2 The Commission's approach

Key points

- The Commission used hospital-standardised mortality ratios (HSMRs) to measure the effectiveness and quality of hospital care.
  - HSMRs describe how effectively hospitals reduce in-hospital mortality relative to how well they are predicted to perform, after accounting for patient and other factors affecting mortality.
- The Commission also assessed hospital performance by measuring the efficiency with which hospitals use their resources to provide services and to contribute to the quality of health care.
- The Commission measured hospital efficiency in three ways:
  - output-oriented technical efficiency — the extent to which a hospital is able to produce more of any output without producing less of some other output or using more of some input
  - input-oriented technical efficiency — the extent to which a hospital is able to reduce the use of any input without producing less of some output or using more of some other input
  - cost efficiency — the extent to which a hospital produces its outputs at least possible cost.
- In calculating the various performance measures, the Commission took into account a range of factors including:
  - patient characteristics that were outside the control of hospitals
  - financial incentives faced by hospitals
  - other hospital characteristics, such as the hospital's teaching status, location, the presence of emergency departments and intensive care facilities, and the presence of specialist facilities.

The Commission benchmarked the performance of public and private hospitals in terms of:

- hospital-standardised mortality ratios (HSMRs), which are measures of hospital effectiveness and quality
- the efficiency with which hospitals provide their services.

The approach to measuring hospital effectiveness and quality is outlined in section 2.1. Key concepts of hospital efficiency are described in section 2.2. The Commission's approach to estimating efficiency is outlined in section 2.3. Other issues for estimating hospital efficiency are outlined in section 2.4. A more detailed description of the methods for estimating HSMRs and efficiency are outlined in appendix C.

2.1 Measuring hospital effectiveness and quality

Patients seek hospital services in order to improve their physical and emotional wellbeing relative to what would otherwise be the case. The effectiveness of a hospital's health care is an important aspect of a hospital's performance because it is a major contributor to the key purpose of a hospital — to heal the sick and injured and to provide for their safety.

Considerable effort, in Australia and overseas, has been put into measuring hospital effectiveness and quality. In Australia, for example, a number of reporting frameworks have been developed (including ACHS 2008; AIHW 2009b and SCRGSP 2009, as well as the National Healthcare Agreement reporting) to assist policy makers and hospital administrators to understand and assess the extent to which they are meeting those objectives.

Some measures of hospital effectiveness and quality

The ideal measure is one which captures the incremental improvement to a patient's health status after an episode in hospital care. Theoretical measures of incremental improvements include the improvements to the disability-adjusted life years or quality-adjusted life years of a patient. However, as noted by the Centre for Health Economics (Monash University) (cited in PC 2009), such measures of patient outcomes are not generally available.

The community also places value on other aspects of hospital services — such as accessibility and waiting times for services. These dimensions are included in many definitions of hospital quality (for example, ACSQHC 2009; Campbell, Roland and Buetow 2000). A number of organisations in Australia and overseas have sought to identify and measure various dimensions of hospital quality (for example, ACHS 2008; AHRQ 2009; AIHW 2009b; Department of Health (Victoria) 2009; Kelley and Hurst 2006). It was not possible, however, to incorporate these measures in the estimation of hospital effectiveness and quality.

A review of Australian and overseas studies of hospital efficiency (appendix B) suggests that potential measures of hospital quality include:

- adverse events
- hospital-acquired infections
- unplanned re-admissions
- in-hospital mortality.

Data for each these variables are included in the National Hospital Morbidity Database (NHMD) in Australia. However, the current reporting of these variables suffer from a number of limitations. In the case of adverse events, there are two problems. First, the categories for classifying adverse events do not fully reflect the true incidence of hospital error. Other classification categories in the NHMD may also report an incidence of adverse events, but these are not routinely collected (AIHW 2009a). For example, published Australian Institute of Health and Welfare (AIHW) data indicate that the rate of adverse events varied between 3.4 and 3.7 per cent for private hospitals and over 5 per cent for public hospitals in 2007-08. However, Wilson et al. (1995) estimated that approximately 8.3 per cent of all separations involved in-hospital adverse events, while Jackson (2008) used Victorian and Queensland data (which at the time were the only data that contained a flag to indicate if a diagnosis arose during hospitalisation) and found that adverse events occurred in 12.3 per cent of separations. This suggests that available data under represent the true incidence of adverse events by an unknown margin.

The second problem is whether the adverse event originated in hospitals or in the wider community. The NHMD classifications for adverse events do not distinguish where the adverse event took place — in a hospital or elsewhere.

Similar problems exist with the NHMD data on hospital-acquired infections. The Commission notes that there is some debate as to whether existing morbidity data can feasibly distinguish between hospital-acquired and community-acquired infections. One view is that the ICD-10-AM codes can be used to distinguish infections (such as *staphylococcus aureus*) that are either *methicillin* or multi-antibiotic resistant, and which are believed to have occurred within a hospital setting. The AIHW advised the Commission that NHMD data from 2007-08 and previous years cannot accurately identify whether an infection arose during hospital stay or in the community (PC 2009).

There are also similar measurement issues with unplanned re-admissions within 28 days of discharge. Re-admission rates are likely to substantially underestimate the true re-admission rate because NHMD data are not linked to individual patients, and so it is not possible to determine if discharged patients are re-admitted to another hospital, which would not be captured by these data.

Which measure to use?

The Commission is of the view that the in-hospital mortality rate is the most reliable indicator in this circumstance. That said, there are two issues that should be recognised. First, in-hospital mortality rates are only a partial measure of patient outcomes. However, there is some evidence that in-hospital mortality is correlated with the processes of care for a range of conditions (chapter 4). Second, random variation in the incidence of mortality may still exist from year-to-year, such as the outbreak of influenza epidemics. Mortality rates would be unsuitable as a quality measure if they varied greatly from year to year in a random fashion. Several authors reviewed the incidence of random variation with in-hospital mortality data (Ben-Tovim, Woodman, Harrison et al. 2009; ACSQHC 2009) and concluded that such variation was not significant.

Standardising in-hospital mortality

To use the incidence of in-hospital mortality as a quality measure, three other limitations to this variable had to be addressed.

First, the incidence can vary between hospitals for reasons that are beyond their control. For example, some hospitals may specialise in treating higher or lower-risk patients, a factor that is likely to impact on observed mortality rates.

Second, mortality rates are likely to vary according to the range of services provided by different hospitals. Hospitals that provide palliative care facilities, for example, are more likely to experience higher mortality rates.

Third, in-hospital mortality rates can raise a problem of *collinearity* when they are used in an estimation of technical or cost efficiency. Since in-hospital mortality is a factor included in the measurement of cost and technical efficiency, and that many of the factors that are thought to influence in-hospital mortality also influence hospital efficiency (such as the services provided by hospitals), there is the likelihood that collinearity will emerge as a problem in the estimation process. When collinearity arises, it is difficult to determine whether in-hospital mortality has a significant bearing on the estimated cost and technical efficiency of a hospital.

These three problems are addressed in this study by risk adjusting mortality rates. Risk adjustment is the process by which a number of explanatory factors (such as the patient's age, gender, degree of comorbidity, and other variables influencing the probability that a patient died or lived in hospital) are used to account for differences in the incidence of in-hospital mortality.

Once predicted values of hospital mortality rates were obtained from the risk adjustment regressions, the Commission then estimated HSMRs for each hospital. Ben-Tovim, Woodman, Harrison et al. (2009) report three variants of the definition of the HSMR. The version adopted for this study, and the only one feasible given the data available to the Commission, is the ratio of the actual number of acute in-hospital deaths to the expected number of in-hospital deaths, for all conditions accounting for in-patient mortality.[1]

The estimates of adjusted mortality rates are included in the subsequent analysis of technical efficiency and hospital costs. Risk adjustment has been used in a number of efficiency studies that have used re-admission rates (Chua, Palangkaraya and Yong 2009), mortality rates (Paul 2002) and both re-admission and mortality rates (Clement et al. 2008).

The Commission used negative binomial regression analysis to risk adjust mortality rates, since only hospital-level data were available to the Commission. The Commission explored several approaches to risk adjusting these data, including Tobit regressions, and weighted logistic regression. The Tobit regression approach is described in PC (2009).[2] The negative binomial approach used for this study and the results are described in detail in chapter 4 and appendix C.

2.2 Concepts of hospital efficiency

The subject of hospital efficiency is about identifying how well hospitals are using their scarce resources (medical and nursing staff, beds, and medical and pharmaceutical supplies, for example) to provide hospital services and improve patient health outcomes and patient safety.

---

1 The second definition is the HSMR calculated on 20 per cent of diagnoses that account for 80 per cent of in-hospital mortalities. The third definition includes the remaining 80 per cent of cases that account for 20 per cent of deaths (Ben-Tovim, Woodman, Harrison et al. 2009). These are not feasible as the Commission does not have access to data on the number of mortalities for different types of cases.

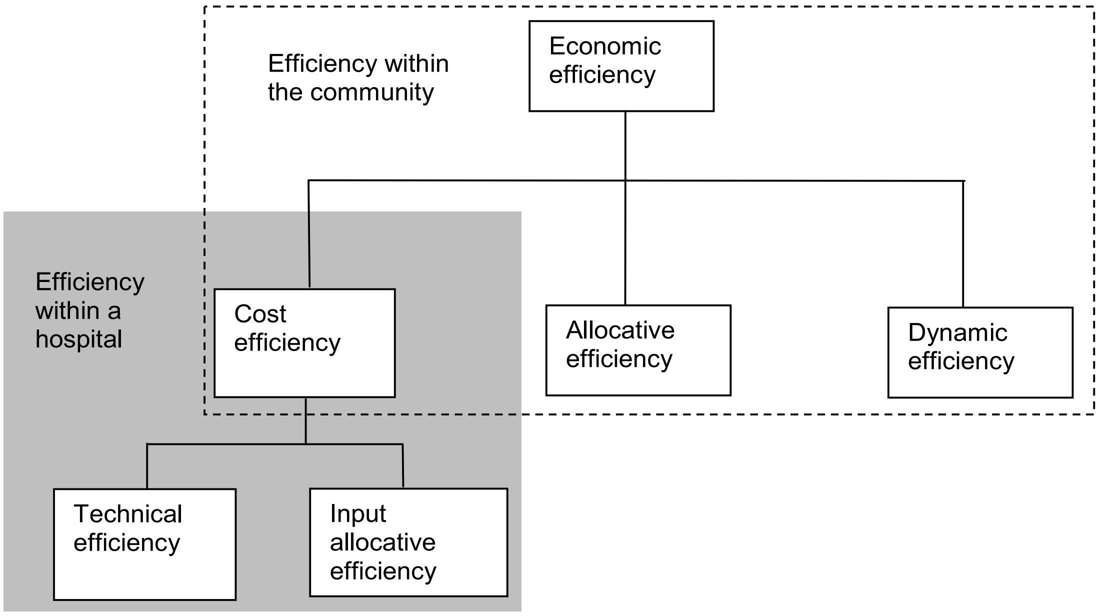
2 The Commission notes that the average risk-adjusted mortality rate using the Tobit regression was found to be approximately 0.55, which is less than the expected average of 1.0.

Definitions of hospital efficiency

Efficiency, in its broadest sense, refers to how well a community's resources are used to improve its wellbeing. If resources are wasted or not put to their best use, community welfare will be not maximised. This broad interpretation is known as 'economic efficiency'. Economic efficiency has three major components:

- cost efficiency (CE) — the degree to which a community's outputs are produced at the least possible cost[3]
- allocative efficiency — how a community's resources are allocated across different uses so as to generate the greatest wellbeing at a given point in time
- dynamic efficiency — the allocation of a community's resources over time, including allocations designed to improve economic efficiency and to generate more resources (PC 2009) (figure 2.1).

Figure 2.1 **Illustration of the components of efficiency**



Cost efficiency itself comprises:

- technical efficiency (TE) — which is usually defined to be the extent to which a hospital is able to produce more of any output or patient health outcome without producing less of some other output, outcome or using more of some input

3 The terminology of cost efficiency used in this study is used extensively in the benchmarking literature and is synonymous with the definition of productive efficiency used in PC (2009).

- input allocative efficiency — the extent to which a hospital is using the appropriate mix of inputs, given the input prices it faces (figure 2.1).

The above definition of technical efficiency is often referred to as 'output-oriented technical efficiency'. Technical efficiency can alternatively be defined as the degree with which a hospital can reduce its use of inputs to produce a given set of outputs or quality.

Efficiency is an important concept for hospitals because any resources saved can be used towards providing additional services elsewhere in a hospital or within the healthcare system, or put towards other useful purposes (including individual consumption or saving), ultimately promoting the wellbeing of the community.

The study does not, however, contend with the broader issue of *allocative efficiency* — the efficiency with which the health sector as a whole is providing the appropriate mix of hospital and other healthcare services. A study on allocative efficiency would need to focus on the pricing of hospital services, recognising the peculiarities of the operation of public and private health insurance and the asymmetries of information within the health sector. A study on the allocative efficiency of hospitals is beyond the dataset and scope of this analysis.

Illustration of the concepts of efficiency

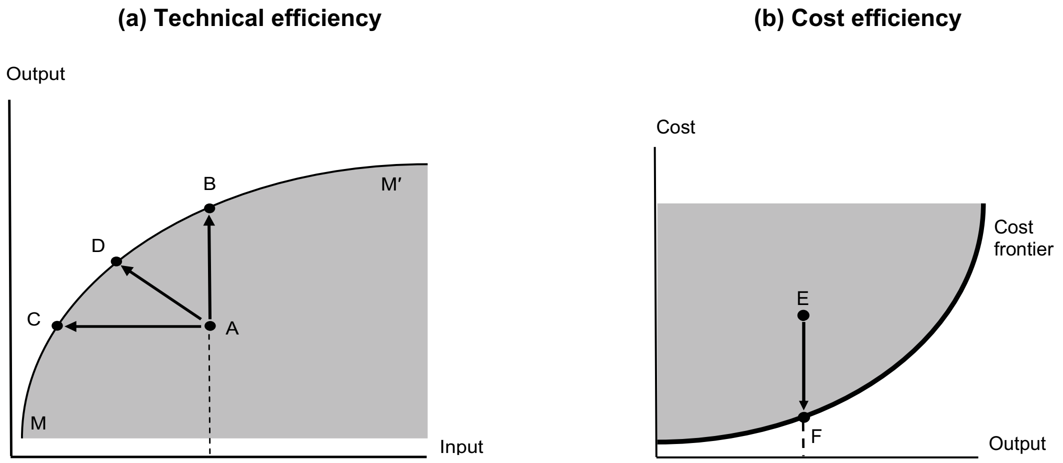
In this study, the Commission assessed hospital performance on the basis of three measures:

- output-oriented technical efficiency
- input-oriented technical efficiency
- cost efficiency.

Each of these approaches are illustrated in figure 2.2. The first approach measures hospital efficiency in terms of how much additional output a hospital must produce to be technically efficient, that is, on the best-practice frontier *MM'*. This is the distance from its position at *A* to its frontier at *B*.

In contrast, the second approach measures the hospital efficiency in terms of how much fewer resources the hospital could employ and still produce the same level of output. This is given as the distance *A* to *C* (figure 2.2).

Figure 2.2 **Illustration of input and output-oriented approaches**



Source: Adapted from Coelli et al. (2005).

While the output and input-oriented technical efficiencies are expected to be similar in aggregate, there is no reason to suggest that for an individual hospital this should be the case. It is also possible to calculate the average of the output and input-oriented approaches. This assumes that hospitals face some degree of flexibility in terms of their choice of inputs and outputs, but are not entirely restricted in either case. In terms of figure 2.2, this would represent the distance from its position at *A* to its frontier at *D*.

Cost efficiency is defined as the extent with which a hospital can reduce its costs and still produce the same level of output. This is given as the distance between hospital *E* and its frontier estimated at *F* in figure 2.2. Cost efficiency is a commonly used to compare hospital performance, even though there are significant problems associated with obtain in comparable measures of prices and costs.

Why distinguish between output and input-oriented efficiency?

The distinction between output and input-oriented technical efficiency was made to account for the different behavioural assumptions of public and private hospitals. One of the challenges is to accurately represent mathematically the motivations and behaviours of public, for-profit and not-for-profit hospitals. It is clear that these groups of hospitals have different motivations, and as a consequence, exhibit different behaviours when providing medical and surgical care. This in turn has implications for how these hospitals are to be modelled.

For-profit hospitals, for example, are mainly concerned with maximising the return to shareholders. They are concerned with jointly maximising their revenues and minimising their costs. This in turn means that they have the incentive to change both the inputs they employ and the type and level of services they offer, although the degree to which they can influence their outputs and inputs depends on their individual circumstances.

Not-for-profit hospitals, like for-profit hospitals, appear to have some freedom to choose their level of input use and outputs too. Even though they do not necessarily maximise their financial returns for the benefit of their owners, they nonetheless have incentives to maximise their level of output and quality, as this is consistent with their charter of providing care to the community (Newhouse 1970). Any returns arising from their service provision are reinvested in order to further increase their capacity and quality of care or donated to charitable causes.

Public hospitals, on the other hand, do not appear to have the same degree of flexibility in choosing their output mix for two reasons. First, under the National Healthcare Agreement, public hospitals are not able to refuse medical treatment to the public. Second, public hospitals tend to be funded for a target level of services — funding allocations are determined at the beginning of each year for, but not beyond, a target level of activity (DHS nd; NSW Health 2008). This funding method applies even in jurisdictions that pay hospitals on a casemix basis.

The distinction between output and input-oriented technical efficiency is useful in this regard. Output-oriented technical efficiency is more likely to be more representative of for-profit hospitals, while the input-oriented technical efficiency is more likely to be representative of public hospitals.

2.3 Techniques for estimating hospital efficiency

A literature review of the techniques used to determine the best-practice frontier suggests that the majority of studies employed data envelopment analysis (DEA) and stochastic frontier analysis (SFA) (appendix B). Twenty-seven studies employed DEA (including those that estimated Malmquist productivity indexes) and twenty-two employed SFA (including the closely related stochastic distance functions). While DEA and SFA are conceptually similar at a broad level — they both establish best-practice frontiers given a set of observations — differences in how they determine those frontiers can lead to noticeable differences in the observed results.

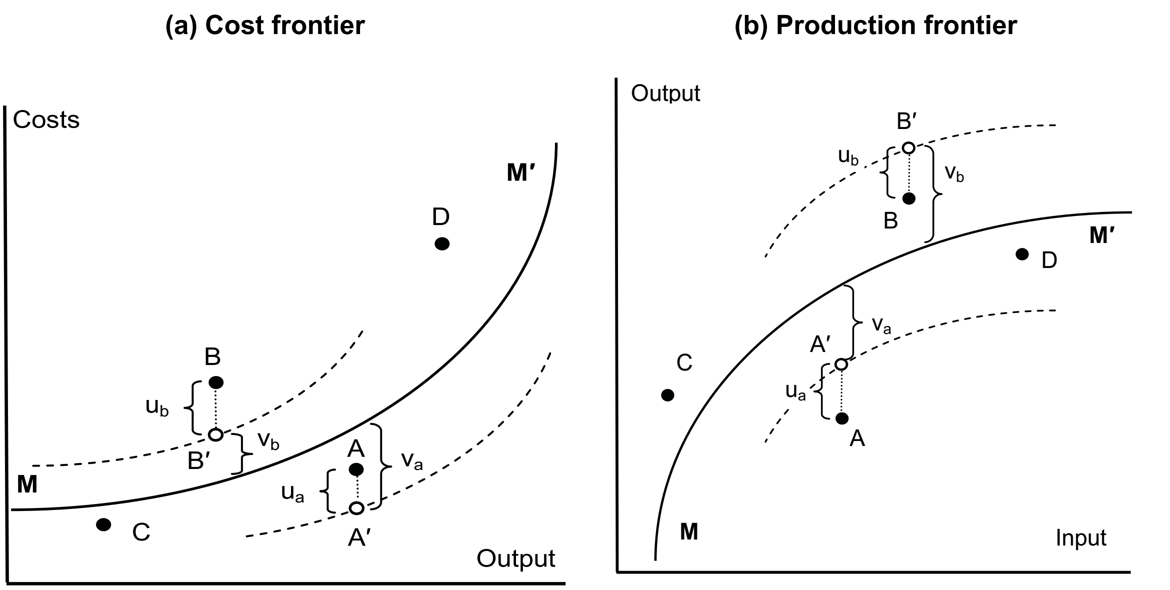
Drawing upon the advice of external referees, SFA appears to be the superior technique in this circumstance because, as an econometric method, it permits the significance of variables to be statistically tested. The technique, with a suitably flexible functional form, is likely to exhibit lower sensitivity than DEA which does not account for outliers and random effects. This in turn is likely to yield more conservative efficiency estimates than if DEA were used.

Graphical illustration

SFA was first developed to study the efficiency of firms. It was originally developed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) and later generalised by Schmidt and Sickles (1984) for use with time-series cross-section panel data.

A good introductory summary of SFA can be found in Coelli et al. (2005) and a more advanced treatment in Kumbhakar and Lovell (2000). SFA frontier estimation can be graphically represented as a two-step regression. In the first step, a regression equation is estimated to pass through the mean of the data, much like OLS (in this example, between points *A*, *B*, *C* and *D*) (figure 2.3). This gives the average function *MM′*. In SFA, unlike OLS, the curve *MM′* is then shifted *for each hospital*. For hospitals *A* and *B*, the curve is shifted by amounts v_a and v_b respectively. These shifts establish the *stochastic frontiers A′* and *B′* respectively, against which the efficiency scores are calculated.

Figure 2.3 **Illustration of SFA approach**



Source: Adapted from Coelli et al. (2005).

Even though SFA resembles classical ordinary least squares (OLS), it differs from OLS in the treatment of the residual. In OLS, the random error term is assumed to be symmetrically distributed. If, however, the residuals are skewed (that is, they are not symmetric), SFA can be used to partition the error terms into a non-normally distributed component and a pure random error term that is normally distributed.

Mathematical expression

Mathematically, inefficiency is measured as a component of the stochastic frontier regression equation:

$$y_i = f(\mathbf{x}_i) - u_i + v_i \quad (1)$$

where y_i is the dependent variable, x_i is the independent variables, v_i is the random error term, and u_i is the inefficiency component, for hospital *i*. The term v_i captures random variations across hospitals reflecting random events that might include:

- measurement error in the variables
- other random events that affect costs or output
- the combined effects of other omitted factors, many of which are not amenable to quantification (Coelli et al. 2005).

In the context of a cost function, u_i is a measure of cost efficiency, and in the case of a multi-input multi-output production function (distance function), it is a measure of technical efficiency.

Both of the terms v_i and u_i are assumed to be independent and identically distributed. It is assumed that the random error (v_i) adopts a normal distribution with a zero mean and a constant variance. In the Aigner–Lovell–Schmidt models, the inefficiency component (u_i) can be assumed to have a half-normal, truncated half-normal, exponential or gamma distribution, with a positive mean.

In the output and input-oriented distance functions, the distribution of the technical efficiency term was assumed to have a half-normal distribution, because it was found to generate the most plausible distribution of efficiency scores. An exponential distribution was adopted in the cost function because it was the distribution that was most likely to solve, and because the reported levels of efficiency scores were less susceptible to change, especially when variables were included in the efficiency effects model.

The choice of distribution for u_i affects the estimated efficiency scores of each observation. However, there is some evidence to suggest that the ordinal ranking of the scores are less sensitive to those distributional assumptions (Kumbhakar and Lovell 2000).

When the estimated function is in natural logarithmic form, efficiency is simply:

$$Index_i = \exp(-u_i) \tag{2}$$

Estimating technical efficiency

When applied to the output-oriented distance (production) function, the equation becomes:

$$-y_{iK} = f(\mathbf{y}_i, \mathbf{x}_i, q_i, \mathbf{z}_i) - u_i + v_i \tag{3}$$

where y_{iK} is the base output of hospital i, which is a function of other outputs (y_i), inputs (x_i), quality (q_i) and z_i (factors outside the control of hospitals). The error term is divided into purely random error term (v_i) and a measure of output-oriented technical efficiency (u_i).

When applied to the input-oriented distance (production) function, equation (1) is specified as:

$$-x_{iM} = f(\mathbf{y}_i, \mathbf{x}_i, q_i, \mathbf{z}_i) - u_i + v_i \tag{4}$$

where x_{iM} is use of the base input of hospital i, which is a function of other inputs (x_i), outputs (y_i), quality (q_i) and z_i (factors outside the control of hospitals). The error term is divided into purely random error term (v_i) and a measure of input-oriented technical efficiency (u_i).

The factors outside the control of hospitals include a range of patient and establishment characteristics. These include a hospital's assigned function in the healthcare system and the type of patients that seek treatment. The inclusion of these control variables in the frontier equation means that the resulting efficiency scores are net of their effect.

Details of all the variables included in both the cost and technical efficiency equations are explained in chapter 3. A more detailed derivation of the equations presented in this section is provided in appendix C.

Estimating cost efficiency

In the context of a cost function, equation (1) is specified as:

$$c_i = f(\mathbf{w}, \mathbf{y}_i, q_i, \mathbf{z}_i) - u_i + v_i \tag{5}$$

where c_i is the total cost of hospital i, which is a function of input prices (w),[4] outputs (y_i), quality (q_i) and z_i (factors outside the control of hospitals). The error term is comprised of purely random error term (v_i) and a measure of cost efficiency (u_i).

A more detailed derivation of these equations is given in appendix C.

Testing for differences in efficiency between hospitals

To test whether any differences in the efficiency scores for different ownership groups are statistically significant, the hospital efficiency scores are regressed as a function of three binary variables:

- *Private / Public or contracted* — to test whether there was a statistically significant difference between all private hospitals (assigned a value of '1') and public and contracted hospitals (assigned a value of '0')
- *For-profit / Not-for-profit* — to test whether there was a statistically significant difference between for-profit private hospitals (assigned a value of '1') and not-for-profit private hospitals (assigned a value of '0')
- *Contract / Other* — to test whether there was a statistically significant difference between public contract hospitals (assigned a value of '1') and all other hospitals (assigned a value of '0').

The coefficient values of these binary variables cannot be reported due to ABS commercial-in-confidence concerns, but their sign and statistical significance can be reported.

The method of the second-stage regression of the efficiency scores is covered in appendix C.

Functional form

The estimation applies a full transcendental logarithmic (translog) function. In addition to first-order values, the translog contains squared and cross-terms of all inputs, output, price and cost variables, as appropriate. An alternative functional form is the Cobb-Douglas model, which only contains the first-order values. Due to its expanded set of variables, the translog function generates a more precise fit of the model than the Cobb-Douglas (Nguyen and Coelli 2009). More details of this functional form is provided in appendix C.

---

[4] Input prices are constant across hospitals in the same sector and jurisdiction (chapter 3).

2.4 Other modelling issues

A large number of multivariate studies of hospital efficiency have been undertaken worldwide, although only a few examined the efficiency of Australian hospitals. For example, O'Neill et al. (2008), in a detailed study of 79 DEA studies did not include any Australian studies in their review. A similar pattern can be gleaned from literature reviews by Butler (1995), Peacock et al. (2001), Hollingsworth (2008) and Hollingsworth and Peacock (2008). Of the small number of Australian studies available, there are thirteen commonly cited studies that have been published since the mid-1990s. These include Butler (1995), SCRCSSP (1997), Webster, Kennedy and Johnson (1998), Yong and Harris (1999), Wang and Mahmood (2000a, 2000b), Paul (2002), Queensland Department of Health (2004), Mangano (2006), Jensen, Webster and Witt (2007), Gabbitas and Jeffs (2008), and Chua, Palangkaraya and Yong (2008, 2009). The Commission's preliminary stage of analysis added to this body of literature (PC 2009).

A comprehensive review of the modelling approaches and findings of previous studies is presented in appendix B, yet a cursory examination of Australian studies suggests that:

- private hospitals are less costly than public hospitals (when medical costs are excluded) (for example, Butler 1995)
- private hospitals give rise to better health outcomes than public hospitals (for example, Chua, Palangkaraya and Yong 2008)
- for-profit private hospitals are more technically efficient than not-for-profit private hospitals (for example, Webster, Kennedy and Johnson 1998)
- metropolitan public acute hospitals are more technically and cost efficient than smaller rural hospitals (for example, SCRCSSP 1997; Wang and Mahmood 2000a) (table 2.1) .

There are, however, several reasons to believe that these conclusions need to be further tested. First, a variety of methods was used to estimate the efficiency scores. For example, several studies employed DEA while others employed SFA for a production function. See section 2.4 and appendix C for a discussion of SFA. DEA is described in Coelli et al. (2005).

Second, the studies often measured different aspects of hospital efficiency which are not always comparable. It is inappropriate, for example, to directly compare technical efficiencies with cost efficiencies.

Third, many of the studies compared different samples of hospitals — some have focused on only a few major hospitals in each state (for example, Mortimer 2002;

Yong and Harris 1999) while others focused the all of the hospitals in a state (for example, Paul 2002), while another focused on all acute private hospitals in Australia (Webster, Kennedy and Johnson 1998).

Table 2.1 **Summary of efficiency scores from selected Australian studies**[a]

Authors and year published	*No. of hospitals and years*	*Measure of efficiency*[b]
Data envelopment analysis — multi-output production		
SCRCSSP (1997)	109 Victorian public hospitals for 1994-95	OTE=0.581 TE=0.775 SE=0.751
Webster, Kennedy & Johnson (1998)	301 private hospitals in 1994-95	OTE=0.282–0.861 TE=0.393–0.898 SE=0.757–0.970
Wang and Mahmood (2000a)	113 NSW public hospitals (in two peer groups – large and small) 1997-98	OTE=0.457 TE=0.834 SE=0.547
Mortimer (2002)	38 Victorian public hospitals in 1993	PTE=0.81
Queensland Department of Health (2004)	Queensland public hospitals for 2000-01 to 2002-03	TE=0.963
Stochastic frontier analysis — single-output production		
Webster, Kennedy & Johnson (1998)	300 private hospitals in 1994-95	TE=0.71–0.79
Mortimer (2002)	38 Victorian public hospitals in 1993	TE=0.80
Mangano (2006)	116 Victorian public hospitals 1992-93 to 1995-96	TE=0.75
Gabbitas and Jeffs (2007)	State-level observations for 1996-97 to 2004-05	TE=0.87
Stochastic frontier analysis — cost function		
Webster, Kennedy & Johnson (1998)	280 private hospitals in 1994-95	CE=0.77–0.96
Yong and Harris (1999)	35 large Victorian acute public hospitals for 1994-95	CE=0.95–0.97
Wang and Mahmood (2000b)	113 NSW public hospitals (in two peer groups – large and small) 1997-98	CE=0.90–0.92
Stochastic distance function — multi-output production		
Paul (2002)	223 NSW public hospitals in 1995-96	TE=0.735

[a] Some studies distinguish overall technical efficiency (OTE) from technical efficiency. OTE is technical efficiency under the assumption of constant returns to scale. In this study, technical efficiency is assumed to be based on variable returns to scale, which means that the derived estimates will be net of any effects of scale economies.[b] Efficiency scores are fractions, that 0.50 represents 50 per cent efficiency. **TE** Technical efficiency, **OTE** Overall technical efficiency, **CE** Cost efficiency, **SE** Scale efficiency.

Finally, there were differences between the studies in the variables used to measure hospital quality and to control for a variety of factors outside the control of hospitals.

These issues are addressed in this study from the use of a single estimation technique (stochastic frontier analysis), the separate estimation of technical and cost efficiency, the use of a consistent sample of public and private hospitals, and a comprehensive treatment of variables for the treatment of quality and factors outside the control of hospitals.

The issues of the measurement of hospital quality and the role of factors outside the control of hospitals, are covered below.

Relationship between quality and efficiency

Hospitals vary significantly in terms of the services they provide. Correctly specifying the measurement of the quality of hospital care is important to ensure that the estimated efficiency scores are as accurate as possible.

A review of Australian and overseas literature on hospital efficiency suggests three broad approaches to take account of variations in the quality of hospital health care (appendix B). The first approach compares a hospital's performance solely in terms of the quantity of (intermediate) outputs provided by the hospital (for example, Dor and Farley 1996; Jacobs 2001; Rosko and Chilingerian 1999; Scott and Parkin 1995; Webster, Kennedy and Johnson 1998). Such services include the number of separations, procedures, emergency department visits, and outpatient department services. The attraction of this approach is that it permits, through the use of casemix-adjustment, a hospital's activity to be differentiated across procedures and diagnoses. Another attraction is that it is comparatively easy to attribute a hospital's resource use to its outputs, whereas attributing cause and effect is far more difficult for health outcomes (Hollingsworth and Peacock 2008).

A disadvantage of only using quantity as a measure of hospital activity is that it assumes that there is no relationship between hospital quality and the level of activity. If a hospital faces a trade-off between quantity of services provided and outcomes and quality of its services, then this modelling assumption would penalise those hospitals that focus on achieving better outcomes and higher quality of services.

The second approach is to compare hospital performance solely in terms of a clearly identifiable patient health outcome, such as unplanned re-admission rates and mortality rates (for example, Chua, Palangkaraya and Yong 2008; Jensen, Webster and Witt 2007). The attraction of this approach is that it provides a clear measure of the resources used to achieve a particular health outcome. Its disadvantage is that it does not provide any information about which we can judge the efficient use of scarce resources in a hospital environment.

A third approach is to compare hospital performance in terms of both the quantity of outputs and partial indicators of patient health outcomes. The attraction of this approach is that it enables researchers to explore the relationship between hospital activity and the quality of its services. The disadvantage of this technique is that its usefulness will depend upon the availability and quality of patient outcome data.

What is the relationship between hospital output and quality?

The approach of representing hospital activity in terms of quantities of services and partial measures of health outcomes and quality gives rise to another question: what is the relationship between input use, outputs and patient outcomes?

A number of overseas studies have found that increases in hospital activity may improve patient outcomes or at worst may have no effect on the quality of patient outcomes (box 2.1). Chirikos, French and Luther (2004) argued that 'learning by doing', scale economies, and comparative advantage can provide an explanation of why increasing output can lead to improved patient health outcomes.

A number of authors cautioned against using the volume of hospital activity as a measure of hospital quality (Gruen et al. 2009; Hewitt 2000). As Halm, Lee and Chassin (2002, p. 517) said:

> The magnitude of the [volume–outcome] relationship varies greatly among individual procedures and conditions. The clinical and policy significance of this finding is complicated by methodological shortcomings of many studies. Even when a significant association exists, volume does not predict outcome well for individual hospitals or physicians.

Other authors have noted that insufficient effort was placed on identifying other contributing factors. For example, Carson (2009, p. 1566) noted:

> The evidence supporting a causal effect — high volume surgeons or institutions lead to better surgical outcomes — is not as conclusive as it may seem from the large number of published studies purporting to show such a connection. Case mix and statistical bias is not accounted for in many studies and when taken into account often minimizes apparent differences.

Even if increases to output were correlated with improvements to quality, it may not be possible for an individual hospital to increase its output and improve its quality simultaneously. High volume surgical hospitals are more likely to have specialised units that provide related non-surgical care. For example, a hospital that performs a large volume of cancer surgery is more likely to also have radiation and medical oncologists and cancer specialist nurses. It is therefore difficult to separate the volume of cancer surgery performed by such a hospital from the effects of the specialised oncology and cancer nursing care (Hogan and Winter 2008) or other structures and processes specific to each hospital (Christian et al. 2005).

Box 2.1 **The ‘volume–outcome’ relationship**

An inverse relationship between surgical volume and mortality was described by Luft, Bunker and Enthoven (1979). Since then, many studies have considered the possibility of a ‘volume–outcome relationship’ for various procedures and treatments, and found better outcomes for patients who are treated by hospitals and/or medical practitioners who conduct a greater volume of that procedure or treatment.

The US Institute of Medicine examined evidence from 88 studies concerning eight conditions and procedures, and found that higher volume (whether assessed by hospital or by physician) was associated with better health outcomes in three quarters of the studies reviewed (Hewitt 2000). No studies found a negative relationship between volume and outcome. Other systematic reviews had very similar findings:

> Twenty years of research have established that, for some procedures and conditions, higher volume among hospitals and physicians is associated with better outcomes. (Halm, Lee and Chassin 2002, p. 517)

> Overall, the studies in this review, when combined, demonstrate a quantifiable and statistically significant inverse association between case volume and mortality. (Gruen et al. 2009, p. 208)

> All other things being equal, a higher volume provider will have a marginally better mortality rate than a lower volume provider. This is probably more significant, both in terms of effect size and clinical importance, for complex procedures. (Campbell et al. 2006, p. 162)

Moreover, Birkmeyer, Dimick and Staiger (2006) demonstrated that the volume-outcome relationship is stable over time. Historical measures of procedural volume can therefore identify hospitals that are likely to have better outcomes in the future.

What is the relationship between input use and quality?

A number of other authors have argued that it is the intensity of input use that determines hospital quality rather than the quantity of services. For example, McCue, Mark and Harless (2003) argued that observed improvements to patient outcomes were due to increased hospital resources and therefore operating costs rather than economies of scale. Conversely, reducing hospital resources are thought to worsen the quality of health care.

A number of studies examined the interaction between staffing, and in particular nursing levels, and patient outcomes. Needleman et al. (2002) drew on the 1997 administrative data of 799 hospitals in 11 US states (covering 5 million medical and 1.1 million surgical discharges). The authors found that:

- among medical patients, a higher proportion of hours per care per day by registered nurses and a greater number of hours of care provided by registered nurses was associated with shorter lengths of stay, lower rates of urinary tract infection, upper gastrointestinal bleeding, pneumonia and cardiac arrest, and lower rates of 'failure to rescue' (which was defined as death from pneumonia, shock, cardiac arrest, upper gastrointestinal bleeding, sepsis or deep venous thrombosis)

- among surgical patients, a higher proportion of care provided by registered nurses was associated with lower rates of urinary tract infections, and a greater number of hours of care provided by registered nurses was associated with lower rates of 'failure to rescue'.

The authors did not find any association between the levels of registered nursing and in-hospital mortality, or between licensed nurses and nurses' aides and the rates of adverse outcomes.

Aiken et al. (2002) examined the relationship between patient-to-nurse ratios and the incidence of dying within 30 days of admission, and failure-to-rescue (defined as deaths following complications) among surgical patients. The authors drew on a survey of 10 184 staff nurses and 232 000 discharged patients from 168 non-federal general hospitals in Pennsylvania in 1998-99. The authors found, after adjusting for patient and hospital characteristics (such as whether it was a teaching hospital, and the available technologies), that an incremental increase in the patient-to-nurse ratio increased the odds-ratio of in-hospital mortality and failure-to-rescue by 7 per cent.

Finally, Kane et al. (2007), in a review of other studies, also concluded that higher registered nurse staffing was associated with less hospital-related mortality, failure to rescue, cardiac arrest, hospital acquired pneumonia, and other adverse events. The effect of increased registered nurse staffing on patients safety was strong and consistent in intensive care units and in surgical patients.

These two views form the bases of hypotheses that can be tested in this study. The output-oriented distance function permits the testing of the volume–outcome relationship and the input-oriented distance function permits the testing of the input–outcome relationship.

Factors outside the control of hospitals

Some of the observed differences in the services that hospitals provide and patients they treated do not reflect decisions by the hospitals themselves, but rather are factors outside their control. There is a risk that hospital efficiency estimates would be biased if any of these 'external' factors are ignored. Worthington (2004), for example, argued that ignoring patient characteristics could result in estimates of hospital efficiency representing differences in patient characteristics rather than the hospital's performance.

Most past Australian studies did not sufficiently account for factors outside the control of hospitals (appendix B). For example, Queensland Department of Health (2004), SCRCSSP (1997), Wang and Mahmood (2000a and 2000b) and Webster, Kennedy and Johnson (1998) did not take into account any such characteristics. Yong and Harris (1999), Mangano (2006) and Paul (2002) accounted for whether the hospital had a teaching status and whether it was in a metropolitan area. Only Jensen, Webster and Witt (2007) and Paul (2002) took into account the socioeconomic status of the patient population and the amount of research undertaken at the hospital. And, only Chua, Palangkaraya and Yong (2009) took into account the effects of competition for hospital services. A similar pattern can be observed for many overseas studies (for example, Färe, Grosskopf and Valdmanis 1989; Maniadakis and Thanassoulis 2000).

Where external factors have been taken into account in Australian and overseas studies, they have tended to include:

- patient characteristics, such as:
  - patient comorbidities (for example, Zuckerman, Hadley and Iezzoni 1994)
  - gender and age profile of patients (for example, Zuckerman, Hadley and Iezzoni 1994)
  - patient socioeconomic characteristics (for example, Jensen, Webster and Witt 2007; Paul 2002)
- financial incentives of hospitals, such as:
  - source of patient revenues — the extent to which a hospital is funded using a prospective payment system or operates under capped budgets (for example, Brown 2003; Dor and Farley 1996)
  - market power of the hospital (for example, Chua, Palangkaraya and Yong 2009; Rosko and Chilingerian 1999)
- hospital characteristics that include geography, roles and functions, such as:
  - hospital location (for example, Granneman, Brown and Pauly 1986; Herr 2008)
  - whether it is a teaching or university hospital, and the extent of research and development (for example, Linna 1998; Yong and Harris 1999)
  - the presence of specialist facilities or technologies (for example, O'Neill 1998; Yaisarwang and Burgess 2006)
  - the extent to which the hospital participates in inter-hospital transfers (for example, Jacobs 2001).