91.36 (78.5–105.8)	12	26	18
A1013	27462	192	217.15	35	47.77	227	266.71	88.42 (76.4–101.8)	73.27 (51.0–101.9)	85.11 (74.4–96.9)	13	6	8
A1014	25276.55	167	188.29	44	40.87	211	232.79	88.69 (75.7–103.2)	107.65 (78.2–144.5)	90.64 (78.8–103.7)	14	38	15
A1015	43771.57	434	481.53	80	108.35	514	599.19	90.13 (81.8–99.0)	73.83 (58.5–91.9)	85.78 (78.5–93.5)	15	7	9
A1016	37992.91	287	318.22	65	60.64	352	385.69	90.19 (80.1–101.3)	107.18 (82.7–136.6)	91.27 (82.0–101.3)	16	37	16
A1017	20106.52	144	158.08	39	38.58	183	200.07	91.09 (76.8–107.2)	101.08 (71.9–138.2)	91.47 (78.7–105.7)	17	31	19
A1018	22186.01	209	228.79	43	56.15	252	281.49	91.35 (79.4–104.6)	76.58 (55.4–103.2)	89.52 (78.8–101.3)	18	9	13
A1019	82247.16	424	462.66	157	133.67	581	576.71	91.64 (83.1–100.8)	117.45 (99.8–137.3)	100.74 (92.7–109.3)	19	51	32
A1020	35695.89	133	144.43	49	49.55	182	193.91	92.08 (77.1–109.1)	98.90 (73.2–130.8)	93.86 (80.7–108.5)	20	28	23

(continued)

Table 16 (continued): Observed and expected number of deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group A1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
A1021	29060.4	176	190.94	42	39.48	218	221.39	92.17 (79.1–106.8)	106.39 (76.7–143.8)	98.47 (85.8–112.4)	21	36	29
A1022	57505.83	273	295.15	66	82.28	339	379.52	92.49 (81.8–104.1)	80.22 (62.0–102.1)	89.32 (80.1–99.4)	22	11	12
A1023	23290.99	140	150.48	24	28.55	164	186.39	93.03 (78.3–109.8)	84.06 (53.8–125.1)	87.99 (75.0–102.5)	23	13	10
A1024	43507.94	510	544.63	105	114.04	615	673.66	93.64 (85.7–102.1)	92.07 (75.3–111.5)	91.29 (84.2–98.8)	24	19	17
A1025	51138.63	422	447.44	78	109.72	500	515.17	94.31 (85.5–103.8)	71.09 (56.2–88.7)	97.05 (88.7–105.9)	25	4	27
A1026	92870.13	451	478.02	146	150.62	597	618.88	94.35 (85.8–103.5)	96.93 (81.8–114.0)	96.46 (88.9–104.5)	26	25	25
A1027	23866.06	182	191.07	40	50.93	222	241.17	95.25 (81.9–110.1)	78.54 (56.1–107.0)	92.05 (80.3–105.0)	27	10	21
A1028	16206.31	145	150.83	31	25.81	176	178.66	96.13 (81.1–113.1)	120.12 (81.6–170.5)	98.51 (84.5–114.2)	28	52	30
A1029	49856.89	536	557.39	115	105.55	651	671.73	96.16 (88.2–104.7)	108.95 (89.9–130.8)	96.91 (89.6–104.7)	29	41	26
A1030	54724.37	391	402.17	107	101.32	498	531.72	97.22 (87.8–107.4)	105.60 (86.5–127.6)	93.66 (85.6–102.3)	30	34	22
A1031	85014.36	775	787.03	215	198.44	990	972.39	98.47 (91.7–105.7)	108.35 (94.3–123.8)	101.81 (95.6–108.4)	31	40	33
A1032	82730.06	542	543.17	146	131.10	688	667.68	99.78 (91.6–108.5)	111.37 (94.0–131.0)	103.04 (95.5–111.0)	32	45	36
A1033	22731.9	179	179.28	33	34.68	212	207.38	99.84 (85.8–115.6)	95.16 (65.5–133.6)	102.23 (88.9–117.0)	33	24	35
A1034	61193.27	535	535.71	119	134.62	654	669.68	99.87 (91.6–108.7)	88.40 (73.2–105.8)	97.66 (90.3–105.4)	34	16	28
A1035	16578.69	91	90.91	30	21.82	121	112.72	100.09 (80.6–122.9)	137.48 (92.7–196.3)	107.35 (89.1–128.3)	35	66	44
A1036	63875.61	505	501.42	144	130.66	649	592.88	100.71 (92.1–109.9)	110.21 (92.9–129.8)	109.47 (101.2–118.2)	36	43	47
A1037	32121.94	212	210.46	57	47.43	269	263.45	100.73 (87.6–115.2)	120.17 (91.0–155.7)	102.11 (90.3–115.1)	37	53	34
A1038	69013.7	551	545.70	122	105.78	673	644.19	100.97 (92.7–109.8)	115.34 (95.8–137.7)	104.47 (96.7–112.7)	38	48	40
A1039	42827.35	414	400.38	91	96.05	505	467.68	103.40 (93.7–113.9)	94.75 (76.3–116.3)	107.98 (98.8–117.8)	39	22	45
A1040	18413.09	68	65.07	30	24.25	98	85.80	104.51 (81.1–132.5)	123.69 (83.4–176.6)	114.22 (92.7–139.2)	40	61	54
A1041	41986	323	308.10	75	60.59	398	371.43	104.84 (93.7–116.9)	123.79 (97.4–155.2)	107.15 (96.9–118.2)	41	62	43

(continued)

Table 16 (continued): Observed and expected number of deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group A1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
A1042	24986.04	138	131.30	31	34.54	169	169.17	105.10 (88.3–124.2)	89.76 (61.0–127.4)	99.90 (85.4–116.1)	42	17	31
A1043	50525.68	450	428.07	112	112.68	562	542.20	105.12 (95.6–115.3)	99.40 (81.8–119.6)	103.65 (95.3–112.6)	43	29	38
A1044	29232.16	292	272.10	53	53.87	345	333.48	107.31 (95.4–120.4)	98.39 (73.7–128.7)	103.46 (92.8–115.0)	44	27	37
A1045	33567.65	178	165.38	58	42.91	236	207.35	107.63 (92.4–124.7)	135.15 (102.6–174.7)	113.82 (99.8–129.3)	45	65	53
A1046	20557.86	155	143.43	35	28.82	190	174.58	108.07 (91.7–126.5)	121.45 (84.6–168.9)	108.83 (93.9–125.5)	46	57	46
A1047	22311.62	301	278.48	59	58.23	360	345.85	108.09 (96.2–121.0)	101.33 (77.1–130.7)	104.09 (93.6–115.4)	47	32	39
A1048	28272.78	303	275.51	62	54.51	365	324.84	109.98 (97.9–123.1)	113.75 (87.2–145.8)	112.36 (101.1–124.5)	48	47	51
A1049	16978.06	131	119.11	25	27.13	156	147.61	109.98 (92.0–130.5)	92.16 (59.6–136.1)	105.68 (89.7–123.6)	49	20	42
A1050	67976.92	530	480.42	180	145.77	710	598.06	110.32 (101.1–120.1)	123.48 (106.1–142.9)	118.72 (110.1–127.8)	50	60	58
A1051	54906.39	586	529.43	120	119.22	706	635.06	110.68 (101.9–120.0)	100.66 (83.5–120.4)	111.17 (103.1–119.7)	51	30	48
A1052	18835.92	173	155.26	38	35.17	211	188.46	111.43 (95.4–129.3)	108.03 (76.4–148.3)	111.96 (97.4–128.1)	52	39	50
A1053	29670.07	215	192.57	36	34.02	251	222.47	111.65 (97.2–127.6)	105.82 (74.1–146.5)	112.82 (99.3–127.7)	53	35	52
A1054	34145.57	402	359.66	62	67.83	464	404.31	111.77 (101.1–123.3)	91.40 (70.1–117.2)	114.76 (104.6–125.7)	54	18	56
A1055	48204.3	336	293.16	105	73.15	441	358.15	114.61 (102.7–127.5)	143.54 (117.4–173.8)	123.13 (111.9–135.2)	55	67	63
A1056	47210.35	362	315.41	99	82.29	461	400.33	114.77 (103.3–127.2)	120.30 (97.8–146.5)	115.15 (104.9–126.2)	56	54	57
A1057	26682.48	256	221.23	30	39.66	286	272.63	115.71 (102.0–130.8)	75.64 (51.0–108.0)	104.90 (93.1–117.8)	57	8	41
A1058	45401.98	516	445.53	101	83.60	617	551.52	115.82 (106.0–126.3)	120.81 (98.4–146.8)	111.87 (103.2–121.1)	58	56	49
A1059	54452.89	382	323.53	101	81.80	483	398.39	118.07 (106.5–130.5)	123.47 (100.6–150.0)	121.24 (110.7–132.5)	59	59	61
A1060	84337.28	652	546.95	170	147.08	822	679.82	119.21 (110.2–128.7)	115.58 (98.9–134.3)	120.92 (112.8–129.5)	60	49	60
A1061	26753.52	224	187.53	55	43.48	279	233.78	119.45 (104.3–136.2)	126.50 (95.3–164.7)	119.34 (105.7–134.2)	61	63	59
A1062	29043.85	247	205.00	67	55.65	314	274.57	120.49 (105.9–136.5)	120.41 (93.3–152.9)	114.36 (102.1–127.7)	62	55	55

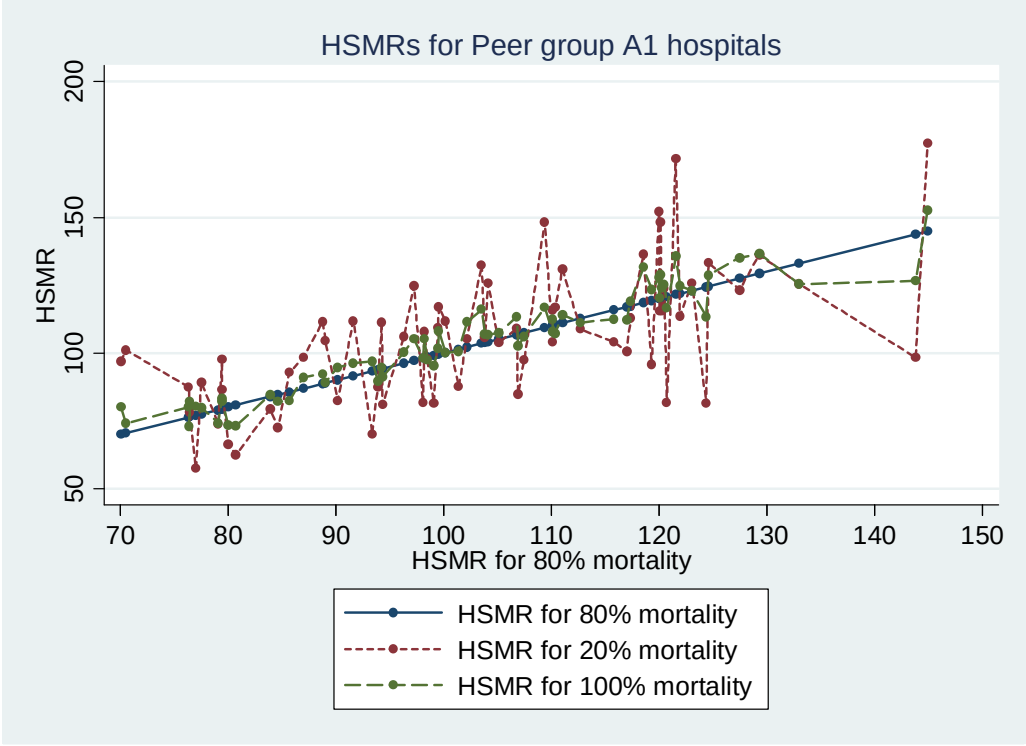
(continued)

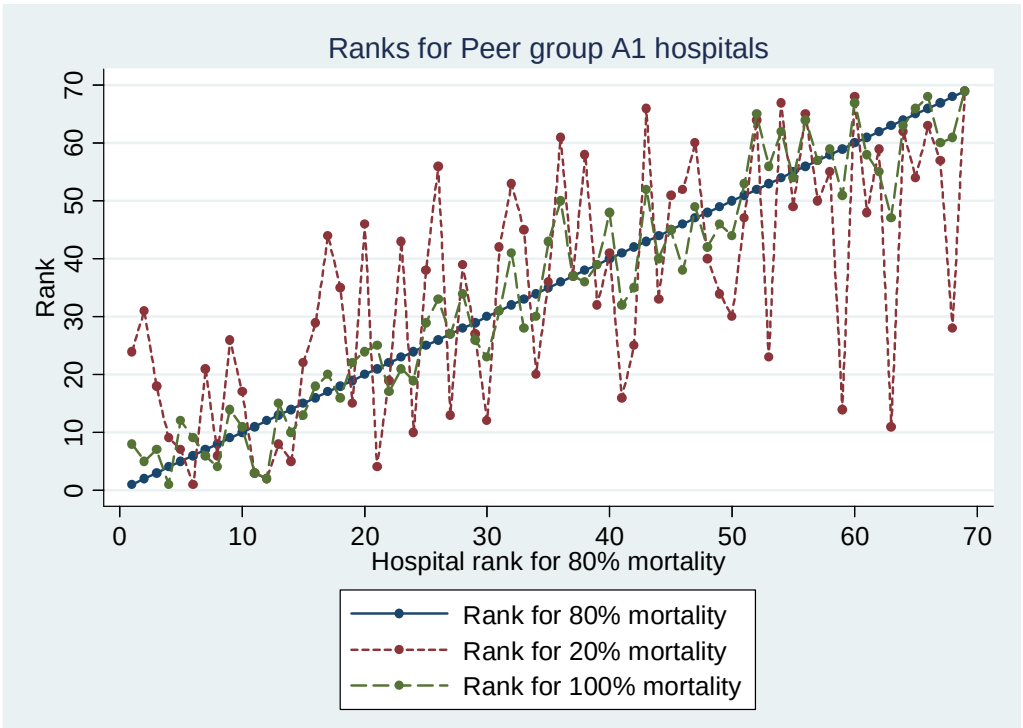
Table 16 (continued): Observed and expected number of deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group A1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
A1063	17672.03	209	173.20	47	29.40	256	199.63	120.67 (104.9–138.2)	159.86 (117.4–212.6)	128.24 (113.0–144.9)	63	68	66
A1064	49841.73	614	501.05	152	118.11	766	629.94	122.54 (113.0–132.6)	128.69 (109.0–150.9)	121.60 (113.1–130.5)	64	64	62
A1065	61723.66	563	452.55	147	126.01	710	554.98	124.41 (114.3–135.1)	116.66 (98.6–137.1)	127.93 (118.7–137.7)	65	50	65
A1066	23069.21	266	207.16	44	46.31	310	249.93	128.40 (113.4–144.8)	95.00 (69.0–127.5)	124.03 (110.6–138.6)	66	23	64
A1067	76698.91	556	430.69	135	110.38	691	519.36	129.09 (118.6–140.3)	122.31 (102.5–144.8)	133.05 (123.3–143.4)	67	58	67
A1068	44580.66	260	200.58	84	52.52	344	257.06	129.62 (114.3–146.4)	159.95 (127.6–198.0)	133.82 (120.1–148.7)	68	69	68
A1069	9835.32	129	86.99	24	21.36	153	113.31	148.28 (123.8–176.2)	112.35 (72.0–167.2)	135.03 (114.5–158.2)	69	46	69

Note: In the table 'cwaseps' refers to the case weighted average number of separations; 'O' = observed number of deaths; 'E' = expected number of deaths; 'LCI' = lower confidence interval; 'UCI' = upper confidence interval.

The figures below provide a graphical representation of the HSMRs and ranks for the three case groups analysed, for peer group A1. The differences in rank were most marked between the analyses of the case groups including, respectively, 80% and 20% of in-hospital deaths. The HSMRs for the lower risk group were the most variable.





Figures 7 and 8: HSMRs and ranks for peer group A1 hospitals

5.6 Caterpillar plots

This section presents examples of the use of caterpillar plots to summarise HSMRs. As before, we have limited presentation to several peer groups, which is sufficient for the purposes of demonstration. In this section, we present plots of the hospitals in four peer groups for the high-risk case group accounting for 80% of all in-hospital deaths.

Figures 9 to 12 display the variation of HSMRs in the peer groups A1, B1, C2 and D1. The 95% confidence interval associated with each point estimate indicates the degree of uncertainty of the point estimate and is dependent on both the observed and expected number of deaths (the larger the observed and expected number of deaths the narrower the confidence intervals). The caterpillar plots allow for a quick visual display of the extent of between-hospital variability, and the degree of precision for each of the estimates using the confidence intervals. Those hospitals in which the confidence intervals do not overlap can generally be assumed to be different in terms of HSMRs.

Differences in the distribution of HSMRs between peer groups might represent true differences in risk, but they might also be due to models and available data allowing incomplete adjustment of risk. It is certainly the case that casemix differs substantially between peer groups. Hence, as for other characteristics of hospitals, comparisons within peer groups may be more meaningful than those between peer groups, even after adjustment.

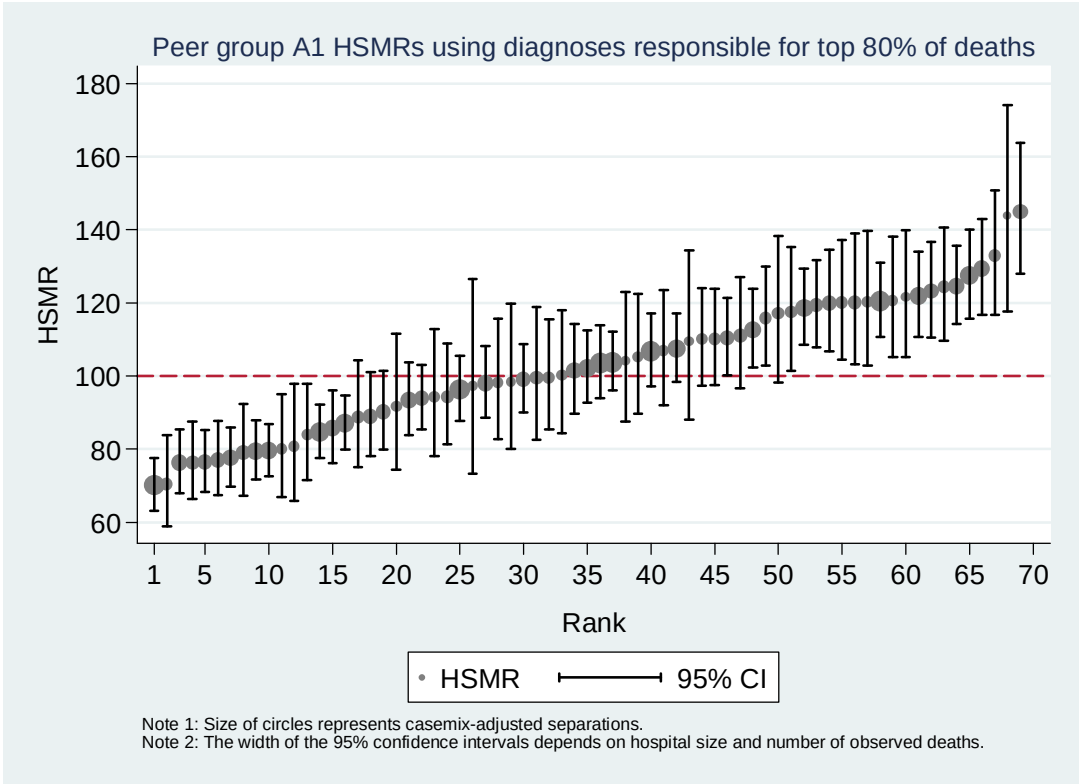


Figure 9: Caterpillar plot of variation in point estimates in HSMR for peer group A1, 80% of in-hospital mortality

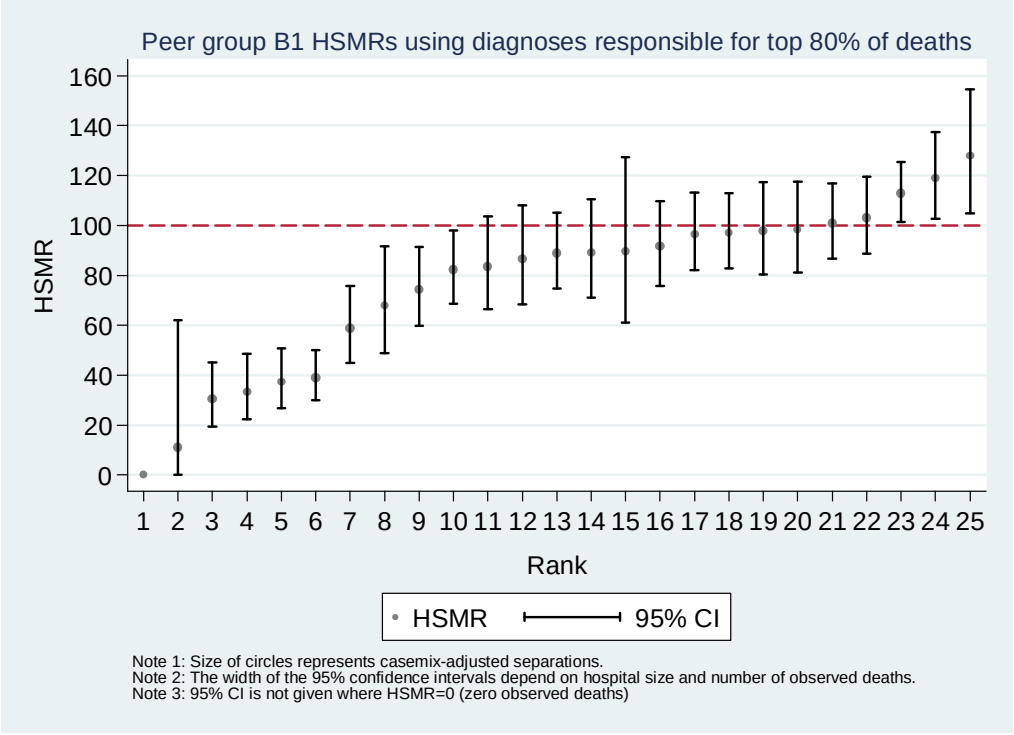


Figure 10: Caterpillar plot of variation in point estimates in HSMR for peer group B1, 80% of in-hospital mortality

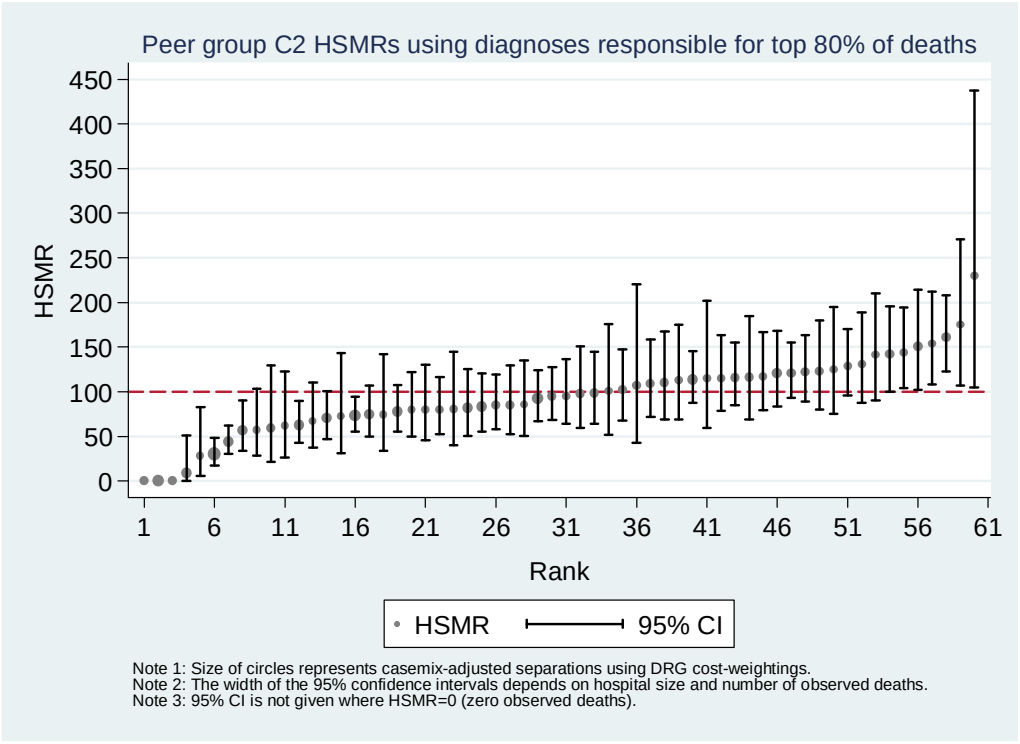


Figure 11: Caterpillar plot of variation in point estimates in HSMR for peer group C2, 80% of in-hospital mortality

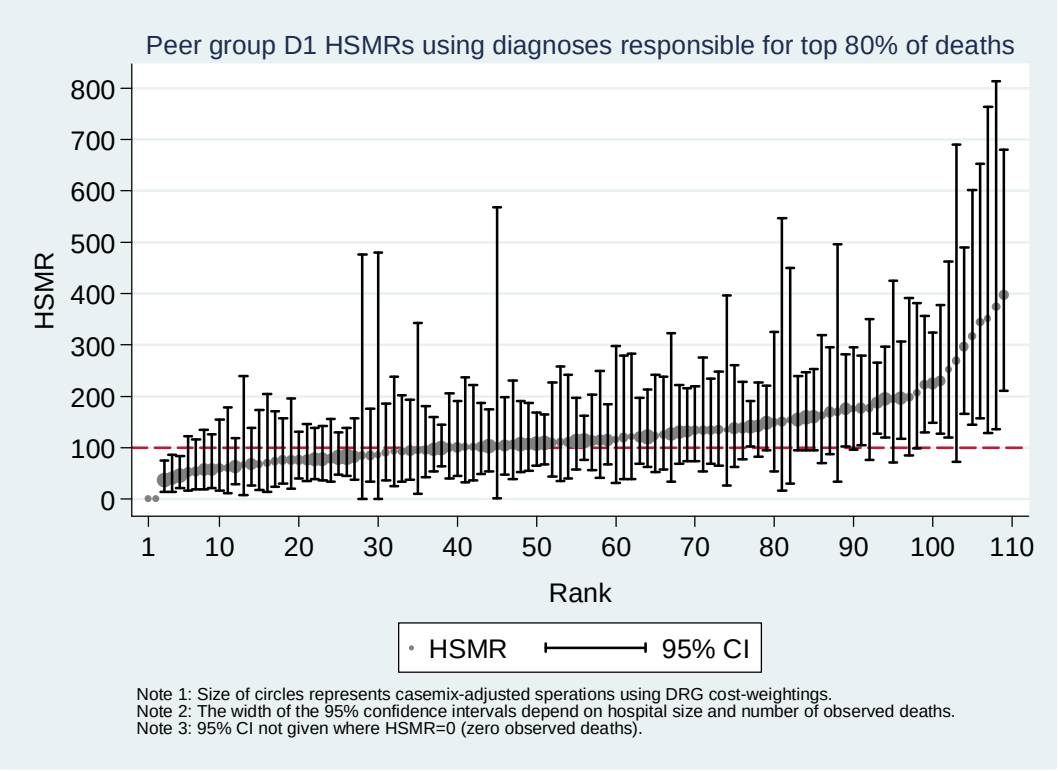


Figure 12: Caterpillar plot of variation in point estimates in HSMR for peer group D1, 80% of in-hospital mortality

5.7 Funnel plots

This section demonstrates the presentation of study data in the form of funnel plots. Compared with tables and caterpillar plots, funnel plots allow graphical information about a large number of hospitals to be presented in only a few figures. We illustrate the approach here by presenting information on peer groups A1, B1 and B2. Funnel plots for other peer groups are provided in Appendix 3.

Figures 13 to 15 display the variation in HSMRs for the A1, B1 and B2 hospitals according to the expected number of deaths and the size of the institution (as assessed by the number of cost-weight adjusted separations). The position of the marker shows the HSMR versus the number of deaths predicted by the model. The size of the marker represents the size of the hospital, measured as casemix-adjusted separations. Each of the figures summarises results for one of the three case sets: high-risk diagnoses accounting for 80% of deaths; the lower risk diagnoses accounting for the remaining 20% of deaths, and all diagnoses.

Funnel plots allow for quick visual detection of 'out-lying' institutions, which are represented as points outside the funnel. More than one peer group is shown in each of the figures, coded by colour.

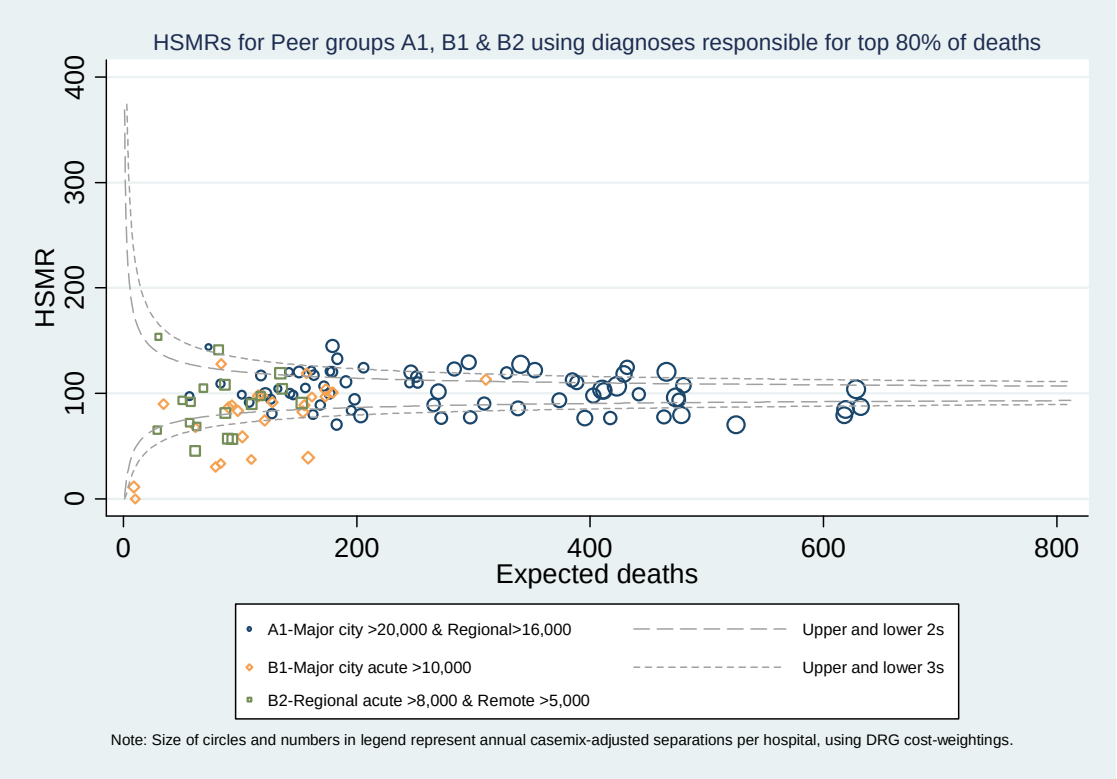


Figure 13: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group A1, B1 and B2, 80% of in-hospital mortality

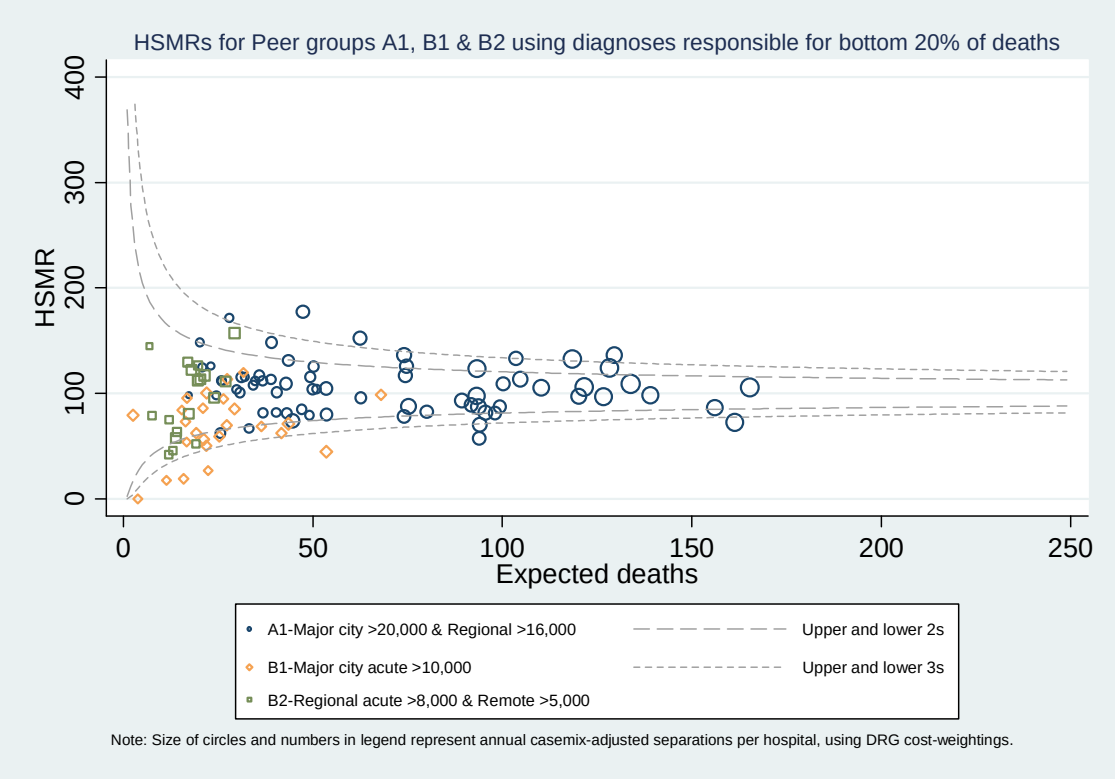


Figure 14: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group A1, B1 and B2, 20% of in-hospital mortality

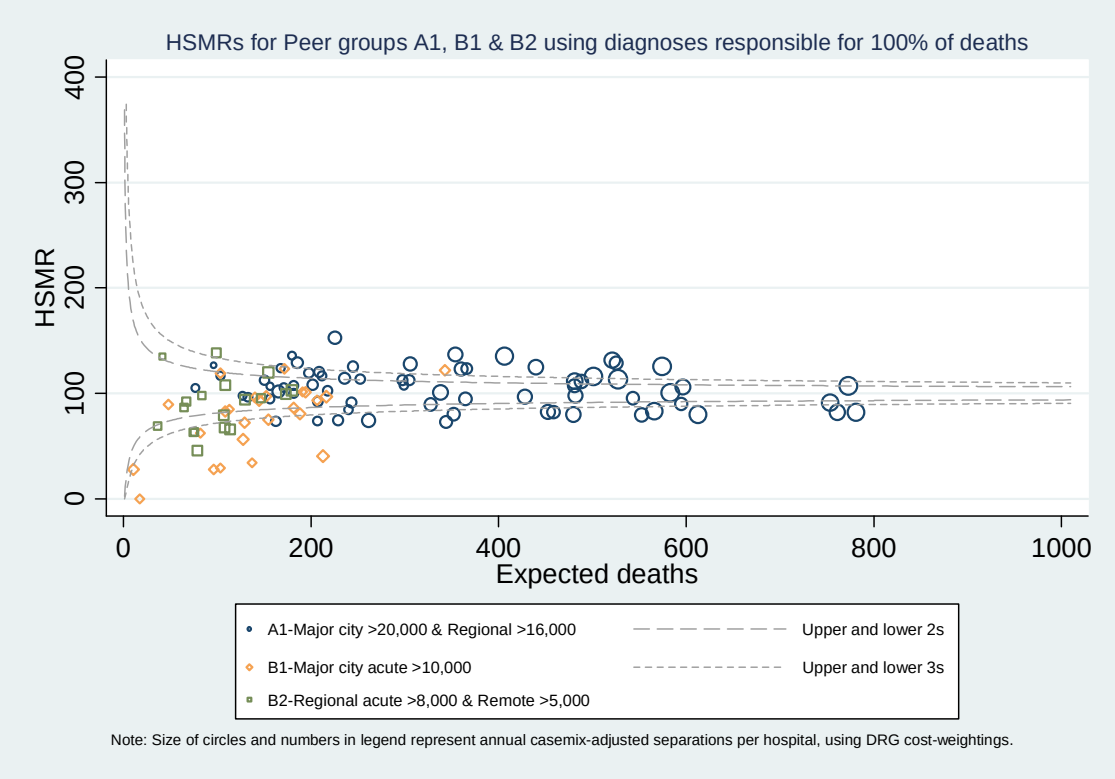


Figure 15: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group A1, B1 and B2, 100% of in-hospital mortality

5.8 Model development

The RACM model only includes untransformed values of variables and main effects. This is not necessarily the best way to model the data (see Section 4.8).

Fractional polynomials suggested the best powers of age for the transformation of age were age (i.e. a linear term) and age cubed. The Akaike information criterion (AIC) reduced from 266865.8 (80 df) to 266183.2 (79df) ($p < 0.001$). Table 17 displays the observed and expected deciles of risk for three different models: the standard RACM model, the full interaction model using the 50% developmental model data set (random sample of 50% of the 2005–06 data) and the full interaction model using the validation data set (with the remaining 2005–06 data).

Table 17: Observed and expected deciles of risk for 3 different models

	Model without interactions			Full model with interactions			Full model applied to 50% sample		
Decile	Obs	Exp	sqrt((obs-exp)^2/exp)	Obs	Exp	sqrt((obs-exp)^2/exp)	Obs	Exp	sqrt((obs-exp)^2/exp)
1	29	67.9	4.7	2	8.6	2.25	4	3.8	0.08
2	64	167.8	8.0	18	32.3	2.52	8	15.0	1.80
3	164	315.5	8.5	72	89.5	1.85	46	41.7	0.67
4	302	558	10.8	185	221.4	2.45	101	105.6	0.44
5	631	926.5	9.7	504	531.5	1.19	251	261.9	0.67
6	1,418	1,560.4	3.6	1,178	1,221.4	1.24	609	611.5	0.10
7	2,654	2,653.1	0.0	2,640	2,500.7	2.79	1,259	1,259.7	0.02
8	4,874	4,491.9	5.7	5,030	4,897.5	1.89	2,573	2,453.7	2.41
9	9,171	8,112.5	11.8	9,511	9,507.4	0.04	4,755	4,746.8	0.12
10	21,918	22,371.4	3.0	22,306	22,435.7	0.87	11,109	11,177.4	0.65
			65.9			17.08			6.96

The model fit for the standard RACM model was $Chi^2 = 65.9$, 8df, $p < 0.001$ and the fit increased substantially with the ERM model using the 50% 2005–06 validation sample data set ($Chi^2 = 6.96$, 10df, $p = 0.73$). Not only does the ERM produce better fit overall, but the residual differences between observed and expected deaths are spread more evenly over risk deciles than when the RACM model is used (Table 17). Figure 16 demonstrates that that the observed and predicted proportions of mortality fit well for all deciles.

HSMRs were calculated for the 80% mortality outcomes for the A1 hospital peer group. For the sake of comparison, the RACM model was re-run, placing the primary diagnoses in risk decile groups but otherwise leaving the model as is. HSMR plots are provided using the ERM model, the modified RACM model, and the RACM model as previously described (Figure 17).

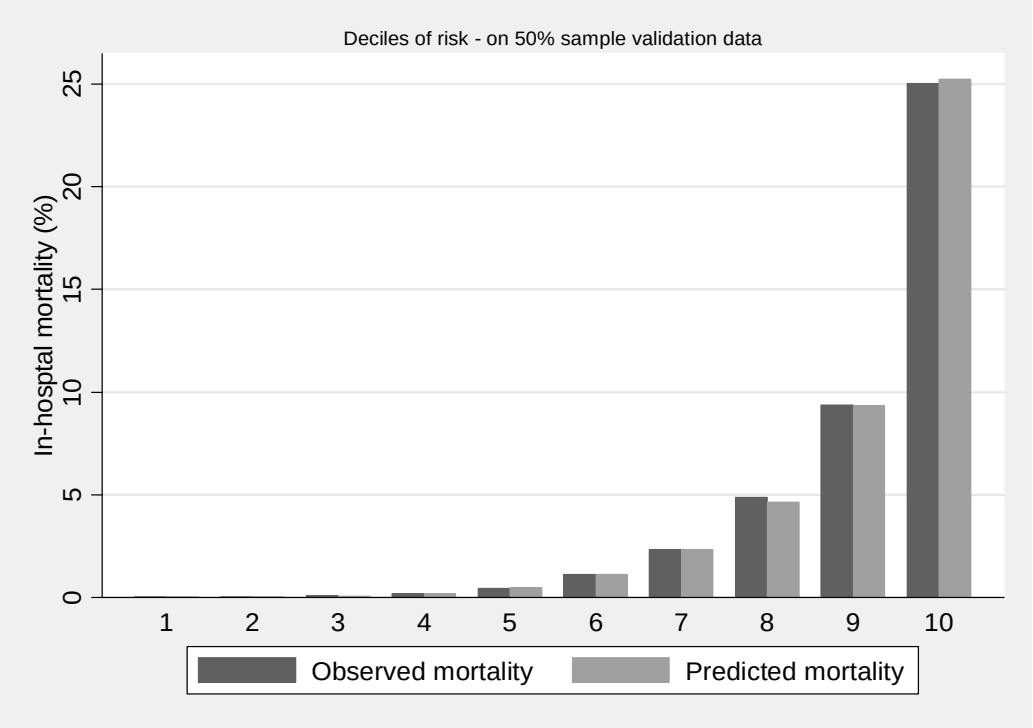


Figure 16: Observed and predicted proportions of mortality by deciles of risk

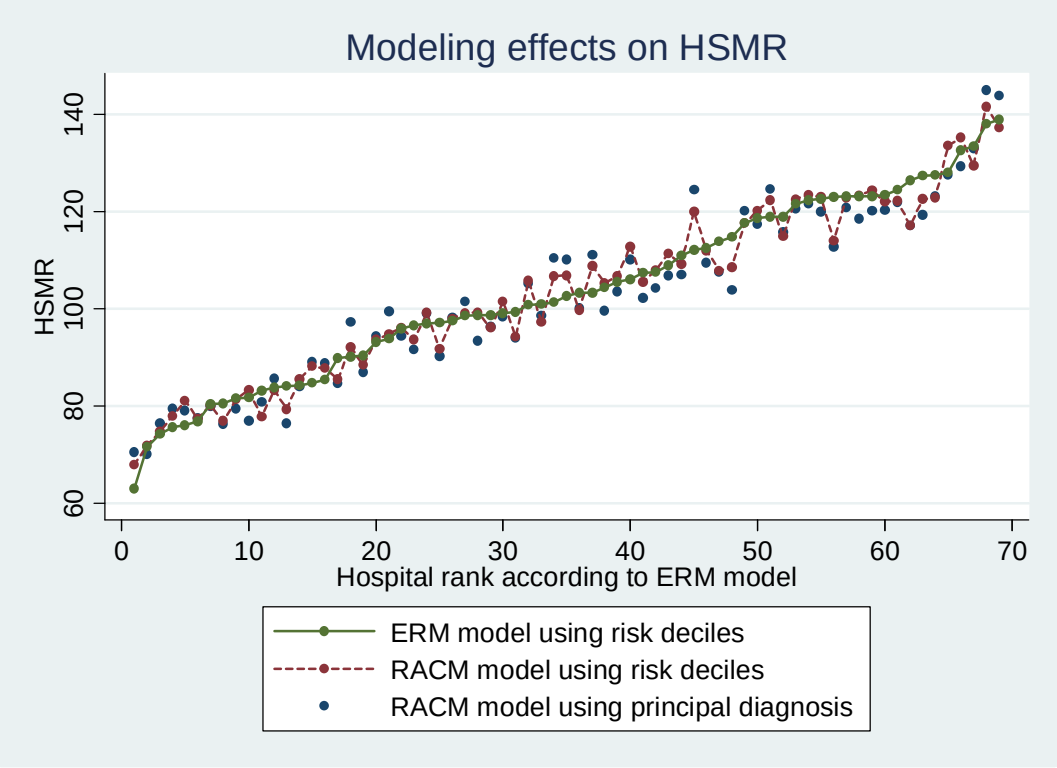


Figure 17: HSMR plots using the ERM model, the modified RACM model, and the RACM model

5.9 Inclusion of SEIFA

When the SEIFA index of socioeconomic status was included as a five-category variable in the standard RACM model, it was found to be a significant predictor of in-hospital mortality (LR test : $Chi^2 = 29.13$, 4df, $p < 0.001$). However, the change in the pseudo R^2 statistic was only marginal (from 0.2459 to 0.2460). The effect of increasing quintiles of SEIFA on the odds of in-hospital mortality compared with the odds for in-hospital mortality for the first SEIFA quintile are shown in Table 18.

Table 18: Effect of increasing quintiles of SEIFA[(a)] on the odds of in-hospital mortality

SEIFA quintile	Odds ratio	Std. Error	z	P	LCI	UCI
Most disadvantaged	1.000	–	–	–	–	–
Second most disadvantaged	1.029	0.016	1.86	0.064	0.998	1.061
Middle quintile	0.992	0.017	–0.46	0.648	0.961	1.025
Second most advantaged	0.971	0.017	–1.71	0.087	0.938	1.004
Most advantaged	0.942	0.017	–3.39	0.001	0.911	0.975

(a) Based on the ABS's SEIFA 2001 Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) score for the statistical local area of the patients area of usual residence (ABS 2004).

5.10 Longitudinal analysis

In addition to applying the RACM and ERM models to a single year of hospital separations data, we undertook a longitudinal analysis of data for that year (2005–06), the year before and the year after. The longitudinal analysis has been undertaken to demonstrate the feasibility of basing this approach on Australian data.

As discussed in the literature review (see Section 2.7.2), longitudinal studies are of considerable importance for confirming the presence of systematic variations in mortality outcomes, and for assessing the extent to which a data source provides information on in-hospital mortality, rather than 'noise'.

Reliance solely on cross-sectional comparisons of performance would miss patterns such as hospitals whose rates remained static although there was a general trend towards improvement, or hospitals whose results improved or deteriorated to an important extent over time, despite the absolute mortality rates for the hospitals not deviating enough form group means to attract attention on cross-sectional study.

This section provides information on the method employed and the results of the analysis of data covering the 3–year period 2004–05 to 2006–07.

We used a method based closely on that reported by Heijink (2008). This is a two-step analysis, outlined here and described fully below.

The first step is logistic regression modelling. As before, this was done to reduce variation among hospitals due to different case profiles (i.e. risk adjustment). We used the same modelling approach used for the single-year study (i.e. RACM).

The second step is two-stage multi-level logistic regression. This was done to explain remaining variation of risk-adjusted HSMRs within and between hospitals—especially variation over time.

Following Heijink, we did this analysis on the high-risk (80%) case group.

5.10.1 Method

Data

This analysis uses data for hospital separations that occurred in Australia from 1 July 2004 to 30 June 2007. As in the single-year analysis, the data were provided by the AIHW from the NHMD.

Institution mapping

A longitudinal analysis of this nature depends on tracking individual hospitals over time. Unfortunately, this is not as simple as it sounds. Hospitals merge, change ownership, change their names, and change from public to private and vice versa. No 'map' was available to track these changes. In the absence of an available map we made one to cover the 3–year period under study.

We obtained from the AIHW website tables that listed, for each data year, hospital names, establishment identifiers and several other characteristics, including average available beds, peer group code and regional designation. We used these tables, in conjunction with establishment identifier codes in the NHMD data, to construct the map. Many hospitals were easy to map: names and establishment IDs remained identical over the 3 years. Many others had some differences, which were assessed carefully. Establishments for which mapping doubt remained were omitted from the analysis. Private hospitals were generally not identified separately in the NHMD, and were not in the tables, and could not be included in this part of the analysis.

Of the 856 hospitals identified in the three data years, 736 were matched across all 3 years and retained for the longitudinal analysis. Each of these hospitals was assigned a study identifier, which was used in this part of the analysis.

Case selection, peer groups and modelling

Exclusion criteria for years 2004–05 and 2006–07 were applied as for the single-year analysis described above (Section 4.5). Records meeting the following criteria were selected from the three annual files:

1. hospital establishment identifier was one of the 736 that were mapped over the 3 years
2. Principal Diagnosis code was one of those in the high-risk group (These codes are listed in Appendix A1.)
3. the hospital was in one of the peer groups A1, A2, B1, B2, C1, C2, D1, D2 or D3.

These exclusions reduced the number of cases for analysis to 2,012,302.

A logistic regression model for in-hospital mortality above was created using the following covariates: age, sex, length of stay, elective/emergency status, principal diagnosis, Charlson index and transfer status. Modelling followed the RACM method described above for the single-year analysis. Model coefficients were determined using the first year of data (2004–05). These coefficients were then applied to each record in each of the data years 2004–05 to 2006–07 to generate a probability of death. The sum of these values for all records belonging to a hospital gave the expected number of deaths for that establishment. This was done separately for each year.

HSMRs for each year were then calculated by dividing the observed number of deaths by the expected number of deaths for each hospital and for each year. An HSMR was calculated for each of the 3 years for 418 hospitals with a peer group of A1, A2, B1, B2, C1, C2, D1, D2 or D3. Overall HSMRs for each of these peer groups were also calculated (Table 19).

Following calculation of annual HSMRs for these 418 hospitals, a two-stage multi-level linear regression model was developed in order to assess any systematic change in HSMRs over time, and also the within-hospital correlation of HSMRs over time.

Multi-level models partition the variance of the data into fixed and random effects. Fixed effects for our models were the overall mean HSMR in 2004–05 and the decrease in HSMR for each of the following 2 years. Random effects were the overall variance in HSMRs across hospitals (denoted in the results as 'random intercept for hospitals'), the variance in the slopes of HSMRs across time ('random slopes for hospitals') and the covariance (i.e. degree of correlation) between the random intercept and the random slopes.

The correlation across time for hospitals was assessed using the intraclass correlation coefficient (ICC), which is defined as the ratio of the (level 2) between-hospital variance (random intercept for hospitals) and the total hospital variance (random intercept for hospitals plus the (level 1) within-hospital variance). A high degree of correlation indicates that compared with between-hospital variation, within-hospital variation across time is small.

Observed and model-predicted HSMRs were also plotted across time to allow visual assessment of the data. The model-predicted HSMRs incorporate the fixed and random effect components of the model, but not the unexplained (level 1) within-hospital variation (i.e. residual variation not explained by the modelling). The model-predicted HSMRs can therefore be thought of as depicting the explained (i.e. systematic) variance in the HSMRs.

5.10.2 Results

The 3–year analysis was done to demonstrate an approach to longitudinal analysis of in-hospital mortality, and to examine the adequacy of Australian hospital morbidity data for this purpose.

The overall HSMRs for the whole data for the first year (2004–05) is, by definition, 100 (95% CI= 99–101). The overall HSMR declined to 98.6 (95% CI= 97–100) for the second year (2005–06) and to 95.5 (95% CI= 94–97) for the third year (2006–07).

The annual mean HSMRs for each peer group are presented in Table 19. Because the logistic regression modelling was built using data from all hospitals combined (rather than being stratified by peer group), the first-year HSMRs are not set to 100—revealing differences between the groups. The effect of applying a model derived from all cases to very different types of hospital is particularly evident for peer group A2, WCHs.

Looking across the rows, it can be seen that there was a tendency for HSMRs to decrease over time for peer groups A1, A2, B1, C2 and D2.

The results of the multi-level modelling of HSMRs are shown in Table 20. Although HSMRs for most groups decreased across time, the only significant decreases in HSMR after 2004–05 were for peer group A1 in 2006–07 (–6.3, 95% CI = –9.9 to –2.6, $p < 0.001$) and for peer group C2 in 2006–07 (–18.0, 95%CI = –35.6 to –0.5).

The ICC values are high for most of the peer groups, indicating that within-hospital variation between the 3 years is small in relation to between-hospital variation.

Table 19: Mean HSMRs (and 95% confidence intervals) by financial year and peer group

	Financial year		
Peer group	2004–05	2005–06	2006–07
A1	104.3 (98.8,109.7)	102.6 (98.0, 107.1)	98.0 (93.0, 103.0)
A2	201.3 (87.8, 314.8)	168.5 (74.3, 262.8)	167.0 (72.5, 261.6)
B1	80.4 (67.2, 93.5)	78.1 (63.9, 92.4)	77.4 (65.0, 89.9)
B2	96.2 (80.4, 112.1)	90.7 (76.7, 104.6)	96.2 (82.6, 109.8)
C1	68.6 (55.3, 81.9)	75.8 (60.4, 91.2)	68.4 (54.1, 82.7)
C2	107.0 (86.5, 127.5)	96.8 (83.9, 109.7)	88.9 (78.3, 99.6)
D1	133.8 (111.7. 156.0)	133.0 (117.3, 148.7)	136.6 (122.0, 151.2)
D2	119.9 (102.8, 136.9)	120.9 (102.3, 139.4)	108.0 (93.5, 122.5)
D3	98.2 (71.0, 125.4)	100.6 (84.1, 117.1)	106.3 (80.5, 132.1)

Another way of presenting this information is provided in Figures 18 to 20.

The pair of charts in each row represents one of the peer groups included in the longitudinal part of the study. The thick line in each chart presents the peer-group mean HSMRs for each year (like the values in Table 19). Each of the dashed lines represents one of the hospitals in the peer group. The chart on the left in each pair ('Observed') shows the risk-adjusted HSMRs as calculated by applying the logistic regression model based on 2004–05 data to this year and to each of the other years. The other chart in each pair ('Predicted') displays the risk-adjusted HSMRs predicted by the multi-level model.

The more linear each hospital line is across the 3 years, the less variation there is within that hospital across time. As a consequence, the relative contribution of between-hospital variation in HSMRs to the total variation is higher and, by definition, the ICC is therefore higher too.

The difference in HSMRs between the two charts demonstrates the amount of residual variation in the HSMRs that cannot be explained by the multi-level models. Note that the vertical scale differs between charts.

These results are generally similar to those reported by Heijink et al. (2008), whose approach we followed. Like them, we found a downward trend in overall risk-adjusted HSMR, and that variation was mostly between-hospitals, not within hospitals.

The main difference between Heijink et al. (2008) and our analysis is their examination of a wider range of covariates as predictors of in-hospital mortality. The satisfactory performance of the method when applied to Australian hospitals data suggests that it will be fruitful to extend our analysis in a similar way. Exact replication is unlikely to be feasible, because some of the covariates used by Heijink et al. may not have direct Australian equivalents, due to differences in health system organisation and health information. However, data on some other potential covariates may exist in Australia.

It should be recognised that that these are results of a demonstration analysis. Although they offer support for the view that Australian hospital morbidity data provide an adequate basis for calculation of indicators of in-hospital mortality, caution should be taken not to over-interpret these results, which have some limitations.

The analysis presented here is based on only 3 years of data. That was enough to allow us to test the extent to which Australian hospitals data provide 'signal' rather than 'noise' in hospital-level HSMRs. Subsequent analyses will benefit from the use of data for a larger number of years.

The analysis presented here is for only one of the three indicators defined in Section 4.5.2: namely the indicator restricted to the group of Principal Diagnoses associated with the highest number of in-hospital death, and which together account for 80% of all in-hospital deaths.

As explained above, the lack of a 'map' led to the omission of some public hospitals. Many private hospitals could not be included, due to the lack of hospital-specific identifiers in the NHMD.

Table 20: Fixed and random effects and intra-class correlation coefficients for the multi-level models

	Hospital peer groups								
	A1	**A2**	**B1**	**B2**	**C1**	**C2**	**D1**	**D2[(a)]**	**D3**
	(*n* = 61)	**(*n* = 9)**	**(*n* = 25)**	**(*n* = 18)**	**(*n* = 27)**	**(*n* = 59)**	**(*n* = 103)**	**(*n* = 80)**	**(*n* = 36)**
Fixed effects									
Constant (group mean for 2004–05)	104.3 (2.6)	201.3 (46.8)	80.4 (6.6)	96.2 (7.2)	68.6 (6.7)	107.0 (10.0)	133.8 (10.2)	119.9 (8.4)	98.2 (12.3)
Year									
2004–2005	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2005–2006	–1.7 (1.6)	–32.8 (21.6)	–2.2 (2.8)	–5.6 (4.9)	7.2 (4.3)	–10.2 (5.5)	–0.8 (9.5)	1.0 (8.2)	2.3 (12.9)
2006–2007	–6.3† (1.9)	–34.3 (22.3)	–2.9 (2.8)	0.0 (6.1)	–0.2 (5.0)	–18.0* (9.0)	2.8 (10.1)	–11.9 (8.2)	8.1 (16.0)
Random effects									
Level 1 variance	71.7 (13.)	2,051.4 (725.3)	99.8 (20.4)	177.0 (60.7)	222.8 (61.8)	379.8 (70.5)	4,444.8 (440.1)	2,662.8 (299.6)	2,459.4 (587.9)
Level 2 variances									
Random intercept for hospitals	435.0 (114.1)	20,259.1 (12,363.7)	1,021.5 (357.3)	1,032.1 (515.5)	1,037.7 (455.4)	10,604.5 (2,140.1)	9,745.1 (2,660.9)	3,027.5 (630.9)	6,350.3 (3198.8)
Random slope for hospitals	17.8 (11.8)	90.3 (223.8)	0.5 (2.0)	78.0 (64.7)	56.9 (56.0)	995.3 (222.9)	379.9 (216.3)	–	1,078.3 (625.1)
Covariance of random slope and intercept	–52.3 (31.1)	–1,352.7 (1,981.8)	–21.7 (49.1)	–181.4 (159.4)	–51.7 (129.6)	–3,048.9 (665.2)	–1,924.0 (792.1)	–	–2,239.1 (1,331.7)
Intraclass correlation coefficient	0.86	0.91	0.91	0.85	0.82	0.97	0.69	0.53	0.72
–2 x log-likelihood	1,471.2	282.9	626.4	469.5	734.0	1,791.8	3,582.6	2,672.2	1,189.9

(a) A random-intercept only model was used for peer group D2 due to non-convergence with a random intercepts and random slopes model.
* $p < 0.05$, † $p < 0.001$ versus 2004–2005. Figures in brackets denote standard errors.

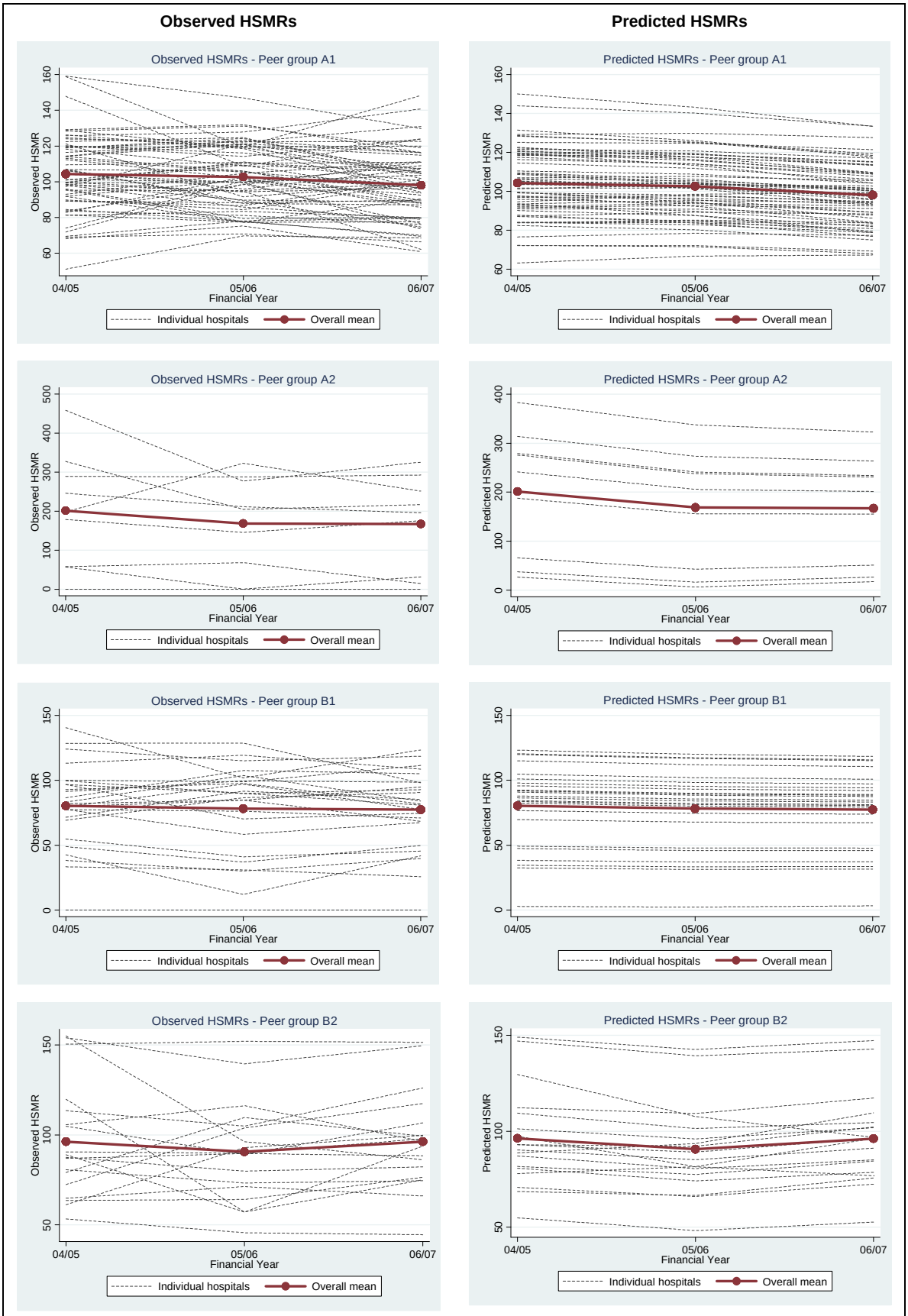


Figure 18: Observed and predicted hospital-specific and group mean HSMRs by financial year and peer group: peer groups A1, A2, B1 and B2

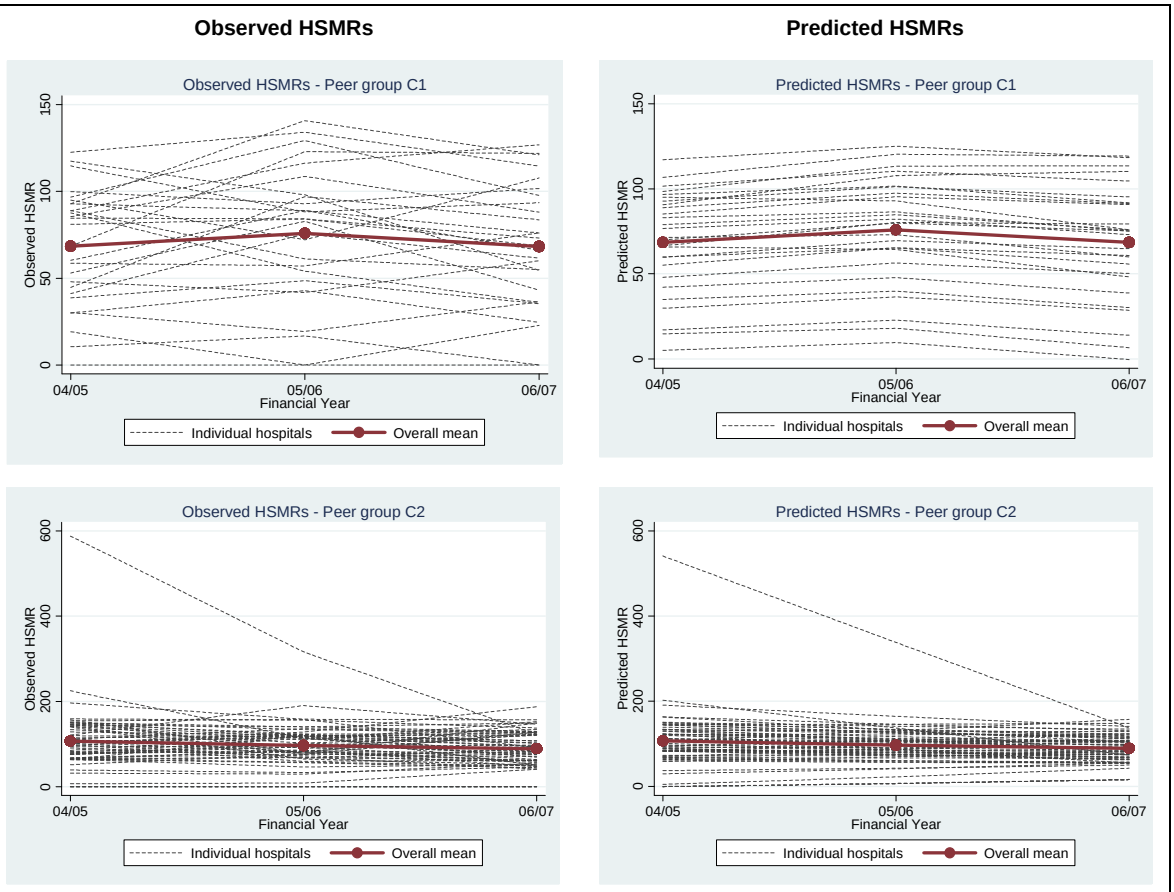


Figure 19: Observed and predicted hospital-specific and group mean HSMRs by financial year and peer group: peer groups C1 and C2

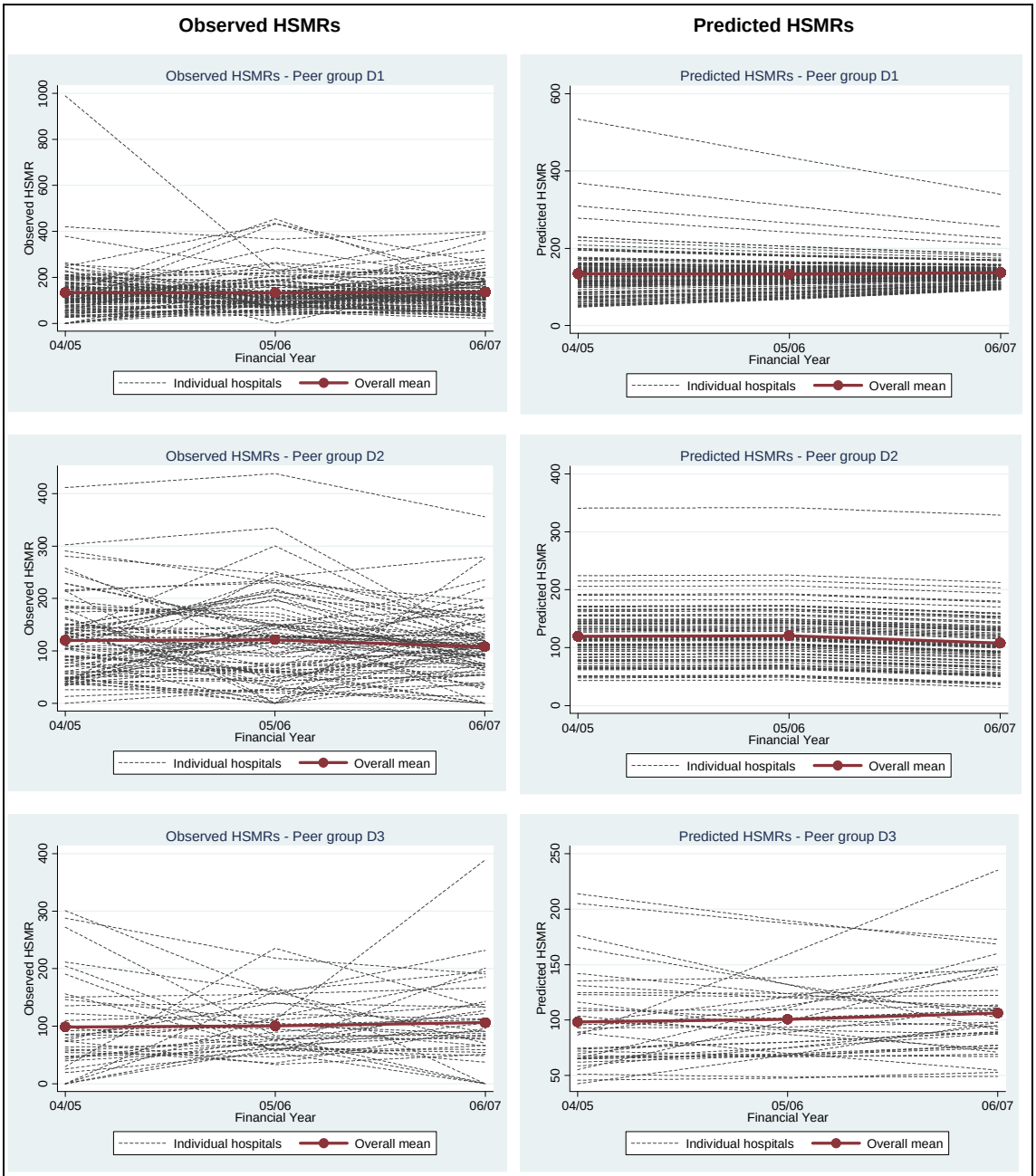


Figure 20: Observed and predicted hospital-specific and group mean HSMRs by financial year and peer group: peer groups D1, D2 and D3

6 Discussion

6.1 Can we produce in-hospital mortality indicators using Australian administrative data?

This study demonstrates an approach to specifying a set of indicators of in-hospital mortality and calculating values for them using currently available Australian administrative data, from the NHMD. The analytic approach was based on findings of a review of relevant literature.

The work demonstrates the technical feasibility of producing indicators of in-hospital mortality now using national data. In particular, the section based on longitudinal analysis of 3 years of data provides support for the position that current Australian morbidity data are largely of adequate quality to support this type of use.

The particular indicators specified here are not the only ones possible. However, they exemplify major types of indicators: namely those focusing on in-hospital mortality among relatively high-risk cases, those focusing on in-hospital mortality among low-risk cases, and an overall group including all cases and in-hospital deaths. They are general-purpose indicators, rather than indicators specific to particular types of diagnosis, treatment or service. They should be applicable to a wide range of hospitals, though probably less so for some (e.g. specialised hospitals with an atypical casemix, such as WCHs).

The present study was based on current holdings of Australian hospital separation data, and this was considered to provide a sufficiently robust basis for the current program of work. Our assessment of the literature, as reviewed in Section 2.4.2, led us to conclude that it may be preferable to include deaths occurring soon after discharge from hospital, and that death within 30 days of discharge is a suitable criterion. At the national level, this data set has not been routinely linked to other major national databases such as the National Death Index, although several jurisdictions have undertaken such linkages at state level demonstrating its feasibility. As we have stated, we were not able to apply this aspect of indicator definition because we did not have access to linked hospital separations and deaths data, but it is likely that this will become feasible in the near future.

The following sections present further discussion of these and other points raised by this project.

Model assessment

The model parameters generated by applying the RACM (the de facto international standard logistic regression model for in-hospital mortality) to Australian data are very similar to those reported in the international literature. The model shows good discrimination (in terms of the c-statistic, 0.87 for the high-risk 80% mortality group). As reported elsewhere, the explanatory power of the model, as indicated by pseudo R^2 values, although seemingly modest (0.24 for the 80% set), were consistent with the international literature and typical for logistic regression models that compare the fitted model with the null model, in which none of the variation is explained. (This differs from the situation with linear regression, where comparison is with the saturated model, in which

100% of variation is explained, and where higher R^2 values are often obtained.) The c-statistic and pseudo R^2 values were higher for the 20% model (0.96 and 0.34, respectively) and the 100% model (0.95 and 0.35, respectively). In these sets, subjects had been grouped according to deciles of mortality risk—based on primary diagnosis. Although this technique guarantees an increase in discrimination and explanatory power, the change in pseudo-R^2 values and c-statistics with the inclusion of the deciles were similar to the changes in these statistics with the inclusion of the primary diagnosis groups for the 80% models.

The Hosmer–Lemeshow test did not demonstrate good fit for any of the RACM models. However, as has been discussed previously, the Hosmer–Lemeshow goodness of fit method is sensitive to the very large sample sizes used here, and the RACM model is not sophisticated. The graphic plots of deciles of observed and expected risks show that the RACM model fit is closer for the deciles of higher risk than for the lower deciles, where the model seems to over-predict expected mortality. This exemplifies the problems of fitting graduated risks in relation to outcomes that can have only one of two values: alive or dead (Chassin et al. 1996).

The unevenness across the deciles of risk is likely to be much less important for large hospitals, with large volumes of patients and larger number of both observed and expected deaths, than for smaller institutions. Indeed, the caterpillar plots for the A1 peer group hospitals show HSMRs that vary steadily across a substantial range, and demonstrated that there are large hospitals with HSMRs with narrow confidence intervals that have mortality rates that are significantly below, or significantly above, the national average for that peer group. But the unevenness, which is more marked for peer groups including hospitals with lower case volumes, further confirms the inappropriateness of simply rank ordering the hospitals from end to end, rather than looking for outlier groups and institutions.

Although the HSMRs for the B1 peer group hospitals are within realistic boundaries, those for the small hospitals in C2 and D1 are hard to interpret. The small number of both observed and expected deaths generates HSMRs in those groups that range from 0 to over 300—some with very wide confidence intervals.

Analysis by peer group

Unadjusted HSMRs should be expected to differ between hospitals because of their different casemix. Adjustment models, such as those presented in this report, do much to overcome differences in casemix, but institutional level differences remain. Hospitals have been divided into peer groups to enable comparisons of like with like. Peer groups may also provide a useful basis for assessment of in-hospital mortality. The analyses presented in this report are based on overall models—based on all hospitals. It would also be possible to make peer-group-specific models, at least for the groups that treat sufficiently large numbers of patients. In any case, interpretation of HSMRs across peer-group boundaries should be undertaken cautiously.

Small hospitals

The dilemmas posed by small hospitals are substantial. In-hospital death is a relatively rare event in many of those settings, and mortality rates are likely to be subject to many extraneous influences related to the casemix of those hospitals, and to opportunities for end-of-life care in rural and remote regions.

The most straightforward way to deal with small hospitals is to exclude them from mortality monitoring: setting some mortality rate criteria (e.g. at least 50 deaths in any one of the three previous financial years, or some other mortality threshold yet to be determined). Other simple approaches to dealing with sparse data include enlarging the reporting period for small hospitals (e.g. calculate HSMRs for a rolling 2– or 3–year period) or reporting HSMRs only for clusters of small hospitals. If none of these is deemed sufficient, then analysis using a Bayesian method that creates shrunken estimates—i.e. estimates of the HSMR which are shifted towards a value obtained from known information about other hospitals (known as a 'prior' probability)—could be developed for consideration by the Commission.

Refined risk-adjustment model (ERM)

This discussion has so far revolved around findings based on the de facto international standard risk-adjustment model (RACM). The modest fit of this model prompted us to consider whether it could be improved. We developed a more refined risk-adjustment model (labelled the ERM model in this report) that allowed for the possibility that some variables, such as age, were not simply linear in relation to mortality risk. The model also allowed for interactions between the modelled variables (we found significant interactions for all the major variables modelled). We acknowledge helpful advice from Professor DW Hosmer in the course of this work.

The ERM model displayed a number of technically more acceptable characteristics. The model fit was a substantial improvement over the RACM model, and the residual differences between observed and expected mortality were spread much more evenly across the risk deciles.

At this point, we have not gone on to analyse all peer groups for every combination of mortality using the ERM modelling, and the ERM analysis is provided for the sake of comparison. Technically, it is a superior model and the improvements in model fit justify its further development. If, however, there is a concern that any Australian study should follow work done internationally, then the Commission may want to continue with the RACM model, despite its poorer performance. On a practical note, the large number of interactions that are computed within the ERM model make major demands on computing power. Interested groups lacking access to powerful desktop computers can expect long processing times to compute the ERM models.

SEIFA and other factors

The finding that a measure of social deprivation (SEIFA) did not add substantially to the discriminatory power of the risk-adjustment modelling is ambiguous. It might reflect somewhat flatter social gradients within the Australian population than in settings in which socioeconomic variables have been found to be influential—at least in relation to access to health care. However, it could also reflect insensitivity of SEIFA to relevant aspects of deprivation or other social determinants of health. Conversely, it could be the case that variables in our model took some account of any such differences. Aboriginal and Torres Strait Islander peoples, as a group, have well-known excess early mortality and other characteristics of poor health status. We did not examine the practicability of examining this subgroup separately in the present study. Although it would be possible to make such an examination, we anticipate that relatively small case numbers and uncertain identification of Indigenous status in the NHMD would be important constraints.

Longitudinal analysis

We demonstrated the feasibility of longitudinal analysis of in-hospital mortality in Australia using NHMD data for 3 years. Analysis showed that most of the variation in HSMRs was between hospitals, not within hospitals, suggesting sufficient data quality and stability in hospital specific HSMRs to provide a basis for indicators.

The results presented for the 3–year analysis show a modest decline in risk-adjusted mortality over a 3–year period. There is some very tentative indication that this kind of pattern may be emerging elsewhere (e.g. Heijink et al. 2008, Kelman and Friedman 2007). Much more detailed work needs to be done to ensure that the trend is not an artefact of a number of different factors; for example, of coding changes (between and within jurisdictions or individual institutions), a reflection of the changing demography of hospital populations (hospital populations are not simply representative of populations as a whole), or an outcome of changing locations of places of death. Analysis using additional years of data will be a stronger basis for assessing trends. But the possibility remains that the trend is real. If so, it might be the case that an increased emphasis on hospital safety is beginning to have a demonstrable effect on hospital mortality, and is possibly of sufficient interest to warrant further study.

Methods of presentation

We have demonstrated three forms of presentation of HSMRs: tables, caterpillar plots and funnel plots. Each has distinct strengths and limitations.

Tables provide ready access to specific values for an institution or a group of institutions. However, the overall pattern of HSMRs is difficult to assimilate from a large table. Also, tabulated data, ranked by HSMR values, encourages unhelpful and statistically meaningless over-interpretation of the rank position of hospitals whose HSMR values do not differ significantly. For this reason, they are not preferred as a method for public dissemination of results.

Caterpillar plots provide a good overview of the range of HSMRs and of the associated confidence intervals. HSMRs for a population of hospitals tend to include many values in a 'middle range': not different from one another to a statistically significant extent (e.g. Figure 12). Caterpillar plots show this property rather clearly, especially if they are drawn in a way that gives at least as much visual emphasis to the confidence intervals as to the point estimates. They also show outliers, if present.

Funnel plots allow the identification of those small numbers of hospitals that are true outliers, with mortality results that are either much worse, or much better, than most hospitals. One limitation is that they do not facilitate comparison of non-outlier hospitals—a matter likely to be of interest to people responsible for each charted institution. Funnel plots are, perhaps, more difficult to interpret than caterpillar plots.

We conclude that although good use can be made of all three methods of presentation, the choice for public reports (if made) should be between the two forms of chart.

We suspect that many members of the public may find caterpillar plots easier to interpret than funnel plots. We are not aware of empirical data on this matter (though a study could certainly be done).

6.1.1 Indicators specified in the project

This report presents the results of a proof-of-concept project on the development of in-hospital mortality indicators based on existing Australian administrative data.

The three indicators specified in this study are intended to represent types of indicator described in the international literature, while also reflecting a pragmatic response to characteristics of the National Hospital Minimum Dataset and to the short time available for this project. The three indicators specified in this project are:

Indicator 1: High-risk group. This was specified as the Principal Diagnoses that accounted for 80% of in-hospital deaths in Australian hospitals, and had the highest number of cases of in-hospital mortality, in 2005–06.

Indicator 2: Lower risk group. This includes all Principal Diagnoses that are not included in the first indicator, and accounted for 20% of deaths.

Indicator 3: Indicator 3 includes all principal diagnoses. Thus, it includes all cases and all deaths.

The first of the three is an example of an indicator focusing on relatively high-risk conditions. Overall, the group of 68 Principal Diagnosis codes included in it account for less than one-fifth of all cases, but the cases selected by this criterion include four-fifths of all deaths in hospital. This type of indicator (i.e. including 80% of in-hospital deaths) is quite common in the literature.

Conversely, the second is an example of an indicator focusing on a lower risk set of conditions—i.e. diagnoses which, as a group, accounted for over 80% of cases, but 20% of deaths.

The third indicator includes all cases and all deaths.

Apart from Principal Diagnosis, we applied a single set of case inclusion criteria throughout the project. These are specified in Section 4.5.1. They are similar to those reported by other recent work of similar type (e.g. CIHI 2007, Heijink et al. 2008).

Table 21 provides a demonstration of how the three generic indicators specified in this report could be applied. In this instance, specific indicators are framed in terms of a generic indicator and a hospital peer group. A similar approach could, in principle, be applied to subsets of separations grouped in other ways. Examples are the types of diagnosis, types of service and types of procedures. However, formal statistical assessment of any such groups is necessary before practical feasibility can be assured. The large number of possible variations goes beyond the scope of this report. A cautionary observation is that this approach is limited by the (fortunately) relatively low probability of most types of admitted cases ending as an in-hospital death.

Although not done in this project, it is technically possible to produce summary HSMR values for regions, jurisdictions, or other groups of cases, in much the same way as peer group summaries were produced in this project (see Table 19).

Table 21: Application of derived indicators to hospital peer group

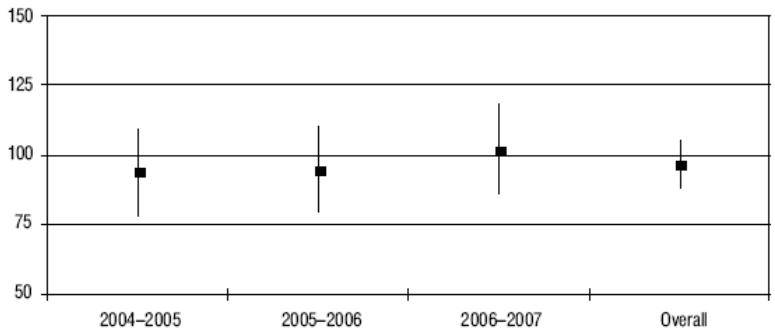
Indicator	Definition	Peer group description
Indicator 1a	Diagnoses that account for 80% of in-hospital deaths in peer group A1 Australian hospitals (high risk)	Major city hospitals with >20,000 acute casemix-adjusted separations and Regional hospitals with >16,000 acute casemix-adjusted separations per annum
Indicator 1b	Diagnoses that account for 80% of in-hospital deaths in peer group B1 Australian hospitals (high risk)	Major city acute hospitals treating more than 10,000 acute casemix-adjusted separations per annum
Indicator 1c	Diagnoses that account for 80% of in-hospital deaths in peer group C1 Australian hospitals (high risk)	Medium acute hospitals in Regional and Major city areas treating between 2,000 and 5,000 acute casemix-adjusted separations per annum, and acute hospitals treating <2,000 casemix-adjusted separations per annum but with >2,000 separations per annum
Indicator 1d	Diagnoses that account for 80% of in-hospital deaths in peer group D1 Australian hospitals (high risk)	Small Regional acute hospitals (mainly small country town hospitals), acute hospitals treating <2,000 separations per annum, and with less than 40% non-acute and outlier patient days of total patient days
Indicator 2a	Diagnoses that account for 20% of in-hospital deaths in peer group A1 Australian hospitals (high risk)	Major city hospitals with >20,000 acute casemix-adjusted separations and Regional hospitals with >16,000 acute casemix-adjusted separations per annum
Indicator 2b	Diagnoses that account for 20% of in-hospital deaths in peer group B1 Australian hospitals (high risk)	Major city acute hospitals treating more than 10,000 acute casemix-adjusted separations per annum
Indicator 2c	Diagnoses that account for 20% of in-hospital deaths in peer group C1 Australian hospitals (high risk)	Medium acute hospitals in Regional and Major city areas treating between 2,000 and 5,000 acute casemix-adjusted separations per annum, and acute hospitals treating <2,000 casemix-adjusted separations per annum but with >2,000 separations per annum
Indicator 2d	Diagnoses that account for 20% of in-hospital deaths in peer group D1 Australian hospitals (high risk)	Small Regional acute hospitals (mainly small country town hospitals), acute hospitals treating <2,000 separations per annum, and with less than 40% non-acute and outlier patient days of total patient days
Indicator 3a	Diagnoses that account for 100% of in-hospital deaths in peer group A1 Australian hospitals (high risk)	Major city hospitals with >20,000 acute casemix-adjusted separations and Regional hospitals with >16,000 acute casemix-adjusted separations per annum
Indicator 3b	Diagnoses that account for 100% of in-hospital deaths in peer group B1 Australian hospitals (high risk)	Major city acute hospitals treating more than 10,000 acute casemix-adjusted separations per annum
Indicator 3c	Diagnoses that account for 100% of in-hospital deaths in peer group C1 Australian hospitals (high risk)	Medium acute hospitals in Regional and Major city areas treating between 2,000 and 5,000 acute casemix-adjusted separations per annum, and acute hospitals treating <2,000 casemix-adjusted separations per annum but with >2,000 separations per annum
Indicator 3d	Diagnoses that account for 100% of in-hospital deaths in peer group D1 Australian hospitals (high risk)	Small Regional acute hospitals (mainly small country town hospitals), acute hospitals treating <2,000 separations per annum, and with less than 40% non-acute and outlier patient days of total patient days

6.2 How might in-hospital mortality indicators be used at different levels in Australia?

Countries that use in-hospital mortality indicators do so at different levels and for different purposes. For example the CIHI publishes HSMR trends by health region and hospital. The results are designed to be used by hospitals and health regions to monitor and understand their trends over time. An example of reporting of HSMRs by region is presented in Figure 21.

Newfoundland and Labrador
Western Regional Integrated Health Authority

Hospital Standardized Mortality Ratio, Excluding Palliative Care, 2004–2005, 2005–2006 and 2006–2007



	HSMR Excluding PC	HSMR Excluding PC 95% CI	HSMR All Cases	HSMR All Cases 95% CI
2004–2005	93	78–109	108	94–123
2005–2006	94	80–110	111	97–126
2006–2007	101	86–118	125	110–141
Overall	**96**	**88–105**	**115**	**106–123**

Hospitals included in the regional result:

- Bonne Bay Health Centre
- Calder Health Care Centre
- Dr. Charles L. Legrow Health Centre
- Rufus Guinchard Health Care Centre
- Sir Thomas Roddick Hospital
- Western Memorial Regional Hospital

Figure 21: Regional HSMR reporting by the Canadian Institute for Health Information (Taken from the report 'HSMR: a new approach for measuring hospital mortality trends in Canada.'CIHI 2007)

In Australia, in-hospital mortality indicators could be presented in a similar manner to that already used by the Canadians and others. The method of presentation will depend on the use to which the HSMRs are to be put. But, at every point, the HSMRs are always best considered as the starting points for further investigation rather than as definitive measures of a hospital's standing.

6.3 Are the in-hospital mortality indicators valid and reliable?

Validity refers to the extent to which a measurement truly measures what it is intended to measure.

If in-hospital mortality, *per se*, is the subject of interest, then the validation of indicators of the type specified in this report is relatively straightforward. Death is usually a well-defined event, though ventilators and other devices can complicate assessment. 'In-hospital death' is amenable to definition, though there is some room for ambiguity (e.g. how to treat cases of people who died while at a hospital, but had not been formally admitted, or cases where a person died before reaching a hospital, but was certified as dead after arrival?). However, the main issues are whether the available data sources are complete and reliable. These are amenable to study.

If hospital quality and safety is the subject of interest then the validation of the indicators is much more complicated. As discussed in Section 2.8.2, safety and (especially) quality are complex abstractions, which are difficult to define and measure.

The specific issue of the adequacy of administrative hospital separations data for risk adjustment could be subjected to formal study, along the lines of Aylin et al. (2007).

Reliability refers to the extent that a measurement method, if applied more than once under the same conditions, will give the same result. Repeated measurement of the same hospitals is emerging as a basis for assessing the reliability of measurements of in-hospital mortality (e.g. Heijink et al. 2008). This is based on the assumption that the true risk of in-hospital mortality in most hospitals is not likely to vary much from year to year, after adjustment for a small set of the characteristics of cases and provided that case numbers are sufficient to prevent small chance fluctuations in number of deaths from dominating results.

In this project, the 3–year analysis of in-hospital mortality, using indicator 1 (a relatively high-risk group of cases) produced the reassuring finding that within-hospital variation of HSMRs over the 3 years accounted for a generally low proportion of the total HSMR variation. In line with expectation, this was most true for groups of relatively large hospitals. Further work should be done to extend such analysis to other years, and other ways of selecting and grouping hospitals and cases.

6.3.1 Limitations

Mapping public hospitals over time

As discussed in Section 5.8.1 for the longitudinal analysis, a map was not available to track individual hospitals over time. We were able to develop a map for the purposes of this project; however, we were not able to include some hospitals (notably private hospitals) or to map some public hospitals. A map will be necessary for further longitudinal studies. Mapping is implicit in ongoing data linkage systems that include hospital separations data.

Problem of private hospitals

Private hospitals were largely excluded from this project, because they are not well-identified in the data source available to us (the NHMD). Although all records are marked as to whether the patient was in a private hospital, in many cases the information does not enable the private hospital cases to be grouped according to hospital, which was necessary for this project.

Limitations of time for project

The present project was undertaken in a short period of time. Although this did not present too great a challenge for the literature review, it did present significant challenges in the modelling aspects of the project. As indicated in Section 6.1, the long processing times to compute the models chosen for application had an impact on our ability to carry out much of the internal validation work necessary for these types of activities. With more time, we could have tried variations of indicators, applied them to hospitals grouped in additional ways, and done further development and evaluation of risk-adjustment models. It also ruled out time-consuming aspects of a more ideal study, such as attempting to obtain person-linked linked hospital and mortality data. Time constraints also had an impact on our ability to check the reliability and validity of the construction of our institutional map. We were also unable to fully explore the potential of longitudinal analysis.

6.4 Presentation and use of indicators of in-hospital mortality

6.4.1 How should in-hospital mortality indicators be presented?

In-hospital mortality indicators were presented in three different ways in the present report: as ranked tables, as caterpillar plots and as funnel plots. It is not possible to state explicitly what the best method of presentation is because any method will be governed by a number of factors including the purpose of the reporting, whether the material will be in the public or private domain, and the intended audience (experts or novices).

The main Australian example of publicly available hospital-specific reports including information on in-hospital mortality is from Queensland, where 'Measured Quality' reports are available via the Internet (e.g. <http://www.health.qld.gov.au/quality/measured_quality/2004/bay_redl.pdf>. These are extensive reports containing a great deal of information on many aspects of hospital

performance. The mortality data in the reports is presented as numerical values in tables, with peer-group values for comparison and use of symbols to indicate differences of statistical significance, and colour to mark values assessed to be outliers.

One of the specified outcomes for the National Indicators project is to 'enable the Commission to report publicly on the state of safety and quality'. From this we are able to assume that at least some in-hospital mortality indicators, if produced, should be presented in the public domain. With this assumption in mind we recommend that primary (see below) national in-hospital mortality indicators be publicly presented, in the main, as caterpillar plots. Caterpillar plots have the advantage of simplicity compared with funnel plots and are also less likely to encourage over-interpretation of small and non-significant differences than presentation in simple 'league tables'.

Caterpillar plots can be constructed and drawn in a range of ways, some of which will be more successful than others in communicating information on in-hospital mortality. We have provided some examples of ways to construct caterpillar plots to enable consideration of this issue (Appendix 4).

Further consideration of the method of presentation will be required when considering levels of disaggregation of HSMR analysis and presentation. Inclusion of numerous hospitals (at least 10 or so; preferably 20 or more) is needed to produce a plot recognisable as a caterpillar plot. Presentation of HSMRs concerning smaller groups of hospitals could follow the methods adopted by the CIHI (2007) (see, for example, Figure 21).

6.4.2 How should in-hospital mortality indicators be used in Australia?

As discussed in the conclusion to the literature review (Section 2.10), we recommend that in-hospital mortality indicators be used as screening tools, rather than being assumed to be definitively diagnostic of poor quality and/or safety. A screening tool is a signalling device. It is intended to signal that a problem may exist and that further detailed investigation is required.

6.5 What are the methodological obstacles to producing mortality indicators in Australia now?

6.5.1 Model checking and refinement

The models used in the project (RACM and ERM) will benefit from further scrutiny and refinement. We think that the general analytic approach is satisfactory, but there is room for improvement in its details. The ERM model demonstrates the possibility of improving on the RACM model. There may be potential to improve on the current ERM model, though we did not have sufficient time to exploit this possibility. Likewise, we have demonstrated the approach when applied to general-purpose indicators, including one (the high-risk set, including 80% of in-hospital deaths) that is now common in the international literature. There has not been an opportunity in the present project to explore the performance of the approach on indicators specified in other ways (there is an almost limitless number of possible ways).

We think that data in the NHMD offer potential to develop models that improve further on the already substantial improvement of the ERM model over the RACM model. For example, probability of in-hospital death is predicted better by some four- and five-character principal diagnosis codes than by their parent three-character codes, as used in this project. The extent of the potential improvement is not yet known, nor whether the gains in model performance would outweigh the added computational burden of analysis. Other potential enhancements include inclusion of additional socio-demographic characteristics (such as Indigenous status) and peer-group specific analysis.

Better understanding of some aspects of the data might also make a useful contribution. For example, 'admission' to hospital is a complex concept, particularly for emergency cases, and there are differences between hospitals in the point at which a patient is recognised as having been admitted. Such differences could influence whether certain cases involving death soon after arrival at a hospital are recorded in the NHMD.

Of great interest is the increased precision that may result from the inclusion in the models of national level coding of variables to show whether secondary diagnoses recorded for a case were present on admission (known as C-codes in Victoria). Risk adjustment is intended to adjust for patient-level variation in risk present at the point of admission, not for adverse events and other problems that occur during hospital stays. The latter should be sought out and analysed—not included in risk adjustment. Comparisons of the precision of risk-adjustment models with and without present-on-admission codes, and the impact of that coding on HSMRs will generate considerable interest locally and nationally, and will be a major contribution to the further development of measures of hospital safety.

6.5.2 Consultation

Consultation concerning indicators of in-hospital mortality is required with technical experts and stakeholders. Engaging key stakeholders in the finalisation of a 'standard Australian method' for producing in-hospital mortality ratios has the potential to improve on the methodological work reported here. An important step is to consult with state agencies and hospital groups: can they provide evidence of jurisdiction-level (or hospital-level) data issues that might influence findings and can be taken into account in models or risk adjustment? For instance, we have excluded records that were designated neither as elective nor as emergency. This third category may have different meanings in different jurisdictions. Because omission of palliative-care cases forms part of the approach that we have taken, possible differences in identification of such cases between jurisdictions or between hospitals would also benefit from scrutiny.

6.5.3 Suggested improvements to data collections

One of the National Indicators project objectives is to 'Enable the Commission to advise Ministers on whether existing reporting processes and collections should be continued, enhanced or replaced.' The NHMD has been demonstrated—at least in the context of this report—to be adequate for producing in-hospital mortality indicators. However, a small number of enhancements of the NHMD would contribute greatly to the usefulness of the NHMD for this, and other, purposes.

Data linkage

There are two aspects of data linkage that are particularly relevant to the production of in-hospital mortality indicators. The first is internal linkage within the NHMD and the second is external linkage of the NHMD with the National Death Index. We consider both forms of linkage to be vital enhancements to the NHMD to enable more valid and reliable in-hospital mortality indicators to be produced in Australia.

Variation in definitions and practices concerning hospital admission also has potential to influence measured in-hospital mortality. An argument akin to that concerning inclusion of deaths soon after discharge could be made for the inclusion of cases in which death occurs at a hospital, but before formal admission. Whether this would have an important effect on results is not known, but warrants investigation.

Internal linkage

At the present time, separations within the NHMD are not internally linked by person. Individuals—some with serious and persisting conditions—are likely to experience more than one episode of in-hospital care within a period covered by a study of in-hospital mortality. Without the ability to link related separations, it is not possible to be sure whether a person whose episode of hospital care ended with transfer to another hospital, or with a 'statistical type change', died during the next episode of inpatient care. Even a person who separates with discharge home might have been re-admitted soon after, with the possibility of fatal outcome of that episode. We are unable to take these factors into account when modelling because of the lack of internal linkage in the National Hospital Minimum Dataset to group the separations belonging to an individual patient.

External linkage

The second role of data linkage relevant to this type of work is linkage between hospital records and death registers (i.e. the National Death Index). This is necessary to enable studies that include deaths soon after discharge (i.e. to assess 30–day mortality).

Timeliness of availability of data

Reasonable expectations for timeliness of national indicators based on hospital inpatient data are not met at present. Although case records are generally processed, coded and accessible at state or territory level within a few months of separation, the NHMD file is released only annually, and records in it are from 1–2 years old by the time they become available for use. This prompts the question: can hospital morbidity data be made available more rapidly and frequently for purposes such as reporting indicators of hospitalised mortality?

Investigations into the feasibility of a more timely release of NHMD data—perhaps quarterly—should be considered.

Validation of coding

Mortality indicators depend on the reliability and quality of coding of hospital records. The most important variable for this purpose is Principal Diagnosis.

The quality of Principal Diagnosis coding is the subject of various coding audits in which a selection of records undergo independent recoding and the results are compared with the codes originally assigned. The Australian Coding Benchmark Audit—a method for auditing the diagnosis codes assigned to separation records in Australia—has been published by the

National Centre for Classification in Health (NCCH 2000). Neither this nor any other auditing method is mandated. The extent of auditing undertaken is difficult to assess; results are usually treated as confidential and are usually not published.

A second type of tool for quality checking is also exemplified by a product of the NCCH. Performance Indicators for Coding Quality (PICQ) (NCCH 2006). PICQ is software that screens coded records for compliance with the Australian Coding Standards for the ICD-10-AM and the Australian Classification of Health Interventions. It flags errors, and probable errors: allowing checking and recoding. This sort of tool can also be used to detect patterns of errors (e.g. a high prevalence of doubtful codes for records from a particular specialty in a hospital), which can be used to prompt investigation and corrective action. As with audits, application of such tools is not mandatory, and the extent of their use is unknown.

Introduction of indicators of in-hospital mortality is likely to heighten interest in the quality of the data on which they are based. Confidence in the indicators is likely to be enhanced by undertaking and publishing data-quality audits. An example of a project and study design that could be adapted for this purpose is the study of the quality of external causes coding in a sample of records from a sample of hospitals in four states, which has recently been undertaken by a team led by Dr Kirsten McKenzie of the Queensland University of Technology (a paper relevant to this point is in preparation but has not yet been published)

6.6 International benchmarking

In order to provide an accurate point of comparison with OECD countries, the model used to calculate in-hospital mortality should be consistent with the models and methods produced elsewhere. As yet, there is no internationally governed or stipulated standard practice for calculating HSMRs; however, the RACM model is consistent with how HSMRs are calculated in a number of different countries. The ERM model makes significant improvements to the RACM and we would suggest that, with proper peer-reviewed scrutiny and replication, it may be suggested as a potential candidate for an International standard.

Alternatively, the best performing model developed on the basis of Australian data could be used for national purposes. Additional analysis using a poorer-performing, but more widely-used, model (i.e. RACM) could be undertaken for the specific purpose of international comparisons.

6.7 Conclusion

The literature review in this report shows an emerging international consensus on best practice for national studies of hospital mortality, concerning a measure (the risk-adjusted HSMR factors to be included in risk-adjustment models, modelling methods, and types of cases to exclude (e.g. palliative-care cases). While discussion continues on the adequacy of administrative data for measuring in-hospital mortality, administrative data from good-quality systems appear to be adequate. In-hospital mortality rates are now reported regularly and publicly in several countries or jurisdictions within countries (United Kingdom, the Netherlands, Canada, and Queensland, Australia).

We applied two models: the most widely used approach, the RACM model, and the better-performing ERM model. This demonstrates that national indictors of in-hospital mortality

can be produced using the Australian NHMD, and that findings have statistical properties similar to those reported elsewhere.

A longitudinal study of 3 years of data—following the approach of a recently-reported national study of in-hospital mortality in the Netherlands—provides evidence suggesting that although some unexplained variation in risk-adjusted HSMRs remains after modelling, Australian administrative data provide a strong 'signal' related to hospital-specific values. The findings were similar to those reported for the Netherlands.

Although further work is required to confirm the findings of this project, to elaborate them (e.g. to review and refine specifications for indicator case inclusion) and to extend them to issues that we could not deal with in this study (e.g. data linkage to include deaths within 30 days), it appears that Australian hospital data—like data from Canada, England and the Netherlands—can be used to measure risk-adjusted in-hospital mortality.

Variations in hospital mortality appear to fulfil the necessary criteria to qualify as a performance measure. The questions that remain are exactly which indicators, used in exactly which ways.

The literature review pointed to the continuing uncertainty concerning the relationship between variations in hospital mortality and other measures of hospital structure and process. This does not argue against the use of mortality-based indicators. In our view, it does mean that variations in hospital mortality measures should be viewed as screening tests. High or rising HSMRs should not be assumed to be definitively diagnostic of poor quality or safety. Nor should low or declining HSMRs be assumed to mean that all is well. Such results produced by a screening tool signal that further investigation is warranted to understand "What goes on here?'

References

ABS (Australian Bureau of Statistics) 2004. Census of population and housing: socio-economic indexes for areas (SEIFA), Australia—technical paper, 2001. ABS cat. no. 2039.0.55.001. Canberra: ABS.

AIHW (Australian Institute of Health and Welfare) 2005. METeOR (Metadata Online Registry). AIHW, Australian Government. Viewed July 2008 <http://meteor.aihw.gov.au>.

AIHW 2007. Australian hospital statistics 2005–06. Number 30. Cat. no. HSE 50. Canberra: Australian Institute of Health and Welfare.

Ansari MZ, Ackland MJ, Jolley DJ, Carson N & McDonald IG 1999. Inter-hospital comparison of mortality rates. International Journal for Quality in Health Care 11(1):29–35.

Auerbach AD, Landefeld CS & Shojania KG 2007. The tension between needing to improve care and knowing how to do it. The New England Journal of Medicine 357(6):608–13.

Austin PC 2008. Bayes rules for optimally using Bayesian hierarchical regression models in provider profiling to identify high-mortality hospitals. BMC Medical Research Methodology 8(1):30–40.

Austin PC, Naylor CD & Tu JV 2001. A comparison of a Bayesian vs. a frequentist method for profiling hospital performance. Journal of Evaluation in Clinical Practice 7(1):35–45.

Aylin P, Bottle A & Majeed A 2007. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. British Medical Journal 334(7602):1044–51.

Berwick DM 2008. The science of improvement. Journal of the American Medical Association 299(10):1182–4.

Berwick DM & Wald DL 1990. Hospital leaders' opinions of the HCFA mortality data. Journal of the American Medical Association 263(2):247–9.

Best WR & Cowper DC 1994. The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients. Medical Care 32(4):390–400.

Birkmeyer JD, Dimick JB & Staiger DO 2006. Operative mortality and procedure volume as predictors of subsequent hospital performance. Annals of Surgery 243(3):411–7.

Birkmeyer JD, Siewers AE, Finlayson EVA, Stukel TA, Lucas FL, Batista I, Welch HG & Wennberg DE 2002. Hospital volume and surgical mortality in the United States. The New England Journal of Medicine 346(15):1128–37.

Bradbury RC, Stearns FE, Jr. & Steen PM 1991. Inter-hospital variations in admission severity-adjusted hospital mortality and morbidity. Health Services Research 26(4):407–24.

Campbell SM, Roland MO & Buetow SA 2000. Defining quality of care. Social Science & Medicine 51(11):1611–25.

Charlson ME, Pompei P, Ales KL & MacKenzie CR 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of Chronic Diseases 40(5):373–83.

Chassin MR, Hannan EL & DeBuono BA 1996. Benefits and hazards of reporting medical outcomes publicly. New England Journal of Medicine 334 (6):394–8.

Chassin MR, Park RE, Lohr KN, Keesey J & Brook RH 1989. Differences among hospitals in Medicare patient mortality. Health Services Research 24(1):1–31.

CIHI (Canadian Institute for Health Information) 2007. HSMR: A new approach for measuring hospital mortality trends in Canada. Ottawa: CIHI.

Coory M & Scott I 2007. Analysing low-risk patient populations allows better discrimination between high-performing and low-performing hospitals: A case study using inhospital mortality from acute myocardial infarction. Quality and Safety in Health Care 16(5):324–8.

Corben P, Fung CS & Lyle D 1994. Hospital-related mortality in NSW—preliminary risk-adjustment. New South Wales Public Health Bulletin 5(7):71–8.

Daley J, Iezzoni LI & Shwartz M 2003. Conceptual and practical issues in developing risk-adjustment methods. In: Iezzoni LI. Risk adjustment for measuring health care outcomes, 3rd edn. Chicago: Health Administration Press, 179–206.

Daley J, Khuri SF, Henderson W, Hur K, Gibbs JO, Barbour G, Demakis J, Irvin G 3rd, Stremple JF, Grover F, McDonald G, Passaro E Jr, Fabri PJ, Spencer J, Hammermeister K, Aust JB & Oprian C 1997. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. Journal of the American College of Surgeons 185(4):328–40.

DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S & Mark DB 1997. Comparing risk-adjustment methods for provider profiling. Statistics in Medicine 16(23):2645–64.

Devereaux PJ, Choi PTL, Lacchetti C, Weaver B, Schunemann HJ, Haines T, Lavis JN, Grant BJB, Haslam DRS, Bhandari M, Sullivan T, Cook DJ, Walter SD, Meade M, Khan H, Bhatnagar N & Guyatt GH 2002. A systematic review and meta-analysis of studies comparing mortality rates of private for-profit and private not-for-profit hospitals. Canadian Medical Association Journal 166(11):1399–406.

DoHA (Department of Health and Ageing) 2007. National Hospital Cost Data Collection (NHCDC) Peer Group Report Round 10 (2005–2006) AR-DRGv5.0. P3–2277. Canberra: Commonwealth of Australia.

Donabedian A 1966. Evaluating the quality of medical care. The Milbank Memorial Fund Quarterly 44(3, Part 2):166–203.

Dr Foster Intelligence 2007. How healthy is your hospital? Special Edition Hospital Guide. Dr Foster Research Limited. Viewed 1 May 2008. <http://www.drfoster.co.uk/hospitalguide>.

Dubois RW, Rogers WH, Moxley JH, Draper D & Brook RH 1987. Hospital inpatient mortality. Is it a predictor of quality? New England Journal of Medicine 317(26):1674–80.

Duckett SJ & Kristofferson SM 1978. An index of hospital performance. Medical Care XVI(5):400–7.

Duckett SJ, Coory M & Sketcher-Baker K 2007. Identifying variations in quality of care in Queensland hospitals. Medical Journal of Australia 187(10):571–5.

Ehsani JP, Jackson T & Duckett SJ 2006. The incidence and cost of adverse events in Victorian hospitals 2003–04. Medical Journal of Australia 184(11):551–5.

Fink A, Yano EM & Brook RH 1989. The condition of the literature on differences in hospital mortality. Medical Care 27(4):315–36.

Fleming ST, McMahon LF Jr, DesHarnais SI, Chesney JD & Wroblewski RT 1991. The measurement of mortality: a risk-adjusted variable time window approach. Medical Care 29(9):815–28.

Fung CH, Lim Y-W, Mattke S, Damberg C & Shekelle PG 2008. Systematic review: the evidence that publishing patient care performance data improves quality of care. Annals of Internal Medicine 148(2):111–23.

Garnick DW, DeLong ER & Luft HS 1995. Measuring hospital mortality rates: are 30–day data enough? Health Services Research 29(6):679–95.

Geraci JM, Johnson ML, Gordon HS, Petersen NJ, Shroyer AL, Grover FL & Wray NP 2005. Mortality after cardiac bypass surgery: prediction from administrative versus clinical data. Medical Care 43(2):149–58.

Gibbs J, Clark K, Khuri S, Henderson W, Hur K & Daley J 2001. Validating risk-adjusted surgical outcomes: Chart review of process of care. International Journal for Quality in Health Care 13(3):187–96.

Gilligan S. & Walters M 2008. Quality improvements in hospital flow may lead to a reduction in mortality. Clinical Governance: An International Journal 13 (1):26–34.

Glance LG, Osler TM, Mukamel DB & Dick AW 2008. Impact of the present-on-admission indicator on hospital quality measurement: experience with the Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators. Medical Care 46(2):112–9.

Goldman RL & Thomas TL 1994. Using mortality rates as a screening tool: the experience of the Department of Veterans Affairs. Joint Commission Journal on Quality Improvement 20(9):511–22.

Goldstein H & Spiegelhalter DJ 1996. League tables and their limitations: statistical issues in comparisons of institutional performance. Journal of the Royal Statistical Society: Series A (Statistics in Society) 159(3):385–443.

Gordon HS, Johnson ML, Wray NP, Petersen NJ, Henderson WG, Khuri SF & Geraci JM 2005. Mortality after noncardiac surgery: Prediction from administrative versus clinical data. Medical Care 43(2):159–67.

Green J & Wintfeld N 1995. Report cards on cardiac surgeons—assessing New York State's approach. New England Journal of Medicine 332(18):1229–33.

Green J, Passman LJ & Wintfeld N 1991. Analysing hospital mortality. The consequences of diversity in patient mix. Journal of the American Medical Association 265 (14):1849–53.

Guy WA 1867. On the mortality of London hospitals: and incidentally on the deaths in the prisons and public institutions of the Metropolis. Journal of the Statistical Society of London 30(2):293–322.

Hadorn DC, Keeler EB, Rogers WH & Brook RH 1993. Assessing performance of mortality prediction models. Final report for HCFA Severity Project. MR-181-HCFA: Santa Monica, CA: RAND Corporation.

Hannan EL, Kilburn H Jr, Racz M, Shields E & Chassin MR 1994. Improving the outcomes of coronary artery bypass surgery in New York State. Journal of the American Medical Association 271(10):761–6.

Hartz AJ, Gottlieb MS, Kuhn EM & Rimm AA 1993. The relationship between adjusted hospital mortality and the results of peer review. Health Services Research 27 (6):765–77.

HDSC (Health Data Standards Committee) 2006. National Health Data Dictionary. Version 13. Cat. no. HWI 101. Canberra: AIHW.

Heijink R, Koolman X, Pieter D, van der Veen A, Jarman B & Westert G 2008. Measuring and explaining mortality in Dutch hospitals; the Hospital Standardised Mortality Rate between 2003 and 2005. BMC Health Services Research 8(1):73–80.

Hibbard JH, Stockard J & Tusler M 2005. Hospital performance reports: impact on quality, market share, and reputation. Health Affairs 24(4):1150–60.

Hofer TP & Hayward RA 1996. Identifying poor-quality hospitals: Can hospital mortality rates detect quality problems for medical diagnoses? Medical
Care 34(8):737–53.

Horn SD 2006. Performance measures and clinical outcomes. Journal of the American Medical Association 296(22):2731–2.

Hosmer DW & Lemeshow S 2000. Applied logistic regression. 2nd edn. New York: John Wiley & Sons Inc.

Iezzoni LI 1997a The risks of risk adjustment. Journal of the American Medical Association 278(19):1600–7.

Iezzoni LI 1997b. Assessing quality using administrative data. Annals of Internal Medicine 127(8 Supplement):666–74.

Iezzoni LI (ed.) 2003a. Risk adjustment for measuring health care outcomes, 3rd edn. Chicago: Health Administration Press.

Iezzoni LI 2003b. Coded data from administrative sources. In: Iezzoni LI. Risk adjustment for measuring health care outcomes, 3rd edn. Chicago: Health Administration Press, 83–138.

Iezzoni LI, Shwartz M, Ash AS, Hughes JS, Daley J & Mackiernan YD (1996a) Severity measurement methods and judging hospital death rates for pneumonia. Medical Care 34(1):11–28.

Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS & Mackiernan YD 1996b. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. American Journal of Public Health 86(10):1379–87.

Iglehart JK 1986. Early experience with prospective payment of hospitals. New England Journal of Medicine 314(22):1460–4.

Jarman B, Bottle A, Aylin P & Browne M 2005. Monitoring changes in hospital standardised mortality ratios. British Medical Journal 330(7487):329–30.

Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, Hurwitz B & Iezzoni LI 1999. Explaining differences in English hospital death rates using routinely collected data. British Medical Journal 318(7197):1515–20.

Jencks SF, Williams DK & Kay TL 1988. Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. Journal of the American Medical Association 260(15):2240–6.

Jha AK, Orav EJ, Li Z & Epstein AM 2007. The inverse relationship between mortality rates and performance in the Hospital Quality Alliance Measures. Health Affairs 26(4):1104–10.

Kahn KL, Rogers WH, Rubenstein LV, Sherwood MJ, Reinisch EJ, Keeler EB, Draper D, Kosecoff J & Brook RH (1990) Measuring quality of care with explicit process criteria before

and after implementation of the DRG-based prospective payment system. Journal of the American Medical Association 264(15):1969–73.

Keeler EB, Kahn KL, Draper D, Sherwood MJ, Rubenstein LV, Reinisch EJ, Kosecoff J & Brook RH 1990. Changes in sickness at admission following the introduction of the prospective payment system. Journal of the American Medical Association 264(15):1962–8.

Kelman S & Friedman JN 2007. Performance improvement and performance dysfunction: an empirical examination of impacts of the Emergency Room wait-time target in the English National Health Service. Faculty Research Working Papers Series RWP07–034. Cambridge MA: John F Kennedy School of Government, Harvard University.

Knaus WA, Draper EA, Wagner DP & Zimmerman JE 1985. APACHE II: a severity of disease classification system. Critical Care Medicine 13(10):818–29.

Knaus WA, Draper EA, Wagner DP & Zimmerman JE 1986. An evaluation of outcome from intensive care in major medical centres. Annals of Internal Medicine 104(3):410–8.

Kohn LT, Corrigan JM & Donaldson MS (eds) 2000. To err is human: building a safer health system 2000. Washington DC: National Academies Press, R729.8.T6.

Krakauer H, Bailey RC, Skellan KJ, Stewart JD, Hartz AJ, Kuhn EM & Rimm AA 1992. Evaluation of the HCFA model for the analysis of mortality following hospitalisation. Health Services Research 27 (3):317–19.

Lakhani A, Coles J, Eayres D, Spence C & Rachet B 2005. Creative use of existing clinical and health outcomes data to assess NHS performance in England: Part 1—performance indicators closely linked to clinical care. British Medical Journal 330(7505):1426–31.

Lemieux-Charles L & McGuire WL 2006. What do we know about health care team effectiveness? A review of the literature. Medical Care Research and Review 63(3):263–300.

Lohr KN 1991. Medicare: a strategy for quality assurance. Journal of Quality Assurance 13(1):10–3.

Manheim LM, Feinglass J, Shortell SM & Hughes EFX 1992. Regional variation in Medicare hospital mortality. Inquiry—Blue Cross and Blue Shield Association 29(1):55–66.

Marshall G, Shroyer ALW, Grover FL & Hammermeister KE 1998. Time series monitors of outcomes: a new dimension for measuring quality of care. Medical Care 36(3):348–56.

Maxwell RJ 1984. Quality assessment in health. British Medical Journal (Clinical Research Edition) 288(6428):1470–2.

Mohammed MA & Deeks JJ 2008. In the context of performance monitoring, the caterpillar plot should be mothballed in favour of the funnel plot. The Annals of Thoracic Surgery 86(1):348.

Montgomery DC 2001. Introduction to statistical quality control, 4th edn. New York: John Wiley & Sons, Inc.

Moses LE & Mosteller F 1968. Institutional differences in postoperative death rates. Commentary on some of the findings of the National Halothane Study. Journal of the American Medical Association 203(7):492–4.

NCCH (National Centre for Classification in Health) 2000. Australian Coding Benchmark Audit (ACBA), 2nd edn. Sydney: NCCH.

NCCH 2004. ICD-10-AM 4th edn. Sydney: NCCH.

NCCH 2006. Performance Indicators for Coding Quality (PICQ). Sydney: NCCH.

NHPC (National Health Performance Committee) 2004. National report on health sector performance indicators 2003. A report to the Australian Health Ministers' Conference, November 2004. Cat. no. HWI 786. Canberra: AIHW.

Nightingale F 1860. Hospital statistics. Programme of the Fourth Session of the International Statistical Congress. London: Eyre and Spottiswoode for HMSO, 63–71.

Nightingale F 1863. Notes on hospitals. London: Longman, Green, Longman, Roberts, and Green.

Omoigui NA, Miller DP, Brown KJ, Annan K, Cosgrove ID, Lytle B, Loop F & Topol EJ (1996) Outmigration For coronary bypass surgery in an era of public dissemination of clinical outcomes. Circulation 93(1):27–33.

Park RE, Brook RH, Kosecoff J, Keesey J, Rubenstein L, Keeler E, Kahn KL, Rogers WH & Chassin MR (1990) Explaining variations in hospital death rates. Randomness, severity of illness, quality of care. Journal of the American Medical Association 264(4):484–90.

Pine M, Norusis M, Jones B & Rosenthal GE 1997. Predictions of hospital mortality rates: a comparison of data sources. Annals of Internal Medicine 126(5):347–54.

Pitches DW, Mohammed MA & Lilford RJ 2007. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. BMC Health Services Research 7:91 (doi:10.1186/1472-6963-7-91)

Polanczyk CA, Lane A, Coburn M, Philbin EF, Dec GW & Di Salvo TG 2002. Hospital outcomes in major teaching, minor teaching, and nonteaching hospitals in New York State. The American Journal of Medicine 112(4):314–5.

Polanczyk CA, Rohde LE, Philbin EA & Di Salvo TG 1998. A new casemix adjustment index for hospital mortality among patients with congestive heart failure. Medical Care 36(10):1489–99.

Powell AE, Davies HTO & Thomson RG 2003. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. Quality and Safety in Health Care 12 (2):122–8.

Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE & Ghali WA 2005. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Medical Care 43(11):1130–9.

Queensland Health 2004. Queensland Measured Quality Hospital Reports 2004. The State of Queensland (Queensland Health), Queensland Government. Viewed August 2008. <http://www.health.qld.gov.au/quality/mq_reports2004.asp>.

Queensland Health 2007. Moving ahead. Queensland public hospitals performance report 2006–07. Brisbane: The State of Queensland (Queensland Health), Queensland Government.

Reed JF 3rd, Olenchock SA Jr, Murphy SA & Garzia FM 2003. Off the shelf or recalibrate? Customising a risk index for assessing mortality. The Heart Surgery Forum 6(4):232–6.

Romano PS & Mutter R 2004. The evolving science of quality measurement for hospitals: implications for studies of competition and consolidation. International Journal of Health Care Finance and Economics 4(2):131–57.

Romano PS, Zach A, Luft HS, Rainwater J, Remy LL & Campa D 1995. The California Hospital Outcomes Project: using administrative data to compare hospital performance. The Joint Commission Journal on Quality Improvement 21(12):668–82.

Rosen HM & Green BA 1987. The HCFA excess mortality lists: a methodological critique. Hospital & Health Services Administration 32(1):119–27.

Rosenthal GE, Baker DW, Norris DG, Way LE, Harper DL & Snow RJ 2000. Relationships between in-hospital and 30–day standardised hospital mortality: implications for profiling hospitals. Health Services Research 34(7):1449.

Rubin HR, Pronovost P & Diette GB 2001. The advantages and disadvantages of process-based measures of health care quality. International Journal for Quality in Health Care 13(6):469–74.

Schuster MA, McGlynn EA & Brook RH 1998. How good is the quality of health care in the United States? The Milbank Memorial Fund Quarterly 76(4):517–63.

Scobie S, Thomson R, McNeil JJ & Phillips PA 2006. Measurement of the safety and quality of health care. The Medical Journal of Australia 184(10 Suppl):S51–S5.

Scott IA & Ward M 2006. Public reporting of hospital outcomes based on administrative data: risks and opportunities. Medical Journal of Australia 184(11):571–5.

Shahian DM & Normand SL 2008. Comparison of 'risk-adjusted' hospital outcomes. Circulation 117(15):1955–63.

Shapiro MF, Park RE, Keesey J & Brook RH 1994. The effect of alternative case-mix adjustments on mortality differences between municipal and voluntary hospitals in New York City. Health Services Research 29(1):95–112.

Shih A & Schoenbaum SC 2007. Measuring hospital performance: The importance of process measures. The Commonwealth Fund: Commission on a High Performance Health System 6.

Shojania KG & Forster AJ 2008. Hospital mortality: when failure is not a good measure of success. Canadian Medical Association Journal 179(2):153–7.

Shwartz M & Ash AS 2003. Evaluating risk-adjustment models empirically. In: Iezzoni L I. Risk adjustment for measuring health care outcomes, 3rd edn. Chicago: Health Administration Press, 231–74.

Silber JH & Rosenbaum PR 1997. A spurious correlation between hospital mortality and complication rates: the importance of severity adjustment. Medical Care 35 (10 Supplement):OS77–OS92.

Silber JH, Rosenbaum PR & Ross RN 1995. Comparing the contributions of groups of predictors: which outcomes vary with hospital rather than patient characteristics? Journal of the American Statistical Association 90(429):7–18.

Smith DW 1994. Evaluating risk adjustment by partitioning variation in hospital mortality rates. Statistics in Medicine 13(10):1001–13.

Smith DW, Pine M, Bailey RC, Jones B, Brewster A & Krakauer H 1991. Using clinical variables to estimate the risk of patient mortality. Medical Care 29(11):1108–29.

Spiegelhalter DJ 1999. Surgical audit: statistical lessons from Nightingale and Codman. Journal of the Royal Statistical Society: Series A (Statistics in Society) 162(1):45–58.

Spiegelhalter DJ 2002. Funnel plots for institutional comparison. Quality and Safety in Health Care 11(4):390–1.

Spiegelhalter DJ 2005. Funnel plots for comparing institutional performance. Statistics in Medicine 24(8):1185–202.

StataCorp 2007. Stata Base Reference Manual, Vol. 2 (I–P). Stata Statistical Software: Release 10. College Station, TX: StataCorp LP, 175.

Stern RS & Epstein AM 1985. Institutional responses to prospective payment based on diagnosis-related groups. Implications for cost, quality, and access. The New England Journal of Medicine 312(10):621–7.

Sundararajan V, Quan H, Halfon P, Fushimi K, Luthi J-C, Burnand B & Ghali WA 2007. Cross-national comparative performance of three versions of the ICD-10 Charlson index. Medical Care 45(12):1210–5.

Tekkis PP, McCulloch P, Steger AC, Benjamin IS & Poloniecki JD 2003. Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data. British Medical Journal 326(7393):786–8.

Thomas JW & Hofer TP 1998. Research evidence on the validity of risk-adjusted mortality rate as a measure of hospital quality of care. Medical Care Research and Review 55(4):371–404.

Thomas JW & Hofer TP 1999. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. Medical Care 37(1):83–92.

Thomas JW, Holloway JJ & Guire KE 1993. Validating risk-adjusted mortality as an indicator for quality of care. Inquiry—Blue Cross and Blue Shield Association 30(1):6–22.

Van Der Weyden MB 2005. The Bundaberg Hospital scandal: the need for reform in Queensland and beyond. The Medical Journal of Australia 183(6):284–5.

Werner RM & Bradlow ET 2006. Relationship between Medicare's hospital compare performance measures and mortality rates. Journal of the American Medical Association 296(22):2694–702.

Wilson RM & Van Der Weyden MB 2005. The safety of Australian healthcare: 10 years after QAHCS. The Medical Journal of Australia 182(6):260–1.

Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L & Hamilton JD 1995. The quality in Australian health care study. The Medical Journal of Australia 163(9):458–71.

Wright J, Dugdale B, Hammond I, Jarman B, Neary M, Newton D, Patterson C, Russon L, Stanley P, Stephens R & Warren E 2006. Learning from death: a hospital mortality reduction programme. Journal of the Royal Society of Medicine 99(6):S303–8.

Zalkind DL & Eastaugh SR 1997. Mortality rates as an indicator of hospital quality. Hospital and Health Services Administration 42(1):3–15.

Appendix 1 Diagnoses accounting for 80% of in-hospital deaths

Cases in the NHMD with data-year 2005–06 and satisfying study inclusion criteria were summarised according to the frequency of deaths in hospital, by three character ICD-10-AM code. The 68 codes listed in Table A1 are the ones with the highest frequency of deaths. Between them, the 68 codes were present in less than 20% of records but 80% of deaths.

Table A1.1: Principal diagnosis codes occurring most frequently among in-hospital deaths in 2005–06

ICD code	Description
A41	Other sepsis
C15	Malignant neoplasm of oesophagus
C16	Malignant neoplasm of stomach
C18	Malignant neoplasm of colon
C20	Malignant neoplasm of rectum
C22	Malignant neoplasm of liver and intrahepatic bile ducts
C25	Malignant neoplasm of pancreas
C34	Malignant neoplasm of bronchus and lung
C45	Mesothelioma
C50	Malignant neoplasm of breast
C56	Malignant neoplasm of ovary
C61	Malignant neoplasm of prostate
C64	Malignant neoplasm of kidney, except renal pelvis
C67	Malignant neoplasm of bladder
C71	Malignant neoplasm of brain
C78	Secondary malignant neoplasm of respiratory and digestive organs
C79	Secondary malignant neoplasm of other sites
C80	Malignant neoplasm without specification of site
C83	Diffuse non-Hodgkin lymphoma
C85	Other and unspecified types of non-Hodgkin lymphoma
C90	Multiple myeloma and malignant plasma cell neoplasms
C91	Lymphoid leukaemia
C92	Myeloid leukaemia
E11	Type 2 diabetes mellitus
E86	Volume depletion
E87	Other disorders of fluid, electrolyte and acid-base balance
G93	Other disorders of brain
I20	Angina pectoris
I21	Acute myocardial infarction
I25	Chronic ischaemic heart disease
I26	Pulmonary embolism
I46	Cardiac arrest
I48	Atrial fibrillation and flutter
I49	Other cardiac arrhythmias
I50	Heart failure
I60	Subarachnoid haemorrhage
I61	Intracerebral haemorrhage
I62	Other nontraumatic intracranial haemorrhage
I63	Cerebral infarction
I64	Stroke, not specified as haemorrhage or infarction
I70	Atherosclerosis
I71	Aortic aneurysm and dissection
J15	Bacterial pneumonia, not elsewhere classified
J18	Pneumonia, organism unspecified
J22	Unspecified acute lower respiratory infection
J44	Other chronic obstructive pulmonary disease
J69	Pneumonitis due to solids and liquids
J84	Other interstitial pulmonary diseases
J90	Pleural effusion, not elsewhere classified
J96	Respiratory failure, not elsewhere classified
K52	Other noninfective gastroenteritis and colitis

(continued)

Table A1.1 (continued): Principal diagnosis codes occurring most frequently among in-hospital deaths in 2005–06

ICD code	Description
K55	Vascular disorders of intestine
K56	Paralytic ileus and intestinal obstruction without hernia
K57	Diverticular disease of intestine
K63	Other diseases of intestine
K70	Alcoholic liver disease
K72	Hepatic failure, not elsewhere classified
K85	Acute pancreatitis
K92	Other diseases of digestive system
L03	Cellulitis
N17	Acute renal failure
N18	Chronic renal failure
N39	Other disorders of urinary system
R55	Syncope and collapse
S06	Intracranial injury
S32	Fracture of lumbar spine and pelvis
S72	Fracture of femur
T81	Complications of procedures, not elsewhere classified

Appendix 2 Summary tables of HSMRs in 2005–06

Table A2.1: Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group B1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
B1001	11443.59	0	10.64	0	3.97	0	19.19	0.00 (0.0–34.5)	0.00 (0.0 92.4)	0.00 (0.0–19.1)	1	1	1
B1002	18551.28	3	11.81	3	2.97	6	13.95	25.41 (5.1–74.2)	101.10 (20.3 295.4)	43.01 (15.7–93.6)	2	23	5
B1003	11878.78	30	85.26	3	12.03	33	100.68	35.18 (23.7–50.2)	24.94 (5.0 72.9)	32.78 (22.6–46.0)	3	3	2
B1004	12964.14	42	113.65	6	24.16	48	143.81	36.95 (26.6–50.0)	24.83 (9.1 54.0)	33.38 (24.6–44.3)	4	2	3
B1005	20944.06	68	181.55	25	59.55	93	241.34	37.45 (29.1–47.5)	41.98 (27.2 62.0)	38.53 (31.1–47.2)	5	4	4
B1006	15766.76	57	90.70	13	17.97	70	111.83	62.85 (47.6–81.4)	72.36 (38.5 123.8)	62.60 (48.8–79.1)	6	12	6
B1007	18426.22	79	116.55	18	23.41	97	140.31	67.78 (53.7–84.5)	76.89 (45.5 121.5)	69.13 (56.1–84.3)	7	15	7
B1008	11577.67	52	68.82	13	19.14	65	90.25	75.56 (56.4–99.1)	67.90 (36.1 116.1)	72.02 (55.6–91.8)	8	11	8
B1009	17108.93	131	162.84	33	53.30	164	224.40	80.45 (67.3–95.5)	61.91 (42.6 86.9)	73.08 (62.3–85.2)	9	8	9
B1010	17342.59	78	96.47	12	20.54	90	116.39	80.85 (63.9–100.9)	58.42 (30.2 102.1)	77.32 (62.2–95.0)	10	7	11
B1011	12139.77	83	98.78	12	16.51	95	115.44	84.02 (66.9–104.2)	72.70 (37.5 127.0)	82.29 (66.6–100.6)	11	13	12
B1012	17108.45	149	174.22	23	30.21	172	202.48	85.52 (72.3–100.4)	76.12 (48.2 114.2)	84.95 (72.7–98.6)	12	14	14
B1013	17052.74	95	109.79	12	24.08	107	142.61	86.53 (70.0–105.8)	49.84 (25.7 87.1)	75.03 (61.5–90.7)	13	5	10
B1014	11474.29	152	170.61	38	34.13	190	202.80	89.09 (75.5–104.4)	111.35 (78.8 152.8)	93.69 (80.8–108.0)	14	24	19
B1015	14171.88	41	45.75	17	20.16	58	65.28	89.62 (64.3–121.6)	84.34 (49.1 135.0)	88.84 (67.5–114.9)	15	17	17

(continued)

Table A2.1 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group B1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
B1016	15302.6	170	189.17	26	39.80	196	232.21	89.86 (76.9–104.4)	65.33 (42.7 95.7)	84.41 (73.0–97.1)	16	10	13
B1017	18521.34	114	125.39	22	23.92	136	153.82	90.91 (75.0–109.2)	91.96 (57.6 139.2)	88.41 (74.2–104.6)	17	21	16
B1018	14139.55	117	126.53	18	22.73	135	149.04	92.47 (76.5–110.8)	79.18 (46.9 125.2)	90.58 (75.9–107.2)	18	16	18
B1019	21415.35	183	194.15	32	49.00	215	247.50	94.26 (81.1–108.9)	65.30 (44.7 92.2)	86.87 (75.6–99.3)	19	9	15
B1020	18994.63	173	181.04	33	33.13	206	218.70	95.56 (81.8–110.9)	99.60 (68.5 139.9)	94.19 (81.8–108.0)	20	22	20
B1021	16109.59	184	190.68	15	27.32	199	204.97	96.50 (83.1–111.5)	54.90 (30.7 90.6)	97.09 (84.1–111.6)	21	6	21
B1022	16485.85	143	144.46	35	30.60	178	172.18	98.99 (83.4–116.6)	114.38 (79.7 159.1)	103.38 (88.8–119.7)	22	25	22
B1023	17483.18	356	349.79	69	76.51	425	395.87	101.78 (91.5–112.9)	90.19 (70.2 114.1)	107.36 (97.4–118.1)	23	19	23
B1024	14126.42	189	165.65	25	27.49	214	185.28	114.10 (98.4–131.6)	90.95 (58.8 134.3)	115.50 (100.5–132.1)	24	20	25
B1025	14412.28	107	90.50	16	18.25	123	111.06	118.23 (96.9–142.9)	87.65 (50.1 142.4)	110.75 (92.0–132.1)	25	18	24

Table A2.2: Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group C2

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
C2001	2841.45	0	2.64	0	1.06	0	5.08	0.00 (0–139.2)	0.00 (0–345.5)	0.00 (0–72.2)	2	3	1
C2002	3206.87	0	0.03	0	0.71	0	0.51	0.00 (0–11735.7)	0.00 (0–519.2)	0.00 (0–715.5)	1	4	2
C2003	5783.27	0	2.07	0	0.91	0	3.17	0.00 (0–177.4)	0.00 (0–401.8)	0.00 (0–115.8)	3	2	3
C2004	4715.84	3	11.35	0	3.10	3	14.97	26.42 (5.3–77.2)	0.00 (0–118.3)	20.04 (4.0–58.5)	4	1	4
C2005	6937.37	19	61.39	2	10.77	21	71.07	30.95 (18.6–48.3)	18.58 (2.1–67.1)	29.55 (18.3–45.2)	5	6	5
C2006	2381.76	6	12.67	2	4.87	8	16.33	47.36 (17.3–103.1)	41.11 (4.6–148.4)	48.98 (21.1–96.5)	6	12	7
C2007	2561.47	11	20.54	3	8.92	14	28.46	53.56 (26.7–95.8)	33.65 (6.8–98.3)	49.19 (26.9–82.5)	7	9	9
C2008	4232.04	50	85.35	8	22.76	58	118.13	58.58 (43.5–77.2)	35.15 (15.1–69.3)	49.10 (37.3–63.5)	8	10	8
C2009	5069.51	31	51.46	1	11.26	32	67.10	60.24 (40.9–85.5)	8.88 (0.1–49.4)	47.69 (32.6–67.3)	9	5	6
C2010	3273.44	7	11.54	7	4.69	14	17.09	60.64 (24.3–125.0)	149.25 (59.8–307.5)	81.93 (44.8–137.5)	10	52	20
C2011	5089.75	29	44.24	5	8.08	34	54.19	65.55 (43.9–94.1)	61.91 (20.0–144.5)	62.74 (43.4–87.7)	11	18	10
C2012	2356.46	10	14.79	4	5.54	14	20.55	67.62 (32.4–124.4)	72.21 (19.4–184.9)	68.12 (37.2–114.3)	12	21	12
C2013	4569.78	29	42.39	10	8.26	39	52.69	68.41 (45.8–98.3)	121.09 (58.0–222.7)	74.02 (52.6–101.2)	13	45	15
C2014	2336.8	17	24.50	5	5.04	22	28.56	69.39 (40.4–111.1)	99.20 (32.0–231.5)	77.04 (48.3–116.7)	14	34	16
C2015	5613.38	56	78.53	6	10.49	62	92.87	71.31 (53.9–92.6)	57.19 (20.9–124.5)	66.76 (51.2–85.6)	15	16	11
C2016	4885.57	39	52.73	8	10.44	47	66.80	73.96 (52.6–101.1)	76.63 (33.0–151.0)	70.36 (51.7–93.6)	16	24	13
C2017	2322.79	21	27.92	5	6.99	26	35.29	75.22 (46.5–115.0)	71.58 (23.1–167.0)	73.68 (48.1–108.0)	17	20	14
C2018	2499.07	16	20.90	5	3.81	21	25.14	76.55 (43.7–124.3)	131.39 (42.3–306.6)	83.52 (51.7–127.7)	18	48	21
C2019	3244.37	34	41.94	4	7.61	38	41.96	81.06 (56.1–113.3)	52.56 (14.1–134.6)	90.56 (64.1–124.3)	19	14	24
C2020	2744.83	11	12.65	2	3.80	13	16.85	86.94 (43.3–155.6)	52.57 (5.9–189.8)	77.13 (41.0–131.9)	20	15	17

(continued)

Table A2.2 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group C2

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
C2021	2123.94	13	14.83	10	6.62	23	21.65	87.66 (46.6–149.9)	151.14 (72.4–278.0)	106.25 (67.3–159.4)	21	53	32
C2022	4969.4	35	39.52	5	11.15	40	50.62	88.57 (61.7–123.2)	44.83 (14.4–104.6)	79.02 (56.4–107.6)	22	13	19
C2023	5284.65	44	49.21	9	12.22	53	63.35	89.41 (65.0–120.0)	73.67 (33.6–139.9)	83.67 (62.7–109.4)	23	23	22
C2024	3327.35	27	28.09	3	7.64	30	38.65	96.11 (63.3–139.8)	39.25 (7.9–114.7)	77.61 (52.4–110.8)	24	11	18
C2025	3324.38	28	28.29	6	6.85	34	37.66	98.98 (65.8–143.1)	87.55 (32.0–190.6)	90.29 (62.5–126.2)	25	28	23
C2026	2665.88	37	36.85	7	6.87	44	44.82	100.42 (70.7–138.4)	101.88 (40.8–209.9)	98.17 (71.3–131.8)	26	35	27
C2027	2405.12	35	34.84	7	5.56	42	37.80	100.45 (70.0–139.7)	125.88 (50.4–259.4)	111.11 (80.1–150.2)	27	47	34
C2028	3083.97	30	29.51	5	8.00	35	36.38	101.68 (68.6–145.2)	62.47 (20.1–145.8)	96.20 (67.0–133.8)	28	19	25
C2029	3920.41	52	49.82	7	7.94	59	51.74	104.38 (77.9–136.9)	88.19 (35.3–181.7)	114.03 (86.8–147.1)	29	30	38
C2030	4165.54	35	33.34	8	9.70	43	44.27	104.98 (73.1–146.0)	82.49 (35.5–162.5)	97.14 (70.3–130.9)	30	25	26
C2031	2288.87	15	14.27	7	4.08	22	19.51	105.08 (58.8–173.3)	171.77 (68.8–353.9)	112.78 (70.7–170.8)	31	56	36
C2032	4269.18	63	59.33	14	14.32	77	68.07	106.19 (81.6–135.9)	97.78 (53.4–164.1)	113.12 (89.3–141.4)	32	33	37
C2033	2917.91	29	27.19	12	8.15	41	34.80	106.67 (71.4–153.2)	147.24 (76.0–257.2)	117.83 (84.5–159.9)	33	51	42
C2034	2423.85	31	28.41	5	6.83	36	34.51	109.13 (74.1–154.9)	73.23 (23.6–170.9)	104.32 (73.1–144.4)	34	22	29
C2035	2381.69	34	31.03	5	5.69	39	37.19	109.59 (75.9–153.1)	87.88 (28.3–205.1)	104.86 (74.6–143.3)	35	29	31
C2036	3878.93	45	40.39	9	10.43	54	51.54	111.40 (81.2–149.1)	86.27 (39.4–163.8)	104.78 (78.7–136.7)	36	27	30
C2037	3079.27	49	43.18	9	7.55	58	50.09	113.49 (84.0–150.0)	119.16 (54.4–226.2)	115.78 (87.9–149.7)	37	43	39
C2038	3731.99	28	24.02	10	10.23	38	35.59	116.59 (77.5–168.5)	97.72 (46.8–179.7)	106.78 (75.6–146.6)	38	32	33
C2039	2419.27	20	16.67	1	4.53	21	20.78	120.00 (73.3–185.3)	22.06 (0.3–122.8)	101.07 (62.5–154.5)	39	7	28
C2040	4793.88	40	33.00	10	8.86	50	44.34	121.19 (86.6–165.0)	112.81 (54.0–207.5)	112.76 (83.7–148.7)	40	39	35
C2041	2354.82	26	21.28	13	6.94	39	27.60	122.17 (79.8–179.0)	187.44 (99.7–320.6)	141.33 (100.5–193.2)	41	58	54

(continued)

Table A2.2 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group C2

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
C2042	2211.37	17	13.53	7	4.39	24	18.30	125.67 (73.2–201.2)	159.51 (63.9–328.7)	131.18 (84.0–195.2)	42	54	50
C2043	2223.58	52	41.06	20	7.76	72	46.41	126.64 (94.6–166.1)	257.72 (157.4–398.1)	155.13 (121.4–195.4)	43	60	57
C2044	2883.28	36	28.36	8	7.01	44	37.91	126.94 (88.9–175.7)	114.13 (49.1–224.9)	116.08 (84.3–155.8)	44	41	40
C2045	4189.45	41	31.33	13	9.31	54	43.73	130.86 (93.9–177.5)	139.65 (74.3–238.8)	123.49 (92.8–161.1)	45	49	46
C2046	2464.02	24	18.25	7	6.51	31	24.44	131.54 (84.3–195.7)	107.59 (43.1–221.7)	126.85 (86.2–180.1)	46	37	47
C2047	2547.32	32	23.98	7	8.45	39	32.40	133.45 (91.3–188.4)	82.83 (33.2–170.7)	120.35 (85.6–164.5)	47	26	43
C2048	2696.55	43	31.87	9	8.80	52	40.13	134.90 (97.6–181.7)	102.24 (46.7–194.1)	129.58 (96.8–169.9)	48	36	49
C2049	3241.3	33	23.73	8	6.65	41	33.25	139.09 (95.7–195.3)	120.34 (51.8–237.1)	123.30 (88.5–167.3)	49	44	45
C2050	3017.74	29	20.77	8	6.82	37	30.32	139.66 (93.5–200.6)	117.38 (50.5–231.3)	122.02 (85.9–168.2)	50	42	44
C2051	3567.16	79	56.39	17	15.34	96	67.29	140.09 (110.9–174.6)	110.85 (64.5–177.5)	142.68 (115.6–174.2)	51	38	56
C2052	3268.31	34	24.11	8	6.59	42	33.03	141.00 (97.6–197.0)	121.34 (52.2–239.1)	127.17 (91.6–171.9)	52	46	48
C2053	2663.86	38	26.68	15	8.62	53	37.19	142.42 (100.8–195.5)	174.05 (97.3–287.1)	142.50 (106.7–186.4)	53	57	55
C2054	3610.11	11	7.63	5	3.13	16	11.81	144.15 (71.9–257.9)	159.90 (51.5–373.1)	135.53 (77.4–220.1)	54	55	51
C2055	2270.93	40	26.94	17	12.02	57	40.92	148.50 (106.1–202.2)	141.40 (82.3–226.4)	139.29 (105.5–180.5)	55	50	53
C2056	2884.4	62	40.08	9	9.82	71	51.56	154.67 (118.6–198.3)	91.63 (41.8–173.9)	137.71 (107.5–173.7)	56	31	52
C2057	3055.64	41	26.17	2	7.10	43	36.84	156.68 (112.4–212.6)	28.17 (3.2–101.7)	116.73 (84.5–157.2)	57	8	41
C2058	2087.2	29	16.21	5	4.41	34	21.13	178.90 (119.8–256.9)	113.48 (36.6–264.8)	160.89 (111.4–224.8)	58	40	58
C2059	2532.19	29	13.63	4	6.66	33	20.41	212.83 (142.5–305.7)	60.04 (16.2–153.7)	161.65 (111.3–227.0)	59	17	59
C2060	2063.29	18	5.28	5	2.60	23	7.73	341.17 (202.1–539.2)	192.44 (62.0–449.1)	297.45 (188.5–446.3)	60	59	60

Table A2.3: Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group D1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
D1001	249.7	0	0.01	0	0.18	0	0.15	0.00 (0.0–51453.2)	0.00 (0.0–2095.3)	0.00 (0.0–2474.4)	1	16	1
D1002	1502.34	6	15.88	1	3.22	7	20.73	37.79 (13.8–82.3)	31.08 (0.4–172.9)	33.77 (13.5–69.6)	2	21	2
D1003	1747.71	8	21.04	1	5.21	9	25.14	38.02 (16.4–74.9)	19.21 (0.3–106.9)	35.80 (16.3–68.0)	3	18	3
D1004	1758.83	10	23.29	2	4.86	12	28.13	42.94 (20.6–79.0)	41.16 (4.6–148.6)	42.66 (22.0–74.5)	4	25	7
D1005	751.57	6	12.15	0	2.92	6	14.63	49.39 (18.0–107.5)	0.00 (0.0–125.7)	41.01 (15.0–89.3)	5	15	5
D1006	1349.86	8	15.05	1	4.06	9	19.39	53.14 (22.9–104.7)	24.64 (0.3–137.1)	46.41 (21.2–88.1)	6	19	9
D1007	682.81	4	7.49	1	2.34	5	11.35	53.40 (14.4–136.7)	42.69 (0.6–237.5)	44.06 (14.2–102.8)	7	27	8
D1008	1130.28	6	10.56	3	3.37	9	15.26	56.79 (20.7–123.6)	89.06 (17.9–260.2)	58.99 (26.9–112.0)	8	46	11
D1009	444.55	3	5.27	0	2.05	3	7.40	56.96 (11.4–166.4)	0.00 (0.0–179.0)	40.54 (8.1–118.5)	9	7	4
D1010	263.29	2	3.25	2	0.93	4	4.37	61.52 (6.9–222.1)	214.14 (24.0–773.2)	91.49 (24.6–234.2)	10	94	30
D1011	1099.99	6	9.67	4	2.82	10	12.19	62.08 (22.7–135.1)	141.95 (38.2–363.4)	82.04 (39.3–150.9)	11	75	20
D1012	494.36	4	6.18	2	1.99	6	7.96	64.75 (17.4–165.8)	100.72 (11.3–363.6)	75.37 (27.5–164.1)	12	55	17
D1013	467.67	5	7.50	2	2.06	7	9.56	66.66 (21.5–155.6)	97.23 (10.9–351.0)	73.19 (29.3–150.8)	13	53	15
D1014	1008.56	9	12.46	2	4.38	11	18.59	72.23 (33.0–137.1)	45.70 (5.1–165.0)	59.17 (29.5–105.9)	14	29	12
D1015	1565.96	11	15.12	1	3.71	12	19.32	72.73 (36.3–130.1)	26.93 (0.4–149.8)	62.10 (32.0–108.5)	15	20	13
D1016	673.17	8	10.54	0	2.78	8	13.94	75.91 (32.7–149.6)	0.00 (0.0–132.1)	57.37 (24.7–113.1)	16	5	10
D1017	2050.58	14	18.36	5	4.08	19	22.05	76.26 (41.7–128.0)	122.56 (39.5–286.0)	86.19 (51.9–134.6)	17	67	24
D1018	530.39	1	1.30	0	0.83	1	2.38	76.80 (1.0–427.3)	0.00 (0.0–443.4)	41.96 (0.5–233.5)	18	14	6
D1019	1836.16	17	22.12	2	5.13	19	28.02	76.87 (44.8–123.1)	38.95 (4.4–140.6)	67.80 (40.8–105.9)	19	24	14
D1020	436.64	4	5.11	3	2.21	7	7.23	78.20 (21.0–200.2)	135.78 (27.3–396.7)	96.79 (38.8–199.4)	20	71	35

(continued)

Table A2.3 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group D1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
D1021	406.16	6	6.99	1	1.25	7	9.43	85.83 (31.3–186.8)	79.81 (1.0–444.0)	74.19 (29.7–152.9)	21	42	16
D1022	738.5	18	20.23	3	3.32	21	22.79	88.96 (52.7–140.6)	90.39 (18.2–264.1)	92.14 (57.0–140.8)	22	49	31
D1023	696.18	10	11.06	4	4.54	14	14.39	90.38 (43.3–166.2)	88.10 (23.7–225.5)	97.30 (53.1–163.3)	23	45	36
D1024	1937.9	26	28.37	4	4.26	30	31.78	91.66 (59.9–134.3)	93.80 (25.2–240.2)	94.39 (63.7–134.8)	24	52	33
D1025	1499.31	13	14.12	2	3.08	15	16.40	92.04 (49.0–157.4)	64.87 (7.3–234.2)	91.46 (51.2–150.9)	25	38	29
D1026	955.85	9	9.68	2	2.73	11	13.79	92.98 (42.4–176.5)	73.39 (8.2–265.0)	79.76 (39.8–142.7)	26	41	18
D1027	489.02	6	6.30	4	1.79	10	9.27	95.30 (34.8–207.4)	223.52 (60.1–572.3)	107.89 (51.7–198.4)	27	99	45
D1028	441.52	2	2.07	0	0.46	2	2.43	96.73 (10.9–349.3)	0.00 (0.0–792.6)	82.30 (9.2–297.1)	28	6	22
D1029	733.61	11	11.36	3	2.93	14	14.31	96.86 (48.3–173.3)	102.35 (20.6–299.1)	97.81 (53.4–164.1)	29	58	38
D1030	1147.8	12	12.28	5	4.59	17	18.82	97.72 (50.4–170.7)	109.04 (35.1–254.5)	90.35 (52.6–144.7)	30	61	28
D1031	1705.78	13	13.30	2	4.61	15	18.29	97.72 (52.0–167.1)	43.39 (4.9–156.6)	82.03 (45.9–135.3)	31	28	19
D1032	1814.43	11	11.18	2	2.49	13	14.63	98.40 (49.1–176.1)	80.46 (9.0–290.5)	88.86 (47.3–152.0)	32	43	27
D1033	903.62	10	10.14	4	2.46	14	13.51	98.57 (47.2–181.3)	162.36 (43.7–415.7)	103.65 (56.6–173.9)	33	84	42
D1034	722.32	9	9.11	1	2.57	10	12.15	98.82 (45.1–187.6)	38.89 (0.5–216.4)	82.31 (39.4–151.4)	34	23	23
D1035	1527.35	17	17.03	2	5.28	19	23.09	99.83 (58.1–159.8)	37.87 (4.3–136.7)	82.28 (49.5–128.5)	35	22	21
D1036	686.64	6	5.94	1	2.01	7	7.89	101.05 (36.9–220.0)	49.79 (0.7–277.0)	88.68 (35.5–182.7)	36	30	25
D1037	437.25	1	0.99	3	0.97	4	1.97	101.06 (1.3–562.3)	308.10 (61.9–900.2)	203.49 (54.7–521.0)	37	107	100
D1038	1830.54	12	11.69	6	4.21	18	16.87	102.70 (53.0–179.4)	142.37 (52.0–309.9)	106.68 (63.2–168.6)	38	76	43
D1039	1848.87	30	28.14	5	4.61	35	28.52	106.62 (71.9–152.2)	108.54 (35.0–253.3)	122.72 (85.5–170.7)	39	60	54
D1040	385.92	6	5.51	1	1.80	7	7.89	108.97 (39.8–237.2)	55.49 (0.7–308.7)	88.69 (35.5–182.7)	40	35	26
D1041	981.45	16	14.62	6	4.27	22	18.69	109.42 (62.5–177.7)	140.64 (51.4–306.1)	117.69 (73.7–178.2)	41	73	51

(continued)

Table A2.3 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group D1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
D1042	490.22	10	9.13	2	2.19	12	12.33	109.56 (52.5–201.5)	91.18 (10.2–329.2)	97.34 (50.2–170.0)	42	50	37
D1043	1988.56	24	21.75	8	5.25	32	29.83	110.36 (70.7–164.2)	152.27 (65.6–300.1)	107.27 (73.4–151.4)	43	79	44
D1044	518.61	8	6.94	0	2.00	8	8.43	115.25 (49.6–227.1)	0.00 (0.0–183.5)	94.84 (40.8–186.9)	44	13	34
D1045	1739.28	23	19.50	5	3.40	28	20.56	117.92 (74.7–176.9)	146.96 (47.4–343.0)	136.19 (90.5–196.8)	45	77	66
D1046	726.86	8	6.78	1	1.73	9	9.70	117.99 (50.8–232.5)	57.90 (0.8–322.2)	92.76 (42.3–176.1)	46	37	32
D1047	1215.85	19	15.88	11	4.56	30	19.52	119.68 (72.0–186.9)	241.23 (120.3–431.7)	153.67 (103.7–219.4)	47	105	81
D1048	691.06	15	12.47	4	2.63	19	16.34	120.28 (67.3–198.4)	152.23 (41.0–389.7)	116.28 (70.0–181.6)	48	78	50
D1049	506.27	9	7.48	3	1.35	12	8.90	120.33 (54.9–228.4)	222.32 (44.7–649.6)	134.91 (69.6–235.7)	49	97	64
D1050	1254.03	15	12.29	4	3.14	19	16.85	122.03 (68.2–201.3)	127.43 (34.3–326.2)	112.79 (67.9–176.1)	50	69	47
D1051	1352.38	13	10.55	6	2.57	19	15.09	123.21 (65.5–210.7)	233.18 (85.1–507.5)	125.90 (75.8–196.6)	51	104	56
D1052	726.51	12	9.65	4	4.32	16	13.88	124.30 (64.2–217.1)	92.49 (24.9–236.8)	115.28 (65.8–187.2)	52	51	49
D1053	767.86	10	8.03	1	1.48	11	9.72	124.60 (59.6–229.2)	67.66 (0.9–376.4)	113.14 (56.4–202.5)	53	40	48
D1054	749.03	7	5.61	3	1.84	10	7.27	124.77 (50.0–257.1)	163.31 (32.8–477.2)	137.52 (65.8–252.9)	54	85	70
D1055	744.58	10	7.95	1	2.39	11	10.90	125.83 (60.2–231.4)	41.93 (0.5–233.3)	100.89 (50.3–180.5)	55	26	40
D1056	798.36	15	11.83	4	2.83	19	14.23	126.75 (70.9–209.1)	141.47 (38.1–362.2)	133.56 (80.4–208.6)	56	74	61
D1057	176.84	3	2.35	1	1.50	4	3.98	127.55 (25.6–372.7)	66.79 (0.9–371.6)	100.62 (27.1–257.6)	57	39	39
D1058	399.49	7	5.37	5	2.89	12	6.93	130.25 (52.2–268.4)	172.72 (55.7–403.1)	173.24 (89.4–302.6)	58	88	90
D1059	331.38	5	3.83	2	1.97	7	5.91	130.67 (42.1–304.9)	101.66 (11.4–367.0)	118.48 (47.5–244.1)	59	57	52
D1060	1454.86	43	32.81	12	7.41	55	39.34	131.05 (94.8–176.5)	161.92 (83.6–282.9)	139.82 (105.3–182.0)	60	83	72
D1061	1413.3	23	17.48	7	4.54	30	22.63	131.57 (83.4–197.4)	154.21 (61.8–317.7)	132.58 (89.4–189.3)	61	82	60
D1062	252.9	6	4.53	1	0.77	7	4.55	132.43 (48.4–288.3)	129.26 (1.7–719.2)	153.77 (61.6–316.8)	62	70	82

(continued)

Table A2.3 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group D1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
D1063	1013.31	15	10.95	5	2.96	20	14.60	136.94 (76.6–225.9)	168.73 (54.4–393.8)	136.97 (83.6–211.6)	63	87	68
D1064	500.97	6	4.32	4	2.62	10	7.28	138.76 (50.7–302.0)	152.41 (41.0–390.2)	137.43 (65.8–252.7)	64	81	69
D1065	902.48	14	10.03	3	2.74	17	12.58	139.55 (76.2–234.2)	109.45 (22.0–319.8)	135.18 (78.7–216.4)	65	62	65
D1066	1052.11	17	12.16	6	2.69	23	15.70	139.83 (81.4–223.9)	223.13 (81.5–485.7)	146.51 (92.8–219.8)	66	98	76
D1067	1857.83	17	11.74	7	4.19	24	17.05	144.75 (84.3–231.8)	167.11 (66.9–344.3)	140.80 (90.2–209.5)	67	86	73
D1068	1665.76	18	12.26	0	3.17	18	16.36	146.78 (86.9–232.0)	0.00 (0.0–115.8)	110.02 (65.2–173.9)	68	12	46
D1069	508.22	3	2.03	1	0.73	4	3.27	147.75 (29.7–431.7)	136.65 (1.8–760.3)	122.23 (32.9–312.9)	69	72	53
D1070	619.42	7	4.72	3	2.43	10	6.09	148.26 (59.4–305.5)	123.65 (24.9–361.3)	164.14 (78.6–301.9)	70	68	88
D1071	617.73	10	6.69	3	2.57	13	10.43	149.44 (71.5–274.8)	116.67 (23.4–340.9)	124.60 (66.3–213.1)	71	65	55
D1072	1298.77	20	13.32	8	3.73	28	17.54	150.20 (91.7–232.0)	214.50 (92.4–422.7)	159.65 (106.1–230.7)	72	95	84
D1073	678.23	17	11.23	8	4.05	25	15.12	151.32 (88.1–242.3)	197.69 (85.1–389.6)	165.31 (107.0–244.0)	73	91	89
D1074	267.13	14	8.96	2	1.74	16	11.75	156.26 (85.4–262.2)	115.07 (12.9–415.5)	136.19 (77.8–221.2)	74	64	67
D1075	915.97	12	7.59	7	3.43	19	11.79	158.14 (81.6–276.3)	204.26 (81.8–420.9)	161.10 (96.9–251.6)	75	92	86
D1076	321.34	5	3.11	0	1.19	5	4.89	160.99 (51.9–375.7)	0.00 (0.0–308.5)	102.33 (33.0–238.8)	76	17	41
D1077	388.42	6	3.71	0	0.61	6	4.61	161.69 (59.0–351.9)	0.00 (0.0–599.5)	130.29 (47.6–283.6)	77	10	58
D1078	1027.32	18	11.05	6	5.28	24	15.56	162.83 (96.5–257.4)	113.63 (41.5–247.3)	154.26 (98.8–229.5)	78	63	83
D1079	1858.28	22	13.48	10	4.33	32	18.23	163.25 (102.3–247.2)	230.89 (110.5–424.6)	175.51 (120.0–247.8)	79	103	91
D1080	1171.41	18	10.85	7	5.94	25	16.81	165.91 (98.3–262.2)	117.91 (47.2–243.0)	148.70 (96.2–219.5)	80	66	77
D1081	647.41	12	7.19	3	2.79	15	10.53	166.97 (86.2–291.7)	107.60 (21.6–314.4)	142.46 (79.7–235.0)	81	59	75
D1082	1144.16	22	12.88	10	2.51	32	16.90	170.77 (107.0–258.6)	397.67 (190.4–731.4)	189.31 (129.5–267.3)	82	108	95
D1083	1193.14	31	17.96	3	5.76	34	22.82	172.60 (117.2–245.0)	52.10 (10.5–152.2)	148.97 (103.2–208.2)	83	32	78

(continued)

Table A2.3 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group D1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
D1084	386.58	10	5.63	2	1.98	12	7.99	177.64 (85.0–326.7)	101.05 (11.3–364.8)	150.15 (77.5–262.3)	84	56	79
D1085	386.97	8	4.42	2	2.24	10	7.42	180.82 (77.9–356.3)	89.26 (10.0–322.3)	134.83 (64.5–248.0)	85	47	63
D1086	394.09	11	6.08	0	2.17	11	8.22	181.01 (90.2–323.9)	0.00 (0.0–168.7)	133.82 (66.7–239.5)	86	9	62
D1087	1552.27	23	12.55	2	3.66	25	18.02	183.34 (116.2–275.1)	54.67 (6.1–197.4)	138.73 (89.8–204.8)	87	34	71
D1088	908.21	12	6.43	0	1.84	12	8.45	186.76 (96.4–326.3)	0.00 (0.0–199.3)	142.00 (73.3–248.1)	88	3	74
D1089	225.69	3	1.60	0	0.58	3	2.34	187.51 (37.7–547.9)	0.00 (0.0–632.6)	128.25 (25.8–374.7)	89	11	57
D1090	1387.25	8	4.01	9	2.13	17	6.45	199.49 (85.9–393.1)	422.45 (192.8–802.0)	263.72 (153.5–422.3)	90	109	106
D1091	1112.18	23	11.33	7	3.69	30	15.52	202.99 (128.6–304.6)	189.76 (76.0–391.0)	193.31 (130.4–276.0)	91	90	97
D1092	302	12	5.87	5	1.65	17	7.14	204.31 (105.4–356.9)	302.16 (97.4–705.1)	238.16 (138.7–381.3)	92	106	105
D1093	265.76	12	5.85	1	1.89	13	8.09	205.09 (105.9–358.3)	53.03 (0.7–295.0)	160.65 (85.5–274.7)	93	33	85
D1094	1449.34	22	10.71	3	3.61	25	15.52	205.39 (128.7–311.0)	83.06 (16.7–242.7)	161.13 (104.2–237.9)	94	44	87
D1095	658.92	18	8.69	2	3.55	20	13.14	207.21 (122.7–327.5)	56.40 (6.3–203.6)	152.26 (93.0–235.2)	95	36	80
D1096	1165.51	28	13.40	11	5.09	39	17.85	208.97 (138.8–302.0)	216.29 (107.8–387.0)	218.49 (155.4–298.7)	96	96	103
D1097	1458.31	35	16.72	11	5.34	46	23.13	209.27 (145.7–291.1)	205.94 (102.7–368.5)	198.87 (145.6–265.3)	97	93	98
D1098	260.31	5	2.34	0	0.81	5	3.79	213.64 (68.8–498.6)	0.00 (0.0–451.4)	131.97 (42.5–308.0)	98	4	59
D1099	871.67	15	6.94	2	1.99	17	9.52	216.23 (120.9–356.7)	100.55 (11.3–363.0)	178.50 (103.9–285.8)	99	54	92
D1100	276.42	10	4.51	1	1.11	11	5.80	221.80 (106.2–407.9)	89.90 (1.2–500.2)	189.69 (94.6–339.4)	100	48	96

(continued)

Table A2.3 (continued): Observed and expected deaths, HSMRs, 95% CIs, and the peer rankings for 80%, 20% and 100% of in-hospital deaths for peer group D1

Study assigned ID	cwaseps	80%		20%		100%		HSMRs			Rank		
		O	E	O	E	O	E	80%(LCI–UCI)	20%(LCI–UCI)	100%(LCI–UCI)	80%	20%	100%
D1101	1058.58	20	8.29	0	2.18	20	11.18	241.36 (147.4–372.8)	0.00 (0.0–168.3)	178.87 (109.2–276.3)	101	2	93
D1102	461.58	4	1.61	1	0.54	5	2.72	249.04 (67.0–637.6)	186.62 (2.4–1038.3)	183.91 (59.3–429.2)	102	89	94
D1103	637.57	15	5.92	6	2.62	21	10.53	253.24 (141.6–417.7)	228.88 (83.6–498.2)	199.50 (123.4–305.0)	103	101	99
D1104	530.33	9	2.94	0	0.83	9	4.27	306.55 (139.9–582.0)	0.00 (0.0–442.6)	210.78 (96.2–400.2)	104	1	101
D1105	601.32	14	4.07	1	1.98	15	6.39	344.37 (188.1–577.8)	50.60 (0.7–281.6)	234.66 (131.2–387.1)	105	31	104
D1106	472.97	13	3.42	2	1.31	15	5.09	380.07 (202.2–650.0)	152.40 (17.1–550.2)	294.46 (164.7–485.7)	106	80	107
D1107	271.58	8	2.00	0	1.50	8	3.67	400.74 (172.6–789.7)	0.00 (0.0–244.8)	218.24 (94.0–430.0)	107	8	102
D1108	746.91	14	2.89	5	2.19	19	4.64	485.07 (265.0–813.9)	228.00 (73.5–532.1)	409.49 (246.4–639.5)	108	100	108
D1109	440.94	17	3.49	3	1.30	20	4.75	486.58 (283.3–779.1)	230.24 (46.3–672.7)	421.45 (257.3–650.9)	109	102	109

Appendix 3 Funnel plots of HSMRs in 2005–06

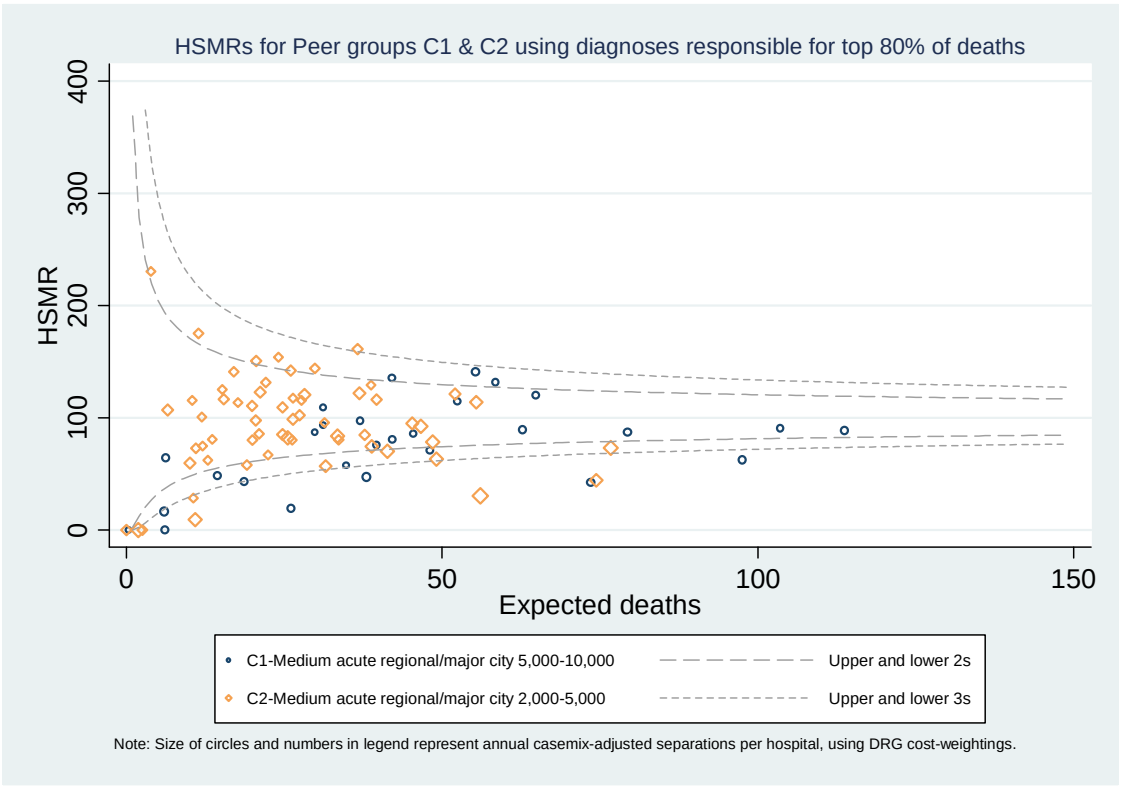


Figure A3.1: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group C1, and C2, 80% of in-hospital mortality

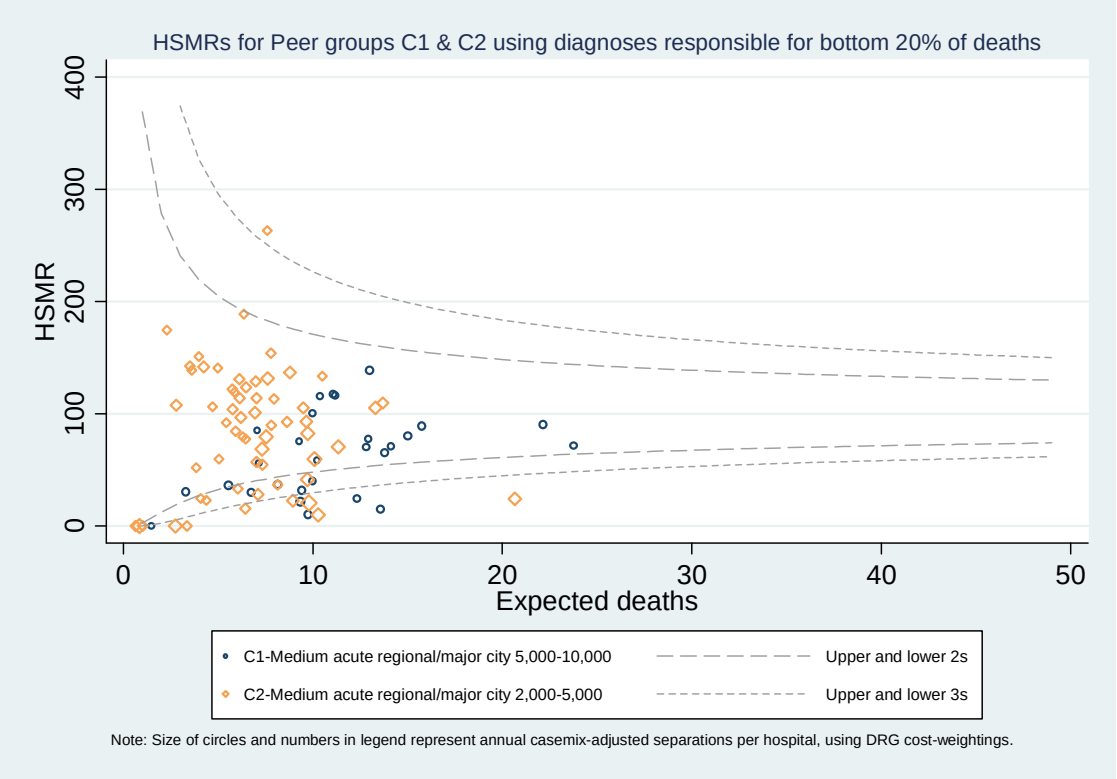


Figure A3.2: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group C1, and C2, 20% of in-hospital mortality

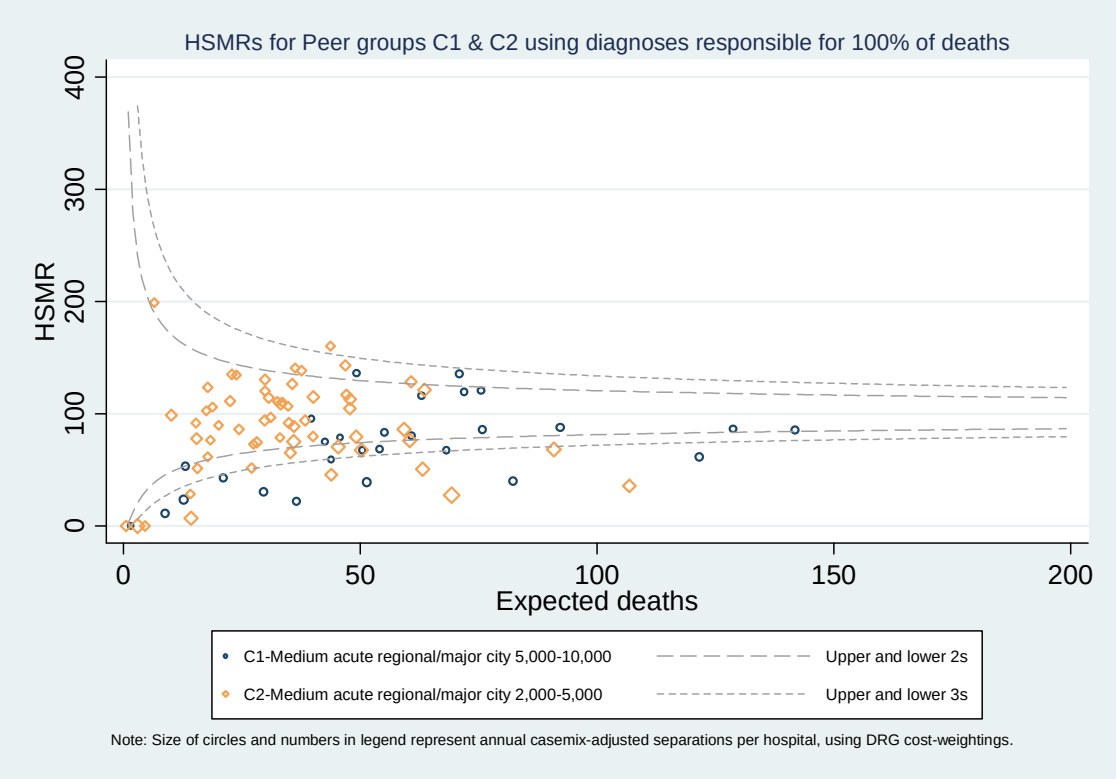


Figure A3.3: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group C1, and C2, 100% of in-hospital mortality

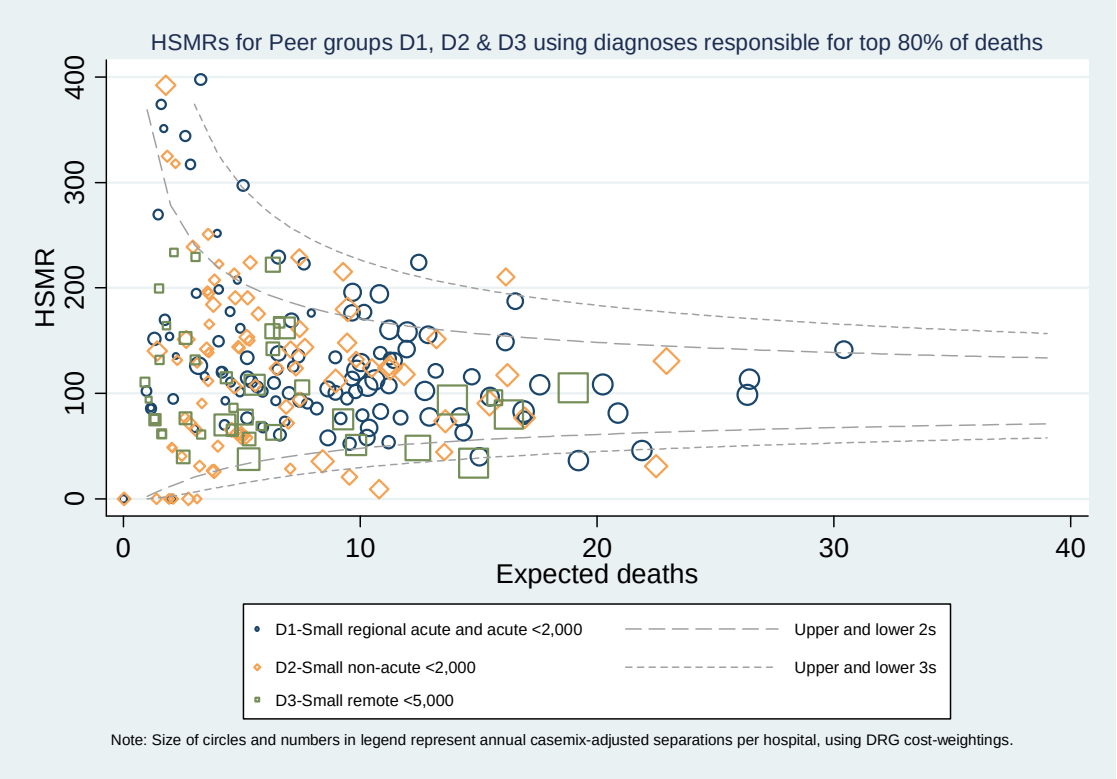


Figure A3.4: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group D1, D2 and D2, 80% of in-hospital mortality

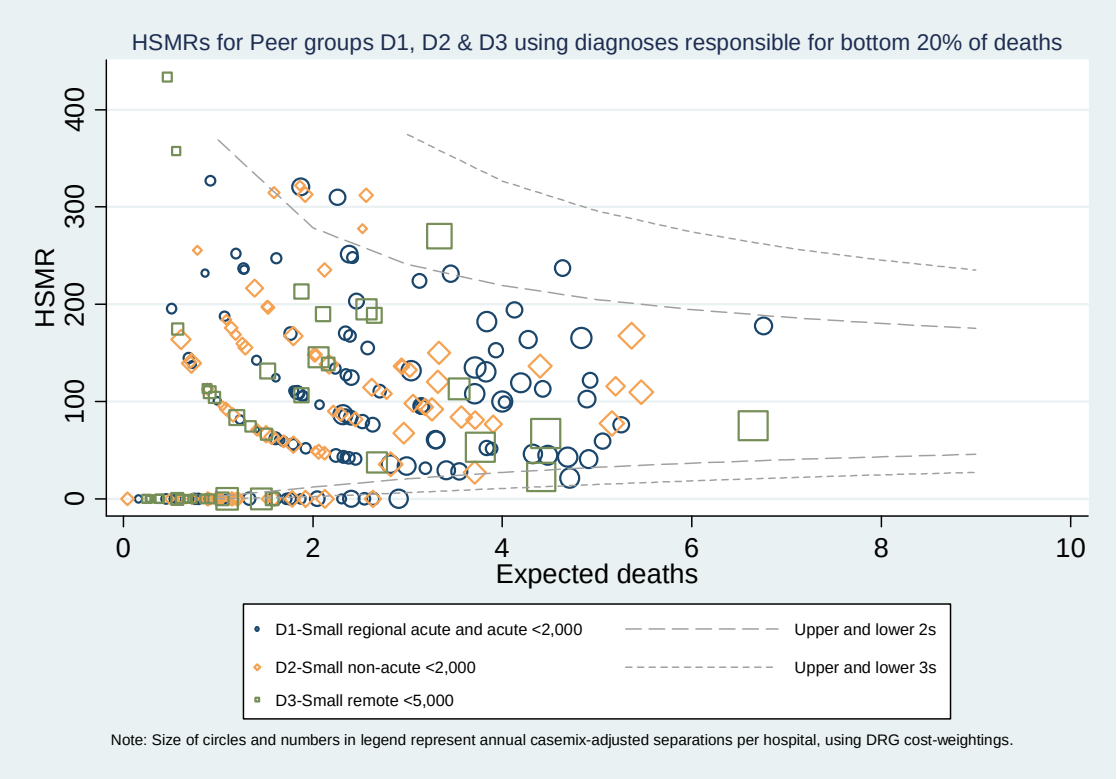


Figure A3.5: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group D1, D2 and D3, 20% of in-hospital mortality

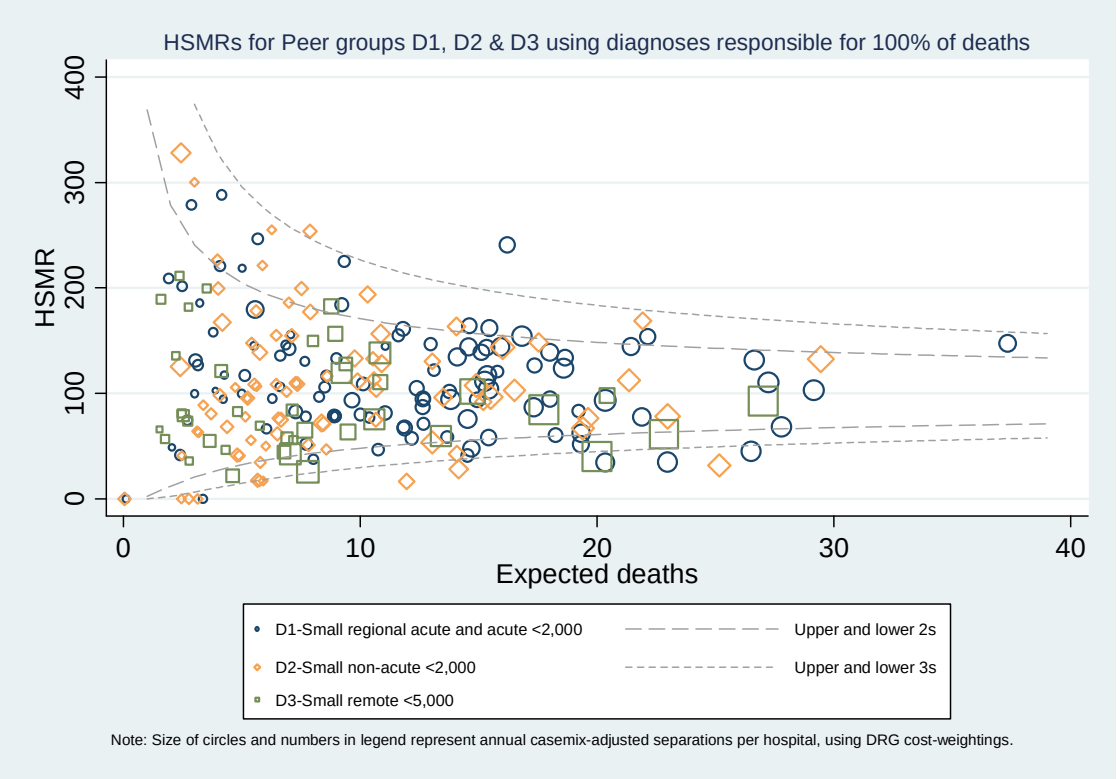


Figure A3.6: Variation in HSMRs according to the expected number of deaths and the size of the institution, peer group D1, D2 and D3, 100% of in-hospital mortality

Appendix 4 Caterpillar plots

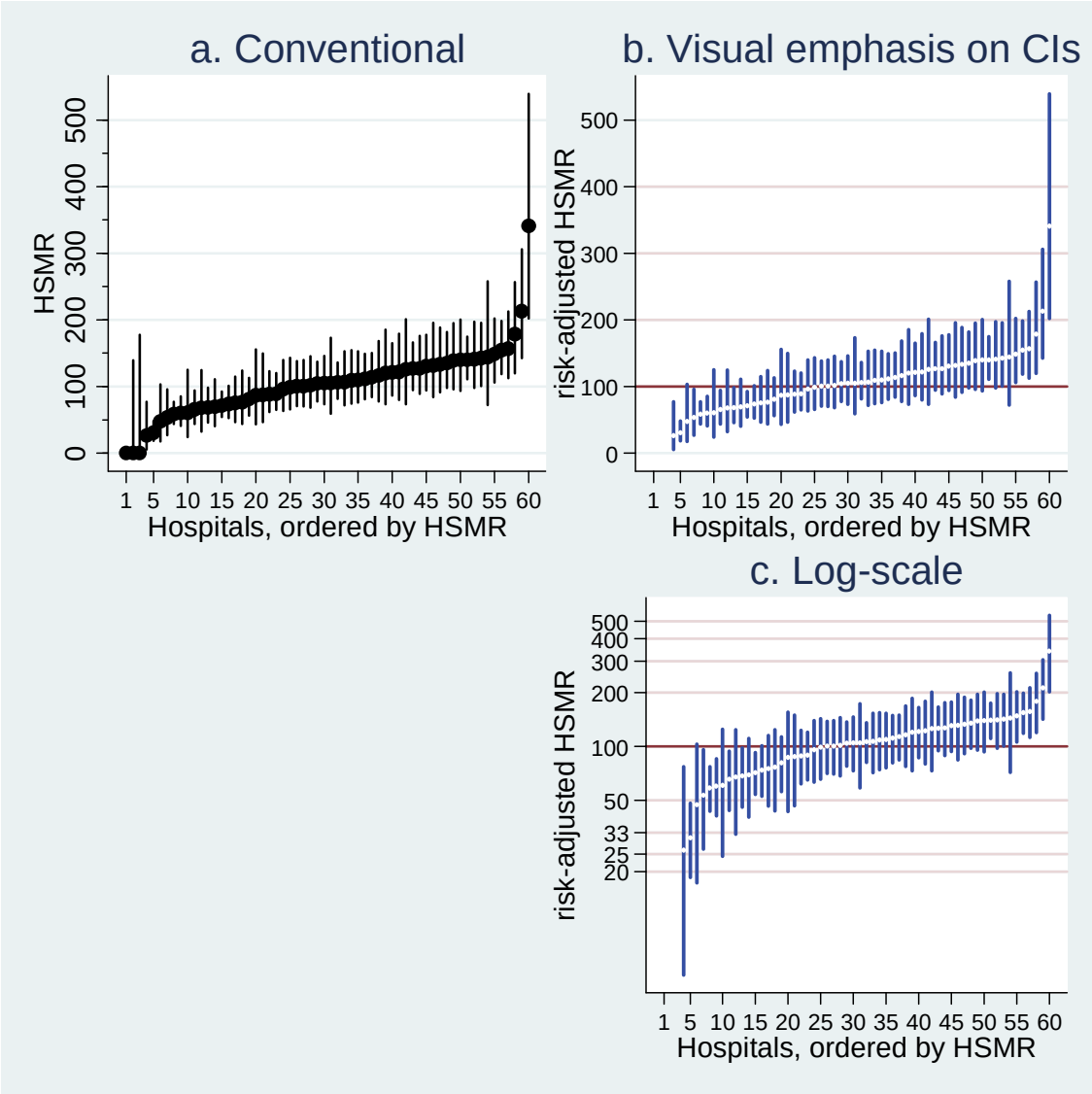


Figure A4.1. Caterpillar plots: variations of format and scaling

Figure A4.1a presents HSMRs for a group of hospitals. The figure is a fairly standard caterpillar plot. The y-axis is linear and the point estimates have more visual emphasis than the confidence intervals.

Figures A4.1b and A4.1c present the same HSMRs as Figure A4.1a. Figure A4.1b differs only in giving more visual emphasis to the confidence intervals, on the grounds that this will reduce a tendency of readers to focus on the point estimates, thus treating the figure like a league table.

Figure A4.2c presents the same data with the HSMRs placed on a log scale. The argument for doing this is that because the HSMR is a ratio it should be presented on a ratio scale. An

HSMR of 50 is half the reference value of 100 and an HSMR of 200 is twice the reference value. Use of a log scale places HSMRs of 50 and 200 the same distance from the line marking 100. (Guide-lines have been placed on Figure A4.2b at points corresponding to half and double 100, one-third and triple, and so on.)

An effect of the transformation is to give more visual emphasis to HSMRs that are below 100, especially those well below it. Conversely, HSMRs that are well above 100 have less visual emphasis.

Figure A4.2 shows the same set of HSMRs in three further ways. Like Figure A4.1c, all of these figures place the HSMRs on a log scale. The main difference is that the presentation has been transposed. The result is somewhat similar to a forest plot. Although not done here, this orientation lends itself to inclusion of hospital names in the plot. The names can be placed in the ordinary orientation for reading and many names can be shown in a figure that will fit on a single page.

The first chart in Figure A4.2, like all of those in Figure A4.1, shows the point estimate of HSMR as a fixed symbol. The second differs by using a circle centred on the point estimate value, the size of which corresponds to the size of the hospital (measured in terms of casemix adjusted separations). Funnel plots include that information. The third has both a small symbol marking the point estimate and a circle indicating hospital size.

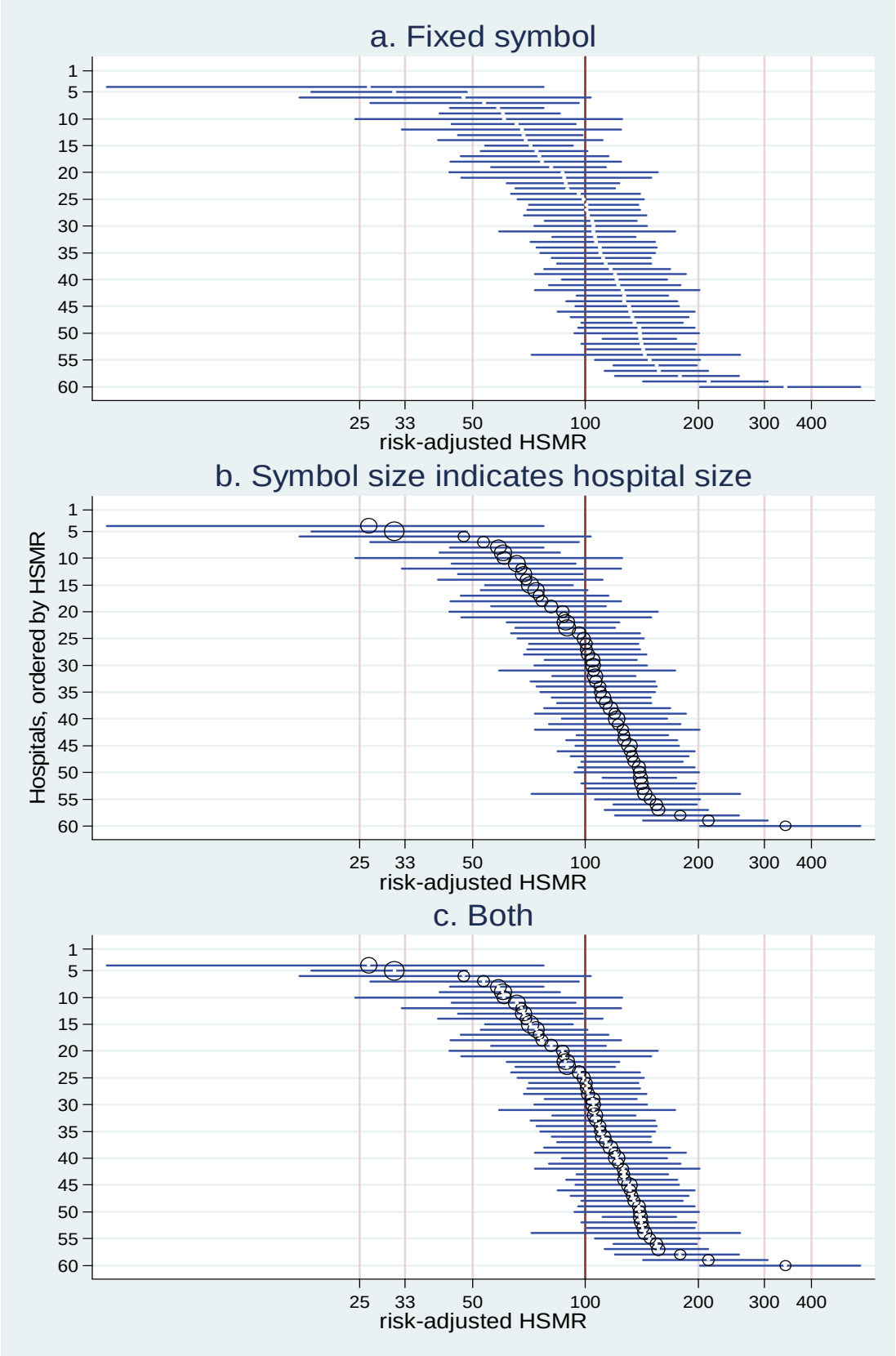


Figure A4.2. Three presentations of transposed log-scaled caterpillar plots

Appendix 5 Data issues

National Hospital Morbidity Database

The hospital separations data were provided by the AIHW, from the NHMD. All data elements within the NHMD used in this study conform to the requirements and definitions set out in the National Health Data Dictionary (HDSC 2006) unless otherwise specified.

Detailed information about individual data elements within the NHMD can be found within the National Health Data Dictionary or online using METEOR (<http://meteor.aihw.gov.au>)—the AIHW's Metadata Online Registry.

For further information about the data used in this project and about the topics below, readers are referred to the AIHW publication Australian Hospital Statistics, 2005–06 (AIHW 2007) and the equivalent publications for data-years 2004–05 and 2006–07.

The following sections are taken from the Appendixes of Australian Hospital Statistics 2005–06 (AIHW 2007) and provide information about categories and classifications used in this report.

Public and private hospitals

Taken from: Australian Hospital Statistics 2005–06 Appendix 2: Hospitals contributing to the report and public hospital peer groups (AIHW 2007: 311)

'Throughout this report, unless otherwise specified:

- public acute hospitals and public psychiatric hospitals are included in the public hospital (public sector) category
- all public hospitals other than public psychiatric hospitals are included in the public acute hospital category
- private psychiatric hospitals, private free-standing day hospital facilities and other private hospitals are included in the private hospital (private sector) category
- all private hospitals other than private free-standing day hospital facilities are included in the other private hospitals category.

There is currently some variation between jurisdictions in whether hospitals that predominantly provide public hospital services, and that are privately owned and/or operated, are reported as public or private hospitals. A selection of these hospitals is listed in Table A2.1 in the AIHW report with information on whether they are reported as public or private hospitals.

Other changes in hospital ownership or management arrangements can also affect whether hospital activity is reported as public or private. For example, between 2003–04 and 2004–05 two private hospitals in Western Australia were purchased by the Western Australian Department of Health and were amalgamated with two existing public hospitals. Hence the activity associated with the former private hospitals is now included in the activity reporting of the two public hospitals. From 2004–05, the Mersey Community Hospital in Tasmania, which previously operated as a private hospital providing predominantly public services on

a contracted basis, merged with the Northwest Regional Hospital and is categorised as a public hospital.'

Public hospital peer groups

Taken from: Australian Hospital Statistics 2005–06. Appendix 2: Hospitals contributing to the report and public hospital peer groups (AIHW 2007: 317)

The AIHW worked with the National Health Ministers' Benchmarking Working Group (NHMBWG) and the National Health Performance Committee (NHPC) to develop a national public hospital peer group classification for use in presenting data on costs per casemix-adjusted separation. The aim was to allow more meaningful comparison of the data than comparison at the jurisdiction level would allow.

The peer groups were therefore designed to explain variability in the average cost per casemix-adjusted separation. They also group hospitals into broadly similar groups in terms of their range of admitted patient activity, and their geographical location, with the peer groups allocated names that are broadly descriptive of the types of hospitals included in each category. Although not specifically designed for purposes other than the cost per casemix-adjusted separation analysis, the peer group classification is recognised as a useful way to categorise hospitals for other purposes, including the presentation of other data.'

The peer group to which each public hospital was assigned for 2005–06 is included in Table A2.2 within the Australian Hospital Statistics 2005–06 publication and is summarised in Table 2 in Section 4.1.2 of this report.

SEIFA

Taken from: Australian Hospital Statistics 2005–06. Appendix 1: Technical notes (AIHW 2007: 301–2)

The 'SEIFA Index of Advantage/Disadvantage was generated by the ABS using a combination of Census data, including variables measuring both advantage and disadvantage. A higher score on the index indicates that an area has attributes that measure advantage, such as a relatively high proportion of people with high incomes or a skilled workforce. It also means an area has a low proportion of people with variables that measure disadvantage, such as low incomes and relatively few unskilled people in the workforce.

Conversely, a low score on the index indicates that an area has a high proportion of individuals with variables that measure disadvantage, such as low incomes and more employees in unskilled occupations, and a low proportion of people with variables that measure advantage, such as high incomes or people in skilled occupations. Hence, the index offsets any disadvantage in an area with advantage.

Separation rates by quintile of advantage/disadvantage were generated by the AIHW by using the SEIFA scores for this study for the SLA of usual residence of the patient reported for each separation. The most disadvantaged quintile represents the areas containing the 20% of the population with the least advantage/most disadvantage and the most advantaged quintile represents the areas containing the 20% of the population with the least disadvantage/most advantage.'

Errors, inconsistencies and uncertainties

NHMD data are generally abstracted from records, entered and coded in hospitals, passed to state and territory health departments, then to the AIHW before being provided to the National Injury Surveillance Unit (NISU). Processing occurs at each of these steps. Errors and inconsistencies can arise due to the large number of people and processes involved in providing the data. Some variations occur in reporting and coding although Coding Standards, National Minimum Data Sets and other mechanisms have reduced this.

Quality of ICD-10-AM coded data

Taken from *Australian Hospital Statistics 2005–06. Appendix 1: Technical* notes (AIHW 2007: 288–9)

'Diagnosis, procedure and external cause data for 2005–06 were reported to the NHMD by all states and territories using the fourth edition of the International statistical classification of diseases and related health problems, 10th revision, Australian modification (ICD-10-AM) (NCCH 2004).

The quality of coded diagnosis, procedure and external cause data can be assessed using coding audits in which, in general terms, selected records are independently recoded, and the resulting codes compared with the codes originally assigned for the separation. There are no national standards for this auditing, so it is not possible to use information on coding audits to make quantitative assessments of data quality on a national basis.

The quality and comparability of the coded data can, however, be gauged by information provided by the states and territories on the quality of the data, by the numbers of diagnosis and procedure codes reported and by assessment of apparent variation in the reporting of additional diagnoses. The comparability of the data can also be influenced by state-specific coding standards.'

List of tables

List of figures