MELBOURNE HEALTH

THE UNIVERSITY OF MELBOURNE

MELBOURNE EPICENTRE

Hospital Mortality Indicator (HMI) Review

DATE: 10 September 2013

PREPARED BY:

Melbourne EpiCentre in collaboration with Fiona Landgren and Carole Staley

Melbourne EpiCentre
c/-The Royal Melbourne Hospital – 7 East, Grattan Street, Parkville 3050
Tel: 03 9342 8772 Fax: 03 9342 8780

Acknowledgements

Thank you to Mr Neville Board and Ms Lisa Wallis, Australian Commission on Safety and Quality in Health Care for their support in scoping the literature review, providing key information sources and providing feedback on report development.

Thank you to the staff at Communicating for Health for their support in developing and formatting the report, and to Wendy Lemaire, Melbourne EpiCentre for her administrative support.

Thank you to the Hospital Mortality Indicator Advisory group members who contributed feedback on the draft report.

Suggested Report Citation

Brand C, Landgren F, Staley C, Tropea J, Liew D, Bohensky M, & Gorelik A. Hospital Mortality Indicator (HMI) Review (2013). Commissioned by the Australian Commission on Safety and Quality in Health Care.

Table of Contents

Abbreviations

ABS	Australian Bureau of Statistics
ACHS	Australian Council on Healthcare Standards
ACSQHC	Australian Commission on Safety and Quality in Healthcare
AE	Adverse Events
AHS	Australian Hospital Statistics
AIHW	Australian Institute of Health and Welfare
AHRQ	Agency for Healthcare Research and Quality
AMI	Acute myocardial infarction
APR-DRG	All-patient Refined Diagnostic Related Groups
AR-DRG	Australian Refined - Diagnosis Related Group
ATSI	Aboriginal and Torres Strait Islander
AusPSIs	Victorian Patient Safety Indicators
CABG	Coronary artery Bypass Grafting
CCI	Charlson comorbidity index
CDC	Centre for Disease Control and Prevention
CEHSEU	Clinical Epidemiology and Health Service Evaluation Unit
CHF	Congestive Heart Failure
CIHI	Canadian Institute for Health Information
CUSUM	Cumulative Sum
DNR	Do not resuscitate
DRG	Diagnostic Related Group
ED	Emergency Department
EWMA	Exponentially-Weighted Moving Average
GP	General Practitioner
HCFA	USA Health Care Financing Authority
HCQI	Health Care Quality Indicators
HMI	Hospital Mortality Indicator
HSMR	Hospital Standardised Mortality Ratio
ICD-9-CM	International Statistical Classification of Diseases and Related Health Problems, Ninth Revision, Clinical Modification
ICD-10	International Statistical Classification of Diseases and Related Health Problems, Tenth Revision
ICD-10-AM	International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification
ICU	Intensive care unit
JCAHO	Joint Commission on Accreditation of Health care Organisations
LMDRG	Low mortality DRG
LOS	Length of stay
MeSH	Medical Subject Heading
NHMD	National Hospital Morbidity Database
NHMRC	National Health and Medical Research Council

NHPA	National Health Priority Area
NHPC	National Health Performance Committee
NHS	National Health Service
OECD	Organisation for Economic Co-operation and Development
PC	Palliative care
RACGP	Royal Australian College of General Practitioners
RCT	Randomised Controlled Trials
SIGN	Scottish Intercollegiate Guidelines Network
SMR	Standardised mortality ratio
SR	Systematic Review
UK	United Kingdom
USA	United States of America
VLAD	Variable Life Adjusted Display
WHO	World Health Organisation

Executive Summary

Purpose and objectives

This review, commissioned by the Australian Commission on Safety and Quality in Health Care (ACSQHC), includes information pertaining to the ACSQHC core in-hospital mortality indicators (HMIs), including;

- Hospital Standardised Mortality Ratio (HSMR)
- Death in low mortality DRGs (LMDRG)
- Condition specific in-hospital mortality for
  - Acute myocardial infarction (AMI)
  - Stroke
  - Fractured neck of femur
  - Pneumonia

The review will be used to inform further development, implementation and evaluation of the HMIs.

Review methods

The structured narrative review has integrated information from peer reviewed literature and Australian and International grey literature. The included literature was identified using structured peer review and grey literature search strategies designed to answer the following questions

1. What are the current approaches to generation and presentation of mortality indicators?
2. What are the measurement attributes of HMIs?
3. What risk adjustment models and statistical issues are associated with use of HMIs?
4. What is the usefulness of HMIs in terms of monitoring, assessing and improving quality of care?

To meet the timeframes of the review and reduce duplication, a number of scoping decisions were made. Key review sources were identified for each HMI group for use as the basis for updating the literature in that area. For HSMR the base source was the 2009 AIHW report ‘Measuring and reporting mortality in hospital patient’ by Ben Tovim and colleagues (AIHW: Ben-Tovim D et al, 2009). For LMDRG, the base source was a review by Mihrshahi et al (2010) entitled ‘Validity of the indicator, death in low-mortality diagnosis-related group for measuring patient safety and healthcare quality in hospitals’ (Mihrshahi, S. et al, 2010). For Condition specific HMI the base source was the previous ACSQHC report (2010), ‘Mortality and unplanned admissions indicator project’ (Australian Commission on Safety and Quality in Health Care, 2010). Further, the review of risk adjustment models and statistical issues focused on HSMR and, for all HMI groups, only articles which included issues related to in hospital definitions of HMI were included. Therefore, the limitations of this review include that articles are relevant to the questions of interest, which may have added insightful commentary on the questions of interest but did not meet inclusion criteria, may have been overlooked.

Peer reviewed studies were critically appraised. The information has been collated and summarized for each HMI group, with accompanying UK, Canada and US case studies providing an overview of jurisdictional activities.

Key findings

Generation of HMI models

- Peer reviewed literature indicated a continuing focus on developing and testing HMI models, especially for general (whole of hospital) HMIs such as the ACSQHC HSMR.
- A key area of interest pertains to definition of mortality with a move in some jurisdictions (UK in particular) towards using 30-day admission or post-discharge models. Whilst there are strong correlations between in-hospital and 30-day models, the impact on outlier status can be significant, especially for condition specific HMIs, stroke and fractured neck of femur. Feasibility of access to linked data for population deaths is the key barrier to 30-day models.
- Inclusion and exclusion of specific patient populations is also a focus of discussion in peer reviewed and grey literature. Unlike the original UK, DR Foster HSMR, the NHS has developed a new SHMI that not only uses a 30-day post discharge definition of mortality but includes all deaths. It provides 'companion' data such as crude mortality data and palliative care rates for support in interpreting results.
- With regard to risk-adjustment, there has been particular interest in exploring issues associated with "The constant risk fallacy" whereby assumed constancy of risk associated with risk adjustment variables such as transfers, urgency of admission and comorbidity may be invalidated by changes in hospital admission and discharge policies and differences in coding practice. There is ongoing debate and several studies suggest there is variation related to inclusion of these variables that can impact on in-hospital mortality rates and outlier status. However Australian and UK data suggests the impacts are modest.
- A key variable of interest in peer reviewed and grey literature is that of 'palliative care status'. Whilst there is a need to account for patients admitted for end of life care, it is also acknowledged that palliative care services are provided also for supportive care where death is not expected. There are also concerns about the potential for coding practice changes and the need to ensure such changes are appropriate. Downward trends in mortality in Canada and the UK, have been attributed in part to increased palliative care coding. Overall, it is agreed that palliative care status can impact on hospital mortality rates and needs to be considered when interpreting changes in rates, however there are difficulties in defining this variable and further investigation is needed. As described above, one approach exemplified by the UK SHMI is to include all palliative care patients and assess changes in mortality rates in association with palliative care status rates.
- There is limited investigation into statistical modelling methods, including use of indirect or direct standardisation and fixed or random effects estimation methods. There are benefits and limitations of each method but decisions about which to use should be driven by the purpose for which the data is to be used.
- A number of studies report comparison of models and adaptation of models for other populations. Overall, HMI models exhibit strong discrimination properties but are frequently associated with differences in measures mortality rates and identification of outlier status. Using another jurisdiction's model as the basis for developing a model is generally reported to be associated with modification and adaptation to meet contextual needs. These finding raise questions about the appropriateness of comparing results between jurisdictions or even between hospitals where different commercial HMI products are being used. All models require testing of performance attributes and need for calibration when used in a population other than that in which the model was developed.
- We failed to identify a strong literature focus on the validity and usefulness of HMIs. A previous systematic review concluded that "the general notion that hospitals with higher risk-adjusted mortality have poorer quality of care is neither consistent nor reliable". We identified only two

peer-reviewed studies that have addressed the relationship between HMIs and quality of care since this review and these did not report an association. A number of studies have noted reduction in hospital mortality rates over time. These changes are frequently large, out of keeping with population-based reductions and whilst improvements in hospital quality of care are likely to be contributing, definitive causation is elusive.

- The case studies have contributed to the review. There is variation across jurisdictions with regard to specifications of mortality models and the maturity of implementation systems. In addition to increasing interest in use of 30-day mortality models, all jurisdictions are promoting use of whole of hospital mortality rates in association with other data, including crude mortality rates, palliative care rates, service data e.g for medical, surgical, ICU and condition specific HMIs. Further, jurisdictions are setting up mechanisms by which individual hospitals can directly access their data and some provide instruction about how to interpret the data, using a stepped approach. For instance, the UK Dr Foster recommends using the data in conjunction with other service specific data and to follow a process whereby; data coding is checked, case-mix coding is checked, hospital structure relative to peers is checked (eg staffing, teams, use of care pathways) and process of care is checked before investigating individuals or specific teams.

In conclusion, this review has built on previous literature to provide an updated picture of the work being undertaken to elucidate the composition, utilisation and utility of general HMIs, condition specific HMIs and the death in LMDRG indicator. There is ongoing interest in improving technical specifications of standardised HMI models and to date a lesser focus on investigating implementation and utilisation issues.

There is a strong move towards models that include 30-day linked death data and towards use of companion indicators to support interpretation of identified variation. There is a need to tailor data analysis and presentation to meet the purpose for which the data is being used and this may require data providers and analysers to generate a number of different models and presentation formats for their potential customers (jurisdictions, hospitals, clinicians, patients and their carers, community).

Overall, there remains a significant lack of information available about the resource implications and allocative effectiveness of these programs, and relative added value to existing high quality patient safety performance data, that requires further study.

1. Introduction

1.1 Project context

In August 2011, reflecting an ongoing national focus on improving quality in healthcare, the Council of Australian Government (COAG) National Health Reform Agreement outlined objectives for national health reform, including

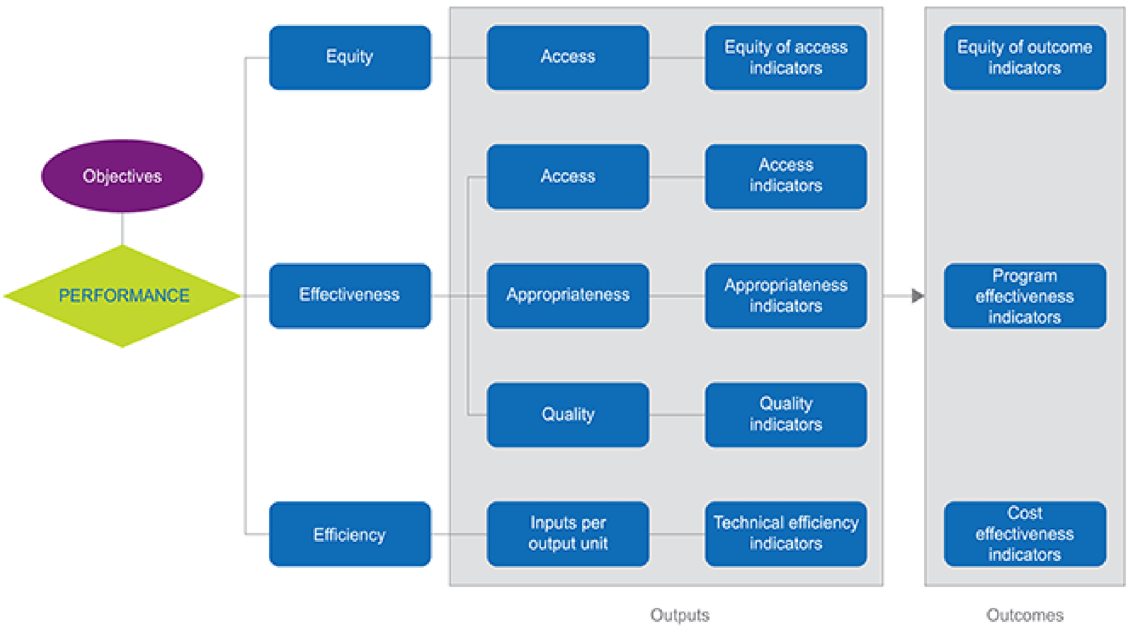
- improving performance reporting through the establishment of the National Health Performance Authority, and
- improving accountability through the Performance and Accountability Framework

The Framework aims to support a safe, high quality Australian health system through improved transparency and accountability. In particular it aims to ensure that:

- members of the public have access to relevant, up to date information about health care providers (both public and private), to allow them to make informed choices;
- Local Hospital Networks (LHNs), hospitals, clinical units, Medicare Locals and health services are provided with comparative clinical performance data to foster continuous quality improvement;
- governments and governing bodies of private hospitals are provided with additional data to ensure robust accountability and oversight of the health system; and
- existing reporting mechanisms are supported and coordinated, such as the COAG Reform Council processes.

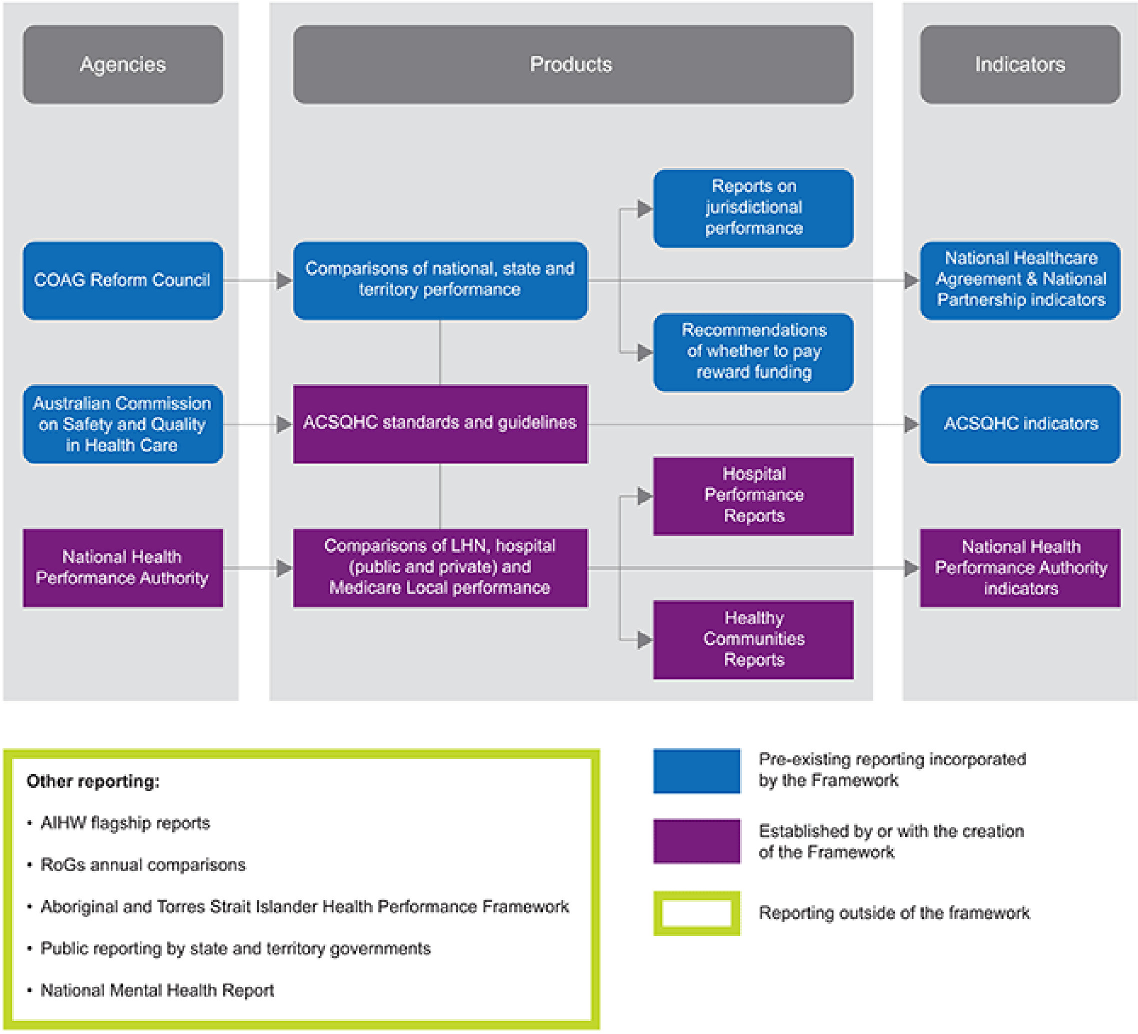
The Framework underpins reporting across three domains – equity, effectiveness and efficiency of service delivery in health care (Figure 1), with indicators drawn from existing indicators such as the Australian Commission on Safety and Quality in Health Care (ACSQHC) indicators and newly developed National Health Performance Authority Indicators (Figure 2).

Figure 1. Indicators for Report on Government Services



Source: National Health Performance Authority, Performance and Accountability Framework, May 2012

Figure 2. Performance and Accountability Framework



Source: National Health Performance Authority, Performance and Accountability Framework, May 2012

Since its inception in 2006, a major body of work for the Australian Commission on Safety and Quality in Health Care has related to the development of quality indicators to support national reporting on the state of safety and quality in health care. This has included hospital level indicators, including hospital mortality and readmission indictors, and has culminated in a draft National core, hospital based outcome indicator specification, provided to the Australian hospital and healthcare sector for trialling and consultation in December 2012. The ACSQHC has made clear the proposed role of the indicators as tools for individual hospitals to use as part of their quality improvement programs (Figure 3).

Work is ongoing to further inform and refine the specifications of the indicators and to inform their implementation, and this current project aims to provide input in this regard, specifically in relation to in-hospital mortality indicators.

Figure 3. ACSQHC Position Statement

ACSQHC Position Statement

These indicators are intended to be generated by jurisdictions and private hospital ownership groups from their admitted patient data collections, and reported back to provider facilities to enable routine comparison of hospital outcomes over time, without initially setting benchmarks or targets. This Consultation Draft is not intended to support performance measurement or pay-for-performance schemes.

The safety and quality value lies in developing the report-review-act cycle, based on the routine supply of timely and targeted data back to hospitals.

Rationale

The rationale for ongoing monitoring and review by hospitals of a set of outcome-based indicators is that significant variance can be a signal for issues of either data quality and consistency, resources, or quality of care.

High outlier rates should be seen as a prompt to further detailed investigation. Learnings may be applied from low outlier rates.

Source: ACSQHC Indicator specification report December 2012

1.2 Project objective

The objective of the project is to provide a structured narrative review pertaining to the usefulness and limitations of Hospital Mortality Indicators (HMIs), including;

- Hospital standardised mortality ratio (HSMR)
- Death in low mortality DRGs
- Condition specific in-hospital mortality for
  - Acute myocardial infarction (AMI)
  - Stroke
  - Fractured neck of femur
  - Pneumonia

Specifically, the project provides;

- Summary of peer reviewed literature focussed on studies published following key reviews on general hospital mortality indicators (AIHW: Ben-Tovim D et al, 2009), condition specific hospital mortality indicators (Australian Commission on Safety and Quality in Health Care, 2010) and mortality in LMDRG (Mihrshahi, S. et al, 2010).
- Presentation of case studies pertaining to issues in implementing and using hospital mortality indicators in three international jurisdictions (Canada, the United Kingdom and the United States of America).

2. Background

There is increasing International interest in measuring outcomes for people who experience an episode of hospitalisation. The purpose for such outcomes measurement varies, and includes at a system level (policy makers, funding providers and community); health system planning, a desire to increase transparency and accountability of healthcare providers and to provide surveillance systems to identify and remedy underperforming organisations and overall to improve population healthcare outcomes. At the level of the healthcare provider, the primary aim of such measurement is to support improvement in healthcare provision by informing quality improvement activities associated with equitable and cost-efficient allocation of resources.

Thus health outcomes measurement is of interest at different levels of the health system, often for multiple reasons that may be preferentially prioritized by different stakeholders, thus influencing how the data is used, for instance; for jurisdiction or international comparison and reporting, for national or jurisdictional public reporting to the community to guide patient choice for health service access, for funding provider feedback to healthcare providers about comparative performance to drive improved performance, for healthcare provider internal review and response activities to drive service improvement, and for research.

Further adding to the complexity of drivers for outcomes measurement are population and organisation based factors that influence what outcomes are measured, how they are measured, how comparisons are made and interpreted and how the data is used and acted upon. An ideal measure will have robust attributes for measuring quality health outcomes. These include strong face value such that it will be considered important and relevant to the population of interest by those who will be using the results or may be judged by the results. Further, the measure or measures should be able to be measured accurately, will be reliable when conditions are constant and when measured across different hospitals and will be sensitive to changes that reflect improvement or worsening in the outcome over time. Of primary importance, if the measure is to be used to judge performance in quality of care, variation identified in the measure between hospitals should be related to quality of care issues, having taken into account random variation and patient population casemix differences. Finally, if these attributes are met, measurement suggesting suboptimal outcomes should enable effective action to be taken by those measuring the outcome. Measurement attributes need to be considered for whole of hospital population health outcomes, such as overall hospital mortality, death in patients associated with low mortality DRGs (LMDRG), as well as for patient health outcomes within specific condition groupings, such as acute myocardial infarction (AMI), stroke, pneumonia or fractured neck of femur.

Three key literature review sources form the basis for this literature update;

1. "Measuring and Reporting Mortality in Hospital Patients" (AIHW: Ben-Tovim D et al, 2009) which addresses issues associated with use of a whole of hospital mortality outcome indicator, the HSMR
2. "Mortality and Unplanned Admissions Indicator Project" report (Australian Commission on Safety and Quality in Health Care, 2010) which documents condition specific mortality indicators for AMI, stroke, pneumonia and fractured neck of femur, and
3. "Validity of the indicator, 'death in low mortality diagnosis-related groups' for measuring patient safety and healthcare quality in hospitals", which reviews peer-reviewed literature pertaining to measurement attributes of the indicator LMDRG (Mihrshahi, S. et al, 2010).

The review has also been informed by a number of reports commissioned by the ACSQHC that investigate technical issues associated with generation of standardised HMIs. Finally, there have been a number of key opinion papers and editorials that have contributed to understanding of issues in development of hospital health outcomes measurement for the purposes of performance measurement and driving improvement associated with the hospital mortality indicators of interest in this review. These will be used to complement the peer-reviewed literature and grey literature case studies where the commentary is relevant.

Overall, these documents indicate that the focus of research literature to date has been on general or general hospital standardised mortality measures, in particular the HSMR. Further, for general and condition specific mortality measures, the research focus has been on technical issues associated with derivation of the measures and testing the strength of their predictive ability for death, rather than on demonstrating construct validity (the accuracy with which the measures identify quality of care sensitive variation), implementation methods or assessing sensitivity to quality related change over time within or between hospitals. The following section summarises in brief existing knowledge based on the primary sources.

2.1 HMI sets and HMI definitions

The ACSQHC report 2010 focused on condition-specific hospital mortality and readmission indicators with the aim of identifying indicator sets across Australian and International jurisdictions and documenting the definitions of the indicators of interest and describing documented rates of these indicators across jurisdictions. The report found that quality outcome indicators for condition specific mortality have been developed across a number of International jurisdictions albeit with considerable difference in purpose, target populations and development methods.

In Australia, relevant indicator sets and indicators included the National Health Performance Indicators (2008-9) and the Queensland Variable Life Adjusted Display, VLAD indicators (2009). It was also acknowledged that the Victorian Patient Safety Indicators (AusPSIs) included the death in LMDRG indicator, which was not specifically addressed in the earlier report but is an indicator of interest in this current review.

2.2 Hospital Standardised Mortality Rate (HSMR)

The history of measuring standardised mortality rates in hospitals has been defined by continuous debate. This debate is driven on the one hand, by fundamental beliefs that monitoring whole of hospital mortality will contribute to quality improvement efforts and on the other by rejection of this position based on technological limitations of the measurement.

As expected, participants in the debate have various stakeholder perspectives and many are also heavily invested; politically, financially and academically. All these factors have influenced the debate thus far as well as the direction of research and application of the HMIs.

The main issues associated with development and use of hospital standardised mortality measures have been summarised by Ben Tovim et al (AIHW: Ben-Tovim D et al, 2009). The issues addressed include;

- Mortality definition (e.g. in-hospital only or including after discharge deaths)
- Defining the population of interest (who to include and exclude in the analysis)
- Issues associated with risk adjustment and statistical methods
- Methods for testing the model performance
- Issues associated with interpreting inter-hospital variation
- Methods of presenting information about hospital mortality

- The role of public or private dissemination of findings
- The relationship between hospital mortality and quality of care

In this current review, these issues will be considered and more recent literature findings will be added to our existing knowledge about HSMR development and application.

2.3 Condition specific hospital mortality indicators

Multiple quality indicators for different condition specific hospital mortality were identified within the indicator sets summarised in the ACSQHC report 2010 and were noted to have varying definitions based on timing of the indicators and descriptions of population inclusion and exclusion criteria.

Information about measurement attributes for application as quality indicators and/or comparative performance measures was limited; however development of most indicator sets appeared to be based on conceptual frameworks for quality improvement. Most detail was provided by the Agency of Healthcare Research and Quality (AHRQ) and the Organisation for Economic Cooperation and development (OECD) groups.

Overall, the findings with regard to application of condition specific HMIs were;

- Many jurisdictions report on health care performance data, using a variety of tabulated and graphical presentation methods.
- There was no information identified within the grey literature review about use of the condition specific HMIs at a local hospital level.
- The peer reviewed literature search found a limited number of research studies pertaining to use of publicly reported hospital data. The majority of studies were observational cross-sectional or cohort studies, focused on AMI and focused on issues relating to indicator definition and risk predictor variables leading the report authors to conclude that “reporting of jurisdictional indicators occurs largely in a vacuum of critical peer reviewed analysis”.
- A number of studies identified improvements in reported indicator rates over time, however reasons for such changes were not systematically addressed and assumptions or hypotheses arising from these reports were not confirmed.

This current review will update the specifications and use of these condition specific HMIs across international jurisdictions.

2.4 Death in low mortality DRG (LMDRG)

This indicator, first included as an indicator of quality of care by the AHRQ Patient Safety Indicator (PSI) set, has considerable face validity as death is not expected for treatment of conditions associated with inherently low risk. Therefore the rationale for the indicator is that patients who die in these low mortality groups are more likely to have had a quality of care issue during their admission.

This position was supported by an early peer reviewed paper in which patients in LMDRG who died were 5.2 (95%CI 2.2-8.4) times as likely as others who died, after risk adjustment, to have had a quality issue associated with their care (Hannan, E. L. et al, 1989). The authors noted a number of methodological limitations of this study, particularly retrospective hindsight bias associated with assessment of records with known death (Caplan, R. A. et al, 1991).More recently, a comprehensive review of literature pertaining to LMDRG measurement attributes has identified a number of concerns related to the validity of LMDRG as a measure of quality of care (Mihrshahi, S. et al, 2010). These issues will be reviewed and the literature updated in this current review.

3. Methods

3.1 Peer reviewed literature methods

3.1.1 Research questions

Four questions informed the literature review methods:

1. What are the current approaches to generation and presentation of mortality indicators?
    a. Reporting periods
    b. Report presentation
    c. Feedback to hospitals
    d. Methods used to distinguish data artefact from quality of care/resource issues
2. What are the measurement attributes of HMIs?
    a. Validity
    b. Reliability
    c. Generalisability
3. What risk adjustment models and statistical issues are associated with use of HMIs?
    a. Variables included
    b. Use of statistical limits to identify outliers
    c. Methods used to distinguish data artefact from quality of care/resource issues
4. What is the usefulness of HMIs in terms of monitoring, assessing and improving quality of care?
    a. Implementation methods
    b. Impacts, outcomes on quality of care

3.1.2 Peer reviewed literature search strategy

A pragmatic search strategy was undertaken. The search terms and search strategy used in the previous ACSQHC report (2010) were reviewed and updated to provide a search strategy that was specific (focused) rather than sensitive (design to capture all) to identify relevant studies.

Search terms were designed to capture;

- Indicators (e.g. performance indicators/health indicators/quality indicators)
- General and condition-specific HMIs of interest
- Organisation level acute care settings (e.g. hospitals, health services)

3.1.3 Search terms

MeSH terms and keyword terms were reviewed and updated (Table 1)

3.1.4 Database interrogation

The search strategy was used to interrogate the following databases; Ovid Medline 1996 to May Week 2, 2013, PsychInfo 2002-2013, the Cochrane Collaboration and the Database of Reviews of Evidence. A snapshot Google search was also used to identify further relevant articles that may not have been indexed to Medline or were not included within the Medline database. A search of PubMed using the terms 'hospital' and 'mortality', 1996-2013 was undertaken as a further means of cross referencing eligible articles. Reference lists of eligible articles were, scrutinised to identify other relevant studies. This multipronged, 'snowball', approach was used to rapidly identify as many eligible articles as possible within the project timelines.

Table 1. Search MeSH and keyword terms

	MeSH terms	Other terms
Outcomes	Hospital mortality Mortality Death	League tables report cards performance Hospital deaths Hospital outcome/s Hospital mortality rate/s Case-fatality rates
Conditions	Myocardial infarction Stroke Pneumonia Hip fracture Femoral neck fracture	Heart attack Acute myocardial infarction, AMI Cardiovascular disease Cerebrovascular accidents Fractured neck of femur
Settings	Hospital Emergency Service, Hospital	Acute care Health service/s hospital-wide In-hospital
HMIs	Quality Indicators, health care Quality of care Benchmarking	Performance indicators Clinical indicators Health indicators Performance measures Quality measures Patient safety Organisation* benchmarking Low mortality DRG Hospital standardised mortality/hospital standardised mortality /HSMR Time-series standardised hospital mortality ratio/s/TSHMR Summary hospital mortality index/SHMI Hospital death/s Mortality data Hospital mortality rates Hospital care

The following journals were additionally searched for relevant articles;

- American Journal of Medical Quality
- Annals of Internal Medicine
- Australian Health Review
- BMC Health Services Research

- British Medical Journal
- Canadian Medical Association Journal
- International Journal of Quality in Healthcare
- Medical Journal of Australia
- New England Journal of Medicine
- Quality management in health care

3.1.5 Criteria for critical appraisal

Studies were considered eligible for critical appraisal if they reported results within the context of a trial study design and

- They reported HMIs based on administrative data definitions
- They reported HMIs at a whole of acute hospital/healthcare level
- They reported variation between acute hospitals/healthcare services
- They reported risk model development or application for use with HMIs
- They reported measurement attributes of HMIs
- They reported use of HMI as an intervention, impact and/or associated implementation issues at a system or hospital level and provide more than detail of trends over time.

Systematic reviews (including meta-analyses), clinical trials, observational studies and case studies (where they involve comparative analysis across multiple hospitals/healthcare services) were included.

Studies identified in the search were considered ineligible for critical appraisal where;

- The study did not present results within the context of a clinical trial
- The study topic was not related to use of HMIs based on administrative data/casemix data definitions such as, based on outcome measures within clinical registries or mixed methods including administrative and clinical data.
- The setting was not whole of hospital/health service level, (such as, department)
- Non English language
- Study settings are limited to one discipline only e.g. paediatric/obstetrics & gynaecology/mental health)
- Studies focussed on interventions designed to reduce mortality but not focused on use of the HMIs primarily (such as, use of bundles to reduce fractured neck of femur mortality)
- Studies were unable to be retrieved within the timeframes of the project.

Where studies did not meet the criteria for critical appraisal but offered interest for discussion issues, potential future relevance to health system performance measurement or expert opinion of issues under debate, these were considered for discretional inclusion in the introduction and/or discussion sections of the report, and have been appropriately referenced. Where a relevant articles (n=1) was identified after the search date, information has been included in the general discussion but the article has not been subject to critical appraisal and full summary.

3.1.6 Article selection and critical appraisal methods

The first pass screen of titles was conducted by one team member and abstracts were reviewed for eligibility by two team members using structured criteria based on the inclusion/exclusion criteria listed above.

Any discrepancies in assessment of eligibility were reconciled by discussion between the team members.

Studies that met inclusion criteria were subject to critical appraisal and were summarised using structured tables that captured elements of the key questions of interest. For studies that focussed on technical issues related to derivation of models for calculating HSMR, Condition specific standardised mortality ratios (SMRs) or LMDRG, a critical appraisal tool was developed based on 3 sources, as there is currently no widely agreed critical appraisal tool available to assess such studies(Krumholz, H.M. et al, 2006, Lang T et al, 2013, Motheral, B. et al, 2003). The key components of the critical appraisal tool included the degree to which authors addressed; definition of outcomes, description of the data sources, setting and patient populations included within the data sources, description of variables of interest, description of statistical methods, reporting of results, including limitations and generalisability to other settings. A shortened version of the appraisal checklist has been included in each study summary (Table 2).

Table 2. Short critical appraisal tool for technical studies

Critical analysis ☐ Good ☐ Adequate ☐ Poor/None	☐☐☐ Clear and explicit definition of the patient and provider sample ☐☐☐ Variables of interest are well defined and summarised ☐☐☐ Mortality outcomes well defined ☐☐☐ Data quality adequately described	☐☐☐ Appropriate analytical approach ☐☐☐ Appropriate model development, validation and performance assessment methods described ☐☐☐ Key results reported well ☐☐☐ Model limitations discussed

For studies that focussed on HMI application, the Scottish Intercollegiate Guidelines Network (SIGN) methodology checklist for cohort studies (SIGN, 2013) for assessing observational studies was used to critically appraise each study (Table 3).

Table 3. Short critical appraisal tool for HMI application studies

Critical analysis ☐ Good ☐ Adequate ☐ Poor/None	☐☐☐ The study addresses an appropriate and clearly focused question ☐☐☐ Clear and explicit definition of the study population and participation rate ☐☐☐ The outcomes are clearly defined ☐☐☐ Data quality adequately described ☐☐☐ Statistical analysis (OR, CI) ☐☐☐ Study limitations discussed

3.2 Grey literature review methods

The grey literature review was designed to complement the peer reviewed literature review in order to provide comprehensive jurisdictional case studies of use of HMI. Three jurisdictions were chosen to form the basis of these case studies (Canada, USA and the UK) as the previous ACSQHC report indicated that the bulk of literature pertained to these jurisdictions. The project team was also aware that other regions, in particular the Netherlands is increasingly active in use of HMIs and where peer reviewed articles fulfilled inclusion criteria, studies from these areas was included in the review.

The grey literature search was tailored to access relevant literature from Canada, USA and UK. Key websites previously identified in 2010 were searched and a Google search using general terms such as ‘hospital mortality indicators, Canada’ were used to identify more recent information or websites.

Information within relevant websites was summarised in a structured way to provide information about; HMIs used within jurisdictions, relevant definitions, population inclusion and exclusion, use of risk models, reporting of HMIs, management of outliers and impacts over time.

The search included websites and organisation listed in (Table 4 below).

Also included in the grey literature search were a number of Australian based reports, commissioned by the ACSQHC and designed to investigate technical aspects of generating the measures and their risk-adjustment models.

3.3 Integration of peer reviewed and grey literature findings

Findings from peer-reviewed and grey literature have been summarised individually (see Appendix 3 for grey literature summary and Appendix 4 for peer reviewed article summary) and key information has been integrated within the main findings of this report to update information about indicator sets and individual indicator specifications across international jurisdictions, and to address key issues relating to development, testing and applying HSMR, LMDRG and condition specific HMIs. In addition, detailed case studies for three jurisdictions; Canada, UK and USA are provided to provide contextual information about how HMI programs are being applied.

Table 4. Grey literature websites

Canada
Canadian Institute for Health Information (CIHI)
Canadian Patient Safety Institute (CPSI)
Statistics Canada
USA
Agency for Health Care Research and Quality (AHCRQ)
Centre for Disease Control and Prevention
Centers for Medicare and Medicaid Services
Commonwealth Fund USA
Institute for Healthcare Improvement (IHI)
International Society for Quality in Health Care (ISQuA)
Joint Commission on Accreditation of Health Care Organisations (JCAHO)
Maryland Hospital Association, Quality Indicator Project
National Committee for Quality Assurance (NCQA)
National Quality Forum
National Quality Measures Clearing House, AHRQ
United Kingdom
Audit Commission Report on Data Quality
Care Quality Commission
CHKS
Dr Foster Intelligence
Health and Social Care Information Centre (HSCIC)
Healthcare Improvement Scotland
ISD Scotland (Information Services Division)
Kings Fund
Mid Staffordshire Report
National End of Life Care Intelligence network
National Health Service
National Institute for Health and Clinical Excellence (NICE)
National Patient Safety Agency (NPSA)
Public Health England, Public Health Observatories
Quality Improvement Hub Scotland
Scotland Performs - NHS
Steering Group National Review of Mortality
Australia
Australian Commission on Safety and Quality in Health Care (ACSQHC)
Australian Institute of Health and Welfare (AIHW)
Australian jurisdictional websites (e.g. NSW Bureau of Health Information)
Clinical Excellence Commission, NSW
National Health Performance Authority(NHPA)
Queensland Health Patient Safety Unit

4. Findings

4.1 Studies and indicators

The PubMed literature search provided greater sensitivity than the Medline search and therefore formed the basis of the search results. It identified 2611 titles, and after title screening 193 abstracts were reviewed and 79 full articles were retrieved. Of these, 34 articles (5 Australian, 11 USA, 5 UK, 4 Canada, 4 Netherlands, 1 Japan, 1 Brazil, 1 Spain, 1 Norway, and 1 Belgium) fulfilled criteria for inclusion and critical appraisal.

These studies and their key objectives are summarised in Table 5. The studies focus on issues associated with model development, testing and application in external datasets. There is increasing interest in issues in application, which was addressed in 11 studies, but only 4 studies specifically addressed the objective of assessing relationships between SMRs and quality of care and none investigated the impact of implementing a program of standardised HMIs on healthcare outcomes.

The indicators included within the review are summarised in Table 6. Those relating to in-hospital mortality only are highlighted in red; these include three general standardised hospital mortality measures (Australia, Canada, UK), three LMDRG indicators (Australia, USA, UK) and 19 condition specific HMIs (Australia 8, USA 4, UK 3, Canada 2, European 2). Of the condition specific indicators there are AMI (6), stroke (6), pneumonia (3) and fractured NOF (3). A condensed summary of included indicators is provided in Appendix 2. Further details on each indicator and links to web-based source data can be found in Appendix 3.

A number of previously identified indicators which extend the definition of mortality beyond inpatient death have not been subject to comprehensive review but are noted in Table 6 for completeness. They are also described in more detail in terms of their definitions in Table 7.

Note also that the indicators sourced and described in detail are generally national or state jurisdictional sets. In the US indicators are commonly developed by private organisations. These are not included in Tables 6 and 7, but reference is made to a small number of examples in the case studies.

Review of jurisdictional indicator sets and individual indicators has demonstrated that there is ongoing revision occurring and reflected in; retirement of indicators, refinement of indicator definitions and changes in provision of public reports. For instance, within the Australian VLAD indicators, heart failure has been dropped. This reflects the general lack of inclusion of this indicator in other indicator sets, except AHRQ.

Findings related to the standardised HMIs of interest have been summarised below within the three major groups;

1. General HMIs
2. Condition specific HMIs
3. Low mortality DRG

Table 5. Summary of included studies and their objectives

Study Author / Year / Country	**Model development and testing**	**Inter-hospital comparison**	**Model comparison**	**Model external application**	**Application/ Implementation methods**	**Validation with quality of care**	**Impacts, outcomes on quality of care**
General HMIs							
Bhat / 2013 / Australia			+	+			
Bottle / 2011 / UK			+				
Cassel / 2010 / USA			+				
Chong / 2012 / Canada	+				+		
Dalton / 2013 / USA	+	+	+				
Girling* / 2012 / UK	+					+	
Gomes / 2010 / Brazil	+	+					
Groene / 2011 / Spain						+	
Jarman / 2010 / Netherlands				+			
Kernisan / 2009 / USA						+	
Kipnis / 2010 /USA	+	+	+				
Kroch / 2010 / USA					+		
Miyata / 2008 / Japan	+			+			
Mohammed / 2009 / UK	+	+					
Mohammed / 2013 / UK					+		
Morsi / 2012 / USA					+		
Popowich / 2011 / Canada					+		
Pouw / 2013 / Netherlands	+	+	+				
Shahian / 2010 / USA		+	+	+			

Study Author / Year / Country	**Model development and testing**	**Inter-hospital comparison**	**Model comparison**	**Model external application**	**Application/ Implementation methods**	**Validation with quality of care**	**Impacts, outcomes on quality of care**
van den Bosch / 2011 / Netherlands	+	+		+			
van den Bosch / 2012 / Netherlands	+	+	+				
van Walraven / 2010 / Canada	+		+	+			
van Walraven / 2011 / Canada	+		+				
Condition specific HMIs							
Aelvoet / 2010 / Belgium		+	+		+		
Borzecki / 2010 / USA			+				
Bradley / 2012 / USA					+	+	
Carretta / 2013 / USA					+		
Clarke / 2010 / Australia					+		
Coory / 2008 / Australia					+		
Drye / 2012 / USA			+				
Kristoffersen / 2012 / Norway		+	+				
Palmer / 2013 / UK			+			+	
Scott / 2008 / Australia	+		+		+		
Low mortality DRG							
Barker / 2011 / Australia				+	+		
Clarke / 2010 / Australia					+		

* Theroretical Mathematical Model

Table 6. Quality indicators for in-hospital mortality 2013 from national and jurisdictional indicator sets

Indicator sets				Condition specific HMI									
	General HMI (in-hospital only)	General HMI (not limited to in-hospital only)	**Low mortality DRG (in-hospital only)**	**Acute myocardial infarction -in-hospital only**	AMI mortality - not limited to in-hospital	**Heart failure – in-hospital only**	Heart Failure mortality - not limited to in-hospital	**Stroke – in-hospital only**	Stroke mortality - not limited to in-hospital	**Pneumonia – in-hospital only**	Pneumonia mortality – not limited to in-hospital	**Fractured neck of femur – in-hospital only**	Fractured neck of femur mortality – not limited to in-hospital
AUSTRALIAN													
National Core Hospital-based Outcome Indictors (ACSQHC)	✓		✓	✓				✓		✓		✓	
Variable Life Adjusted Display Indicators (VLAD)				✓		X		✓		✓		✓	
Victorian Department of Health, Patient Safety Indicators (AusPSI)			✓										
USA													
In-patient Quality Indicators (AHRQ)			✓	✓		✓		✓		✓		✓	
National Hospital Quality Measures (CMS)					✓		✓				✓		
CANADA													
Health Indicators Canada (CIHI)	✓			✓				✓					
UNITED KINGDOM													
Compendium of Indicators (NHS)					✓				✓				✓
Quality accounts (NHS)		✓											
Quality Accounts-Patient Safety (Dr Foster)	✓		✓	✓				✓				✓	
Clinical indicators for National Health Services (NHS SCOT)		✓			✓				✓				✓
EUROPEAN													
Danish National Indicator Project (DNIP)									✓				✓
Health Care Quality Indicators (OECD)				✓				✓					

LEGEND:

✓ (shaded)	Indicator included in the review (in-hospital mortality only)
✓	Indicator not included in the review (see Table 7)
X	Indictor included in previous review (2009) and since removed from indicator set

Table 7. Summary of excluded HMIs

Jurisdiction / Indicator set	Indicators
USA National Hospital Quality Measure	• Death from any cause within 30-days of admission for AMI • Death from any cause within 30-days of admission for pneumonia
England (NHS)– Compendium of Indicators	• Deaths in hospital and after discharge between 0 and 29 days (inclusive) of an emergency admission to hospital with AMI • Deaths in-hospital and after discharge between 0 and 29 days (inclusive) of an emergency admission to hospital with stroke • Deaths occurring in-hospital and after discharge between 0 and 29 days (inclusive) of an emergency admission to hospital with fractured proximal femur
England (NHS) – Quality Accounts	• Summary Hospital-level Mortality Indicator (SHMI)
Clinical Indicators for National Health Services Scotland	• Percentage of patients surviving for 30-days after emergency admission with principal diagnosis of AMI • Percentage of patients surviving for 30-days after emergency admission with principal diagnosis of stroke • Percentage of patients surviving for 30 and 120 days after admission with principal diagnosis of hip fracture • Hospital Standardised Mortality Ratio (HSMR) (30-days from admission)
Danish National Indicator Project	• Proportion of patients who die within 30-days of admission for acute stroke • Proportion of patients who are alive 30-days after admission with hip fracture

4.2 Key issues addressed

Overall, the peer reviewed studies included in the current updated literature review did not identify new issues in relation to deriving and applying standardised HMIs, however they extend the body of knowledge in a number of areas.

For each HMI group, the following questions will be considered, based on information derived from both peer reviewed and grey literature;

- Which jurisdictions are using the indicators?
- How is mortality defined?
- How are the relevant populations included in HMI groups defined?
- What issues are associated with standardised HMI model development and testing?
- Does residual variation in HSMR identify quality of care issues?
- How are hospitals using HMIs?
- Does use of standardised HMIs impact on health outcomes?

The information pertaining to each HMI group can be read in association with more detailed description of the relevant indicators (Appendix 3), peer reviewed studies (Appendix 4), and case studies about use of the HMIs in Canada, Australia and the USA.

4.3 General HMIs

There were 23 articles reporting individual studies relating to general HMIs (Bhat, S.K. et al, 2013, Bottle, A. et al, 2011, Cassel, J.B. et al, 2010, Chong, C.A.K.Y. et al, 2012, Dalton, J.E. et al, 2013, Girling, A. J. et al, 2012, Gomes, A.S. et al, 2010, Groene, O. et al, 2011, Jarman, B. et al, 2010, Kernisan, L.P. et al, 2009, Kipnis, P. et al, 2010, Kroch, E.A. et al, 2010, Miyata, H. et al, 2008, Mohammed, M.A. et al, 2009, Mohammed, M.A. et al, 2013, Morsi, E. et al, 2012, Popowich, J. et al, 2011, Pouw, M.E. et al, 2013, Shahian, D.M. et al, 2010, van den Bosch, W.F. et al, 2011, van den Bosch, W.F. et al, 2012, van Walraven, C. et al, 2010, van Walraven, C. et al, 2011). Of these, 19 studies focused on technical issues associated with developing and/or testing HSMR and 4 studies described issues in implementing, interpreting data at the local hospital level or impacts associated with implementing, interpreting and responding to the data. The majority of studies were undertaken in the USA, UK and Canada.

Ben-Tovim et al report the results of a literature review and they also undertook an analysis of Australian data and reported results of three HSMR models based on adaptation and modification of the Canadian HSMR technical methods (AIHW: Ben-Tovim D et al, 2009).

The issues identified in the Ben-Tovim review and the results of their analysis are summarised in brief in Table 8. The investigators demonstrated that the Australian data is robust and that HSMRs can be developed using models that are similar in performance to those developed in other International jurisdictions. The decisions made in relation to modification of Canadian methods to derive HSMRs reflect in part different views about appropriate statistical methods, but also the need to consider local contextual health system issues with regard to defining mortality indicators and the population to which the indicators will be applied. Such decisions which are reflected in the more recent studies in which existing models are applied in different jurisdictional settings raises questions as to the degree to which inter-jurisdictional comparisons will be relevant, reliable or accurate.

Table 8. Summary of issues reported by Ben-Tovim et al associated with development and application of HSMIs, in general and within Australia (AIHW: Ben-Tovim D et al, 2009)

Issue	Summary of Literature	Analysis of Australian data
Defining hospital mortality		
In-hospital timing	Limiting measures to in-hospital mortality may be affected by hospital discharge policies	In-hospital mortality as there was no 30-day data available
30-day post discharge	• 30-days has been chosen as a cut-off time as death after this period may not relate to hospital care. • The limitation in using a 30-day death definition lies in timely access to data linkage to death data post discharge.	
Defining the population		
	• A broader sample of in-hospital deaths may provide a more representative population for analysis. • If all deaths are included some way of aggregating deaths is needed because the numbers of diagnoses are too large to include individually within the risk adjustment model. • Studies frequently include diagnostic groups accounting for 80% of deaths. • Management of palliative care (PC) deaths	• 68, 3-character ICD codes accounted for 80% in-hospital deaths (4/5 records) • Excluded - neonates - palliative care - patients discharged against medical advice - private hospitals

Issue	Summary of Literature	Analysis of Australian data
	relies on inclusion of a PC type, which is not always included in administrative sets. Some exclude primary diagnoses of cancer. • Restricting analysis to a small number of conditions may be useful when linking mortality outcomes to specific process measures.	
Risk adjustment variables		
	• Age, sex, principal and secondary diagnoses • Ethnicity where available • Other data may include emergency admission status, referral source, address • Illness severity - this is not available in administrative datasets. • It is not easily or uniformly assessed across conditions. Hospital level variables • Size of hospital, type of hospital. • Risk adjustment best undertaken at level of the patient. • Multi-level modelling techniques can be used to consider the influence of variables at multiple levels (e.g. at patient level and hospital level).	• Age, sex • Principal/secondary diagnoses • High risk 80% mortality • Low risk 20% mortality • All deaths • Demographic information • Admission (emergency/elective) • LOS (6 categories) • Added social deprivation, SEIFA • Peer group (based on size, ATSI status, teaching and research status, proportions of acute, rehabilitation, palliative care, non-acute patients treated)
Statistical modelling methods	• Hospital mortality is the dependent variable. • Logistic regression (LR) allows development of a linear equation for the log (odds) of a positive outcome. • The log (odds) increase by the magnitude of the coefficient for each unit increase in the dependent variable. • The exponentiated coefficients are interpreted as the change in the odds of a positive outcome for a unit increase in the associated independent variable. • The coefficients from LR are applied to create a predicted probability of death for each individual in the dataset which can then be used to create a set of reference weights for a hospital. • Each individual hospital is then standardised to a reference hospital population • The standardised mortality is a mortality rate adjusted for casemix. • Indirect standardisation occurs when the observed mortality rate is compared to the expected mortality rate based on all	• Most analyses followed the RACM model but in addition incorporated a 2-stage method in which LR was followed by multi-level modelling. • A second model (Elaborated risk-adjusted mortality model or ERM) was also used that included a different approach to variable preparation and inter-action analysis in the R model. • There was variation in HSMR between hospitals within and between peer groupings demonstrated using caterpillar and funnel plots. • The HSMRs for the lower risk group were most variable. • A longitudinal analysis was performed over 3 years data (2004—2007). • Using multi-level modelling there was a tendency towards a decrease in HSMR over time which was significant only for peer group A1 - within hospital variation (based on (ICC) between the 3 years was small compared to between hospital variation. • R^2 values = 0.24 for the 80% group

Issue	Summary of Literature	Analysis of Australian data
	patients within all hospital populations.	which is modest and in keeping with previous literature; R^2 =0.34 for 20% group, R^2= 0.35 for 100% group. • Good discrimination. • c-statistic for the modified RACM model; 0.87 for the 80% high risk group, 0.96 for the 20% model, 0.95 for the 100% model. • Hosmer-Lemeshow did not demonstrate good fit for any models.

4.3.1 Defining hospital mortality

There are concerns about limiting studies of variation in hospital mortality rates to in-hospital mortality as the impacts of hospital care may become apparent after discharge and that some hospitals might change discharge policies in order to discharge patients with a poor prognosis to community based palliative care or other services. Similar results for in-hospital and 30-day mortality rates have been reported, however Ben-Tovim et al suggest there remained a potential for discharge policies to influence mortality rates. They also indicated that the decision about choosing the definition of mortality in relation to timing was likely to be determined by pragmatic issues such as access to linked death data. In their own analysis of Australian hospital data they focused on in-hospital HSMR but noted the current and future capability for data linkage in various Australian states.

Bhat SK, 2013, Australia

A more recent study has also addressed the issue of mortality definition for HSMR (Bhat, S.K. et al, 2013). This Australian study compared the UK (Jarman's method) derived HSMR (in-hospital death only) to a local linkage derived cumulative mortality ratio (CMR – death within 30-days of admission) using the Linked Western Australian morbidity and mortality register. They reported greater capability of the CMR method to accurately identify deaths, leading to fewer unmatched deaths. The HSMR derived using the inpatient Jarman method was significantly higher for metropolitan teaching hospitals (1.02, 95% CI 0.98-1.07) than the CMR method (0.81, 95%CI 0.77-0.85).

The study was limited by use of extremely old data (1980-1995) and changes in coding and practices may have occurred since this time.

Bottle A, 2011, UK

Bottle et al reported a good correlation between regular HSMR and those based on deaths within 30-days of admission but noted that there were some marked differences with hospitals with more post discharge deaths being affected by this modification (Bottle, A. et al, 2011).

Australian reports

- If linkage to population death registry data and cause of death are desired, there is a current 2 year time delay to consider (Lujic, S. et al, 2012).
- Average in-hospital HSMRs are higher than for models based on 30-day mortality definitions (Lujic, S. et al, 2012).
- Correlations between HSMR and 30-day mortality models was high (c statistic 0.88-0.89), however agreement on outlier status were moderate to good and higher for private hospitals

(kappa=0.64) than public hospitals (kappa=0.5) when HSMR was compared to the 30-day post-admission model and lower when examined by peer group (Lujic, S. et al, 2012).

- Categorization by performance groups (below average, average and above average) remained the same for 72% using the 30-day post-admission model and 71% for the 30-day post-discharge (SHMI) model (Lujic, S. et al, 2012).
- Whilst 30-day mortality may be a preferable measure, it is not always feasible to do this and the gain in precision needs to be considered in relation to the greater complexity required in the process (AIHW: Ben-Tovim D et al, 2009).

Case studies

- In England a new general mortality indicator has recently been introduced, the Summary Hospital-level Mortality Indictor (SHMI), that includes in-hospital death as well as death occurring within 30-days post discharge. Introduction of the indicator has been driven by concerns about transparency and variability in the results obtained from different mortality indicators. The indicator is reported at a national level through the NHS. At this stage the in-hospital HSMR continues to be publicly reported by Dr Foster Intelligence, which now also reports the SHMI in its annual report, Dr Foster Hospital Guide. The report shows some variability in the indicators for a number of trusts.
- The Scottish NHS general mortality indicator includes deaths occurring within 30-days of admission, there being a preference to associate the patient outcome with decisions made at the point of admission. The appropriateness of indexing the record at point of admission has been identified as an issue for consideration by the group advising on indicator development, there being particular concern around selecting the main diagnosis from records sourced from short-stay admission units.

4.3.2 Issues in defining patient populations for deriving the HSMR

The issue of defining the population is important. Firstly, in deriving standardised mortality measures it is necessary to define the risk of certain types of 'people episodes', a process undertaken by designating a certain number of diagnostic groups. There is no 'ideal' number of groups, however as Ben-Tovim et al noted, there is a need to aggregate diagnoses at some level as "inclusion of all possible individual diagnoses is not feasible for the process of risk adjustment".

There is also a need to consider how inclusion or exclusion of certain people/episodes may influence HSMR measurement. Ideally the included population would be highly representative of the hospital being measured. However, it is also important to ensure the population does not include a significant number of death episodes where those deaths are clearly not related to quality of care issues (for example expected deaths on admission), as variation in HSMR between hospitals may be interpreted incorrectly as reflecting quality of care issues leading to inappropriate blame and shame and waste of valuable quality improvement investigation resources.

The issues that have been addressed in the literature and need to be considered in relation to defining patient population are therefore;

- Types of population; for instance; neonates, persons under 18 years, persons with cancer, persons who leave against medical advice
- Types of episodes; for instance, multiple episodes (readmissions), episodes that include inter-hospital transfer, episodes designated as palliative
- Numbers of episodes, for instance where there are few death "events"

Ben-Tovim et al noted that there has been a general lack of discussion within the published literature about the rationale for choosing whether to narrowly or broadly define the population of interest for various HSMR models. In their own analysis of Australian data they group patient data into three groups; (i) those for whom the diagnoses contribute to 80% deaths which reflects common practice for HSMR development in other settings, (ii) those with diagnoses contributing to the remaining 20% and (iii) a group including all diagnoses, 100%. The HSMR model performance (discrimination based on the c-statistic) for all groups was good. Similarly, in a recent study Bottle et al reported that including all admissions, rather than just those accounting for 80% of deaths had a modest impact on HSMR, however four hospitals were flagged as ‘high’ that previously were flagged as ‘average’ (Bottle, A. et al, 2011).

Inclusion and exclusion criteria for patient records differ between HSMR models and countries. Whilst Ben-Tovim et al based their Australian HSMR model on the Canadian RACM model, Australian inclusion/exclusion population criteria were developed based on accessible data and contextual reasons. They excluded patients discharged against medical advice, palliative care patients and neonates, the latter being excluded as the authors noted difficulties in Australia with coding of qualified and unqualified newborn babies. The authors did not discuss the issues associated with readmission episodes and hospital transfers.

The various HSMR methodologies are summarised in Table 9. For completeness the table includes HSMRs for inpatient mortality as well as the more recently developed UK SHMI which also includes 30-day post discharge mortality. It also includes the Scottish HSMR (30-day mortality from admission). This table will be referred to in this and subsequent sections relating to methodological issues.

Table 9. Summary of HSMR methodology – including in-hospital and other models

	ACQSHC (AIHW)	Canadian Health Indicators	Dr Foster Quality Accounts	SHMI England, Wales	SHMR Scotland
Definition	In-hospital mortality	In-hospital mortality	In-hospital mortality	In-hospital and 30-day post discharge mortality	30-daye post admission mortality
Diagnoses	Top 80% (high risk) Bottom 20% (low risk)	Top 80%	Top 80% (56 diagnostic groups)	100%	
Other inclusion	• Care type6:Acute care, geriatric evaluation and management and maintenance care • 29 days ≤ Age ≤ 120 • 1 ≤ LOS ≤ 365 days • Elective or emergency	• Discharge between April 1 of a given year and March 31 of the following year • Acute care admission • 29 days ≤ Age ≤ 120 • Sex recorded as male or female • LOS ≤ 365 consecutive days • Admission category is elective or emergent/urgent • Canadian resident			
Exclusions	• Discharged at own risk • Palliative care • Neonates ≤ 28 days on admission • Missing admission mode, sex	• Cadavers • Stillborns • Sign-outs (that is, discharged against medical advice) • Patients who did not return from a pass • Neonates (age of admission less than or equal to 28 days)	• Day only admissions	• To be agreed on	

	ACQSHC (AIHW)	Canadian Health Indicators	Dr Foster Quality Accounts	SHMI England, Wales	SHMR Scotland
		• Records with brain death as most responsible diagnosis code • Records with palliative care as most responsible diagnosis code			
Variables					
Age	✓ continuous	✓ continuous	✓ categorical – 5 years	✓ Categorical 5 years	✓ continuous
Sex	✓	✓	✓	✓	✓
LOS	✓ categorical (1, 2, 3-9, 10-15, 16-21, 22-365)	✓ categorical (1, 2, 3-9, 10-15, 16-21, 22-365)			
Diagnosis group	✓	✓	✓	✓	✓
Comorbidity	✓ Charlson score	✓ Charlson score	✓ Charlson score	✓ Charlson score	✓
Interaction between age on admission and Charlson score			✓ New 2011		
Admission mode (e.g.elective / emergency)	✓	✓	✓	✓	
Inward transfer	✓	✓			
Deprivation score			✓	✓	✓
Previous emergency admissions			✓	✓	✓
Palliative care			✓		

	ACQSHC (AIHW)	Canadian Health Indicators	Dr Foster Quality Accounts	SHMI England, Wales	SHMR Scotland
Ethnicity			Deleted 2011		
Year of discharge			✓	✓	
Month of admission					
Source of referral			✓		✓
Inpatient / day case					✓
Specialty					✓
Reference year for coefficients (expected deaths)		2009 /10	2010/11	Quarterly	Annually

van den Bosch, WF, 2011, Netherlands & van den Bosch, WF, 2012, Netherlands

Two studies from the Netherlands report results of investigations into the impact of readmissions on HSMR (van den Bosch, W.F. et al, 2011, van den Bosch, W.F. et al, 2012). The first study included data from 6 large Dutch non-university teaching hospitals with a spread of HSMRs from low (HSMR 65) to high (HSMR 114). HSMRs were derived using the previously developed Dutch HSMR model (DHM-2008), which in turn was based on the UK, Dr Foster HSMR model with some differences; definition of the population included (included day cases), use of fewer comorbidity groups, and differences in risk adjustment (no adjustment for palliative care, source of admission and previous number of emergency admissions). The data was grouped in two ways for admissions; according to all first admissions (Admission view) or according to admission frequency (patient view). There were a significant number of readmissions with 31% admitted more than once and readmissions accounting for 60% of total admissions.

The authors reported that DHM-2008 predicted a reduction in mortality per admission, indicating that patients admitted more frequently experience a lower risk of death per admission and they recommend an additional adjustment variable be included for admission frequency whilst acknowledging the practical issues associated with implementing such a change. In their second study (van den Bosch, W.F. et al, 2012), the investigators use data from 89 Dutch hospitals to compare the UK, Dr Foster HSMR model with a second model that includes adjustment for frequency of admission. They reported that the second model resulted in different HSMR results and demonstrated better model performance attributes. Using model 2, 64 hospitals (19.5%) were not higher than expected based on model 1. The authors noted that the Dr Foster model uses a 1 year review period and the 3 year period they had to look back for readmissions was advantageous.

The issues of recurrent admissions has been considered by investigators associated with the UK, Dr Foster HSMR model (Bottle, A. et al, 2011). They observed that multiple admissions per patients did increase the number of funnel plot outliers and suggested that a patient-based measure might be preferable. An unresolved issue is which admission to include.

Bhat SK, 2013, Australia

In addition to investigating the definition of mortality, the previously mentioned Australian study by Bhat SK, et al (2013) also compared the UK, Dr Foster HSMR model to their CMR model and reported that the linked method was more accurate in identification of transfers.

Bottle A, 2011, UK

In this study, the HSMR derived from nine variant models are compared to the reference model (Dr Foster method) using 4 years of admissions (2005/6 to 2008/9) (Bottle, A. et al, 2011). The nine variant models used are based on criticism of the casemix adjustment used in the Dr Foster derived HSMRs. Therefore, two models tested using the patient's first or last admission (to test the criticism that multiple admissions per patient inflates the denominator), one did not include adjustment for palliative care (to test the criticism that many unpreventable deaths are included), one did not include adjustment for comorbidity (to test the criticism that depending upon secondary coding of diagnoses introduces bias due to variable coding practice), one excluded unplanned and same-day admissions (to test the criticism that these inflate the denominator), one defined mortality as 30-days from admission (to test the criticism that variable discharge policies may introduce bias), one included all deaths (to test the criticism that too much hospital activity is excluded) and 1 included 5 specific diagnoses including; AMI, stroke, fractured neck of femur (to test comorbidity bias). For one model there was no adjustment for comorbidity or palliative care and unplanned admissions with zero LOS were excluded.

There were a number of findings reported. Firstly, the regular Dr Foster model had good discriminative performance with a c-statistic of 0.87 and there was good correlation between HSMR and 30-day mortality (0.84). Compared to regular HSMR a small number of hospitals moved from ‘average’ to ‘high’.

Amongst comorbidity groups the c-statistic varied from 0.66 (senility and organic mental disorders) to 0.95 (breast cancer). The most important variable explaining variation in patient-level mortality was age (35 of 56 models) and method of admission (14 models). In four models CCI was the most important variable and in only one it was palliative care. The second most important variable was CCI (27 models), age (10 models) or palliative care (8 models).

Using the patient’s last admission in the four years resulted in a larger number of deaths compared with using the first admission. Excluding zero day unplanned stays ending in live discharges had high correlation with the regular HSMR (0.99) and had the least impact on HSMR, followed by using all admissions. Using the last admission significantly impacted on one hospital and similarly two Trusts were especially affected by adjustment for palliative care as they had higher than average recording of palliative care. It was more common for hospitals to change from ‘high’ to ‘average’ when using first admission or just the five diagnosis groups. Using the last admission led to more deaths being included and reduced the width of the control limits.

The authors note the variable impact of the different models on HSMR measurement but that many criticisms of the regular model resulted in little impact on HSMR measurement. They also noted the marked changes in HSMR associated with palliative care coding, the potential unreliability and bias that could be introduced by this variable and the possibility of gaming that would need careful surveillance. Finally, the impact of multiple admissions was discussed and the authors suggested a patient-based measure would be preferable with more investigation required to decide which admission should be used.

Cassel JB, 2010, USA

This American study explored how hospital mortality rates were computed by four entities and how hospice and palliative care are recognised and handled (Cassel, J.B. et al, 2010). The four entities, Centres for Medicare & Medicaid Services “Hospitals Compare”, USA, News & World Report “Best Hospitals”, Thomson-Reuters “100 Top Hospitals” and Health Grades all use Medicare data to derive risk-adjusted all-cause mortality. However, the methodology used to calculate mortality rates varies considerably including handling of cases that involve hospice care or palliative care. One entity excludes cases with prior hospice care and another excludes those discharged to hospice at the end of the episode of care. Two entities exclude some or all cases that were coded with V66.7 “Palliative Care Encounter” ICD-9-CM diagnosis code.

The authors identified difficulties defining palliative care and hospice cases as the rationale for the different methodologies including: inconsistent use of V66.7 Palliative Care Encounter Code; obstacles obtaining hospice enrolment data; caution re excluding hospital deaths too liberally; excluding cases that involve hospice only at the end of an admission may create an incentive for hospitals to use hospice as a way to hide problems with quality of care earlier in the admission.

Chong C, 2012, Canada

This study by Chong et al explored whether palliative care coding had changed in Canada since 2007 when HSMRs were published publicly (Chong, C.A.K.Y. et al, 2012). Using the Canadian Institute for Health Information (CIHI) database HSMRs (excluding cases coded a palliative care) were recalculated. A hospital standardised palliative ratio (HSPR) was constructed using a binary logistic regression model. The HSPR is interpreted as the observed number of cases with palliation as the most responsible diagnosis/expected such number x 100. A HSMR which included palliative cases (HSMR+palliative) was also calculated using three approaches.

The HSMR rates were significantly lower in 2008-2010 than in 2004-2006 by 8.55 points ($p<0.001$). The HSPR has been increasingly and jumped 48.83 points n 2008-2010 compared with 2006-2008 ($p<0.001$). The HSMR strongly correlates negatively with the HSPR (Pearson $r=-0.862$, $p<0.001$). Estimating the effect of including palliative care cases in the HSMR (HSMR+palliative) under the three scenarios, the HSMR would have decreased by at most 6.35 points ($p<0.001$) and may have potentially even increased by 2.30 points ($p=0.028$), in contrast to an observed HSMR decrease of 8.55.

The authors conclude that palliative care coding rates have increased dramatically since the public release of HSMR results and may have contributed to the observed national decline in HSMR. However, the timing of the public release of HSMRs coincided with the release in 2006 of new guidelines for palliative care coding, which is likely to have improved coding practices. Changes in coding practices have a clear impact on HSMR and comparisons made overtime need to take into account any such changes.

Australian reports

- Utilisation of palliative care status varies across states, however the relative position of states did not differ significantly after removal of palliative care types from the analysis indicating that trends in mortality rates are not a consequence of deaths in coded palliative care services (AIHW, 2011a).

Case studies

- The issue of the inclusion or exclusion of palliative care patients has been ongoing in the UK, with some hospital trusts raising concerns that they were unfairly penalised under the new SHMI methodology of not excluding palliative care. Findings of an investigation conducted in 2013 found coding to be a major issue, with organisations interpreting palliative care in different ways and some organisations having wide variation in coding over the course of three publications of the data. It was noted that there is difficulty in establishing a consistent definition of palliative care, and that there is a risk of gaming. The report also found that applying a model that adjusted for palliative care did not have a significant effect on the majority of trusts, and that the contextual indicators reported with the SHMI supported understanding of the casemix.

4.3.3 HSMR risk adjustment and statistical modelling methods

Considerable debate continues in relation to variables that should be included within risk adjustment models for HSMR, which method of standardisation should be used (indirect or direct) and what statistical modelling methods (fixed or random effects, simple or hierarchical models) should be used. There is increasing interest in examining interactions between variables. With regard to HSMR model testing, there is general consensus that assessment of model performance should consider a number of attributes, including calibration, discriminatory power and goodness of fit.

Risk adjustment variables

According to current HSMR specifications (Appendix 2 – indicators condensed summary) and summarised in Table 8, the Australian, Canadian and UK, HSMR models include age and sex in the risk adjustment process. Comorbidity has been reported to be the next most important variable in risk models after age and method of admission (Bottle, A. et al, 2011). Australia, Canada and the UK include comorbidity based on the Charlson Comorbidity Index (CCI) however it is not entirely clear if the CCI is used in the same way for all models.

Further, the Canadian model uses only preadmission diagnoses and the diagnosis most responsible for admission in the CCI calculations, but it is not clear if the same approach has been adopted in the UK or for Australia. The inclusion of a diagnostic timing flag in Australian administrative datasets, the 'C-Code', would enable such an approach to be taken.

Dalton JE, 2013, USA

In a recent well designed study which supports use of this approach, Dalton et al (USA) investigated the comparative performance of in-hospital mortality risk indices when only diagnoses present on admission (POA) and procedures (primary/principal procedures and secondary procedures where the latter were dated before the primary procedure) were included in the model compared to a model using all diagnostic codes (Dalton, J.E. et al, 2013). Both models (POARisk and AllCodesRisk) were also tested against a modified existing risk stratification index (RSI), which had been derived from a high risk (Medicare) population.

A number of lessons can be elicited. Firstly the initial models derived from 80% of California discharges (2004-2008) that included a procedure did not calibrate well when applied to the remaining randomly selected 20% but performed better after correction and application to another year (2009) of data. This reinforces the need for multi-step development of risk models.

Secondly, the AllCodeRisk model predicted mortality outcomes better, which can be understood as the outcome is in-hospital mortality and in-hospital diagnoses (some of which will be complications) are potentially contributing to that death. It is a reminder that the key issue is to develop risk models that will accurately account for predictive factors for mortality that are not quality sensitive with regard to hospital management.

Thirdly and importantly, when performance was compared in relation to rank based outcomes, over a third of hospitals had a different category classification when the different models were applied. This is likely to adversely influence the ability to compare hospitals across jurisdiction when difference risk models are applied. The authors acknowledge a number of limitations with using the POARisk model, particularly the quality of the POA indicator coding which may be less reliable in regions new to using the codes and may be subject to regional coding practice differences.

The Australian, Canadian and UK HSMR models include admission urgency but there are also some differences in choice of variables. The Australian and Canadian specifications indicate LOS and transfer status are included. These are not clearly documented for the UK model but it does include a social deprivation factor, number of previous emergency admissions, year of discharge, palliative care status, month of admission and source of admission, the latter possibly including transfer status. Since 2011, the UK, Dr Foster HSMR model has included examination of interactions between age on admission and the CCI.

Differences in the choice of variables and the way they are measured may inhibit HSMR comparisons between countries. Ben-Tovim et al reported that social deprivation measured using SEIFA did not influence the Australian HSMR model, which may explain its exclusion. Further the UK. Dr Foster model uses a different social deprivation Index (The Carstairs Index) and as other countries use different indices, it may be preferable not to include a social deprivation score for jurisdictional comparison. On the other hand, if social deprivation is demonstrated to influence HSMR calculations, it may be an important explanatory variable, that could influence policy and planning decisions and should then be considered for inclusion when HSMR is being used for internal jurisdictional mortality performance review.

Another issue raised in relation to choice of risk adjustment variables in HSMR models relates to the stability of inherent risk in chosen variables which when unstable can lead to the "constant risk fallacy".

Mohammed MA, 2009, UK

The issue of constant risk fallacy is addressed by Mohammed et al who explores the validity of casemix adjustment by examining the consistency of relationships between risk factors and mortality across 4 UK hospitals with a wide range of publicly reported HSMRs (Mohammed, M.A. et al, 2009). This paper focuses on exploration of Nicholl's description of 'Constant Risk Fallacy', that, "if risk relations are assumed to be constant, but in fact are not, then casemix adjustment may be more misleading than crude comparisons" (Mohammed, M.A. et al, 2009).

The authors point out that two key mechanisms lead to non-constant risk relations; differential measurement error and inconsistent proxy measures of risk. They then focus on examining patterns between, and over time within, 4 purposively selected UK hospitals in use of coded variables including, casemix (Charlson Comorbidity Index, CCI) and 'emergency/non-emergency' status. They demonstrate that there is stability in interactions between age, deprivation and hospitals indicating no constant risk fallacy. However there are interactions for other variables including casemix and emergency status. They demonstrate that the depth in coding for CCI does vary between hospitals and does change over time and that this change influences the interaction between the variable casemix and hospital (the interaction becomes smaller as depth increases). They also demonstrate inconsistencies in relationships between hospital casemix levels and crude mortality rates, emergency status and LOS that suggests systematic differences in coding and in-hospital policies about admission. One example provided relates to changes in admission policies associated with introduction of the ED 4 hour targets, such that there are more potentially more admissions of less severely ill patients – thus the inherent risk of being an emergency patient is reduced.

The authors suggest that it would be wise to avoid inclusion of variables that are not proven to be reliable and are likely to be influenced by hospital policy impacts. Further, interpretation of variance in SMRs needs to consider the maturity, consistency and constancy of casemix coding, particularly when these surveillance systems are first introduced.

Australian reports

- A number of analyses have been undertaken to assess the impacts of including and excluding populations and risk variables, including age, palliative care status, transfers in and out, urgency of admission, and reference year for risk-adjustment coefficients.
- There are jurisdictional level differences in the magnitude of the effect of risk adjustment on mortality rates, the absolute effect being modest (AIHW, 2011a).
- Age can be used as a continuous variable (AIHW, 2011c).
- Excluding transfers in (6%) and out (15%) impacts on the denominator population for HSMR.
- Using previous year coefficients for risk adjustment generally reduces in-hospital mortality rates and decreases the number of hospital flagged as outliers. The relative position of hospitals was highly correlated when using different coefficients.

Case studies

- In England, the model used to predict the expected number of deaths for the new SHMI calculation is built on fewer risk adjustment variables than the Dr Foster SHMR and fewer than originally proposed by the Technical Review Group that advised about indicator development. Using more risk adjustment variables may improve the predictive power of the models, but can also introduce more data quality issues. It was found that using only age, Charlson comorbidity index, admission method and gender provided a simple and stable model as recommended by the School of Health and Related Research in their final evaluation report.

Statistical modelling methods

Using statistical methods to adjusting data to account for differences between hospitals, or within hospitals over time, that are not due to quality sensitive issues is integral to using HSMRs. Not surprisingly, there is no single ideal method, and in general the process requires a number of steps, all of which have been subject to investigation and debate.

Methods of standardisation of data include direct and indirect methods, which are described in the table below (Table 10.).

Table 10. Attributes of direct and indirect standardisation methods (Adapted from Pouw M.E., et al 2013)

Direct Standardisation	Indirect Standardisation
• Direct standardisation involves standardisation of casemix • The directly standardised mortality rate of a hospital is based on the same casemix as that of other hospitals; that is, to a reference population. • The mortality rate for each group of patients is calculated • The mortality probabilities for each group for each hospital are applied to the same reference population to obtain the reference population number of expected deaths, then these are summed across all groups to reach total expected deaths. • The directly standardised mortality ratio is the ratio between the number of expected and actual deaths	• Indirect standardisation involves standardisation of the mortality rate • The mortality of the casemix is standardised to a reference mortality rate. • Expected probabilities of death are calculated for hospital population categories using logistic regression, with in-hospital mortality as the outcome (dependent variable) and using data from all available/included hospitals • The expected probabilities are then applied to a specific hospital to calculate the expected deaths for that hospital and these are summed across categories to provide the total expected death rate • The ratio between observed deaths in each hospital and expected deaths provides the indirectly standardised mortality ratio

The direct standardisation method means rates are comparable as casemix differences are removed, however it is not possible to directly standardise where there are small numbers in subcategories. Indirect standardisation can be used in this situation, however it may be limited by lack of comparability between ratios if there is interaction between variables, for instance casemix and admission urgency.

Pouw ME, 2013, Netherlands

In a somewhat complex scenario-based Dutch study, Pouw and colleagues test the impacts of different comorbidity risk variables ‘urgency status’ and Charlson Comorbidity Index (CCI) on changing HSMR categories across 61 Dutch hospitals (Pouw, M.E. et al, 2013).

They reported that significant casemix changes, based on different hospital distributions (in some instances within a hospital over time), were associated with changes in HSMR categories. However changes did not occur with smaller casemix differences. There were interactions identified between hospitals and two casemix variables – ‘urgency’ of admission (19 of 50 prediction models) and CCI (7 of 50 prediction models).

The authors suggest that there needs to be caution in interpreting differences in HSMR between hospitals and within hospitals over time where casemix may be subject to change but also acknowledge the limitations of direct standardisation as being “practically impossible when multiple predictors are included in the adjustment model”

Of interest, the authors draw on previous commentary in suggesting “indirectly standardised HSMR provides the mortality rate from a societal perspective as it is based on the population the hospital actually serves, not the national reference population, while a HSMR based on direct standardisation is more relevant to informing patient choice” (Heijink, R. et al, 2008, Pouw, M.E. et al, 2013). The issue of standardisation methods remains of interest to investigators and is influencing changing HSMR technical specifications as reflected by inclusion of interaction testing in the Dr Foster model since 2011.

A further area of interest in the literature relates to estimation methods used in making comparisons between hospitals in relation to application of fixed or random effects statistical models to account for the hierarchical (or multi-level) nature of data; data clustered at the patient level and data general at the level of hospital (Table 11).

Table 11. Attributes of random and fixed effects estimation methods (Adapted from Kipnis P. et al, 2010)

Fixed effects models	Random effects models
• Assumes interest in the specific hospital	• Considers hospitals as a random sample of centres within a larger population
• Assumes each hospitalisation is independent from others	• Focuses on the whole population rather than the observed hospitals
• Ignores the hierarchical structure of the data	• May better reflect hierarchical nature of data
• Considers only one form of variation – random variation of hospitalisation outcomes within a hospital	• Uses the overall error (mean squared error) which is a function of both bias and the variance of the estimator, as the criterion for defining a best estimator for a given parameter.
• May under-estimate overall variance	• The estimate of the true observed mortality rate at a given hospital is a weighted average of the overall mortality rate across all hospitals and the hospital-specific mortality rate. • More weight is placed on the overall mortality rate of smaller samples, thereby “shrinking’ the observed mortality rate towards the overall mean.

Kipnis P, 2010, USA

Kipnis et al, examined the influence of statistical estimation methods for deriving hospital SMRs (Kipnis, P. et al, 2010). Using record data from 17 Kaiser Permanente hospitals in Northern California and simulations, they examined the relative attributes of fixed effects models (n=3) compared to random effects models (n=3). They reported that all models correlate highly with hospital log (SMRs) and hospital ranks, although random effects models exhibit higher specificity for identifying mortality rates than fixed effects models. In contrast, fixed effects models exhibit higher sensitivity for small changes in SMR. The model with best sensitivity and specificity attributes was the aggregate-level fixed effects method.

The authors suggests that both methods offer different opportunities, for instance use of random effects models to identify outlier status and aggregate-level fixed effects model to examine mortality rate changes over time. The authors point out a number of limitations of their study; a limited number of USA hospitals, use of physiological data in their risk variables and the fact that model (random effects model) attributes might be altered by reduced numbers of admissions per hospital.

Australian reports

- Australian data analyses undertaken since the Ben Tovim report (AIHW: Ben-Tovim D et al, 2009) have generally used the indirect standardisation method. The ACSQHC specifications for HSMR do not include detailed discussion about statistical methods of standardisation or use of fixed or random effects models.

Case studies

- The case studies did not identify a focus on issues associated with statistical modelling, other than in relation to inclusion/exclusion of risk adjustment variables and use of regression analysis.
- There is variable information provided within different jurisdictions, the most detailed descriptions being provided by the English SHMI technical specifications report (HSIC Technical Specification SHMI April 2013).

4.3.4 Model comparisons and testing

As demonstrated in the prior sections, there is a good deal of interest in testing different HSMR models in order to assess the relative influence of different aspects of HSMR, ranging from mortality definitions to choice of included patient populations, risk adjustment variables and statistical standardisation and modelling methods. In this section are summarised a number of studies that have developed and tested the performance of a model or multiple models, including assessment of performance in external populations from which the original model was developed. In some instances, the investigators develop multiple models as they continue to explore issues in risk adjustment and statistical modelling as described above.

As Ben-Tovim et al discussed, the technical performance of an HSMR model can be assessed in a number of ways, as summarised in Table 12 below, the key questions being; how well does the model predict mortality?, how well does the model account for variation in patient-level risks?, To what extent can the model be fitted to a different, albeit similar, patient population?

Table 12. Statistical tests used to assess goodness of fit of HSMR models. (Adapted from AIHW: Ben-Tovim, 2009)

Performance test	Description
• C-statistic	• The c-statistic, or area under the receiver-operator curve, is used to assess discrimination • It calculates the probability of assigning a greater risk of death to a randomly selected patient who died compared with a randomly selected patient who survived. • There is no single way of assigning level of performance, however in general; >=0.5 and < 0.6 – no discrimination >=0.6 and <0.7 – poor discrimination >=0.7- and <0.8 – good discrimination >=0.8 and <0.9 – excellent >=0.9 and <1 - outstanding • It has been shown that risk prediction with discrimination comparable with that obtained from clinical databases is possible using routinely collected administrative datasets (Aylin, P. et al, 2007).
• Hosmer-Lemeshow	• An important attribute is the predictive ability of the model which is usually assessed by considering a range of different risk groups, for instance deciles of risk and applying the Hosmer-Lemeshow method (Hosmer, DW. et al, 2013) • This test may not be suitable for large sample sizes, as was demonstrated by Ben-Tovim's analysis of Australian data; and should be used cautiously if a model is developed in a small sample and then applied to a large sample as there can be overdispersion of data.
• R^2 statistic	• This statistic is used to investigate the strength of the relationship between the model and the data compared to the play of chance. • There are limitations in using this method as it was designed for linear regression correlations rather than logistic regression. • Ben-Tovim et al (2009) suggest it is better used to assess the impact of adding or subtracting variables in a logistic regression model rather than providing a single statistic against which to measure model fit.
• Calibration	• This can be undertaken using a variety of methods to correct for differences between populations in which the HSMR model may have been developed and others in which it is applied. • The study design can contribute to reducing 'overfitting' of data to one population by using derivation and validation (confirmatory) sets when developing models. • Ben-Tovim et al (2009) note the lack of empirical studies in this area.

Shahian, 2010, USA

Shahian and colleagues report on the similarities and differences between four commonly used general standardised hospital mortality risk adjustment algorithms using the same administrative dataset (Shahian, D.M. et al, 2010). The four methods varied in their inclusion and exclusion criteria (patient, diagnoses, hospital type), and this resulted in the proportions of discharges included by each method varying from 28-95%. The 3M method was more inclusive, including data from all 83

(100%) hospitals and 95% of discharges in their analysis. The UHC-Premier algorithm included the smallest number of hospitals and discharges, namely 81 (98%) hospitals and 28% of discharges.

81% of hospitals were included in all four methods, but only *22%* of all discharges were included across all methods. These 22% of discharges were older, had higher in-hospital mortality, longer LOS, and more respiratory and circulatory system diagnoses than patients overall.

Hospital-performance categorisation varied substantially across models; 43% of hospitals classified as a higher than expected mortality for 1 method had a lower than expected mortality when classified by one or more of the other methods. All four methods identified three hospitals in 2006 and one hospital in 2007 as having higher than expected mortality, and one hospital in 2007 as having lower than expected mortality.

Overall, this study demonstrated that there was large variation in HSMR estimates according to which risk adjustment algorithm was used – three of which are used widely. The selection criteria, statistical methods used, as well as inadequacies of the data source such as the inconsistent reporting of palliative care and DNR patients, and the way in which each method deals with these factors, are discussed as potentially being related to the differences in mortality results.

The authors conclude that disagreement across the models used suggests that all methods are not reflecting the same underlying construct, and use of hospital-wide mortality rates to evaluate quality of care must proceed cautiously.

Miyata H, 2008, Japan

This study reports use of previously reported HSMR model development methods to develop a Japanese HSMR prediction model (Miyata, H. et al, 2008). The model was tested with and without the LOS variable and with varying depths of comorbidity diagnoses. LOS did not influence model performance greatly and models exhibited high levels of accuracy, with C statistic varying between 0.869 (95% CI 0.860-0.879) for Model 1 with LOS and 0.841 (95%CI 0.830-0.852) for Model 2 without LOS. Models with more comorbidities included performed better than those with lower numbers.

The study did not investigate differences in derived HSMRs between hospitals in the dataset and did not compare the developed models against other existing models.

Gomes AS, 2010, Brazil

This study which described development of a model to derive an HSMR model in a large jurisdiction in Brazil does not offer a great deal more than other studies in this area (Gomes, A.S. et al, 2010). However, the issues associated with choice of model risk variables highlights the potential jurisdictional (and possibly hospital level) differences in data collection/documentation/coding that can influence the derived SMRs and model performance. It also highlights the need to separate consideration of clinico-pathologic variables that are associated with death risk variation not amenable to system control from system factors that influence variance between hospitals, may not currently be amenable to control, e.g. ICU access, at a hospital level but fundamentally represent a quality of care issue at a system level. It could be argued that these factors should not be included within risk models but could be used to interrogate identified variation for explanatory reasons.

Jarman B, 2010, Netherlands

The objective of this study was to use the HSMR as a tool for Dutch hospitals to analyse their death rates by comparing their risk-adjusted mortality to the national average (Jarman, B. et al, 2010). The study used routinely collected hospital data in the National Medical Registration dataset (ICD-9) between 2005-2007. The model drew on methods used in the UK, Dr Foster model. The calculations included adjustments for non-average hospitals with different casemix using ‘scaling’ processes.

There were a number of findings reported. Fifteen hospitals had to be excluded, as they did not meet national standards for data quality. Dutch HSMRs varied widely among hospitals with the chance of dying in hospitals with the highest HSMR being 2.3 times that of dying in hospitals with the lowest HSMR. The c-statistic of the Dutch HSMR model was 0.91 and similar to reported values for similar models in other countries.

The data were presented within a funnel plot and there were a significant number of high and low outliers identified with use of either 95% or 99.8% confidence intervals which raises questions as to whether the residual variation is related to inadequate adjustment for confounding factors. The authors also note that inclusion of LOS and procedure group within casemix adjustment is debatable as they relate both to patient illness and treatment and could therefore determine quality.

van Walraven C, 2010, Canada and, van Walraven C, 2011, Canada

Of these two studies undertaken by the same group, the first cross-sectional study (van Walraven, C. et al, 2010), was undertaken at the Ottawa Hospital and involved validation of the Kaiser Permanent Inpatient Risk Adjustment Methodology (KP-IRAM) to a dataset including records between January 1998 and April 2002 which were coded according to ICD-9-CN. The dataset was divided equally into derivation and validation groups. The KP methods were duplicated except for 2 changes; diagnoses were based on previous hospitalisations and diagnoses for the current admission classified as ‘chronic’ (COPS score), the laboratory based acute physiology score (LAPS) used a different ‘troponin’ measure. As well as validating the main model, the authors also tested 3 other different versions based on Elixhauser or CCI instead of their COPS score and a different system of parameter estimation calculated from the data in this study.

The findings reported were that there were some differences in model discrimination between the four versions. The KP model c-statistic was good at 0.894 (CI 0.891-0.898). The second model in which there were data driven parameter estimate methods performed better with a c-statistic of 0.915 (CI 0.912-0.918) and also was associated with better calibration based on the Hosmer-Lemeshow statistic. Methods of measuring comorbidity did not significantly alter model discrimination. Of interest, lack of data that required modification of the LAPS score, did not significantly impact on model performance.

In the second study, involving the same Ottawa hospital but a different time period (1/4/2004 to 1/4/2009), the investigators considered the role of procedures more fully in model development and tested a previously validated model, the Kaiser Permanente Inpatient Risk Adjustment Model (KP-IRAM) to a model including the procedural Index for Mortality (PIMR) (van Walraven, C. et al, 2011). As there was access to physiological data, an illness severity score was also included in risk adjustment and interaction terms between age, severity score and comorbidity score were also included. Comorbidities were measured using Elixhauser and were measured based on the day of procedure. The dataset was randomly split into derivation (50%) and validation (50%) groups.

The investigators reported that the model performed well based on discrimination performance (c-statistic 0.88), calibration and Hosmer-Lemeshow statistic (p=0.66). There were 16.4% admissions associated with at least 1 PIMR procedure and the strongest association with death included; cardiac resuscitation, ventriculectomy, pericardial drainage and pelvic irradiation. The total PIMR score changed the expected risk of death beyond that estimated by the KP-IRAM. Model discrimination improved from 0.929 [0.926-0.932] to 0.938 [0.935-0.941]. Model calibration did not change

Whilst the study does indicate improvement in the previous model, the discrimination of that model was already high and development of the PIMR required significant manipulation of the coded data to group procedures and provide sufficient power to undertake the analysis. The utility of the PIMR for practical application is therefore uncertain.

Australian reports

- Comparison of the ACSQHC HSMR model with 30-day admission and 30-day post-discharge models identified that (Lujic, S. et al, 2012);
    - average HSMRs were higher than 30-day mortality models
    - there was strong correlation between the different models
    - agreement on outlier status was approximately 70%

Case studies

- In England, the introduction of the new SHMI has provided an opportunity to compare the outcomes of this model with the Dr Foster model. In the latest Dr Foster Hospital Guide both indicators are reported alongside the mortality in low mortality DRGs and surgical mortality.26 trusts show higher than expected mortality according to SHMI, and 20 for the HSMR indicator. Only nine trusts with a higher than expected HSMR were amongst the 26 with a higher than expected SHMI.

4.3.5 Validity of HSMR for measuring quality sensitive outcomes

Currently HSMR is being used for at least two purposes; as a quality improvement tool that 'flags' potential quality of care issues for individual hospitals to review and address and as a measure of hospital performance that is publicly reported. The latter demands use of a tool that accurately and reliably identifies average, good or bad performance to avoid inappropriate public blame and shame. However, as a quality improvement tool it has been suggested that HSMR requires less stringent measurement attributes as it should be viewed as a screen or flag for a potential quality of care issue, not as an absolute measure of performance. Leaving aside the obvious tension between these two objectives for the same measure, there remains, even for a quality improvement tool, the question of allocative efficiency. Are hospital resources being appropriately and efficiently allocated and does use of HSMR improve or adversely affect such resourcing? This issue has not been adequately addressed in the literature to date. The question has been largely addressed from a technical perspective – to what extent is variation in HSMR explained by differences in model design and performance? If a large amount of residual variation is not due to quality sensitive issues then the measure may have little utility and could distract attention and resource from more deserving areas of endeavour.

Jarman defends use of HSMR on the basis that "it just measures mortality, not preventable mortality" (Jarman, B., 2008). However, if we just wanted to measure mortality, a standardised measure would not be needed. Standardisation is being undertaken to allow comparative analysis, either over time within a hospital or between hospitals. Therefore it is the variation between points in time or between hospitals that is the focus of interest in using HSMR as the assumption is that; after considering the play of chance (random variation) and adjusting for potential systematic bias that the residual variation does indeed measure, if not preventable mortality at least quality sensitive issues amenable to intervention and improvement. Increasingly proponents of HSMR are suggesting the measure not be used in isolation but in association with other condition specific indicators. A further question following on from this position has not yet been addressed, namely; does HSMR add value to condition specific HMIs for identification and remediation of quality sensitive issues?

Ben-Tovim et al, reviewed the evidence for an association between hospital mortality data and quality of care, noting the variable way in which quality of care is defined and the heterogeneous study design applied to research associations with quality of care. In summary the authors concluded that the relationship between process measures and mortality outcomes, whether assessed using implicit or explicit review, is inconsistent and that further work should be monitored in this area.

A systematic review undertaken by Pitches et al in 2008 included 36 studies and reported on 51 relationships between process of care and risk adjusted mortality (Pitches, D.W. et al, 2007). The authors found a positive correlation between better quality of care and risk-adjusted mortality in 51% relationships, no correlation in 31% and a paradoxical relationship in 31% leading to the conclusion that “the notion that hospitals with higher risk adjusted mortality have poorer quality of care is neither consistent nor reliable”.

Overall, there have been few high quality studies that address the relationship between HSMR and quality of care. However, these studies are difficult to do for a number of reasons. There is no true gold standard for quality of care and studies must define implicit or explicit proxies/surrogate markers such as processes of care linked to mortality. Further, where comparisons are made with the current ‘gold standard’, medical record review there are considerable costs involved and issues associated with subjectivity due to hindsight bias and variable, often low inter-rater reliability (Girling, A. J. et al, 2012) about ‘preventability’ of adverse events (Thomas, E. J. et al, 2003).

Girling AJ, 2012, UK

This theoretical mathematical modelling study provides some insight in to the ongoing issue of the relationship between hospital standardised mortality rates and quality of care related to preventable mortality (Girling, A. J. et al, 2012). The authors, who have published extensively on the technical limitations of HSMR, constructed a mathematical model based on a number of defined parameters using data derived from literature reports. They acknowledge a key study parameter to be the proportion of deaths that are preventable and a key study limitation to be a lack of scientific literature that has addressed this issue.

An approximate figure for preventability is therefore used, based on literature reporting that 6% deaths (95% CI 3.4%-8.6%) were associated with clinical error (Hayward, R. A. et al, 2001). Despite this limitation, the findings are thought provoking as, unless preventable mortality is at least 15%, whereby the positive predictive value of a SMR identifying a hospital performing within the worst 2.5% is 0.3 (30%), at 6% the PPV is only 0.08 (8%). Within this scenario hospitals identified as performing poorly 10/11 times would not be performing poorly and equally 10/11 poorly performing hospitals would not be identified.

There are clear issues associated with the study related to the number of assumptions and compromises that were required to apply their mathematical model, however their hypothesis has been strengthened by a recent study by Hogan et al, where a medical record review of 1000 patient episodes across 10 acute hospitals in the UK resulted in trained reviewers identifying problems in care for 13.1% deaths, of which they judged 52 (5.2%, 95%CI 3.8-6.6) deaths to be preventable (Hogan, H. et al, 2012). As has been discussed earlier there was only moderate inter-rater reliability for assessment of preventability with kappa=0.49 (95% CI 0.2-0.8). The issue is not that the problem is small and therefore not worth the effort but rather, how can problems and preventable deaths within health be best identified in a timely manner?

Sorinola et al (2012) adapted the study methods of Hogan et al and others and reviewed a convenience sample of 400 patients who had died in 2009 to obtain a more accurate estimate of the number of preventable problems contributing to death at a general hospital in the UK (Sorinola, O.O. et al, 2012). Trained reviewers identified 34 (8.5%) patients who had experienced a problem in their care that contributed to death and their quality of care was more likely to be judged poor. Inter-rater reliability of the review's key variable, 'determination of a problem in care', was good as determined by the Cohen *k* coefficient (k =0.75). Fourteen deaths (3.5%) were considered preventable. All cases considered as preventable deaths were blindly reviewed by a second group of reviewers, the inter-rater reliability for assessment of preventability was not provided. The authors went on to use their findings to implement a range of strategies to monitor mortality and improve care delivery generally. They noted mortality review is important but it is unwise to limit safety monitoring to this small group of patients, as the majority of problems in care may result in morbidity and disability rather than death.

Proponents of HSMR argue the value of HSMR based on published examples of improvement associated with its application. In his defence of HSMR, Jarman refers to positive impacts of using HSMR and references a 'before and after' study by Wright et al of a large acute Trust in the UK where there was a reduction in HSMR over a three year period (Wright, J. et al, 2006). The authors attributed the considerable fall (from HSMR 94.6 to 77.5) to strategies for change which included; surveillance and review of mortality, introduction of a modified early warning score to identify patients with deteriorating condition, use of a palliative care team to support end of life guidelines and a nursing home education program to encourage nursing homes to let patient die there rather than in-hospital, revision of the hand hygiene policy and targeting of a number of medication errors. However, the study design was poor, there is little evidence to support a direct relationship between the strategies employed and mortality improvements and the change in admission of patients with palliative care could be considered a likely major contributor to the change.

This study reflects the many difficulties related to methods and data interpretation in assessing relationships between HSMR and quality of care, in particular in studying the impact of remedial strategies on patient health outcomes.

Groene O, 2011, Spain

A study by Groene et al attempted to explore associations between the 'maturity' of hospitals' quality improvement systems and hospital wide quality and patient safety outcomes, including HSMR (Groene, O. et al, 2011). The measure of hospital quality used in this study was the Maturity Index, a measure derived from a European project Methods of Assessing Response to Quality Improvement Strategies (MARQuIS). Within the MARQuIS model, quality is defined as 'the application of quality policies and procedures, quality governance structures, and quality activities that close the gap between current and expected levels of quality. It is a self-reported questionnaire comprising 113 items across seven domains: policy; planning & documentation; leadership; structure; general quality improvement activities; specific quality improvement activities; patient involvement; and accountability. The domain scores are combined in a mean overall score for each hospital to provide the Maturity Index. The authors matched data from 43 Spanish hospitals that participated in the MARQuIS study with performance data from IASIST 20 Top Hospitals, a voluntary, benchmarking initiative that uses administrative hospital data. A range of statistical methods was used to explore associations between the Maturity Index and hospital level indicators. The study found hospitals with a more mature quality improvement system had higher mortality rates than other hospitals, but not a significant association. The authors discuss this finding from the perspective of the 'methodological challenges' that limit the use of HSMR including low signal noise ratio and subsequent problems of risk adjustment such as the casemix adjustment fallacy or constant risk fallacy. There are a number of limitations with this study: the sample size represents less than 50% of Spanish hospitals, with higher representation of larger and public hospitals; data quality for the hospital level indicators is not established; and the nature of self-reporting quality maturity is subjective.

Kernisan LP, 09, USA

An American study exploring the association between hospital-reported Leapfrog Safe Practice Scores and inpatient mortality also demonstrates the limitations of making comparisons between quality surveys and mortality rates (Kernisan, L.P. et al, 2009). Kernisan et al aimed to determine the relationship between 155 hospitals' Leapfrog Safe Practice Scores (SPSs), a voluntary annual hospital survey assessing 27 practices, and mortality rates calculated using the National Inpatient Sample administrative dataset(Kernisan, L.P. et al, 2009). Hierarchical logistic regression was used to determine the relationship between quartiles of Safe Practice Score and risk-adjusted inpatient mortality, after adjusting for hospital discharge volume and teaching status. The authors found that the 2006 Safe Practices Survey, survey scores were not significantly associated with risk-adjusted inpatient mortality. The distribution of survey scores for this study was skewed, with most hospitals scoring above 770 (of a possible 1000) and there was considerable discussion about the validity of the Leapfrog Safe Practice survey and changes required to better measure quality practices.

Further application issues and impacts of using HSMR are discussed in the next section.

Australian reports

- Issues of validity of HSMR with quality of care have not been systematically investigated in Australian reports.

Case studies

- Issues of validity of HSMR with quality of care have not been systematically investigated in the other jurisdictional reports.

4.3.6 Implementing and utilising HSMRs

The implementation and use of HSMRs is occurring at both a jurisdictional and hospital level. Institutions are using HSMR to better understand contributors to in-hospital mortality by interpreting and validating the data at a hospital level. Various ways of presenting the HSMR are being explored to make it meaningful by flagging issues as they arise.

Chong C, 2012, Canada

The study by Chong et al (previously described) demonstrates that jurisdictions are monitoring aspects of HSMRs and coding over time (Chong, C.A.K.Y. et al, 2012). In this instance the exploration of the impact of palliative care coding on HSMRs since their public release in Canada. The authors found that palliative coding rates had increased, which was likely to have contributed to the decreasing HSMR nationally, demonstrating the impact of changes to coding practices on the HSMR.

Kroch E, 2010, USA

A study, utilising administrative and electronic record data, at a general hospital in America, explored the inclusion of Do Not Resuscitate orders (DNR) and palliative care (PC) status in mortality risk-adjustment models (Kroch, E.A. et al, 2010).

The authors undertook a medical record review to confirm patients who had received PC and/or had a DNR order in place and matched the data with the CareScience database to explore the relationship between DNR, palliative care and how these two variables relate to inpatient mortality using a general multivariate model.

Electronically extracted data showed that all clients receiving palliative care were not coded as V66.7 due to coding practice that stated a written order alone for palliative care was not coded.

The key findings were that patients with care limiting orders have higher risk of mortality than the general inpatient population, however most DNR patient survive the episode (65%) whereas most PC patients do not (73%). The later in the hospital stay that the DNR order is written, the higher the risk of death (27% for orders made on day 1 to 59% for orders after day 5). Mortality rates for patients with PC/DNR orders are higher than expected - the 'mortality rate-risk gap' and is much higher for PC (42%) than for DNR only patients (8%) i.e. PC enhances risk models especially. There were differences in PC and DNR status levels across conditions and services.

The study identified a wide range of issues with care limitation orders, especially DNR documentation and palliative care coding practices, which limits their use in estimating mortality risk. The data supports the need to tailor mortality risk adjustment models to the indicator of interest and target population.

Popowich J, 2011, Canada

The 'If High, Why' program implemented at Alberta Health Service (Canada) is an example of HSMR application (Popowich, J. et al, 2011). Popowich et al outline how two hospitals used raw mortality data and medical record review processes in order to better understand the intricacy of the HSMR measure. Monthly HSMR measures were generated for each hospital (as well as surgical, medicine and intensive care unit HSMRs in addition to the regional aggregate) through the Canadian Institute of Health Information (CIHI) portal. A review of site-based raw mortality data allowed for understanding of patient trends and if findings were inconsistent with the HSMR results, then medical record reviews were undertaken in the targeted areas. A key component of the initiative is the HSMR committee, which supports a standardised process for the analysis and review of the raw monthly mortality data and the corresponding HSMR, with an aim to minimise unnecessary chart review. The medical record reviews are undertaken using the Healthcare Improvement (IHI) Global trigger Tools (GTT) and during the initial phases of the HSMR program led to the implementation of the 'Safer HealthCare Now' bundles of care. The 'If High, Why' program formalises these steps into a process that can be adopted by other hospital sites.

This project was supported by CIHI, who are unable to provide same level of support for developing reports and accessing raw data to all individual sites. There is an expectation that each hospital, province or region avail themselves of the training available on the CIHI e-portal. The forming of a HSMR committee supports the learning about the data and can continue to drive the process. A finding also made by Sorinola et al who established a Mortality Group as a surveillance system for the quality improvement programme(Sorinola, O.O. et al, 2012).

Mohammed M, 2013, UK

Mohammed and Stevens provide another example of how health services are validating the HSMR at a health service/hospital level (Mohammed, M.A. et al, 2013). In 2008/09 the Dr Foster HSMR for Shropshire and Telford NHS Trust Hospitals was 99, but in 2009/10 this jumped to 118 (19% increase). The Trust is made up of two hospitals, the Royal Shrewsbury Hospital and the Princess Royal Telford. The authors deconstructed the HSMR to better understand whether it was an increase in observed deaths (numerator) or a fall in expected deaths (denominator) that was contributing to the increased HSMR and whether the change could be attributed to either of the hospitals. Using the admissions data from the Dr Foster Real Time Monitoring system, the authors plotted the observed and expected deaths using simple run charts to visualise the data. The run charts showed that observed deaths were stable in Royal Shewsbury Hospital and Princes Royal Hospital but expected deaths, especially at Princes Royal Hospital, had fallen. Exploration of the fall in expected deaths revealed the clinical coding process at Princes Royal Hospital was producing a lower depth of coding resulting in lower expected mortality.

Morsi E, 2012, USA

An American study by Morsi et al provides a different insight into the implementation of HSMR and their utility(Morsi, E. et al, 2012). Morsi et al explored the use of publically available mortality data by primary care physicians. A web-based survey was distributed to primary care physicians (92 participants / 47% response rate) associated with three acute care hospitals in Massachusetts to ascertain factors that influence their referral choices. Factors considered very important were: familiarity with the hospital, patient preference and admitting arrangements with the hospitalist group. Publicly available quality measures were reported as not at all important by 42% of the respondents. Findings regarding the limitations of publicly reported performance data include 42% “agreed” that risk-adjusted methods were inadequate to compare hospitals fairly, 76% “agreed” that mortality rates were an incomplete indication of quality of hospital care, 62% “agreed” that hospitals could manipulate data, and 72% “agreed” that the ratings were inaccurate for hospitals with small caseloads. Whilst the findings of this study cannot be generalised to the Australian setting due to differences in the health care systems, an important implication is the who is the target audience for publicly available HSMR and are they really interested.

Whilst some of these studies noted the need to limit the burden of undertaking resource intensive medical record reviews, none mentioned cost implications of collecting and validating data, nor resources required to generate mortality indicators and validate the findings. There was also no reference made to how clinicians were involved in the process, however a Victorian Department of Health initiative noted the difficulties engaging with clinicians and concerns about diverting clinicians away from clinical work to spend time on validating data (Clarke, A.L.L. et al, 2010). These studies did not comment upon the long-term impact of implementing and using HSMRs upon safety and quality of care.

Australian reports

- Issues relating to implementation and utilisation of HMIs in Australian contexts have not yet been published.

Case studies

- Data providers, including jurisdictions and private providers, support hospitals in understanding and actioning their mortality data. Dr Foster for example provides guiding information and a checklist for participating hospitals to understand and investigate results that are outside the expected range. Their Real Time Monitoring product also provides prompts and drill down tools to facilitate investigation.
- Most jurisdictions have dedicated health service web portals for accessing data, as well as software to support local analysis. The Canadian online system for health services is sophisticated and is supported by detailed guidance for understanding and managing results, including making comparisons with appropriate peer groups.
- The new SHMI indictor is reported in association with a number of supporting indictors to aid interpretation.

4.3.7 Summary of HSMR key findings

- HSMR models have been developed across many international jurisdictions that exhibit good performance attributes in relation to measuring variation in death rates.
- All studies report wide variation in HSMR when comparisons are made across hospitals, even within similar hospital groups.
- Different models perform differently when applied to external populations to the one in which the model was developed and all jurisdictions have modified model specifications to suit local contextual needs
- The definition of HSMR varies in relation to timing of death. The focus of this review was on in-hospital HSMR, however there are indications that, where there is ready access to linked data, us of 30-day models from admission or postdischarge are being used.
- There is ongoing debate about variables to include in risk adjustment models due to potential variability in coding practices and hospital admission/discharge policies that may lead to the 'constant risk fallacy'.
- There is consensus that further consideration needs to be given about inclusion/exclusion of multiple admissions, palliative care status and interhospital transfers, all of which impact on HSMR.
- There is ongoing debate about the nature of residual variation associated with use of HSMR and the relationship between high/low outlier status and quality of care. There is a lack of good quality studies investigating the relationship between HSMR and quality of care.
- There is a lack of information about implementing and evaluating HSMR programs, particularly with regard to cost benefit and allocative efficiency.

4.4 Condition specific HMIs

A number of jurisdictions utilise inpatient-only condition specific indicators (Australia, US (AHRQ), UK (Dr Foster) and OECD, but 30-day mortality indicators are also featured in English (NHS), Scottish (NHS), US (CMS) and Danish indicator sets.

Of the 30-day indicators, the English UK indicators measure inpatient and 30-day post-discharge mortality. The others measure mortality 30-days from admission.

4.4.1 Studies and indicators

Ten studies (3 Australian, 1 Belgian, 1 UK, 4 USA and 1 Norwegian) reported on condition specific HMIs. Of these 3 focused on issues associated with the definition of hospital mortality, 1 investigated explanatory structural hospital variables associated with HMIs, 5 investigated issues associated with HMI application/interpretation, and 1 reported on validation of stroke HMI with quality of care measures.

4.4.2 Defining hospital mortality

Kristoffersen DT, 2012, Norway

This is one of the few studies to address condition specific HMIs; for first AMI, stroke and hip fracture in Norwegian hospitals. Given each Norwegian resident has a unique PIN, the study demonstrates the advantage of easy linking between hospital and other National datasets to investigate various definitions of mortality. Three key definitions were tested; in-hospital 30-day mortality (IH30D), 30-day-in-and-out of hospital deaths for patients treated in only one hospital (S30D) and similar 30-day mortality but including weighting for transfers according to the LOS in each hospital (W30D).

The key findings were firstly, that the proportion of deaths within 30-days differed by condition especially for hip fracture, which was lower than for AMI or stroke. Secondly there were major shifts in-hospital rankings when comparing different mortality measures, especially for hip fracture when comparing in-hospital mortality to other 30-day mortality measures.

All mortality measures exhibited strong predictive attributes measured by the C statistic. The way in which weighting was applied for transfers may be questioned as it assumes constant daily risk. The study also highlights the potential condition specific issues that might influence choice of mortality measure for comparative purposes.

Borzecki AM, 2010, USA

In this cross-sectional study of Veteran Affairs (VA) hospital data, investigators compared in-hospital and 30-day post-admission standardised mortality rates for 6 medical conditions, including AMI, CHF, stroke, gastrointestinal haemorrhage, hip fracture and pneumonia (Borzecki, A.M. et al, 2010).

All medical conditions had higher observed 30-day mortality rates than inpatient mortality rates. There was a strong correlation between in-hospital mortality for all conditions except hip fracture and measures of agreement for outlier status were at least moderate agreement ($k>0.40$) except for hip fracture ($k=0.12$) and stroke ($k=0.22$).

The mean number of facilities that changed outlier status was 18 (range 12-23), with the highest number of sites changing status being for pneumonia, one of which changed from high to low.
The conclusion was that in general in-hospital and 30-day mortality rates are comparable except for conditions associated with post acute care rehabilitation such as stroke and hip fracture where there

may be bias due to differential discharge policies and access to rehabilitation care. As the study was focused on VA patients who are primarily white males with a high comorbidity load, generalisation of the results to other settings is uncertain.

Drye EE, 2012, USA

This cross-sectional study also investigated the impact of mortality definition for AMI, heart failure and pneumonia in a 3 year study of data relating to Medicare patients 65 years or more who received treatment in over 3000 non-federal acute care hospitals (Drye, E.E. et al, 2012).

The investigators reported variation between the conditions in relation to LOS (trend to shorter for AMI) and numbers of transfers (most related to AMI). The mean risk adjusted mortality rate (RSMR) differences between 30-day mortality and in-hospital mortality were; AMI 5.3 % (SD 1.3) points, 6.0% (SD 1.3) points for CHF and 5.7% (SD1.4) points for pneumonia.

In-hospital models resulted in different performance classifications for hospitals for 8.2% AMI, 10.8% CHF and 14.7% pneumonia.

Hospital transferred out rates for AMI were negatively associated with in-hospital RSMRs. With 30-day mortality as the gold standard, sensitivity (sn) and specificity (sp) for identifying ‘better’ hospitals for AMI (sn 38.7%, sp 98.3%), for CHF (sn 34.4%, sp 98.5%) and for pneumonia (sn 43.0% sp 97.3%) and for identifying ‘worse’ hospitals for AMI (sn61.7% sp96.8%), for CHF (sn 50.5% sp 96.2%) and for pneumonia (sn 66.6% sp 93.8%). In-hospital mortality measures were associated with more between-hospital variation and using in-hospital mortality favoured hospitals with shorter LOS, especially for pneumonia.

This study supports others in the potential impact of LOS and discharge policies on risk adjusted SMRs and is limited in generalisation only by the restriction of data to persons 65 years or more.

Australian reports

- The indicator with the highest transfer out rates is fractured neck of femur (67%) (AIHW, 2011b).
- Excluding transfers out would result in lower mortality rates for most hospitals, with the greatest difference being for stroke and fractured neck of femur.
- Excluding transfers in generally decreases the number of in-hospital deaths by between 9-10% for most indicators except fractured neck of femur which decreased by 16%.

Case studies

- In the US, the data element “Present on Admission” has been implemented by the AHRQ to provide a means of distinguishing pre-existing comorbidities from complications that occur during the hospitalisation in question. The element is applied in the risk adjustment.

4.4.3 Validity of condition specific HMI for measuring quality sensitive outcomes

One study (Palmer W.L. et al, 2013) was identified that examined the relationship between condition specific HMI (stroke) and quality of care measures.

Palmer WL, 2013, UK

The key objectives of this well designed study were to evaluate system variation in 30-day in-hospital stroke mortality rates and other indicators of good quality care in English NHS hospitals between 1/4/2009 to 31/3/2010, to investigate coding related causes of bias and to investigate convergent validity between chosen measures (Palmer, W.L. et al, 2013).

The investigators chose six indicators that were accessible from an administrative data source, based on a literature review about their relationship to quality of care. These included; same-day scanning, thrombolysis rates, aspiration pneumonia rates, 30-day in-hospital mortality, discharge to usual residence, 30-day readmission rates.

Stroke episodes and indicators were defined by ICD-10 codes and were assigned to clinical subgroups groups. Where there were multiple episodes, these were bundled into a 'superspell' (FCE) and transfers were scored against the referring hospital. Logistic regression methods were used and data was adjusted for demographic variables, comorbid diagnoses (Charlson score), urgency of admission, stroke type and month of admission. Explanatory variables included use of brain scanning and thrombolysis process measures.

The results documented that 69.7% were scanned within 1 day of admission, 2.6% received thrombolysis, 5.3% had aspiration pneumonia, 72.8% were discharged to usual place of residence and fewer than 13.7% were transferred. All indicators, except readmissions were associated with hospital outlier status being identified which was beyond what would be expected by chance alone. Of 25 hospitals identified as outliers at 99.8% level, 20 (80%) were again flagged as outliers when coding practice was included in the regression. (Conversely 1 in 5 identified outliers are no longer outliers once coding practice is considered). There was negligible correlation between use of I64 (unspecified stroke) and 30-day mortality even though the code was used variably (between 0.2-42.6%). As expected, there was a correlation (weak r=-0.17) between I64 and same day scanning rates. There were six pairs of indicators that had a statistical correlation.

The authors concluded that the administratively based indicators offered potential for measuring variation in quality and safety, however cautioned that the current data should instigate debate rather than be applied without further research. In particular, they note that residual variation cannot be attributed to quality of care issues as there was no information about important variables such as stroke severity and pre-stroke function. They suggested the need to correlate such indicators with robust clinical data such as that collected in clinical audit. Of interest, there were no correlations identified between 30-day in-hospital mortality and other stroke indicators.

4.4.4 Implementing and utilising condition specific HMIs

Three of the five studies associated with application / interpretation were related to AMI (Aelvoet, W. et al, 2010, Bradley, E.H. et al, 2012, Scott, I.A. et al, 2008), and two looked at a range of HMIs (Carretta, H.J. et al, 2013, Clarke, A.L.L. et al, 2010). In addition, the study by Coory et al (2008) compare the performance of funnel plots and control charts for AMI mortality, and has been summarised in section 4.6.

Aelvoet W, 2010, Belgium

A study by Aelvoet et al explored inter-hospital comparisons of standardised in-hospital AMI case fatality rates (AMI-CFR) across 109 Belgium hospitals to evaluate the extent to which discharge records allow the assessment of quality of care in the field of AMI and to identify starting points for quality improvement (Aelvoet, W. et al, 2010). This was a complex study using administrative data (Belgian Minimal Clinical Data [MCD]), which was compared to a cardiovascular disease registry (Monitoring of Trends and Determinants in Cardiovascular Disease [MONICA]) to evaluate the quality of the administrative data, for the period 2002-2005.

The authors identified a number of data quality issues associated with use of the MCD in the form of variable coding and case management practices, which contributed to issues with both the numerator and denominator for the AMI mortality rate.

Issues contributing to data quality included uncertainty around diagnostic criteria for AMI due to the absence of national guidelines, leaving choice of diagnostic criteria to individual clinicians; 50% of fatal ST-segment elevation myocardial infarction (STEMI) cases occurred within the first two hours after onset of symptoms and were under-registered in the MCD, probably due to lack of coding in the emergency department; and important information such as symptom-onset to needle-time / initiation of treatment and pharmacological treatment details are lacking in the administrative data.

Despite data limitations, the authors observed sizable differences in the AMI –CFR between hospitals and within hospital types and identified non-conformities to guidelines for treatment, specifically time to re-perfusion based on transfers to hospitals able to perform angioplasty, suggesting opportunities for quality improvement.

Disappointingly, although the authors referred to feedback to the College of Physicians and to individual hospitals, there is no presentation of results in relation to this feedback in the paper.

Bradley EH, 2012, USA

In this study, Bradley et al examined statistical associations between AMI mortality ratios and hospital strategies for reducing mortality associated with AMI, across 533 acute care hospitals over a two year timeframe (2008 – 2009) (Bradley, E.H. et al, 2012). Participants completed a retrospective web-based survey identifying strategies pertaining to organisational values and goals, senior management involvement, staff expertise, communication and coordination among staff, and problem solving and learning.

Standardised mortality ratios for AMI (30-day) were calculated using the Centres for Medicare & Medicaid Services (CMS) methodology. Multivariate regression models were used to examine associations between specific strategies and SMR-AMI.

The authors identified several strategies that were significantly associated with lower SMR-AMI and hospitals that implemented a greater number of effective strategies tended to have a lower SMR-AMI although this observation did not reach statistical significance. Effective strategies included holding monthly meetings to review AMI cases between hospital clinicians and staff who transported patients to hospital, having cardiologists always on site, not cross-training nurses from intensive care units for the cardiac catheterization laboratory, and having physician and nurse champions rather than nurse champions alone. The authors noted fewer than 10% of hospitals reported using at least 4 of these 5 strategies, indicating opportunities for quality improvement.

Whilst this study had a good sample size and excellent response rate it was a retrospective survey with only one respondent completing the survey for each hospital. Survey respondents are likely to have been aware of their hospital's SMR-AMI as this was data in the public domain. A number of the survey questions were subjective and poorly defined. The validity of the SMR-AMI was accepted on face value, with not consideration given to data quality or coding differences.

Carretta HJ, 2013, USA

This study examined hospital and patient structural characteristics associated with AMI, stroke, CHF and pneumonia mortality quality indicators and 30-day post discharge mortality, in Florida (USA) acute hospitals during 2008 (Carretta, H.J. et al, 2013).

The study demonstrates the potential utility of condition specific indicators for examining care characteristics. The key finding that supports previous literature was that hospital volume is associated with mortality outcomes; smaller condition specific case volume being associated with increased risk of AMI, CHF, stroke and 30-day post discharge mortality. However, at the level of whole of hospital, larger hospital size was associated with higher CHF, stroke, and 30-day post discharge mortality.

Hospital and patient characteristics that were significantly associated with mortality varied across the diagnostic groups. For instance, hospital system affiliation was associated with reduced hospital mortality for CHF, whilst 'for profit' hospitals had a higher in-hospital CHF mortality but lower 30-day mortality. Similarly there were higher mortality rates in teaching hospitals for CHF, but lower 30-day mortality. Overall, hospital characteristics had more impact on CHF and stroke than for pneumonia or AMI outcomes.

Generalisation to the Australian setting is uncertain as some of the hospital and patient factors such as health system affiliation and Medicaid/Medicare/Self-pay and Hispanic ethnic background do not necessarily apply. The study did not examine all structural characteristics, for instance staffing structure.

Clarke A, 2010, Australia

A narrative study by Clarke et al reports the experience of the Victorian Department of Health in seeking clinician engagement across 20 health services to test 11 quality of care indicators including in-hospital mortality for low mortality diagnostic groups (LM-DRGs), stroke, heart failure, AMI, pneumonia and fractured neck of femur (Clarke, A.L.L. et al, 2010). The initiative was based on the Queensland Health approach of using routinely collected administrative data to produce variable life-adjusted display (VLAD) control charts.

The intention was to drive improvement at a local level by supporting clinical ownership for safety and quality of the care. An introductory forum was held to engage with senior clinicians and administrators, a reference group was formed who sought advice from relevant peak bodies and benchmarking data was provided to participating health services.

Participating health services reported limited involvement of clinicians, partly due to lack of confidence in the data source; a higher than anticipated level of strain on resources; and too great a delay between incident and report at the hospital level.

The authors noted that "diverting clinicians from patient care may itself create an environment that compromises safety and quality". This issue has not been further explored as yet.

Scott IA, 2008, Australia

Scott et al undertook a retrospective, cross-sectional study comparing variable life-adjusted display curves (VLAD) derived using administrative or clinical prediction models applied to 467 consecutive patients admitted to a single tertiary hospital in Queensland, Australia between 1st July 2003 and 31st March 2006, with a primary diagnosis of AMI (Scott, I.A. et al, 2008).

The authors found the two prediction models generated almost identical VLAD curves, showing a steadily increasing excess mortality over the study period, culminating in an estimated 11 excess deaths. The clinical model showed slightly better discrimination (C=0.89; 95% CI: 0.85 – 0.94) than the administrative model (C=0.84, 95% CI: 0.81 – 0.88).

Following consultation with senior hospital clinicians, sensitivity analyses were undertaken and adjustments made to exclude patient populations with high risk of death that was not under the control of the hospital including; misclassified cases, out of hospital cardiac arrests, deaths within 30 minutes of presentation, transfers in from other hospitals and patients whose care had a palliative intent. The replotting of curves reversed the mortality trend, a net gain of seven or 10 lives.

The authors concluded that appropriate patient selection is more important than choice of dataset or risk-prediction model when statistical process-control methods are used to flag unfavourable mortality trends. The authors identified VLADs are useful for real-time monitoring of outcomes over time within a single institution but cautioned against their use for inter-hospital comparisons. They noted that VLADs do not provide definitive proof of, or explanations for, lower quality care but rather are useful as flags for stimulating further investigation by clinicians of both data quality and process of care issues. Restricting patient population selection can reduce the unnecessary investigation of false positive flags.

Australian reports

- Issues relating to implementation and utilisation of HMIs in individual Australian hospital contexts have not yet been published.

Case studies

- In various jurisdictions there is evidence that the use of condition-specific mortality indicators is being used to drive improvement, with linkages being made between process of care indicators and the mortality indicators.
- Some jurisdictions (e.g AHRQ) have developed composite indicators which combine process of care and outcome indicators to form a performance snap shot with comparisons across States.

4.4.5 Summary of key findings for condition specific HMIs

- The definition of mortality has a varying impact on condition specific HMI outcomes.
- There are generally strong correlations between in-hospital and 30-day mortality models, except for fractured neck of femur and stroke where the differences are more likely to be associated with changes in-hospital rankings.
- Patient population selection, for example inclusion/exclusion of transfers influences HMI outcomes
- Excluding AMI patient populations at high risk of death not amenable to hospital control can improve utility of HMIs and reduce unnecessary investigation of false positive flags when process control charts such as VLADs are implemented.

- Engaging senior hospital clinicians in data interpretation may be difficult at a jurisdictional level, however it can be value adding in determining appropriate selection of patient populations for the HMI denominator and gaining trust in use of administrative data.
- Data quality issues can be identified when administrative datasets are compared to clinical registry data, however the variation between hospitals of similar type identified using administrative models, indicate the potential of such data for screening and further investigation of outlier status.
- Condition specific HMIs can be used to investigate associations between hospital structural characteristics and mortality outcomes. The evidence to date supports a strong consistent relationship with case volume and better outcomes. There is also a potential differential relationship between various HMIs and hospital characteristics that warrants further investigation.

4.5 Low mortality DRG (LMDRG) HMI

The indicator LMDRG is currently included in three international jurisdictions; Australia (ACSQHC), USA (AHRQ) and UK (Dr Foster) (Table 4). A summary of these indicators is provided in Appendix 2 and more detailed information can be found in Appendix 3.

4.5.1 Defining LMDRG

All three jurisdictions define the included population as DRGs in which there is <0.5% mortality, include patients aged 18 years or more and all exclude admissions associated with trauma, cancer or immunocompromised state. There are differences in approach to risk adjustment. There is no adjustment for the Australian indicator; however data is stratified for peer group. The AHRQ has a risk adjustment model that includes demographic data, some comorbidities and transfers into the hospital. Dr Foster provides a LMDRG crude rate and relative risk ratio.

The jurisdictions report LMDRG differently. Australia specified the proportion of separations with death in LMDRGs, AHRQ reports deaths/1000 LMDRG discharges and Dr Foster reports relative risk, where RR 100 is equal to the national average. Both AHRQ and Dr Foster report the data publicly and member hospitals can access their online systems to use the data.

4.5.2 Developing and testing LMDRG: measurement attributes

The rationale for using LMDRG as an indicator has already been addressed. However, the measurement attributes and use of this hospital-level indicator by individual hospitals for quality improvement or for benchmarking performance between hospitals have more recently been questioned. Mihrshahi and colleagues, in a narrative review, have summarized 12 peer reviewed clinical studies and identified a number of issues relating to use of LMDRG (Mihrshahi, S. et al, 2010). Considerable variation was identified in definition of the indicator across jurisdictions related to use of different or modified versions of the ICD coding classification system.

Assessment of construct validity (Table 13) failed to identify a strong relationship between LMDRG and other measures of quality of care, however the studies all had methodological limitations particularly related to inadequate adjustment for casemix and illness severity. Further, the choice of quality measures used for comparison, such as accreditation and external rating scores could be strongly criticized as these measures are not proximal to actual treatment processes of care for the conditions associated with death. Of some concern, where LMDRG was one of a number of PSI investigated, the results for LMDRG were inconsistent and in some instances exhibited an inverse relationship to other quality measures (Isaac, T. et al, 2008).

Table 13. Summary of studies addressing construct validity attributes of LMDRG (Adapted from Mihsrahi et al, 2010)

Measurement Attribute	Population	Main findings	Limitations
Miller MR, 2004, USA	5.7 million paediatric discharges (data from 2000)	Increased LOS Increased hospital costs	All PSI events were in high technology, teaching hospitals. Likely incomplete adjustment for casemix, severity of illness, access to care.
Rosen AK, 2005, USA	281,423 Veterans	Correlated with ‘failure to rescue’ but no other PSI	Potential lack of generalisability to other populations.

Measurement Attribute	Population	Main findings	Limitations
		Increased LOS Increased hospital charges	
Miller MR, 2005, USA	2071 USA hospitals	No association between LMDRG and accreditation scores	There was low variation in accreditation scores yet wide variation in PSI scores. The distance between process of care leading to death and overall accreditation assessment may reduce likelihood of seeing any correlation between the two.
Isaac T, 2008, USA	Patients 65 years or more treated in 4415 hospitals (data from 2003)	Inconsistent weak and often inverse relationships between LMDRG and 3 measures of quality; Hospital Quality Alliance (HQA) process metrics, in-hospital risk adjusted mortality rates for common conditions, US0NEWS rating systems.	HQA quality measures related to only 3 medical conditions and were reported at a later time period than PSI data. US-NEWS score was provided for only 75/4415 hospital assessed. Findings limited to older populations.
Hutchinson A, 2009, UK	148 UK hospitals (data from 2004-2005)	No correlation between 3 PSI (Including LMDRG) and healthcare incident reporting to the National Reporting and Learning System (NRLS)	Technical difficulties associated with implementation of the NRLS may have impacted on study.
Thornlow DK, 2006, USA	995 USA hospitals (data from 2000)	No association between LMDRG and hospital type (teaching/non teaching), ownership (government/private), geographical (rural/urban) status, unlike other PSIs	Statistical methods did not consider hierarchical data issues and assumed PSIs were independent of each other.

4.5.3 Issues in applying LMDRG

Two studies were identified by Mihsrahi et al in which LMDRG was used to specifically compare the quality of care between organisations (Mihrshahi, S. et al, 2010). Rosen et al, reported higher risk adjusted LMDRG for Veterans compared to the National Inpatient Sample but lower than for Medicare patients (Rosen, A. K. et al, 2005). The adequacy of risk adjustment was questioned. A second Veteran's study in the USA by Weeks and colleagues found a higher rate of risk adjusted LMDRG for Veterans than for patients treated outside the Veterans Affairs system (Weeks, W. B. et al, 2008). However, all other PSIs tested were equivalent or lower within the VA system, questioning potential coding practice differences between the two health systems.

Trends in LMDRG over time were tested in two studies (1 USA and 1 UK), both of which reported decreasing trends over time. It is difficult to interpret these trends in relation to changes in quality of care as the denominator for contributing hospitals changed in the first study and in the second study trends in other PSIs either stayed the same or increased.

Two further studies were identified in the literature search relating to LMDRG

Barker AL, 2011, Australia

A study by Barker and colleagues (Barker, A.L. et al, 2011) adds further knowledge to the measurement attributes of LMDRG. This retrospective cohort study examined the frequency of deaths in LMDRGs and the patient and hospital characteristics associated with them. The authors used the Victorian Admitted Episodes Dataset, including 122 public hospital episodes over a 2 year period between July 2006 and June 2008. The AHRQ patient safety indicator specifications for LMDRG, translated by the Victorian Department of Health were applied and a hierarchical logistic regression model was used to test associations of LMDRG with patient and hospital characteristics.

The main findings were that LMDRG episodes were infrequent and ranged from 0-20 deaths per hospital. Over 50% hospitals did not experience a LMDRG episode over the period of study. There was high variability in DRGs with no single DRG diagnosis, procedure or complication being reported in more than 10% of cases. A large number of cases (40%) were in older people (83 years or more) and in episodes of 1 days or less (39%). A high proportion (74%) were emergency and medical in type. The numbers of LMDRG episodes associated with patient transfers varied between the two year period under study (20% in 2006/7 and <12% in 2007/8). Patient characteristics associated with LMDRG were; older age, male, comorbidity level, unplanned admission and transfer from hospital or aged residential care facility. Hospital characteristics were; metropolitan location but not teaching hospital.

The authors examined outlier status with and without risk adjustment (15 versus 59, $p<0.05$) and reported significantly fewer outliers were identified when data was risk adjusted. In view of the low frequency of LMDRG events, confidence intervals were wide across all hospital. The authors came to a number of conclusions and recommendations;

- The low frequency and variability of LDRG diagnoses suggests the indicator will be insensitive to variations in quality of care.
- It is likely that there are many false positives associated with current LMDRG specifications due to unadjusted casemix and that LMDRG should be applied using risk adjustment.
- Associations with low volume hospitals and transfers from hospitals requires further investigation.

Clarke A, 2010, Australia

This narrative study reports efforts by an Australian jurisdiction to engage clinicians in using HMIs, including LMDRG (Clarke, A.L.L. et al, 2010). In brief, this was found to be a difficult process due to concerns of clinicians about data quality, resource implications associated with investigating potential areas of concern and concerns about diversion from clinical service activities. Of interest, the author observes there is a devolved governance approach whereby clinicians are themselves responsible for the safety and quality of the care they provide.

Australian reports

- Issues relating to implementation and utilisation of HMIs in Australian contexts have not yet been published.

Case studies

- The Agency for Healthcare Research and Quality (US) indicator is reported separately and as part of the State Snap Shot which expresses a composite comparative measure of performance.
- Dr Foster reports the LMDRG indicator alongside other mortality indicators including HSMR, SHMI and surgical mortality. Three out of 145 trusts had a higher than expected rate in the 2012 public report, while 26 had a higher than expected SHMI.

4.5.4 Summary of key findings for LMDRG

- There is a paucity of recent literature pertaining to LMDRG attributes or impact.
- The single article reporting application of LMDRG to an Australian context concludes there is a need for risk adjustment of the data. This would be in keeping with practice in the USA and UK.
- The relationship between LMDRG and quality of care remains uncertain.

4.6 Report presentation

The presentation and feedback of HSMRs is intended to disseminate findings about observed and expected deaths and encourage critical thinking about the meaning of variances. The presentation of HSMRs needs to be targeted at the audience in a manner in which it will be understood and useful, whether this is from a funder, health service or consumer perspective. The HSMR may be reported at a hospital level as a single value representing the HSMR for a certain timeframe, or used in comparison with other hospitals' ratios. Ben-Tovim et al (AIHW: Ben-Tovim D et al, 2009) describe three main methods used for presenting HSMRs when making comparisons between hospitals as being league tables, caterpillar plots and funnel plots. Table 14 describes these methods and their attributes. An example of a funnel plot (Figure 4) and an example of a caterpillar plot (Figure 5) are provided below.

Table 14. Summary of report presentation methods

Method	Definition	Attributes
League tables	Ranking of results from highest to lowest, may include confidence intervals to indicate the uncertainty / variability of the result.	• 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. • Improved by use of confidence intervals
Caterpillar plots	HSMRs and their confidence intervals are represented in a graphical plot, with the institutions ranked by HSMR along the x-axis and the HSMR values on the y-axis. The difference between institutions is important when lower confidence limit of the estimate for any 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.	• Useful for identifying outlying units or hospitals. • Lack sensitivity for differentiating between hospitals with different case loads
Funnel plots	A graphical method used to assess variation in data and used to compare different hospitals' HSMR over a single time period. Each hospital's HSMR is shown within the control limits, which form a funnel where the wide base of the funnel represents those hospitals with a lower number of separations and where the HSMR measure is less precise. Data points falling within the control limits are consistent with random or change variation and are said to display "common cause" variation. For data points falling outside the control limits, chance is an unlikely explanation.	• Avoids the ranking of hospitals from highest to lowest and takes into consideration variations in hospital size. • Useful for identifying outlying units or hospitals. • The chart is generally easy to interpret but less useful for directing individual review processes • Computation of funnel plots is quite complex

Figure 4. SMR funnel plot (ANZICS CORE, 2009)

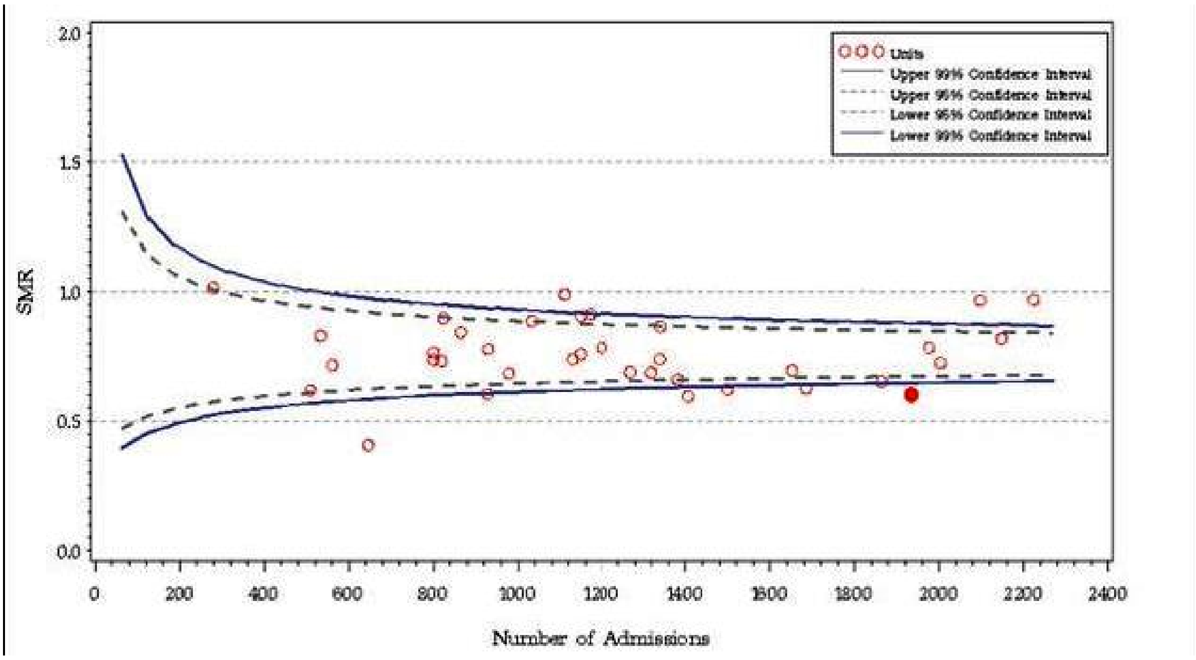
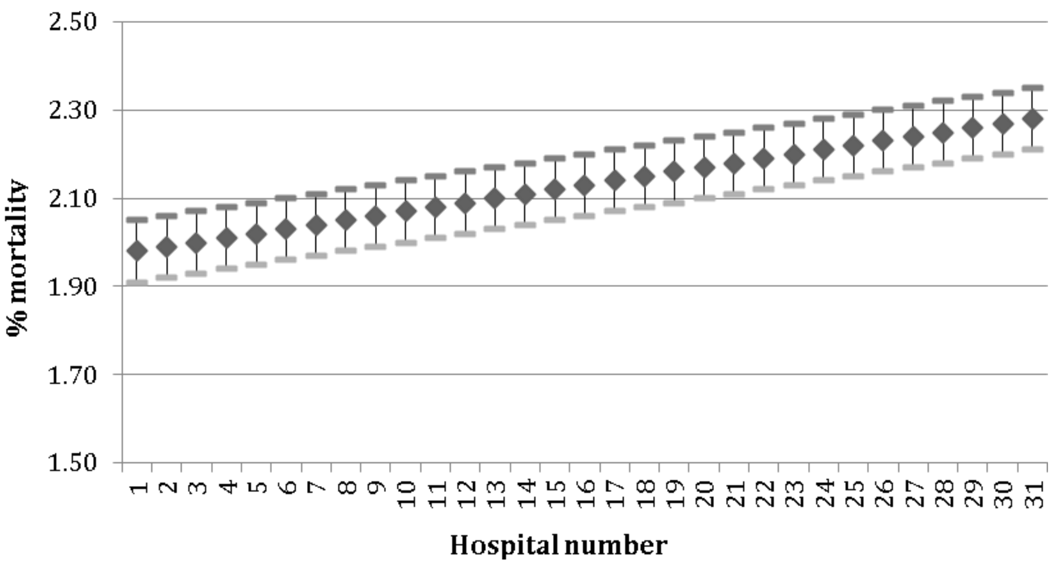


Figure 5. Hypothetical caterpillar plot example (courtesy of Megan Bohensky)



The majority of studies reviewed addressing the application of HSMRs were at a hospital level, rather than between hospitals. Two condition specific studies and one LMDRG study described the presentation and feedback of SMR as comparisons between hospitals (Aelvoet, W. et al, 2010, Barker, A.L. et al, 2011, Coory, M. et al, 2008).

Coory M, 2008, Australia

In this good quality cross-sectional study, the investigators used 30-day in-hospital mortality data to compare the performance of funnel plots with CUSUM control charts to identify variation between hospitals in Queensland for the financial years 2003-4 and 2004-5 (Coory, M. et al, 2008).

The data source was the Queensland Hospital Admitted Patients Data collection and mortality outcomes were risk adjusted using variables based on previous reported literature and evidence based relationship to AMI outcomes. Outlier status using Funnel plots were specified at the 2-sigma (95%) and 3-sigma (99.8%) confidence levels. For CUSUM plots pre-specified hypotheses for outliers were based on relative risk increased or decreased triggers of 30%, 50% and 75%.

Using the funnel plots, no hospital flagged at the 3-sigma level in either financial year. At the 2-sigma level there were 2 low outliers in 2003-4 and 1 high outlier in 2004-5. Using the CUSUM charts, 5/18 (28%) hospitals flagged an increase in relative risk of 75%.

The authors considered that the CUSUM control charts offered greater potential as a starting point for quality improvement as they detect problems early allowing more timely investigation. In contrast they considered that most of the variation in cross-sectional data formats such as funnel plots was due to statistical noise. The study highlights the need to tailor data analysis and presentation methods to the purpose for which the data is to be used.

A Belgian study by Aelvoet aimed to provide feedback to hospitals to enable improvement of care and feedback to the College of Physicians to identify hospitals with “strong” evidence of either superior or inferior quality (Aelvoet, W. et al, 2010). Feedback to hospitals was provided in a graphical format displaying the hospital’s SMR-AMI as the departure from the rate and trend of the other hospitals; and an anonymous and tabular representation of these departures including an indication of statistical evidence. Feedback was provided to the College of Physicians grouping hospitals as having an “average” SMR-AMI or as an “outlier”. Outliers were defined as having either a statistically significant high or low SMR-AMI warranting external audit. There was no commentary regarding how the feedback was received or used.

An Australian study by Barker et al examining deaths in low-mortality diagnosis related groups (LM-DRG), presented the risk-adjusted LM-DRG mortality in rank order for 122 Victorian hospitals (Barker, Anna L et al, 2011). Results were plotted with 95% confidence intervals demonstrating wide confidence intervals reflecting the low frequency of deaths in LM-DRGs and suggesting the indicator is likely to be insensitive to true variations in the quality of hospital care.

A number of hospital-based studies noted the importance of simple to understand representation of the HSMR and the benefits of reviewing a range of mortality data to flag the requirement for further investigation. The study by Sorinola et al used a mortality scorecard including both crude mortality, HSMR and departmental mortality reports to assist with identifying target areas for more in depth investigation through medical record review and case meetings (Sorinola, O.O. et al, 2012).

The Canadian Institute of Health Information e-portal provides a suite of data represented in many formats, as used by Popowich et al, including cumulative sum charts (CUSUM) (Popowich, J. et al, 2011). In this study, HSMRs were provided monthly to each site as an aggregate and also for surgical, medicine and ICU, enabling interpretation within the context of care. If the HSMR was inconsistent with raw data, further investigation in the form of medical record review was undertaken. The importance of timely data and being able to disaggregate the HSMR for specific diseases or areas assisted with better understanding of the raw data and contributors to a changing HSMR whether it is data quality or quality of care issues.

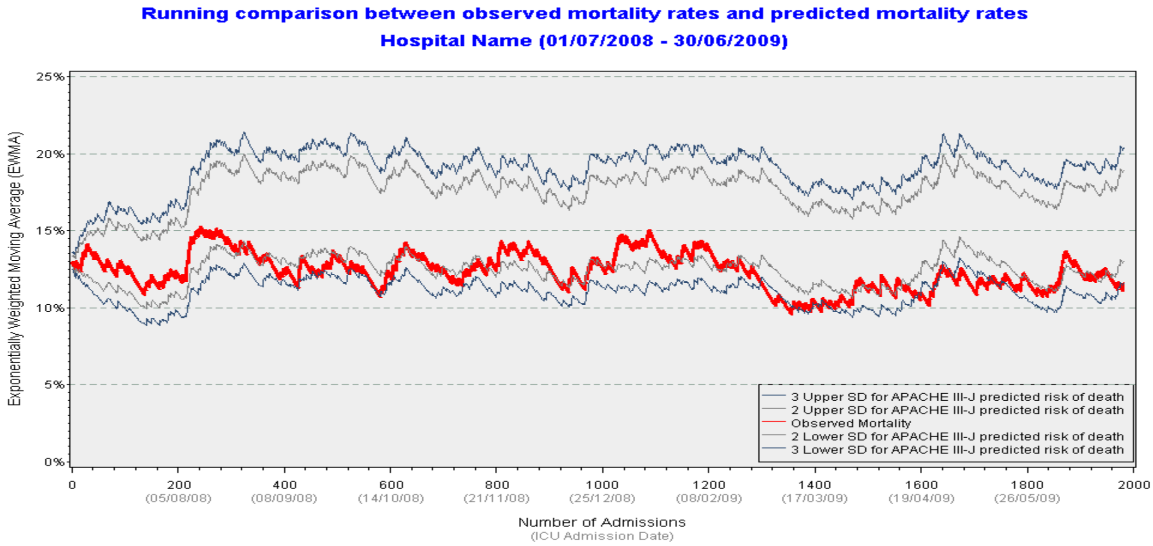
The study by Mohammed and Stevens (2013) demonstrates the value of disaggregating the HSMR, in this instance understanding the numerator and denominator that makes up the HSMR for both the health service overall and the individual hospitals (Mohammed, M.A. et al, 2013). The use of run charts to plot the observed and expected deaths identified the rise in the overall HSMR could be attributed to a fall in expected deaths at one hospital. Investigation revealed the fall in expected deaths was related to a change in coding depth.

HSMR are often calculated on an annual basis and as such their sensitivity to fluctuation over time and their usefulness for individual review in a timely fashion are limited. The HSMR is an aggregate figure further limiting their utility with respect to quality improvement. An alternative is to monitor fluctuations in mortality continuously over time using statistical control processes and time series charts in the form of cumulative sum (CUSUM), exponentially weighted moving average (EWMA) and variable life-adjusted display (VLAD) charts. The benefits and limitations of these charts are displayed in Table 15. Continuous monitoring, with alerts set for when control limits are breached, allows timely investigation and review of data and processes to reduce the risk of mortality.

An invited essay by Penfold et al recommends the use of the CUSUM as a better approach for monitoring in-hospital deaths over the HSMR (Penfold, R.B. et al, 2008). Penfold et al cite the main advantage of the CUSUM as allowing administrators to focus on a subset of patients within a narrow timeframe in which deaths are occurring at a higher than expected rate when alerted as per control limits.

Pilcher et al recommends the exponentially weighted moving average (EWMA) chart above the CUSUM, suggesting the EWMA is more "understandable" to a clinical audience (Pilcher, D.V. et al, 2010). The authors tested whether using a EWMA chart would have been useful in monitoring outcomes of patients admitted to an ICU unit between November 2000 and December 2005. Data submitted to the Australian and New Zealand Intensive Care Society Adult Patient Database (ANZICS) were used to develop a EWMA chart displayed in Figure 6. The EWMA chart showed periods when observed mortality appeared higher than predicted and would not have been detectable by analysing data with a standardised mortality ratio. The authors concluded continuous monitoring of outcomes using an EWMA chart might have advantages over other techniques. A limitation of the study was the quality of the data, which had been submitted at variable intervals and highlights the reliance upon accurate and timely data to monitor trends effectively.

Figure 6. Exponentially weighted moving average (EWMA) chart for an ICU for 2000-2005 (ANZICS CORE, 2009)



NOTE: Control limits for predicted mortality rates are derived using APACHE III-J; Lines shown are calculated using an exponentially weighted moving average with a weighting of 0.005.
Report provided by ANZICS Centre for Outcome and Resource Evaluation (www.anzics.com.au). Data Source:Data submitted to ANZICS CORE Adult Patient Database as at October 2, 2009.

Sequential admissions and patient number are plotted on the horizontal axis for the period from 1 July 2008 – 30 June 2009. Observed mortality (red line) and 95% (light grey) and 99% (dark grey) confidence limits for APACHE III-J predicted risk of death are plotted on the vertical axis.

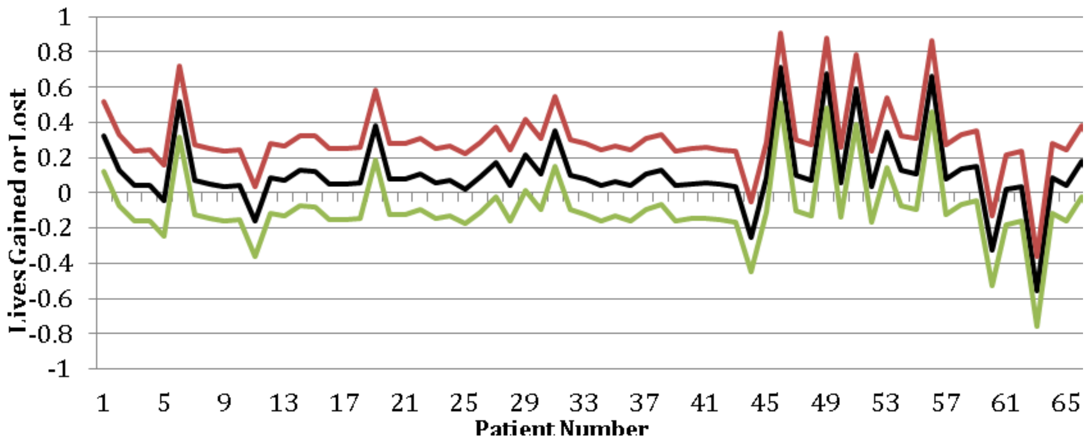
Table 15. Comparison of different types of time series display charts (courtesy of Megan Bohensky)

Chart	Benefits	Limitations
CUSUM (cumulative sum)	• As deaths occur in an organisation, the difference between the number of deaths and the bench mark rate is calculated, and this is cumulatively summed (CUSUM) up over time. If the mortality rate approximates the bench mark rate over the time period, measurements do not deviate significantly from control limit (i.e. higher and lower death rates averaged each other out) and the CUSUM value should vary narrowly around the bench mark level. If the mortality rate is abnormally high or low, measurements will progressively depart from the bench mark. • Has been widely used in healthcare, so is more well-known than other methods.	• Statistically complex, which may limit its interpretability and utility. • Setting control thresholds is not straightforward
EWMA (Exponentially-Weighted Moving Average) chart	• The EWMA chart is unique from the CUSUM and VLAD charts because it provides a weighting factor to deaths in a geometrically decreasing order. This means that the most recent deaths are weighted most highly (i.e. recent performance counts more highly), while the most distant deaths contribute very little (i.e. historical performance is less important). This makes the EWMA more sensitive to small shifts in performance. • Used by Australian organisations (e.g. Australian and New Zealand Intensive Care Society) for reviewing indicators. • Weighting factor allows greater adaptability of the chart to the data and stakeholders needs. • Can detect relatively small shifts in trend with	• Statistically complex, which may limit its interpretability and utility. • If large changes in mortality performance occur, the weighting factor needs to be increased to accurately detect it.

Chart	Benefits	Limitations
	greater emphasis on more recent observations.	
VLAD (Variable Life Adjusted Display) chart	• The VLAD calculates a sum of the number of deaths in an organisation mines the expected number of deaths based on the risk-adjusted prediction score. When a patient dies, the plot decreases by their predicted probability of survival (1-predicted probability of death) and when a patient survives, the chart increases by the predicted probability of their death. The resulting line is the number of net lives saved, if trending upwards, or lives lost, if the line trends down. • Accounts for casemix in a straightforward way, so they may be more readily understood among non-technical audiences. • Used within Australian jurisdictional health departments (e.g. QLD Health), so adoption of this method would allow for greater consistency across health departments.	• Setting control thresholds is not straightforward. • Not originally intended for hypothesis testing but modifications can allow this use • Detection of alarms may not be as sensitive in smaller organisations.

Two Australian studies reported the use of variable life-adjusted displays (VLADs) (Clarke, A.L.L. et al, 2010, Scott, I.A. et al, 2008). Scott et al used an adaption of the graphical variable life-adjusted display (VLAD) to compare two risk prediction models using administrative and clinical data (Scott, I.A. et al, 2008). The authors found the two prediction models generated almost identical VLAD curves. The VLAD using administrative data is displayed in Figure 7 as an example of VLAD charts.

Figure 7. Variable life-adjusted display derived from the administrative model of risk prediction (hypothetical)



The black line shows the cumulative risk-adjusted difference between observed and expected deaths in a hypothetical sample of patients (n = 67) from the administrative risk-prediction model. Red and green lines represent upper and lower control limits, respectively. Lives gained are shown by positive values on the x axis and lives lost by negative values.

The authors support the use of VLADs, to detect runs of favourable or adverse outcomes, as being feasible on a large scale and at a low cost when using readily available administrative data. Accuracy of administrative data is often questioned and the early experience of Queensland Health has shown many flags indicating that review is required have related to data coding problems and inadequate risk adjustment. The authors suggest the accuracy of risk adjustment of administrative data may be improved by adding a small number of readily available laboratory results or clinical variables to administrative datasets (Scott, I.A. et al, 2008).

Scott et al identified VLADs are useful for monitoring outcomes over time within a single institution but cautioned against their use for inter-hospital comparisons. Noting that VLADs do not provide definitive proof of, or explanations for, lower quality care but rather are useful as screening tools for stimulating further investigation by clinicians of both data quality and process of care issues.

The focus of the narrative by Clarke et al was upon clinician engagement in testing quality of care indicators by the Victorian Department of Health (Clarke, A.L.L. et al, 2010). Readily available routinely collected administrative data was used to measure clinical performance and were presented using VLADs. The authors concluded they had experienced significant challenges in engaging with clinicians due to concerns about the quality of the data and resource requirements for investigating variances.

Case studies

- The case studies illustrate the variety of presentation and reporting approaches described in the literature including funnel plots, caterpillar charts etc. Examples and links are included in the detailed case studies.
- It is common for jurisdictions and private providers of data to provide health service level access and public access, with data presented appropriate to the audiences.
- In the US where much of the quality improvement work is decentralised and managed through private providers, the software made available through AHRQ enables health services and health service groups to use and report their data for their particular purpose, including performance related payments. Users also have the ability to change the risk adjustment and reporting to suit their needs.
- In the UK, where transparency and accountability have been particular issues flowing from recent investigations, there is a high level of public access to data. In fact a supplementary report on persistent outliers is available, covering the past five SHMI publications and including contextual indicators recalculated for the outliers and any known issues identified in the past five SHMI publications.
- Frequency of reporting is commonly annually, however the new SHMI indicator in England is reported quarterly.
- With private providers there is a trend towards providing “real time” data to support organisations in addressing potential quality issues in a timely manner. This includes the provision of “alerts” when data approaches or exceeds expected limits.
- Many jurisdictions have sophisticated web-based public reporting with interactive features that enable searching and reporting for regions and individual hospitals for various indicators. A common approach is to colour-code the performance in relation to the average. Some (Dr Foster (UK), ACHQ (US)) also aggregate various indicators to provide an overall performance report, although the technical basis of the aggregate scores are not clear.
- Jurisdictions also include mortality indicators in “scorecard” reports, providing a context for the mortality indicators and aiding interpretation.
- The hospital reporting is also increasingly sophisticated. Canada’s Canadian Hospital Reporting Project (CHRP) analytical highlights include reports that identify notable trends, variations and comparisons across hospitals, peer groups and jurisdictions over time. The aim of the reports is stated to help foster peer learning by identifying other hospitals to learn from in a selected group of indicators.

- In the recent Dr Foster public report, four mortality indicators are reported in a “wheel” type chart which shows the performance of all trusts (rate as expected, above expected or below expected). The indictors include HSMR, SHMI, surgical mortality and death in LMDRG. The Dr Foster report also includes a chart combining efficiency and mortality scores for trusts.
- Given the complexity of the data, many jurisdictions endeavour to routinely communicate key findings through the media and through summary reports. The Canadian program seems to do this well.

5. Discussion

The issues that are being addressed in relation to generating mortality indicators are similar for general HMIs, for condition specific HMIs and for the LMDRG indicator. The focus remains on definition of the indicators and risk adjustment and statistical methods for model development. There is a lesser focus on implementation and utilisation. Interest in the technical issues related to HMI generation, discussed within the peer reviewed literature, are reflected in similar interests manifest in the grey literature and, in particular, the UK and Canadian case studies. There is greater emphasis identified within case studies on implementation of the HMIs, especially at jurisdictional levels with an emerging discussion about how individual hospitals might use the data. Overall, there remains a remarkable gap in robust investigation of the relationship between use of HMIs and improvement in quality of care in peer reviewed and grey literature sources.

Model generation and testing

The definition of mortality remains a frequent focus of discussion, due to concerns about changing mortality rates, particularly improvements, potentially reflecting hospital discharge policies rather than reflecting changes in quality of care. The key barrier to using 30-day post discharge mortality across jurisdictions, including Australia, is lack of timely access to linked hospital episode and complete population based death data that includes cause of death. For instance, in the recent analysis of NSW data, it was noted that there is a 2 year delay between death and registration and coding of cause of death within the ABS dataset (Lujic, S. et al, 2012).

In general, peer reviewed studies and grey literature information have documented good to high correlations between in-hospital mortality (traditional HSMR) and other in-hospital definitions (30-day post admission mortality) and 30-day post-discharge mortality, with standardised in-hospital mortality rates being higher than rates calculated using other definitions (Bhat, S.K. et al, 2013, Bottle, A. et al, 2011). There are exceptions for some condition specific HMIs with poorer correlation for stroke and fractured neck of femur (Borzecki, A.M. et al, 2010).

However, if the reason for measuring standardised mortality rates is to compare differences in rates across peer groups or within a hospital over time, then the more pertinent question relates to whether changing the definition of mortality changes results in relation to outlier status. It has been noted that using 30-day mortality definitions, there is less variation documented between hospitals (Drye, E.E. et al, 2012), and that confining the definition to in-hospital death can favour hospitals with shorter LOS. In this review a number of studies were identified that reported changes in hospital performance categories based on the definition of mortality. While such changes have been described by some authors as “small to medium” (Bottle, A. et al, 2011), the impact on a hospital identified as a high outlier, when such data is publicly reported can be significant due interpretation of outlier status as “poor performance”. Even if the data is used as a ‘flag’ for potential quality sensitive issues, false positive flags may require substantial investigational resource. The Australian NSW report documented a change in performance category for nearly 30% hospitals when different general hospital mortality definitions were used (Lujic, S. et al, 2012). Similarly for condition-specific HMIs, major shifts in hospital rankings when comparing in-hospital mortality to other 30-day mortality measures have been reported, especially for hip fracture (Borzecki, A.M. et al, 2010, Kristoffersen, D.T. et al, 2012) and Drye et al. report performance category changes for hospitals for 8.2% AMI, 10.8% CHF and 14.7% pneumonia (Drye, E.E. et al, 2012). We identified no studies or reports that have investigated the burden on hospitals of investigating outlier flags.

Overall, the case studies indicate an increasing trend towards use of 30-day mortality measures, a notable example being the UK where, despite the development of the in-hospital HSMR, there has now recently been introduction by the NHS of a new general HMI, the SHMI in which mortality is defined as 30-day post discharge (Campbell, M. J. et al, 2012).

Patient population selection and risk adjustment, particularly for general HMIs is another issue that is debated within the peer reviewed literature and is being examined across jurisdictions. The key population characteristics that are discussed are; completeness of included deaths, palliative care status and how to manage transfers and people who have multiple admissions. Other issues in choice of risk adjustment variables relate to bias due to case-mix and constant risk fallacy related to differential depth and quality of coding of diagnoses and impacts of changing hospital admission policies on ‘urgency’ of admission.

There are wide variations in determining what population (wider and narrower definitions) should be included in the general HMI denominator. At the level of hospital, specialist women’s and children’s hospitals and, in some cases specialist cancer hospitals, are not included in HMI calculations. It is generally agreed that there needs to be some form of aggregation of diagnostic groups that form the basis for developing risk adjustment coefficients and calculation of the general HMI. The most common method, exemplified by the Dr Foster HSMR, is inclusion of the diagnostic groups that contribute 80% of deaths. However, there is discussion about using a more inclusive approach, Australia being the only example identified where the top 80% diagnoses and the lowest 20% diagnoses have been investigated (AIHW: Ben-Tovim D et al, 2009), with the lowest 20% being associated with greater variation between hospitals. In the recently developed SHMI, the NHS has reviewed these selection criteria and for the SHMI has decided on a more inclusive approach, including all diagnoses.

Other factors under discussion that are related to population inclusion and exclusion and risk adjustment include; palliative care status, hospital patient transfers, multiple admissions and zero days stay patients. For other variables, such as age there is consensus about inclusion but on(AIHW, 2011c)going investigation about how the variable is incorporated; as a categorical or continuous variable . Of interest for Australia, the death in LMDRG indicator is not risk adjusted although a recent Victorian based analysis of this indicator found that risk adjustment significantly influenced outlier status (Barker, A.L. et al, 2011).

The inclusion or exclusion of and risk adjustment for palliative care patients remains unresolved, although it is clear that adjustment for palliative care can impact on calculated HSMR scores (Bottle, A. et al, 2011). The key issues to consider are, firstly how to identify patients whose death in hospital is part of end of life care and, secondly how to account for such deaths in a way that is not associated with perverse coding practice changes.

Several studies and reports highlight the issues associated with operationalising a definition of palliative care status (AIHW, 2011a, Cassel, J.B. et al, 2010). Increasingly, palliative care includes, particularly for those with chronic conditions, not only end of life care but also supportive care. The Australian report notes that, from a coding perspective, there are various ways in which palliative care services are documented; as the coded care type, by participation in a palliative care program or by services provided by a palliative care physician. Further, palliative care can be coded as a secondary or additional diagnosis (Z51.5). There is no coding of admissions for ‘end of life’ care in Australia. In keeping with reporting of variation between UK trusts in coding of palliative care (Flowers, J. et al, 2010), the Australian report noted variation in utilisation of palliative service coding across states but relative stability over three time points of study between 2006 and 2009. Kroch et al explored inclusion of ‘DNR’ orders and palliative care status in a medical record review and found that most patients with DNR orders survived the admission whilst most with palliative care status did not. It is likely that different international jurisdictions will need to consider ways of managing end of life care patient populations differently according to contextual differences in application of palliative care and end of life care definitions and coding practice.

Further, there is evidence to support the premise that changes in trends in standardised mortality rates over time may be significantly influenced by changes in coding practice for palliative care (Chong, C.A.K.Y. et al, 2012). This may be a particular problem when HMIs are first introduced and reported, especially with public reporting as hospitals recognise deficiencies in coding for such patient populations. Stability over time might be expected, however monitoring for perverse coding practices would need to be considered, especially where palliative care end of life status changes after the first few days of admission. An alternative approach, and one that is posed by the NHS, is a broader patient population inclusion approach whereby palliative care patients are included in the denominator but a 'companion' indicator for palliative care rates is provided with SHMI scores for interpretive support.

There is considerable variability in the way in which transfer status is assigned within different models including; no adjustment, inclusion in transfers in, inclusion of transfers out, and deaths scored equally across hospitals or to the hospital transferring out. In Australia, use of linked data may provide more accuracy in identifying transfers (Bhat, S.K. et al, 2013). AIHW reports recognize that the issue of transfers requires further investigation (AIHW, 2011a). In a draft report undertaken by the AIHW in June 2011 (AIHW, 2011b), it is noted that current specifications include adjustment only for transfers in, yet transfers out make up a larger proportion (15%) of separations than transfers in (6%). If transfers out were to be excluded this would be associated with 15% decrease in the population available for HSMR calculation, and if transfers in were excluded, an 11% difference. There would be even greater impact on populations included for condition specific HMIs with an increase of 3-fold in the included population if transfers out were included in the denominator for fractured neck of femur.

Studies from the Netherlands have reported a reduction in mortality per admission for patients admitted more frequently and that a longer 'look back' period can improve the accuracy of readmission identification (van den Bosch, W.F. et al, 2011, van den Bosch, W.F. et al, 2012), whilst UK investigators reported a significant impact of using the patient's last admission in calculation of deaths compared to using the first admission (Bottle, A. et al, 2011). Overall, there has not been a great deal of investigation of this aspect of defining patient populations and probably further investigation is warranted.

The debate of issues associated with comorbidity adjustment continues. Which comorbidities or comorbidity scores to use, how many and should all or only those present on admission be included for risk adjustment? There are many differences in the choice of included comorbidities across HMI models and often the application of these is poorly described. The 'Charlson score' is most commonly used, but often with no information about the way in which it was applied. Whilst in-hospital outcomes are better predicted by inclusion of all comorbidities (Dalton, J.E. et al, 2013), in-hospital diagnoses may indicate a potential quality issue. Therefore, there is increasing interest in using diagnoses present only on admission, an opportunity accessible in Australia where the diagnostic timing code "the C-code" differentiates in-hospital from on admission diagnoses. It remains uncertain as to the quality of documentation of these codes. Where there is no access to diagnosis timing codes, some authors have chosen to use 'chronic condition' comorbidities. Depth of coding has been reported to be associated with the 'constant risk fallacy', described by Mohammed as "if risk relations are assumed to be constant, but in fact are not, then case mix adjustment may be more misleading than crude comparisons" (Mohammed, M.A. et al, 2009).In simple terms, the AIHW have described this as "the constant effect assumption of each variable"(AIHW, 2011a). Mohammed et al (2009) demonstrated variation in depth of coding between hospitals and over time suggesting variable coding practice. Bottle et al (2011) also note some variability in impact on HSMR calculations but note that the impacts are small, and similar findings were reported in an Australian analysis (AIHW, 2011a). In the Australian context, the report noted the reliance on hospitals identifying and coding secondary conditions and that while the quality of coded data in the National Hospital Mortality Database (NHMD) is subject to auditing, that there is no national standard for auditing at the time of report writing (March 2011) and that diagnosis data is subject to state-specific coding standards. A previous report had identified variation in reporting of additional diagnoses (AIHW, 2010). The more recent report tables evidence for relative stability of HSMRs for Australian states over three time periods from 2006-7, 2007-8 and 2008-9 suggesting a constant effect was operating for comorbidity (using the Charlson Comorbidity Index) within the risk adjustment model.

Similar concerns about the constant risk fallacy have been raised in relation to use of the urgency of admission status. Emergency admission is included in many HMI models, including Dr Foster, the Canadian model and the ACSQHC specifications. Mohammed et al demonstrated clear interactions between comorbidity status and 'emergency' status. Emergency status is associated with an inherent increase in mortality risk, however changing admission policies such as those associated with reducing ED waiting times could potentially change that inherent risk if more patients with lower severity conditions are being admitted.

Other areas of interest related to generating HMI models relate to choice of statistical methods, such as use of direct or indirect standardisation, use of fixed or random effects models and choice of reference data for generating risk-adjustment coefficients. It is suggested that the choice of statistical method/s be chosen to reflect the purpose for which data is used. Indirect standardisation is the most commonly employed method of standardising HMIs. Therefore, HMIs based on indirect standardisation are useful for comparisons with the national average and allow for inclusion of categories with small numbers but may be less useful for comparing between groups due to reported paradoxical results, particularly if there are interactions between variables such as case-mix and urgency status. Stated another way, "indirectly standardised HSMR provides the mortality rate from a societal perspective as it is based on the population the hospital actually serves, not the national reference population, while a HSMR based on direct standardisation is more relevant to informing patient choice" (Heijink, R. et al, 2008, Pouw, M.E. et al, 2013)

There is also debate about the relative value of using hierarchical estimation models to account for clustering of variables at different levels of the health system. Both methods fixed or random have their benefits and limitations and again it is suggested that purpose should drive choice of method. For instance, Kipnis et al (2010) suggest use of random effects models to identify outlier status and aggregate-level fixed effects models to examine mortality rates over time.

A further AIHW report (AIHW, 2011d) investigated use of in-hospital mortality risk-adjustment coefficients and also suggested choice of methods being directed by data use intentions. The authors suggest that, if coefficients remain the same (that is, based on the same data reference period), then this will support more accurate comparisons of mortality rates over time for individual hospitals. In contrast, if coefficients are updated regularly (such as each year or more frequently) then this will more accurately compare amongst hospitals within that time period.

A development in the literature over the last 5 years has been the increase in number of studies in which address HMI model comparisons and testing or adaptation within patient populations other than those in which they were developed. Thus, there has been development of HSMR models in Japan (Miyata, H. et al, 2008), Brazil (Gomes, A.S. et al, 2010) and the Netherlands (Jarman, B. et al, 1999). The adapted models generally exhibit high discrimination performance attributes as measured using the c-statistic. In all examples, there were adaptations made from the original model methods to allow for contextual differences or perspectives. This need for adaptation, discussed by Ben Tovim et al (AIHW: Ben-Tovim D et al, 2009), in relation to the Australian HSMR development, may limit the ability to compare across jurisdictions.

In addition differences between models may limit comparisons between hospitals within a jurisdiction. Cassel et al (2010) report the differences when the same data is provided to 4 different entities in the USA, largely related to differences in the way patient populations were defined for inclusion in HSMR calculation. Similarly Shahian et al (2010) demonstrated marked variation in hospital performance categorisation when 4 different methods were applied to the same administrative dataset. If a number of different commercial entities are involved in providing HSMR measurement services, it will be important not only to provide standard specifications for HMIs but also to potentially audit the way in which these specifications are applied if across hospital comparisons are to be made at jurisdictional or national levels in Australia.

Utility and validity of HMIs

This review identified few studies that specifically addressed issues of utility and validity, although many studies refer to identified variation of HMIs across hospitals or jurisdictions as demonstrating potential utility for flagging potential quality of care issues. The ongoing debate, largely confined to use of general HMIs and the indicator death in LMDRG is at times quite emotive.

Critics of the measures point to the technical shortcomings of the measures, lack of evidence for a relationship between reported variation in HMI scores and quality of care and adverse impacts of unfair blame and shame associated with public reporting of false positive outlier status. Proponents of the measures point to the case studies (such as Mid Staffordshire in the UK) where outlier status was subsequently followed by very negative review findings, defend the technical attributes of the measures, concede the possibility that there is unmeasured variation not related to quality of care but reinforce the role of such measures as 'flags' for further investigation.

There is merit on both sides and a surprising lack of investigation focused on addressing the two key questions. Firstly, does variation between hospitals/areas in HMI scores indicate better or poorer quality of care and secondly, does implementation of a HMI monitoring and reporting system improve quality of care? Designing studies to answer these questions are not easy to perform as there is no gold standard against which to measure quality of care and cluster randomised controlled trials testing the efficacy and effectiveness of HMI programs would be difficult to undertake.

With regard to investigating the relationship between variation in HMI scores and quality of care, investigators have variously chosen to measure /benchmark quality of care, using explicit or implicit methods, against process or other outcome measures or aggregate measures based on self-report or external appraisal. Not surprisingly, given the heterogeneous nature of study objectives and methods Pitches et al (2008) in a systematic review reported a near fifty-fifty chance that there was a positive correlation between risk-adjusted mortality and better quality of care. They concluded that the "general notion that hospitals with higher risk-adjusted mortality have poorer quality of care is neither consistent nor reliable". Two of three newer studies included in this review that investigated the relationship between general HMis and quality of care reported no association (Groene, O. et al, 2011, Kernisan, L.P. et al, 2009). However both were of limited methodological quality. The third and most interesting study was the mathematical modelling exercise undertaken by Girling et al, in which the authors formed the hypothesis that, unless preventable mortality is 15%, hospitals would be unlikely to be correctly classified as high or low outliers. They were limited in their analysis by a lack of data in the literature about the proportion of preventable deaths, however a recent UK study reported 5.2% deaths in a sample of 1000 were potentially preventable. There are major issues associated with subjectivity of assessing preventability and low reported inter-rater reliability (IRR) for such assessments, the IRR in this study being moderate at k=0.49 (Hogan, H. et al, 2012), however the results warrant further investigation. Although it has been strongly proposed that HSMRs do not measure avoidable or preventable death, but simply identify variation that needs to be investigated when statistically 'extreme', unless the residual variation reflects quality of care issues, then 'further investigation' will utilise constrained quality resources that could have been allocated elsewhere.

For condition specific HMIs, one study reported differences in standardised AMI HMI according to conformity to guidelines for re-perfusion although this was poorly described (Aelvoet, W. et al, 2010). A further study (Bradley, E.H. et al, 2012) reported an association between AMI HMIs and hospital management strategies for reducing mortality although the results need to be reviewed with caution as they were based on a self-reported survey by a single organisational individual. The strategies, such as monthly review meetings and having cardiologists, having physician champions alone or with nurse champions but not nurse champions alone, should drive discussions about the capacity of current service models to support high quality care, particularly in view of other evidence supporting a clear association between condition volume and care outcomes (Carretta, H.J. et al, 2013).

Standardised HMIs can be implemented at a number of levels, including jurisdictional and individual hospital. Case studies have focused on the implementation at the jurisdictional level. Some of the key findings include; ongoing investigation of technical issues relating to adaptation of models to allow for local contextual issues which has been discussed in some detail earlier, using HMIs to measure trends over time and setting up of systems for reporting HMI scores back to hospitals and to the general public.

There is increasing interest in reporting trends in HMIs over time, particularly within grey literature. Of interest, a recent peer reviewed paper (Jacques, R. M. et al, 2013) was identified (post-dating the search timeframes for this review). The authors noted that the Dr Foster HSMR fell by 42% from 115 in2002 to 67 in 2011. In order to review mortality gains and validate the Dr Foster HSMR, they used the more recently developed SHMI by analysing NHS data between April 2005 and September 2009. Over the time period they documented an 8% increase in total number of admissions per year and a 5% fall in deaths in-hospital or up to 30-days post-discharge. Between 2005-6 and 2010-11, there was a reduction of 24% (from 112 to 85) in SHMI, with reductions in observed deaths and increases in the denominator. Excluding comorbidity in the model was associated with an 18% falls in SHMI. The estimated quarterly reduction in SHMI varied widely between hospitals, however using a funnel plot all were within 99.9% confidence intervals except Mid Staffordshire which had a large quarterly decrease. The authors reviewed observed deaths up to 30-days post-discharge and noted a 15% decrease, commenting that the reasons could be multi-fold including better hospital care, more community deaths and improved population health. There was also an increased 15% expected deaths that could be related to a small increase in age of patients (50.5y to 51.6y), increasing numbers admitted as emergencies (75% to 78%) and a large increase in recording of comorbidities (26% to 35%). In association with these data, population mortality rates have improved by 10% over the same period. The authors conclude that the falling SHMI reflects a number of factors, including improvements in care which is in keeping with reported improved mortality associated with AMI, stroke and fractured neck of femur and reductions in prevalence of clostridium difficile and MRSA. It was not possible to perform a direct comparison with Dr Foster HSMR as annual recalibration would mask changes over time. This paper provides good insight into the many issues associated with interpreting trends in general HMIs over time, particularly if the only data available is regularly recalibrated. It also draws attention to different data sources, some of which are high quality such as clinical registry data and well conducted clinical audits which can provide accurate information about system performance over time.

In another jurisdictional level study, undertaken in Victoria, a pilot implementation of 11 AHRQ adapted indicators (AUS-PSIs) (Clarke, A.L.L. et al, 2010), using the Queensland VLAD methods was described albeit in limited detail. Key findings were that hospitals reported difficulties engaging clinicians in the process of HMI review for a number of reasons including lack of trust in the data, lack of timeliness of the data and "a higher than anticipated level of strain on resources". The strain was not quantified. The lack of clinician engagement underscores the need for high quality measures. In a study of use of publicly available mortality data, primary care physicians reported that these measures were not at all important for 42% respondents with 76% agreeing that they were an incomplete indication of the quality of care (Morsi, E. et al, 2012). Whilst this study introduces a raft of different issues related to use of publicly reported data beyond the scope of this review, it serves to highlight potential issues associated with implementing such systems, and the difficulty of engaging all relevant stakeholders in the process.

At the level of individual hospitals, others are setting up systems to respond to HMI data (Popowich, J. et al, 2011), for instance, using general HMI data in conjunction with associated data (raw mortality data) and medical record reviews (using IHI Trigger tool methods (Institute for Healthcare Improvement (IHI))) within a 'stepped' investigational process to reduce the burden of unnecessary record review. "Companion" indicator review is now increasingly recommended. This reflects concerns relating to the limitations of using general HMIs alone, in being able to respond and 'action' identified variation over time or between groups. Using associated indicators, including crude mortality rates, palliative care rates and other condition specific HMIs may help increase the utility of measuring HMIs, however this hypothesis has not been investigated in any systematic manner. For instance, there is no evidence to support the utility of condition specific HMIs with or without general HMI scores.

The way in which data is presented may be an important determinant for hospitals wanting to use such data. Only one study (Australian) was identified during the period of the literature search that specifically addressed this issue (Coory, M. et al, 2008). The study, which was well conducted again highlighted the need for purpose to drive process. For hospitals to respond to HMI measures the data must be timely and actionable. Dr Foster makes the point that HSMR is different to CUSUM chart driven mortality monitoring (Dr Foster Intelligence, 2011). Or alert system. Indeed, the former is used to assess mortality rates relative to others over a designated time period whereas the latter is based on analysis of internal performance in real time. In the only study we identified that addressed the relative performance of different methods of data analysis and presentation, the control chart CUSUM method performed better than cross-sectional funnel plot presentation for identifying outlier variation for standardised AMI HMI (Coory, M. et al, 2008).

Overall, there are few who question the need for deaths to be scrutinised as part of a hospital's ongoing quality monitoring activities. For instance, there is increasing attention being paid to individual screening and structured review of deaths in hospitals (Brand, C.A. et al, 2012), that is mortality audit. However, there is little attention being paid as to how the various methods of measuring and assessing the quality of deaths can best be structured to provide the most effective and efficient mortality review program for an individual hospital, for a health group or Trust or for jurisdictions.

In summary, this review has built on previous literature to provide an updated picture of the work being undertaken to elucidate the composition, utilisation and utility of general HMIs, condition specific HMIs and the death in LMDRG indicator. There is ongoing interest in improving technical specifications of standardised HMI models and to date a lesser focus on investigating implementation and utilisation issues. There is a strong move towards models that include 30-day linked death data and towards use of companion indicators to support interpretation of identified variation. There is a need to tailor data analysis and presentation to meet the purpose for which the data is being used and this may require data providers and analysers to generate a number of different models and presentation formats for their potential customers (jurisdictions, hospitals, clinicians, patients and their carers, community). Overall, there remains a significant lack of information available about the resource implications and allocative effectiveness of these programs, and relative added value to existing high quality patient safety performance data that requires further study.

6. References

Aelvoet, W., Terryn, N., Molenberghs, G., De Backer, G., Vrints, C. and van Sprundel, M. (2010) 'Do inter-hospital comparisons of in-hospital, acute myocardial infarction case-fatality rates serve the purpose of fostering quality improvement? An evaluative study', *BMC Health Serv Res,* 10, 334.

AIHW (2010) *Hospital statistics 2008-2009.* , Health services series no. 34 Cat. No. HSE 84 Canberra AIHW.

AIHW (2011a) *Draft report on Hospital Standardised Mortality Ratio indicator: Preliminary investigation to identify possible artefactual causes of non-random variation*.

AIHW (2011b) *Hospital transfers in Core, hospital-based outcome indicators: Draft Report for the Australian Commisssion on Safety and Quality in Health Care*

AIHW (2011c) *Treatment of age in Core, hospital-based outcome indicators: Draft report for the Australian Commission on Safety and Quality in Health Care*.

AIHW (2011d) *Use if in-hospital mortality risk-adjustment coefficients: Report for the Australian Commission on Safety and Quality in Health Care*.

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

ANZICS CORE (2009) *ANZICS CORE Adult Database*.

Australian Commission on Safety and Quality in Health Care (2010) *Mortality and Unplanned Admissions Indicator Project*, ACSQHC, Sydney.

Aylin, P., Bottle, A. and Majeed, A. (2007) 'Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models', *BMJ,* 334(7602), 1044.

Barker, A. L., Brand, C. A., Evans, S. M., Cameron, P. and Jolley, D. J. (2011) '"Death in low-mortality diagnosis groups": frequency and the impact of patient and hospital characteristics', *Medical Journal of Australia,* 195(2), 89-94.

Barker, A. L., Brand, C. A., Evans, S. M., Cameron, P. A. and Jolley, D. J. (2011) '"Death in low-mortality diagnosis-related groups": frequency, and the impact of patient and hospital characteristics', *Med J Aust,* 195(2), 89-94.

Bhat, S. K. and Malla, S. (2013) 'Validation of Jarman's method of calculation of hospital standardised mortality ratios', *Aust Health Rev,* 37(2), 147-51.

Borzecki, A. M., Christiansen, C. L., Chew, P., Loveland, S. and Rosen, A. K. (2010) 'Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions', *Med Care,* 48(12), 1117-21.

Bottle, A., Jarman, B. and Aylin, P. (2011) 'Hospital standardized mortality ratios: sensitivity analyses on the impact of coding', *Health Serv Res,* 46(6pt1), 1741-61.

Bradley, E. H., Curry, L. A., Spatz, E. S., Herrin, J., Cherlin, E. J., Curtis, J. P., Thompson, J. W., Ting, H. H., Wang, Y. and Krumholz, H. M. (2012) 'Hospital strategies for reducing risk-standarized mortality rates in acute myocardial infarction', *Annals of Internal Medicine,* 156(9), 618-221.

Brand, C. A., Landgren, F., McKenzie, L., Jones, C., Irving, L. and Judson, R. (2012) *Mortality and Morbidity Review in Victoria: Literature Review and Benchmarking Survey. Commissioned by Victorian Management Insurance Agency (VMIA)*.

Campbell, M. J., Jacques, R. M., Fotheringham, J., Maheswaran, R. and Nicholl, J. (2012) 'Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years', *BMJ,* 344, e1001.

Caplan, R. A., Posner, K. L. and Cheney, F. W. (1991) 'Effect of outcome on physician judgments of appropriateness of care', *JAMA,* 265(15), 1957-60.

Carretta, H. J., Chukmaitov, A., Tang, A. and Shin, J. (2013) 'Examination of hospital characteristics and patient quality outcomes using four inpatient quality indicators and 30-day all-cause mortality', *American Journal of Medical Quality,* 28(1), 46-55.

Cassel, J. B., Jones, A. B., Meier, D. E., Smith, T. J., Spragens, L. H. and Weissman, D. (2010) 'Hospital mortality rates: how is palliative care taken into account', *Journal of Pain and Symptom Management,* 40(6), 914-924.

Chong, C. A. K. Y., Nguyen, G. C. and Wilcox, M. E. (2012) 'Trends in Canadian hospital standardised mortality ratios and palliative care coding 2004-2010 a retrospective database analysis', *BMJ Open,* 2(e001729), doi:10.1136/bmjopen-2012-001729.

Clarke, A. L. L., Shearer, W., McMillan, A. J. and Ireland, P. D. (2010) 'Investigating apparent variation in quality of care: the critical role of clinician engagement', *Medical Journal of Australia,* 193(8), S111-S113.

Coory, M., Duckett, S. and Sketcher-Baker, K. (2008) 'Using control charts to monitor quality of hospital care with administrative data', *Int J Qual Health Care,* 20(1), 31-9.

Dalton, J. E., Glance, L. G., Mascha, E. J., Ehrlinger, J., Chamoun, N. and Sessler, D. I. (2013) 'Impact of Present-on-admission Indicators on Risk-adjusted Hospital Mortality Measurement', *Anesthesiology,* 118(6), 1298-1306.

Dr Foster Intelligence (2011) 'Understanding HSMRs: A Toolkit on Hospital Standardised Mortality Ratios', [online], available: http://drfosterintelligence.co.uk/wp-content/uploads/2011/10/HSMR-Toolkit-Version-6-October-2011.pdf [accessed 5 August 2011].

Drye, E. E., Normand, S. L., Wang, Y., Ross, J., Schreiner, G. C., Han, L., Rapp, M. and Krumholz, H. M. (2012) 'Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: an observational study with implications for hospital profiling', *Ann Intern Med,* 156(1 Pt 1), 19-26.

Flowers, J., Abbas, J., Ferguson, B. and Jacobsen, B. (2010) 'Dying to know: how to interpret and investigate hospital mortality measures.', [online], available: http://www.apho.org.uk/resource/view.aspx?RID=95780 [accessed

Girling, A. J., Hofer, T. P., Wu, J., Chilton, P. J., Nicholl, J. P., Mohammed, M. A. and Lilford, R. J. (2012) 'Case-mix adjusted hospital mortality is a poor proxy for preventable mortality: a modelling study', *BMJ Qual Saf,* 21(12), 1052-6.

Gomes, A. S., Kluck, M. M., Riboldi, J. and Fachel, J. M. (2010) 'Mortality prediction model using data from the Hospital Information System', *Rev Saude Publica,* 44(5), 934-41.

Groene, O., Mora, N., Thompson, A., Saez, M., Casas, M. and Sunol, R. (2011) 'Is the maturity of hospitals' quality improvement systems associated with measures of quality and patient safety?', *BMC Health Serv Res,* 11, 344.

Hannan, E. L., Bernard, H. R., O'Donnell, J. F. and Kilburn, H., Jr. (1989) 'A methodology for targeting hospital cases for quality of care record reviews', *Am J Public Health,* 79(4), 430-6.

Hayward, R. A. and Hofer, T. P. (2001) 'Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer', *JAMA,* 286(4), 415-20.

Heijink, R., Koolman, X., Pieter, D., van der Veen, A., Jarman, B. and Westert, G. (2008) 'Measuring and explaining mortality in Dutch hospitals; the Hospital Standardized Mortality Rate between 2003 and 2005', *BMC Health Serv Res,* 8, 73.

Hogan, H., Healey, F., Neale, G., Thomson, R., Vincent, C. and Black, N. (2012) 'Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study', *BMJ Qual Saf*.

Hosmer, D., Lemeshow, S. and Sturdivant, R. (2013) *Applied Logistic Regression, Wiley Series in Probability and Statistics,* New Jersey: John Wiley and Sons Inc.

Institute for Healthcare Improvement (IHI) 'IHI Global Trigger Tool for Measuring Adverse Events', [online], available: http://www.ihi.org/knowledge/Pages/Tools/IHIGlobalTriggerToolforMeasuringAEs.aspx [accessed 5 August 2013].

Isaac, T. and Jha, A. K. (2008) 'Are patient safety indicators related to widely used measures of hospital quality?', *J Gen Intern Med,* 23(9), 1373-8.

Jacques, R. M., Fotheringham, J., Campbell, M. J. and Nicholl, J. (2013) 'Did hospital mortality in England change from 2005 to 2010? A retrospective cohort analysis', *BMC Health Serv Res,* 13(1), 216.

Jarman, B. (2008) 'In defence of the hospital standardized mortality ratio', *Healthc Pap,* 8(4), 37-42; discussion 69-75.

Jarman, B., Gault, S., Alves, B., Hider, A., Dolan, S., Cook, A., Hurwitz, B. and Iezzoni, L. I. (1999) 'Explaining differences in English hospital death rates using routinely collected data', *BMJ,* 318(7197), 1515-20.

Jarman, B., Pieter, D., van der Veen, A. A., Kool, R. B., Aylin, P., Bottle, A., Westert, G. P. and Jones, S. (2010) 'The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care?', *Qual Saf Health Care,* 19(1), 9-13.

Kernisan, L. P., Lee, S. J., Boscardin, W. J., Landfeld, C. S. and Dudley, R. A. (2009) 'Association between hospital-reported leapfrog safe practice scores and inpatient mortality', *JAMA,* 301(13), 1341-1348.

Kipnis, P., Escobar, G. J. and Draper, D. (2010) 'Effect of choice of estimation method on inter-hospital mortality rate comparisons', *Med Care,* 48(5), 458-65.

Kristoffersen, D. T., Helgeland, J., Clench-Aas, J., Laake, P. and Veierod, M. B. (2012) 'Comparing hospital mortality--how to count does matter for patients hospitalized for acute myocardial infarction (AMI), stroke and hip fracture', *BMC Health Serv Res,* 12, 364.

Kroch, E. A., Johnson, M., Martin, J. and Duan, M. (2010) 'Making hospital mortality measurement more meaningful: incorporating advance directives and palliative care designations', *American Journal of Medical Quality,* 25, 24-33.

Krumholz, H. M., Brindis, R. G., Brush, J. E., Cohen, D. J., Epstein, A. J., Furie, K., Howard, G., Peterson, E. D., Rathore, S. S., Smith, S. C., Jr., Spertus, J. A., Wang, Y. and Normand, S. L. (2006) 'Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation', *Circulation,* 113(3), 456-62.

Lang T and Altman D (2013) 'Basic statistical reporting for articles published in clinical medial journals: the SAMPL Guidelines' in Smart, P., Maisonneuve, H. and Poldermann, A., eds., *Science Editors' Handbook*, European Association of Science Editors.

Lujic, S., Jorm, L. and Randall, D. (2012) *Deaths after admission to hospital: a NSW population-based data linkage study*.

Mihrshahi, S., Brand, C., Ibrahim, J. E., Evans, S., Jolley, D. and Cameron, P. (2010) 'Validity of the indicator 'death in low-mortality diagnosis-related groups' for measuring patient safety and healthcare quality in hospitals', *Intern Med J,* 40(4), 250-7.

Miyata, H., Hashimoto, H., Horiguchi, H., Matsuda, S., Motomura, N. and Takamoto, S. (2008) 'Performance of in-hospital mortality prediction models for acute hospitalization: hospital standardized mortality ratio in Japan', *BMC Health Serv Res,* 8, 229.

Mohammed, M. A., Deeks, J. J., Girling, A., Rudge, G., Carmalt, M., Stevens, A. J. and Lilford, R. J. (2009) 'Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals', *BMJ,* 338, b780.

Mohammed, M. A. and Stevens, A. J. (2013) 'A simple insightful approach to investigating a hospital standardised mortality ratio: an illustrative case-study', *PLoS One,* 8(3), e57845. doi:10.1371/journal.pone.0057845.

Morsi, E., Lindenauer, P. K. and Rothberg, M. B. (2012) 'Primary care physicians' use of publicly reported quality data in hospital referral decisions', *Journal of Hospital Medicine,* 7(5), 370-375.

Motheral, B., Brooks, J., Clark, M. A., Crown, W. H., Davey, P., Hutchins, D., Martin, B. C. and Stang, P. (2003) 'A checklist for retrospective database studies--report of the ISPOR Task Force on Retrospective Databases', *Value Health,* 6(2), 90-7.

Palmer, W. L., Bottle, A., Davie, C., Vincent, C. A. and Aylin, P. (2013) 'Meeting the ambition of measuring the quality of hospitals' stroke care using routinely collected administrative data: a feasibility study', *Int J Qual Health Care,* 25(4), 429-36.

Penfold, R. B., Dean, S., Flemons, W. and Moffatt, M. (2008) 'Do Hospital Standardized Mortality Ratios measure patient safety? HSMRs in the Winnepeg Regional Health Authority', *HealthcarePapers,* 8(4), 8-24.

Pilcher, D. V., Hoffman, T., Thomas, C., Ernest, D. and Hart, G. (2010) 'Risk-adjusted continuous outcome monitoring with an EWMA chart: could it have detected excess mortality among intensive care patients at Bundaberg Base Hospital', *Critical Care and Resuscitation,* 12(1), 36-41.

Pitches, D. W., Mohammed, M. A. and Lilford, R. J. (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 Serv Res,* 7, 91.

Popowich, J., Zaborowski, J. and Bellows, M. (2011) 'Hospital Standardized Mortality Ratios: a tale of two sites. Lessons learned from the United Kingdom; Canada catches up', *Healthcare Quarterly,* 14(2), 42-48.

Pouw, M. E., Peelen, L. M., Lingsma, H. F., Pieter, D., Steyerberg, E., Kalkman, C. J. and Moons, K. G. (2013) 'Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization', *PLoS One,* 8(4), e59160.

Rosen, A. K., Rivard, P., Zhao, S., Loveland, S., Tsilimingras, D., Christiansen, C. L., Elixhauser, A. and Romano, P. S. (2005) 'Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data?', *Med Care,* 43(9), 873-84.

Scott, I. A., Thomson, P. L. and Narasimhan, S. (2008) 'Comparing risk-prediction methods using administrative or clinical data in assessing excess in-hospital mortality in patients with acute myocardial infarction', *Medical Journal of Australia,* 188(6), 332-336.

Shahian, D. M., Wolf, R. E., Iezzoni, L. I., Kirle, L. and Normand, S. L. (2010) 'Variability in the measurement of hospital-wide mortality rates', *N Engl J Med,* 363(26), 2530-9.

SIGN (2013) 'Methodology Checklist 3: Cohort studies', [online], available: http://www.sign.ac.uk/guidelines/fulltext/50/checklist3.html [accessed 5 June 2013].

Sorinola, O. O., Weerasinghe, C. and Brown, R. (2012) 'Preventable hospital mortality: learning from retrospective case record review', *Journal of the Royal Society of Medicine Short Reports,* 3(77), DOI: 10.1258/shorts.2012.012077.

Thomas, E. J. and Petersen, L. A. (2003) 'Measuring errors and adverse events in health care', *J Gen Intern Med,* 18(1), 61-7.

van den Bosch, W. F., Kelder, J. C. and Wagner, C. (2011) 'Predicting hospital mortality among frequently readmitted patients: HSMR biased by readmission', *BMC Health Serv Res,* 11, 57.

van den Bosch, W. F., Spreeuwenberg, P. and Wagner, C. (2012) 'Variations in hospital standardised mortality ratios (HSMR) as a result of frequent readmissions', *BMC Health Serv Res,* 12, 91.

van Walraven, C., Escobar, G. J., Greene, J. D. and Forster, A. J. (2010) 'The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population', *J Clin Epidemiol,* 63(7), 798-803.

van Walraven, C., Wong, J., Bennett, C. and Forster, A. J. (2011) 'The Procedural Index for Mortality Risk (PIMR): an index calculated using administrative data to quantify the independent influence of procedures on risk of hospital death', *BMC Health Serv Res,* 11, 258.

Weeks, W. B., West, A. N., Rosen, A. K. and Bagian, J. P. (2008) 'Comparing measures of patient safety for inpatient care provided to veterans within and outside the VA system in New York', *Qual Saf Health Care,* 17(1), 58-64.

Wright, J., Dugdale, B., Hammond, I., Jarman, B., Neary, M., Newton, D., Patterson, C., Russon, L., Stanley, P., Stephens, R. and Warren, E. (2006) 'Learning from death: a hospital mortality reduction programme', *J R Soc Med,* 99(6), 303-8.