GRATTAN Institute

Technical Appendix to *Controlling costly care*

Quantifying variation in Australian acute-care costs

By Ben Weidmann and Stephen Duckett

Table of contents

About this document

This document was prepared as part of the Grattan report *Controlling costly care*. Its purpose is to present the data and methodology used in the report, and to explore the robustness of the analysis.

The *Controlling costly care* project itself is motivated by three factors. First, existing Australian research suggests that even after a variety of 'legitimate and unavoidable'[1] cost-drivers have been considered, substantial variation exists in the cost of providing hospital care.[2]

The second motivation is data availability. Although existing evidence is strongly suggestive that there are substantial differences in how efficiently Australian public hospitals operate, quantifying these differences in a robust way has been hampered by a lack of nationally-comparable patient level data. Grattan's access to the Australia-wide National Hospital Cost Data Collection therefore represents a valuable opportunity to advance our understanding of this potentially important issue.

Last, this work is motivated by concern about the growing pressure on government budgets. Hospital expenditure is the largest single contributor to the growth in Australian public spending.[3] Given the worrying medium-term outlook for government budgets,[4,5] this is an area of expenditure that may well be scrutinised for potential savings. Understanding the scope and nature of cost variation may help inform whether and how policy responds.

This document has six parts. The first lays out our conceptual framework. The second describes the data. The third explores the extent to which variation in hospital costs can be explained by factors beyond hospitals' control. The fourth builds on this, and discusses our approach to benchmarking performance at both the state and hospital level; results of the main benchmarking models are also presented. The fifth section analyses the robustness of our analysis to some methodological choices (such as the definition of outliers) and the stability of estimates over time. The final section concludes.

[1] "Legitimate and unavoidable' is the language used in the *National Health Reform Act 2011* to describe cost drivers that are beyond providers' control and which should therefore be factored into funding arrangements. See page 4 for a detailed discussion of what we define to be legitimate causes of cost variation.

[2] While the scale of variation depends on the methodology used to assess hospitals, research using a variety of approaches illustrates that there may be material differences – from the perspective of overall health spending – between the best and worst performing hospitals. For examples in the Australian context see Productivity Commission (2010); SCRCSSP (1997); Webster*, et al.* (1998); Wang and Mahmood (2000); Yong and Harris (1999); Paul (2002)

[3] Daley*, et al.* (2013), p.15

[4] Ibid.

[5] PWC (2013b)

1. Conceptual Framework

1.1 Introduction

At its core, our analysis is an exercise in benchmarking hospital performance. This immediately raises the prickly question of how performance might be defined and measured. It's unavoidably the case that no two public hospitals are exactly alike. Each serves a different population, provides a different mix of services, has a different set of relationships with other parts of the health system, and so on. This heterogeneity makes performance benchmarking difficult. In particular, it means that simple comparisons of average costs are neither fair nor meaningful. Adjustments need to be made to reflect the reality that different hospitals are required to perform different tasks.

The primary analytical challenge is therefore to define and adjust for drivers of cost that are beyond hospitals' control. We call these costs 'unavoidable and legitimate', with any variation in costs that remains being 'unexplained'. If, after purging the data of unavoidable and legitimate costs, some hospitals *systematically* spend less than others, we interpret this as being indicative of high performance. At the broadest possible level, we conceptualise costs as being a function of 3 factors:

$$C = C(X, Z, E)$$

"Unavoidable and legitimate" cost drivers

X = patient characteristics (e.g. disease severity)

Z = hospital characteristics (e.g. hospital scale)

Other systematic cost variation

E = a hospital's (or state's) ability to control costs. Elsewhere in the literature this is described as "effort" [e.g. Gutacker (2011)]. In *Costly Care* we describe this concept as 'unexplained variation', but it's important to note that this is not the same as 'measurement error' or a 'residual'.

In short, we define 'performance' as an ability to control variations in unexplained costs.[6] Clearly, this definition is far from perfect: most obviously, we are unable to rigorously measure variation in patient outcomes. This issue is discussed at more length in sections 1.3, and in *Controlling costly care*.[7] Given its importance, however, it's worth noting here that using available measures (patient-level data on adverse events[8] and mortality) we find no evidence of a material relationship between hospitals that perform well on 'controlling costs' and quality of care. In other words, it is unlikely that the results presented in this document would significantly change if 'quality' (captured with the best data currently available) were included.[9]

[6] This general approach is applied in a variety of economic activities, often in an effort to estimate the 'effort' of non-market or regulated agents (see Schleifer (1985)). For examples in the health context see Hauck, *et al.* (2003); Gutacker, *et al.* (2011) and Kristensen, *et al.* (2010).

[7] See Section 2 of *Controlling costly care.*

[8] "Adverse events' are the best measure of quality that is recorded in all hospitals and for all patients. They happen when people develop new health problems in hospital, for example due to a fall or infection.

[9] This is in line with recent work on the tangled relationship between cost and quality. Of particular relevance to our work is Gutacker, *et al.* (2011) which added Patient Reported Outcome Measures (PROMs) as a novel and enhanced quality measure. This measure did not substantially alter estimates of cost performance, suggesting the possibility that there may not be large differences between 'efficiency' and an ability to control potentially-avoidable costs.

1.2 Defining costs that are 'unavoidable and legitimate'

In defining unavoidable and legitimate costs, we looked for variables that satisfied three criteria. Factors that were: a) likely to impact costs, b) beyond hospitals' control and c) available to us at patient or neighbourhood (SLA) level. Very broadly, we broke these up into two categories, discussed in turn:

- patient factors (which determine the health profile of the patient, and are beyond hospitals' control)
- provider factors (characteristics of the provider which drive costs, but are not determined by the hospital)

Patient Factors

Direct indicators of patient complexity

Different hospitals are required to treat groups of patients that have differing medical needs. These differences are almost entirely beyond public hospitals' control, making them the biggest driver of legitimate cost variation across providers. By design, the best single predictor of resources required to treat a patient is their Diagnosis Related Group (DRG). DRG assignment is a function of: a patient's disease type and severity; whether surgery is required; the patient's age[10]; whether a patient's stay is 'sameday', and so on.

We then turn to factors which may indicate patient complexity and are linked to cost, but which are not well captured by the DRG system. These are often demographic characteristics, and include age, sex, the presence of comorbidity (as measured by the Charlson index), whether a patient lives in a regional, rural or remote area, and whether a patient is Indigenous. As public hospitals have very limited control over whom they treat, we define any costs associated with these demographic characteristics as beyond providers' control, and therefore legitimate.

We also define a set of conditions that require 'specialised' care, and define the extra cost associated with providing this care as legitimate and unavoidable. The rationale here is that if, for example, a hospital has a specialised cancer unit, then of all the patients in a jurisdiction who are assigned to DRG 'D60A' (ear, nose and throat cancer with catastrophic or severe complications or comorbidities) the specialised hospital will most likely end up treating a greater percentage of the very difficult and resource intensive cases *in this group*. In other words, they may unavoidably and legitimately use more resources than is required to treat the average D60A patient. Evidence in the UK suggests that this phenomenon can have a significant impact on costs, with the result being that providers of specialised care spend substantially more than the casemix average.[11] A similar logic applies with respect to specialised paediatric care, which – like specialised treatments more broadly – we define as an unavoidable and legitimate cost driver.

Next, we include separation and admission 'mode'. This indicates whether, for example, a patient was transferred from another hospital, or discharged to an aged-care provider. It also records in-patient mortality. The mode of a patient's arrival and departure can influence cost, and is not captured by the DRG system. Moreover this is generally beyond the hospital's control

[10] Only relevant for some groups, and generally only in binary terms (e.g., whether a patient is or isn't over the age of 70)
[11] Daidone and Street (2013)

and depends on a provider's position in the system. If, for example, a hospital is a provider of last resort in a jurisdiction, then they may receive an above-average level of complicated transfers. Similarly, if a hospital happens to be located in close proximity to a large residential aged care facility, this may allow the hospital to discharge older patients more quickly, reducing costs.

We also include hours of mechanical ventilation as a proxy for very severe and complex health needs (as the variable ICU hours was not available).

Less direct indicators of patient complexity

Beyond the factors already described, different hospitals may still find themselves treating populations that exhibit characteristics associated with more expensive care. While we did not expect that many of these factors were likely to make a large difference in terms of hospital rankings, in the interests of quantifying unexplained cost differences with as much accuracy and fairness as possible, we sought to include statistical controls that met our three criteria.

In contrast to some work in the hospital efficiency literature[12], we decided (in advance of our analysis) not to define length of stay as a driver of legitimate and unavoidable variation. Although this variable may provide useful information as to the health condition of the patient upon presentation, our view was that:

a) much of this will be captured by other controls; and

b) variation in length of stay could well be driven by variation in costs that are unrelated to the provision of care (e.g. poor patient flow). Controlling for this variable may therefore hamstring our ability to identify hospitals that perform well on managing these avoidable costs.

[12] e.g. Gutacker, *et al.* (2011)

Table 1 – list of patient-related factors, beyond hospital control, that drive costs

Variable	Description	Examples in the literature
Direct indicators of patient complexity		
DRG	Diagnosis Related Groups	Hvenegaard *et al.* (2011) [among many others]
Age	Patient age (either in years, or grouped into age buckets)	Peltola (2012) [among many others]
Sex	Patient's gender	Bestawros A *et al.* (2005) Dormont and Milcent (2004)
Charlson score	A statistical measure of comorbidity	Macario *et al.* (1997)
Regional/Rural/ Remote	The location of a patient's home address	IHPA^ Productivity Commission (2010)
Indigenous Status	Whether or not patient is indigenous	Malyon *et al.* (2013), IHPA
Specialisation	Treatment requiring highly complex care, not well defined by DRG groups	Diadone and Street (2013)
Paediatric specialisation	A special case of specialisation	IHPA
Separation and admission mode	Describes how the patient was admitted and discharged (e.g. transferred from elsewhere). Includes mortality.	Gutacker, N. *et al.* (2011) Productivity Commission (2010)
HMV	Number of hours a patient spent on mechanical ventilation	Wilke and Grube (2010) IHPA
Less direct indicators of patient complexity		
Access to GP	Access to GP in a patient's medicare local area	Breadon and Duckett (2013)
Smoking*	Likelihood of patient being a smoker	Warner (2006)
SEIFA*	Socio-economic Index of Advantage/Disadvantage	Gaughan *et al.* (2012) Cookson and Laudicella (2011)
Physical inactivity*	Likelihood of patient being physically inactive[13]	**
Alcohol usage*	Likelihood of patient having 'harmful' level of alcohol use[14]	**

^Independent Hospital Pricing Authority [these are areas where the Authority's analysis has suggested that the variable in question does indeed influence cost]
*defined at the neighbourhood (SLA) level. See section 2 for a full discussion of data.
**these variables did not meet our criteria in that we didn't find good evidence that they were linked to acute-care costs, but we included them in an effort to be as thorough as possible in accounting for legitimate and unavoidable drivers of cost.

Provider Factors

[13] According to the Social Health Atlas of Australia's definition. See discussion in the data section for references.
[14] Again, we follow use the Social Health Atlas of Australia's definition. See section 2 for a description of data.

Scale and scope

In addition to patient-centred factors, in principle there may be some provider characteristics associated with higher costs that fall under the banner of 'unavoidable and legitimate'. It's possible, for example, that scale effects may exist enabling a large hospital to provide comparable care at a lower cost than a small hospital, regardless of any factors within the respective hospitals' control. Were this the case, interpreting cost differences between large and small hospitals in terms of 'performance' would not be justified, and scale ought to be included as a legitimate and unavoidable driver of cost. Similarly, the extent to which a hospital is focussed (i.e. has a limited scope of services) in providing particular types of care may enable them to channel greater resources on refining processes and developing expertise such that they can deliver care at a lower cost.

Whether scale does in fact systematically drive costs is not clear, with some research suggesting that – once very small hospitals have been removed from analysis - the relationship is weak and non-linear. In a study of NSW hospitals, for example, Wang and Mahmood (2000) found that scale economies did exist for small hospitals, but that large hospitals suffered from diseconomies. In more recent work looking at English hospitals, Gaughan *et al.* (2012) found that across a wide range of treatments hospital characteristics (including both scale and scope) were not significant predictors of cost variation.[15]

Although the evidence of a systematic relationship does not appear to be strong, in the first instance, we err on the side of both fairness to providers and conservatism in our estimates of unexplained cost, and include both scale and scope in the definition of legitimate and unavoidable cost drivers.[16]

Caveat: input price variation

Input prices may vary from hospital to hospital, especially with respect to wages. To the extent that these prices are externally determined, these differences should be classified as 'unavoidable and legitimate'.[17] While we acknowledge that this is important[18], we were not able to match hospital level information on input prices to our main database of patient costs.

We were able to analyse state-level wage information, which is included in *Controlling costly care* (see section 2). However, variation in wages at the state level was not defined as 'unavoidable and legitimate' as states have significant influence in setting these prices.

1.3 The issue of quality and health outcomes

[15] See Gaughan*, et al.* (2012). Note that there were a couple of exceptions. For example, hospitals that had a broader scope of services tended to have lower appendectomy costs.
[16] Note that we couldn't get good data on whether or not a hospital was a 'teaching hospital', nor was there comprehensive data on whether hospitals had specialist units.
[17] In the UK, this factor is explicitly accounted for in funding formulae by the Market Forces Factor adjustment (which discounts total costs in some hospitals by up to 29%, with an average discount of about 5%). See NHS (2012) p.6
[18] Indeed, some research suggests that if any provider-level factor is significant in driving costs, it may be differences in input prices. Kristensen*, et al.* (2010) report that after patient variables had been controlled for, the strongest determinant of cost-variation in diabetes treatment was a factor price index.

In the absence of good information regarding the health outcome of each patient, one possibility open to us was to use the presence of an adverse event as a proxy.[19] Inclusion of adverse events in our cost models would have shifted the analysis away from an examination of 'unexplained cost control' and closer to a focus on 'efficiency'. We decided against this approach for 3 reasons.

First, and most importantly, the relationship between adverse events and cost is not known.[20] One might expect that reducing errors, for example by investing in strategies to limit the likelihood of falls, may be costly, suggesting an inverse relationship.[21] However, it's also the case that errors need fixing, implying that more adverse events result in higher costs. Given these ambiguous effects, it's likely that the relationship is complex.[22] By including adverse events in the analysis, the result may be that high-cost, high-adverse-event providers are evaluated as being more 'efficient' than they would be if adverse events were not considered. In other words, if at some point on the cost-curve the relationship between costs and adverse events is positive, then of two otherwise-identical hospitals, the one with higher levels of adverse events would be judged as 'more efficient'.

Second, the presence of an adverse event is quite a poor measure of health outcome, and provides a limited amount of information about the quality of care a patient received. Moreover, only around 9% of patients in the NHCDC experience an adverse event, which means that this 'quality' proxy treats care received by 91% of patients as identical. Given this, it may be a mistake to oversell the benefits of including this measure in an effort to shift the focus of the analysis to efficiency, rather than controlling costs.

Third, this dataset may well be of use in contributing to our understanding of how adverse events and costs relate. If adverse events were included as a control in our estimates of unexplained cost variation, then this would make a subsequent investigation of how the two relate more difficult, both conceptually and in terms of communication.

1.4 Defining levels of analysis

We focus on three levels of analysis.

- **States & Territories**
  Given that each State and Territory has different health policy settings (for example with respect to funding arrangements and wage rates), otherwise unexplained variation may well change systematically across states.

- **Hospitals**
  Much of the work on patient-level costs has looked at specific service areas (e.g. obstetrics).[23] This is partly due to concerns that analysis at higher levels of aggregation makes it more likely that production processes in different organisations will not be 'strictly

[19] For the purposes of this discussion, 'adverse events' and 'errors' are synonymous.
[20] Productivity Commission (2010), see review on p.126 for a relatively recent summary. Some research has linked increased hospital spending with better patient outcomes [Schreyögg and Stargardt (2010)] while other work has linked cost reductions to improved patient outcomes [McKay and Deily (2008)]. There is also evidence to suggest that both very high and very low-cost care may be associated with lower quality [Hvenegaard, *et al.* (2011)]. Meanwhile, researchers at the Dartmouth Institute in the US found large variations in hospital spending with no evidence that spending is connected to quality of care [Hossain (2009)].
[21] Gutacker, *et al.* (2011)
[22] Ibid.
[23] We call these "service related groups' or SRGs.

comparable'.[24] However, we see a clear policy benefit to including aggregate analysis (i.e. at the hospital level) – as this makes it possible to see an overall picture of the variation that exists within each state.

- **Service groups (which we call 'departments')**
  These represent particular areas of practice (e.g. orthopaedics), within a hospital. For the rest of this document, we refer to these pairings of a service group and a hospital as 'departments'.[25] This is an important level of clustering, especially with respect to our risk-adjustments. It's possible, for example, that the marginal cost impact of treating an indigenous patient may be different in dialysis relative to mental health. This is an issue we return to in the section 6.

[24] Olsen and Street (2008), p.672
[25] The mapping from DRG to these service groups is based on Victoria's Major Clinical Related Group (MCRG) system. There are 40 different groups. MCRGs are equivalent to Service Related Groups, discussed here http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129543826

2. Data

This section describes the datasets and transformations used in our analysis, along with how outliers were defined.

Overview of data selected for analysis

The core data for the analysis comes from the National Hospital Cost Data Collection (NHCDC). The Independent Hospital Pricing Authority [IHPA] granted Grattan access to 3 years' of data, covering 2008-09 to 2010-11. These data had not previously been released outside of government. In *Controlling costly care* we focus on the final year, 2010-11, as we believe that these data are the most relevant to current hospital practices, and have the highest degree of consistency across jurisdictions.[26]

The data cover public hospitals only, and represent a total of 4,114,065 episodes of acute admitted care[27], around 78% of the total during 2010-11.[28]

From this initial sample of 4,114,065 separations, the data set used in the analysis contains 3,930,109 (96% of the original sample). We defined 4 different types of exclusions:

1. very small hospitals, i.e. those with fewer than 2,000 weighted units of activity in 2010-11[29] (n=129,907)
2. patients with a total cost of less than $20 (n=7,127)[30]
3. cases where the recorded hours of mechanical ventilation were greater than the hours in a year (n=12)
4. patients with 'outlier' total costs. Whether or not a separation was an outlier was determined based on their total cost relative to other patients in the same DRG. Given the absence of a compelling rationale in the cost-efficiency literature to identify outliers we adopted two different procedures, and tested their impact on our final results (reported in section 5). These two approaches are outlined below. Note that the primary results reported in *Controlling costly care* use approach 1.

   ### Approach 1: outliers defined by ±3 standard deviations from the DRG mean cost

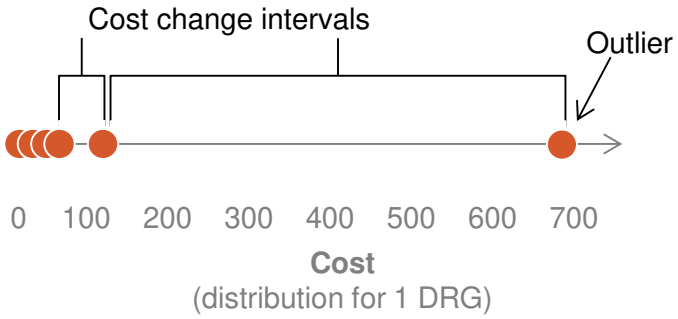
   This follows the method used by Gaughan *et al.* (2012). For each DRG the mean and standard deviation of the cost distribution is calculated. Trim points are defined as being beyond 3 standard deviations of the mean. Any episode with a cost beyond the trim point is defined as an outlier, and excluded from the analysis (n=46,910)

   ### Approach 2 (used in sensitivity analysis only): outliers defined by a variant of IHPA's outlier procedure

[26] See section 5.1 for analysis of how hospital estimates changed from 2009-10 to 2010-11.
[27] The data does **not** include other types of episodes which are nonetheless included in the NHCDC, e.g. emergency department separations, outpatient episodes, and sub-acute care (rehabilitation, palliative care and so on).
[28] Total number of public hospital separations came from AIHW (2013), Table 2.7
[29] We use NWAU12 figures from the Independent Hospital Pricing Authority. To be on the safe side with respect to small-volume hospitals, we also remove three hospitals that have fewer than 2,000 separations but more than 2,000 units of weighted activity.
[30] This follows Independent Hospital Pricing Authority (2013) p.12

This approach adopts the technique IHPA uses in defining outliers for National Efficient Price determination, as illustrated in Figure 1. If the 'cost change interval' – i.e. the gap between a patient's cost and the nearest cost of another patient within the same DRG – is a jump of greater than 500%, then IHPA defines the patient as an 'outlier'.

Figure 1 – illustration of IHPA's outlier identification procedure



Our analytical purpose clearly differs from IHPA's, in that rather than setting a system-wide price, our goal is to benchmark the variation in costs across hospitals and, to a lesser extent, states. In this context, we were concerned that the IHPA definition was too strict, and that hospitals' performance may be overly influenced by the random appearance or otherwise of a very costly patient. This was especially true in light of our decision to focus on the most recent year of data (rather than using all three years'). **As such, we loosened the criterion, such that if the jump was greater than 100%, the separation was defined as an outlier.** In total, this represented 235 separations. Even then, in the interests of conservatism in our estimates and fairness to providers, this approach to outlier determination was secondary, and not used in the analysis reported in *Controlling costly care.*

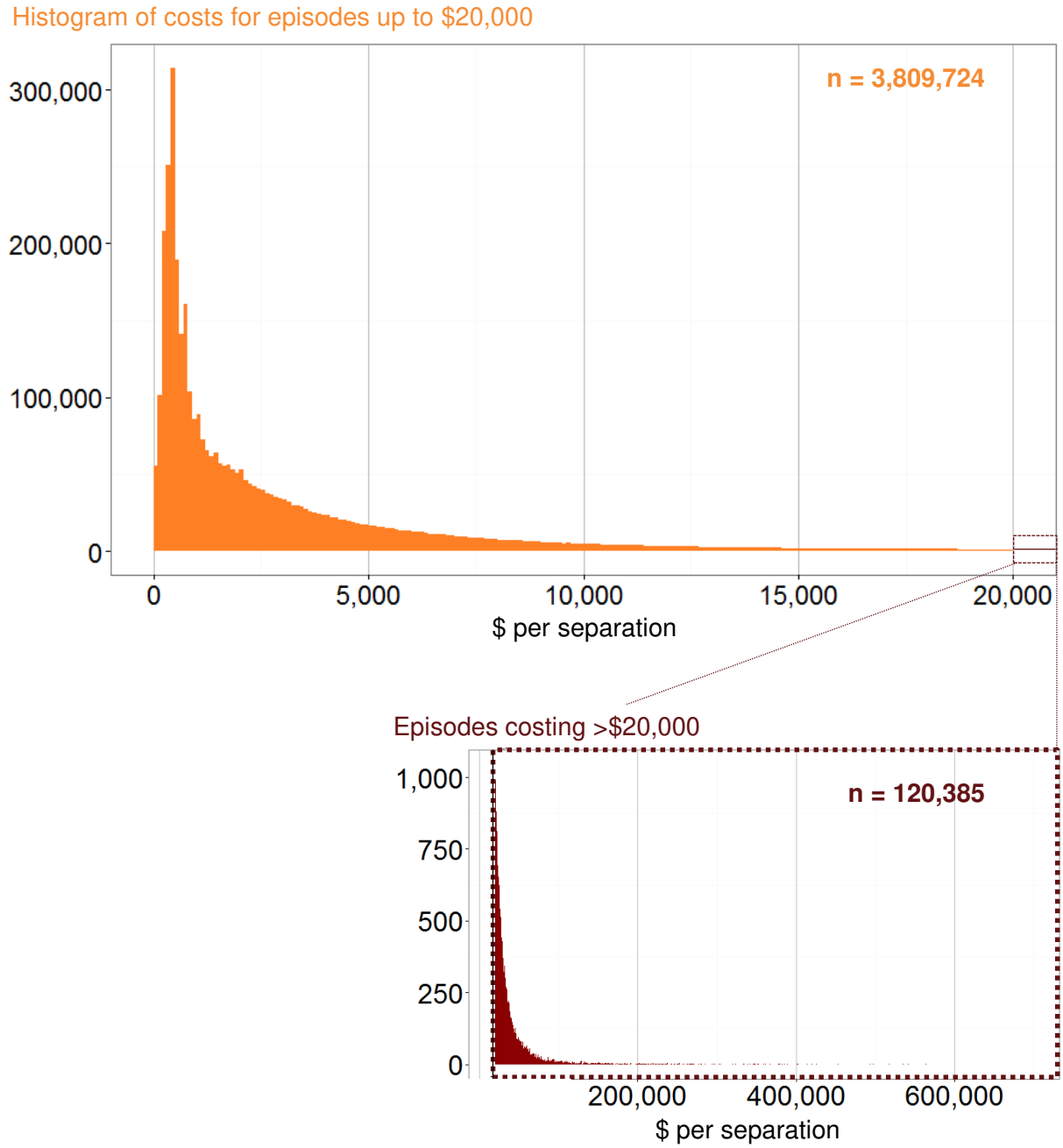
Cost data

The NHCDC provides patient-level information for various cost types – nursing, imaging, and so on.[31] These costs are generated by computerised systems that generate a cost for each episode of care and are subjected to quality audits.[32] All costs buckets, with the exception of depreciation (which was treated differently across states and therefore excluded from the analysis), were summed to yield a total cost for each episode of care. The mean cost was $3,972 and the s.d. was $9,299.

As is typical with health cost data, the distribution of total costs exhibits a strong positive skew. This is illustrated in Figure 2, which presents a histogram of total costs (with the data partitioned into two panels so that the full distribution can be viewed).

[31] For a full list, and constituent elements of each cost bucket, see Australian Government Department of Health and Ageing (2007), chapter 8.
[32] Jackson (2000)

Figure 2 – histogram of total costs (split into two panels)



Positive skewness was present in the cost distribution of all DRGs to a greater or lesser degree.[33] The effect this might have on our modelling is discussed in section 3.1.

Direct indicators of patient complexity

In addition to the cost data, the NHCDC includes information about patients' diagnoses and care. The main measure we use here is a patient's Diagnosis Related Group (6th edition).

[33] There was one exception among the 705 DRGs, namely "Z60C' (sameday rehabilitation) which had 10 patients coded to it in 2010-11.

Given the large number of DRGs (705), rather than include a dummy for each group, we calculate an index **(drg_index)**. The index is equal to a DRGs average cost divided by the overall mean total cost.[34]

The NHCDC data also has up to 30 ICD-AM-10th edition diagnosis codes and we use these to generate a weighted **Charlson** index for each patient. We follow the practice outlined in Quan *et al.* (2011) and exclude primary diagnosis codes as well as those that are flagged as being 'hospital acquired'. In other words, we focus on conditions that, to the best of our knowledge, were comorbidities upon presentation.[35]

The ICD-10 diagnostic codes are also used to create a variable that flags whether an admission required **specialised** care. A list of ICD-10 codes associated with highly complex care was sourced from the NHS Specialised Services National Definitions.[36] The NHS definitions consist of two categories: 'specialised' and 'maybe specialised'. We opted for the more inclusive list, which consisted of 727 ICD-10 codes, across 6 areas of care.[37] A dummy variable was created, (0 = non-specialisation, 1 = specialisation) for any admissions in which one of the listed specialised ICD-10 codes were present.[38]

For paediatric specialised care, we followed IHPA's approach[39] of adjusting only for:

- patients under 17
- not in a newborn or neonate DRG
- being treated in a specialist women's and children's hospital.

We created a numeric variable for **paed_adjustment.** If separations satisfied the above three conditions *and* they were in a DRG that IHPA lists in its paediatric adjustments[40], then our paediatric specialist adjustment took on the value of IHPA's payment adjustment factor.

To account for demographic factors affecting acute-care costs, we use NHCDC information on sex and age. Because of the potential non-linearities in how age might affect the cost of treatment, we divide the variable into 11 categories: two for childhood ('<1 year', '1-16') and

[34] Hvenegaard*, et al.* (2009). Note that this is essentially identical to including dummy variables for each of the 705 DRGs.
[35] In terms of mapping conditions (e.g. 'congestive heart failure') to ICD-AM 10th edition codes, we use the list presented in Quan*, et al.* (2005). Note that we considered using the newer list from Sundararajan*, et al.* (2007) which was developed with Australian data – but Sundararajan's paper reports that this algorithm is outperformed by Quan*, et al.* (2005). In terms of weightings, given the likely changes in health care since the 1980s – especially with respect to chronic disease management, changing demographics, and improvements in treatments and technology – we favour the updated weightings as reported in Quan*, et al.* (2011). While this update was based on Canadian data, some external validation was done in 5 other countries, including Australia.
[36] NHS (2013) [see http://www.specialisedservices.nhs.uk/info/specialised-services-national-definitions]. Note that we only coded ICD-10 diagnostic codes, and not OPCS 4.5 codes (i.e. surgical codes) as we were unable to find a correspondence for the latter and the procedure codes present in the NHCDC.
[37] Cancer, spinal, neurosciences, cystic fibrosis, rheumatology, colorectal, infectious diseases. Unfortunately we were unable to find a mapping between the surgical codes listed by the NHS as specialised, and the ACHI codes in our data.
[38] We also investigated the possibility of including additional variables to reflect other elements of specialisation (e.g. ICU hours) but data here were incomplete. In the case of ICU hours (which is an IHPA adjustment) the data was not present in 2008-09 and 2009-10, and was not available for Victoria, WA and Tas in 2010-11. This being the case we opted for hours of mechanical ventilation as a substitute.
[39] Independent Hospital Pricing Authority (2012) p.9
[40] Ibid.

blocks of 10 years thereafter.[41] We also make use of the NHCDC variables indicating how patients were admitted (e.g. via a transfer from another hospital) and their mode of separation (e.g. discharge home, mortality).[42]

Less direct indicators of patient complexity

The NHCDC has information on patients' statistical local area (SLA). Using this geographic marker, Grattan generated a number of other indicators that may affect the average complexity of care required. The primary data source was the Public Health Information Development Unit's *Social Health Atlas of Australia.*[43] Data was gathered on 3 variables:

1. **smoking** rate (estimated rate per 100 people over the age of 18, in 2007-08);
2. harmful use of **alcohol** (estimated rate per 100 people over the age of 18, in 2007-08);
3. **physical inactivity** (estimated rate per 100 people over the age of 15, in 2007-08).

We also make use of the ABS's Index of Relative **Socio-Economic Advantage and Disadvantage** (for 2011), defined at the postcode level.[44]

Last, we generate a variable that measures **access to primary care**, defined by the number of Full-time Work Equivalent GPs in a given Medicare Local (i.e. the main unit of organisation for primary care in Australia).[45]

Provider-variables

The **scale** of each hospital is defined by the number of episodes in 2010-11.[46] In the results reported in this document, scale is generally defined in terms of thousands of episodes so that coefficients are more easily interpretable.

Scope was defined as using an Information Theory Index (ITI), which measures how similar a hospital's mix of patients is compared to the system as a whole.[47] We calculate this in terms of how patients are distributed across different departments (e.g. orthopaedics, rheumatology etc.). If there are N different departments, the scope index for provider h is calculated as:

$$scope_h = \sum_{j=1}^{N} p_{jh} \ln\left(\frac{p_{jh}}{\varphi_j}\right)$$

[41] This is common in the literature; see Gaughan, *et al.* (2012) for an example. The cut-off for 'childhood' was based on IHPA's threshold for paediatric payments, Independent Hospital Pricing Authority (2012) p.9. The remaining groupings had no clinical meaning, but were chosen in an effort to balance granularity and model parsimony.
[42] For more information on the definitions of these variables, see the Meteor website (http://www.aihw.gov.au/meteor/).
[43] PHIDU (2013)
[44] ABS (2011)
[45] Breadon and Duckett (2013)
[46] Note that during the course of our modelling we also defined scale at the department level, but found that this proved to be an even less useful predictor of cost variation.
[47] The index was originally designed to measure gains in information, but the index it has a reasonably long history as a measure of hospital specialisation (dating back to at least the late 1980s). See Kobel and Theurl (2013) for discussion.

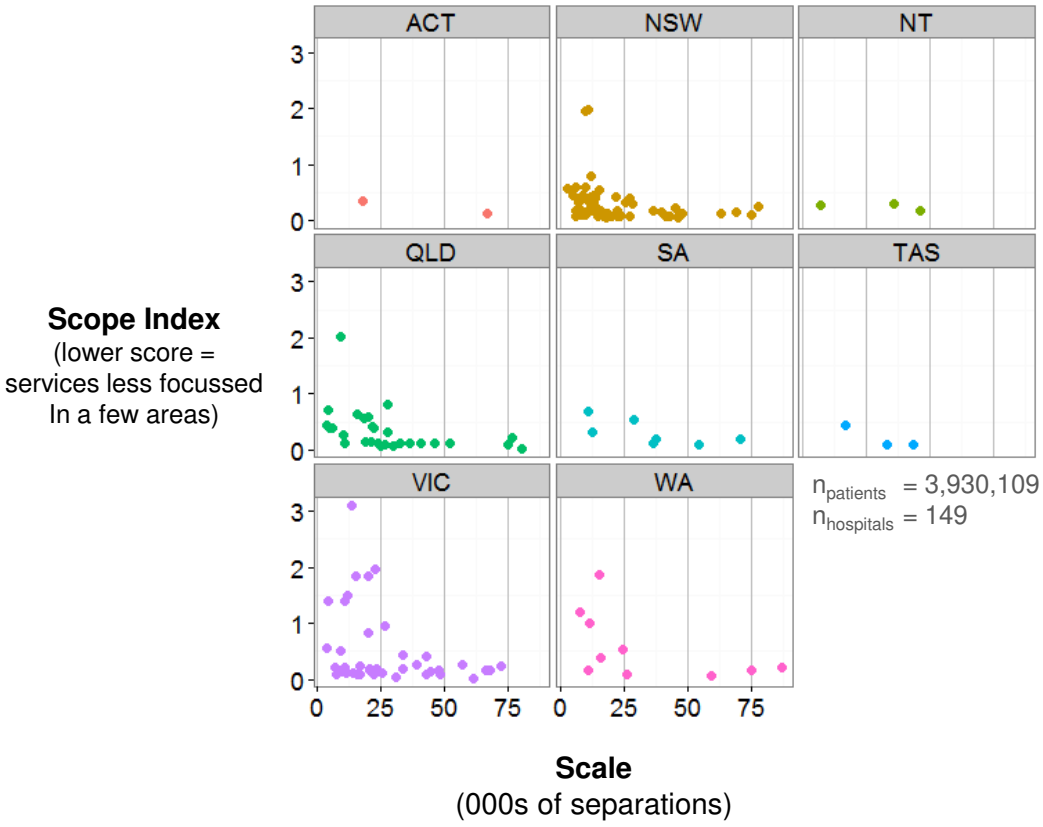
Where:

- p_{jh} is the proportion of provider h's patients that are in department *j*,
- φ_j is the proportion of all patients that are treated in service area *j*.

A provider with a scope index of zero has the exact same mix of patients as a summation of all public hospitals contributing to the NHCDC data. As the concentration of services increases, so does the scope index. The index has no upper bound but in practice even the most focussed hospitals (for example providers in which 95% of the separations are coded to dialysis) have a scope index of around 3. Figure 3 provides an overview of scope scores relative to hospital scale.[48]

Overall summary statistics for our control variables are presented in Table 2.

Figure 3 – overview of scope (relative to scale)



[48] Note that given the skewed distributions presented in Figure 3, we did consider taking the log of this measure to better differentiate small distinctions in specialisation. However, in the modelling this proved to be a poorer predictor of cost variation.

Table 2 - Summary and descriptive statistics of 'legitimate and unavoidable' cost drivers

Direct indicators of patient complexity (patient-level)	Mean	s.d.
Admission mode:		
Adm_transfer = 1 if transferred from another hospital	0.00	0.05
Adm_stat = 1 if 'change of episode type'	0.04	0.19
Age (in years)	51.71	24.37
Charlson (weighted Charlson score)	0.42	1.43
DRG_index	1.00	1.85
Hours of mechanical ventilation (HMV)	0.65	14.82
Indigenous status (= 1 if either Aboriginal or Torres Strait Islander)	0.05	0.23
Rurality		
Inner_reg = 1 if living in 'inner regional Australia'	0.18	0.39
Outer_reg = 1 if living in 'outer regional Australia'	0.08	0.28
Remote = 1 if living in 'remote Australia' or 'very remote Australia'	0.02	0.15
Separation mode		
Sep_transfer = 1 if transferred to another hospital, aged care, or other healthcare provider	0.05	0.22
Sep_stat = 1 if statistical discharge - type change	0.01	0.11
Sep_died = 1 if patient died	0.01	0.09
Sep_other = 1 for other seps apart from discharge to regular accommodation (e.g. left against medical advice)	0.01	0.09
Sex (= 1 if female, 0 if male)	0.50	0.50
Specialisation (=1 if ICD code on NHS specialisation list)	0.04	0.19
Paediatric Specialisation (adjustment factor, based on IHPA)	0.03	0.19
Less direct indicators of patient complexity (neighbourhood level)		
Access to primary care (Full Time GP's per 1,000 people)	99.52	13.81
Alcohol over-use rate (per 100 people)	5.32	1.51
Physical inactivity rate (per 100 people)	35.05	5.04
Smoking rate (per 100 people)	20.86	3.42
Index of relative advantage/disadvantage	987.38	76.22
Provider factors		
Scale (of hospital; seps per year)	42,171	23,132
Scope (Information Theory Index of hospital)	0.27	0.37

The correlation between the variables is illustrated over the page in Table 3. As might be expected, the strongest correlates are the cost-drivers defined at the neighbourhood (rather than patient) level. The coefficients estimating the marginal impact these variables have on acute-care costs should be interpreted with caution. There's also a strong relationship between the DRG_index and hours of mechanical ventilation (HMV), which again is expected given that they are both extremely strong predictors of cost.

Table 3 – correlation of unavoidable and legitimate cost drivers (n=3,933,267)

	Adm Stat	*Adm Transfer*	*Age*	*Charlson*	*DRG Index*	*HMV*	*Indig. Status*	*Inner reg*	*Outer reg*	*Rem ote*	*Sep Died*	*Sep Other*	*Sep Stat*	*Sep Transfer*	*Sex*	*Special.*	*paed spec.*	*Scale*	*Scope*	*ACCESS*	*Alcohol*	*Physical inactivity*	*Smoke*	*SEIFA*
Adm Stat		-0.01	0.02	0.01	0.03	0.00	0.00	0.00	0.00	0.01	0.01	0.22	0.02	0.00	0.00	0.02	-0.01	0.01	-0.01	-0.01	0.00	0.00	0.00	0.00
Adm Transfer	-0.01		-0.03	0.01	0.15	0.06	0.00	0.02	0.03	0.02	0.15	0.04	0.03	0.01	-0.02	0.03	0.01	0.01	0.01	-0.03	0.07	0.02	0.05	-0.03
Age	0.02	-0.03		0.14	0.00	-0.01	-0.09	0.00	-0.02	-0.05	0.10	0.10	0.08	-0.04	-0.07	0.03	-0.31	0.10	-0.10	0.03	-0.02	-0.02	-0.05	0.01
Charlson	0.01	0.01	0.14		0.10	0.02	-0.03	0.01	0.00	-0.01	0.04	0.07	0.11	-0.01	-0.01	0.09	-0.02	0.05	0.03	-0.01	-0.02	-0.02	-0.01	0.01
DRG Index	0.03	0.15	0.00	0.10		0.53	-0.03	0.01	0.00	-0.01	0.15	0.13	0.16	0.01	0.00	0.17	0.00	0.04	0.00	0.01	0.02	0.00	0.00	0.00
HMV	0.00	0.06	-0.01	0.02	0.53		0.00	0.00	0.00	0.00	0.05	0.03	0.12	0.00	-0.01	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Indigenous Status	0.00	0.00	-0.09	-0.03	-0.03	0.00		-0.01	0.16	0.40	-0.02	-0.02	-0.01	0.05	0.03	-0.02	-0.01	-0.02	-0.04	-0.18	0.15	0.04	0.20	-0.21
Inner reg	0.00	0.02	0.00	0.01	0.01	0.00	-0.01		-0.14	-0.07	0.01	0.00	0.00	-0.01	0.00	-0.01	-0.03	-0.27	-0.11	0.00	0.40	0.09	0.26	-0.19
Outer reg	0.00	0.03	-0.02	0.00	0.00	0.00	0.16	-0.14		-0.05	0.00	0.00	0.00	0.00	0.00	0.00	-0.02	-0.10	-0.08	-0.20	0.33	0.18	0.34	-0.13
Remote	0.01	0.02	-0.05	-0.01	-0.01	0.00	0.40	-0.07	-0.05		0.01	-0.01	-0.01	0.04	0.01	0.00	-0.01	-0.02	-0.01	-0.21	0.11	0.01	0.22	-0.28
Sep Died	0.01	0.15	0.10	0.04	0.15	0.05	-0.02	0.01	0.00	0.01		-0.03	-0.02	-0.02	0.00	0.06	-0.03	-0.02	-0.02	0.00	0.01	-0.01	-0.01	0.00
Sep Other	0.22	0.04	0.10	0.07	0.13	0.03	-0.02	0.00	0.00	-0.01	-0.03		-0.01	-0.01	0.01	0.09	-0.02	0.01	-0.02	-0.01	0.00	-0.01	-0.02	0.01
Sep Stat	0.02	0.03	0.08	0.11	0.16	0.12	-0.01	0.00	0.00	-0.01	-0.02	-0.01		-0.01	-0.01	0.04	-0.01	0.01	-0.01	0.01	0.00	0.00	-0.01	0.00
Sep Transfer	0.00	0.01	-0.04	-0.01	0.01	0.00	0.05	-0.01	0.00	0.04	-0.02	-0.01	-0.01		-0.01	0.00	-0.01	0.00	-0.01	0.00	0.01	0.01	0.01	-0.02
Sex	0.00	-0.02	-0.07	-0.01	0.00	-0.01	0.03	0.00	0.00	0.01	0.00	0.01	-0.01	-0.01		0.00	-0.02	-0.06	0.07	0.00	0.00	0.00	0.01	-0.01
Specialisation	0.02	0.03	0.03	0.09	0.17	0.05	-0.02	-0.01	0.00	0.00	0.06	0.09	0.04	0.00	0.00		0.03	0.03	-0.01	0.01	-0.01	-0.01	-0.01	0.01
Paediatric Specialisation	-0.01	0.01	-0.31	-0.02	0.00	0.00	-0.01	-0.03	-0.02	-0.01	-0.03	-0.02	-0.01	-0.01	-0.02	0.03		-0.13	0.11	0.00	-0.02	-0.01	-0.04	0.04
Scale	0.01	0.01	0.10	0.05	0.04	0.02	-0.02	-0.27	-0.10	-0.02	-0.02	0.01	0.01	0.00	-0.06	0.03	-0.13		-0.35	-0.02	-0.20	-0.11	-0.19	0.14
Scope	-0.01	0.01	-0.10	0.03	0.00	0.00	-0.04	-0.11	-0.08	-0.01	-0.02	-0.02	-0.01	-0.01	0.07	-0.01	0.11	-0.35		0.01	-0.14	-0.06	-0.11	0.07
Access	-0.01	-0.03	0.03	-0.01	0.01	0.00	-0.18	0.00	-0.20	-0.21	0.00	-0.01	0.01	0.00	0.00	0.01	0.00	-0.02	0.01		-0.20	0.28	-0.02	-0.11
Alcohol	0.00	0.07	-0.02	-0.02	0.02	0.00	0.15	0.40	0.33	0.11	0.01	0.00	0.00	0.01	0.00	-0.01	-0.02	-0.20	-0.14	-0.20		-0.01	0.44	-0.16
Physical inactivity	0.00	0.02	-0.02	-0.02	0.00	0.00	0.04	0.09	0.18	0.01	-0.01	-0.01	0.00	0.01	0.00	-0.01	-0.01	-0.11	-0.06	0.28	-0.01		0.69	-0.66
Smoking	0.00	0.05	-0.05	-0.01	0.00	0.00	0.20	0.26	0.34	0.22	-0.01	-0.02	-0.01	0.01	0.01	-0.01	-0.04	-0.19	-0.11	-0.02	0.44	0.69		-0.69
SEIFA	0.00	-0.03	0.01	0.01	0.00	0.00	-0.21	-0.19	-0.13	-0.28	0.00	0.01	0.00	-0.02	-0.01	0.01	0.04	0.14	0.07	-0.11	-0.16	-0.66	-0.69	

3. How much cost variation is due to unavoidable and legitimate cost drivers?

This section explores the extent to which cost variation can be explained by factors that are beyond hospitals' control.

Estimating the impact of unavoidable and legitimate cost drivers

To understand the extent to which cost variation can be explained by unavoidable and legitimate cost drivers, we estimated equation 2:

$$c_i = \boldsymbol{X_i'}\alpha + \boldsymbol{Z_h'}\delta + \varepsilon_i \qquad (2)$$

Where

- c_i is the cost of patient i
- **X** is a set of patient factors controlling for unavoidable and legitimate cost differences related to morbidity and treatment (defined at both the patient and neighbourhood levels; see Table 1 for full list)
- **Z** is a set of provider factors, representing hospital scale and scope.
- ε_i is a normally distributed error term; $\varepsilon_i \sim N(0, \sigma_e^2)$.

The results are outlined in Table 5 (on page 21). Unavoidable and legitimate drivers of cost variation account for around 67% of total variation in costs between patients.

In general, the coefficients have the expected signs. Cost increased with: DRG_index, co-morbidity (as measured by Charlson), hours of mechanical ventilation, rurality (increasing in cost from inner regional through to remote), specialisation of care (both adult and paediatric), age (with the exception of children less than a year old, the most expensive were 77-86), alcohol use, and physical inactivity. Admissions were cheaper if the patient lived in an area with more GPs, and if the separation ended in death.

There were a couple of surprises:

- the coefficient on indigenous status was small, negative and insignificant
- cost decreased as the likelihood that a patient smoked increased

Also of note were the coefficients on the 'provider factors': scale and scope. Both the coefficients went against our (weak) prior expectations.

In the case of scale, an increase in hospital size was associated with more expensive care, albeit by the modest figure of $6 for an increase of 1,000 separations per annum. This seems to suggest that, for the hospitals in our sample, diseconomies of scale may characterise the relationship between size and efficiency. Further analysis suggested that although these

diseconomies may especially be present in larger hospitals, the non-linear effects probably aren't substantial.[49]

We also examined the hypothesis that there could be an interaction between scale and scope. Specifically, we sought to understand whether small hospitals in rural areas providing a wide range of services had systematically higher costs (and might therefore be disadvantaged by not including an interaction term to recognise this legitimate cost driver). This did not appear to be the case. In modelling where an interaction term was included, the coefficient on this variable was negative, indicating that diseconomies of scale were in fact largest in full-service hospitals.[50]

As for the coefficient on scope, a 1 unit increase in the scope index (which corresponds to a substantial increase in the extent to which a hospital's services are focussed in particular areas) was associated with a $129 increase in cost per separation.

In both cases, while it's possible that these results reflect genuine cost increases associated with greater scale or increased focus on particular areas of care. It's also possible that the coefficients may reflect shortcomings in patient-level controls. Specifically, larger hospitals may be associated with higher costs as their position in the health system requires them to deal with higher-complexity cases (even after all other observable patient-factors have been accounted for). The same may be true for hospitals that offer a narrower range of services.

How much cost variation can be explained in different areas of care?

We might expect that patient and hospital characteristics may have differential marginal effects on costs depending on the area of care. The extra cost of treating a patient with comorbidities, for example, may be greater in the case of transplants compared to dialysis. To better understand how risk-adjustments differ across different care types, equation 2 was estimated for each of the 40 service groups in our data. In each case, Table 4 illustrates the level of cost variation explained by legitimate factors across different service groups.

[49] In models where we included a scale^2 variable, the coefficient had a positive and significant coefficient. That said, the magnitude was small (around 2 cents increase in cost per patient associated with a unit increase in squared^2.
[50] We checked this again after hospital benchmarking, but it did not appear to be the case that hospitals with a broad scope and low scale had systematically low measured performance. See chart in Appendix A

Table 4 – what proportion of costs in different service areas are explained by legitimate and unavoidable drivers?

Service group	Adjusted R^2	n	s.d. of cost
Extensive Burns	75%	3,169	$28,938
Interventional Cardiology	68%	68,342	$11,366
Transplantation	67%	1,109	$61,499
Qualified Neonate	66%	50,200	$23,559
Colorectal Surgery	65%	36,916	$13,552
Orthopaedics	62%	256,805	$8,228
Haematology	61%	93,069	$10,459
Neurosurgery	59%	35,278	$15,800
Gynaecology	59%	120,437	$2,858
Plastic & Reconstructive Surgery	59%	75,273	$7,329
Breast Surgery	58%	15,927	$4,240
Cardiothoracic Surgery	58%	11,701	$23,560
Ear, Nose & Throat	57%	82,892	$3,368
Obstetrics	56%	250,268	$4,009
Urology	55%	97,333	$4,344
Diagnostic GI Endoscopy	55%	135,661	$3,719
Upper GIT Surgery	54%	52,748	$7,600
Vascular Surgery	53%	33,546	$13,409
Non Subspecialty Surgery	52%	242,406	$5,762
Head & Neck Surgery	52%	10,048	$5,999
Respiratory Medicine	51%	182,085	$8,148
Oncology	48%	53,731	$8,696
Tracheostomy	46%	8,447	$75,329
Clinical Cardiology	45%	191,307	$4,452
Renal Medicine	44%	34,679	$6,803
Unallocated	44%	3,693	$18,129
Neurology	43%	133,454	$6,635
Immunology & Infections	40%	78,369	$6,226
Rehabilitation Acute	38%	200	$10,169
Dermatology	37%	18,732	$3,161
Gastroenterology	37%	69,522	$4,452
Non Subspecialty Medicine	36%	200,495	$3,939
Rheumatology	35%	25,687	$3,964
Drug & Alcohol	34%	45,409	$4,549
Psychiatry	27%	85,370	$12,351
Endocrinology	26%	40,451	$4,841
Ophthalmology	24%	77,496	$2,035
Dentistry	8%	16,665	$1,422
Dialysis	7%	866,703	$246
Chemotherapy & Radiotherapy	4%	124,486	$1,110

Note that service areas where a low level of cost variation is explained by our legitimate and unavoidable cost drivers, there tends to be much less cost variation and very few unique DRGs. 'Chemotherapy and radiotherapy', for example, has only two separate DRGs, while dentistry has one.

Table 5 – Results of equation 2. Effect of unavoidable and legitimate variation

	Equation 2	
Direct indicators of patient complexity (patient-level)	Estimate	(s.e.)*
Admission mode:		
Adm_transfer = 1 if transferred from another hospital	$920	($152)
Adm_stat = 1 if 'change of episode type'	-$1,367	($482)
Age (base category is <1 years)		
1-16	-$482	($136)
17-26	-$354	($156)
27-36	-$281	($154)
37-46	-$380	($158)
47-56	-$354	($159)
57-66	-$297	($161)
67-76	-$206	($161)
77-86	-$195	($165)
87-96	-$390	($181)
97-116	-$905	($223)
Charlson (weighted Charlson score)	$51	($16)
DRG_index	$3,599	($89)
Hours of mechanical ventilation (HMV)	$94	($8)
Indigenous status (= 1 if either Aboriginal or Torres Strait Islander)	-$31	($45)
Paediatric specialisation	$543	($99)
Rurality		
Inner_reg = 1 if living in 'inner regional Australia'	$92	($67)
Outer_reg = 1 if living in 'outer regional Australia'	$129	($103)
Remote = 1 if living in 'remote Australia' or 'very remote Australia'	$657	($129)
Separation mode		
Sep_transfer = 1 if transferred to healthcare provider (e.g. another hospital)	$158	($134)
Sep_stat = 1 if statistical discharge - type change	$2,854	($350)
Sep_died = 1 if patient died	-$2,403	($225)
Sep_other = 1 for other seps apart from discharge to regular accommodation	-$514	($100)
Sex (= 1 if female, 0 if male)	$87	($18)
Specialisation (=1 if ICD code on NHS specialisation list)	$1,940	($101)
Less direct indicators of patient complexity (neighbourhood level)		
Access to primary care (Full Time GP's per 1,000 people)	-$6	($3)
Alcohol over-use rate (per 100 people)	$130	($24)
Physical inactivity rate (per 100 people)	$30	($8)
Index of relative advantage/disadvantage	$0	($0)
Smoking rate (per 100 people)	-$32	($11)
Provider factors		
Scale (of hospital; 1,000s seps per year)	$6	($2)
Scope (Information Theory Index of hospital)	$189	($96)
Summary		
n = 3,930,109		
Residual s.e. = $5,333		
Adj R^2 = 0.671		

*Cluster robust standard errors

3.1 Cost variation and functional form

Given the skewness present in hospital cost data it is relatively common in the literature to find analyses where cost data have been transformed, usually by modelling the log of costs.

Notwithstanding this common practice, Lumley *et al.* (2002) argue that in the presence of sufficiently large sample sizes, linear models with normally distributed errors may perform as well as models with transformed dependent variables even if the dependent variable exhibits very strong skewness.[51] With around 4 million records, our dataset quite comfortably meets the indicative threshold for 'sufficiently large' in Lumley *et al.* (2002).[52] Moreover, keeping costs in dollar units also has the advantage of easy interpretability.

To check that we are not compromising predictive accuracy for convenience, we compared the model specified by equation 2 with an equivalent model where log cost is the dependent variable (i.e. model 2b).

$$\log(c_i) = \boldsymbol{X_i'}\alpha + \boldsymbol{Z_h'}\delta + \varepsilon_i \qquad (2b)$$

The core result is presented in Table 6. As has previously been observed in comparison of hospital cost models with log transform (e.g. Deb and Burgess (2003), and Daidone and Street (2013)) the log model performed less well than the OLS in terms of explaining cost variation. This encouraged us to continue on with an un-transformed measure of cost.[53]

Table 6 – comparing models of raw costs and log costs

	Raw estimate (equation 2)	Log transform (equation 2b)
Adjusted R^2	0.6711	0.6238

[51] Lumley, *et al.* (2002). See also Deb and Burgess (2003)
[52] Lumley, *et al.* (2002) suggest that as few as 500 observations may be enough to justify using a linear model in the presence of a skewed dependent variable.
[53] As a final check, we also performed the cross-validation procedure laid out in Deb and Burgess (2003). Again, we found that the linear model out-performed the log transform. Results are available on request.

4. Benchmarking performance

This section has three parts:

- the first lays out our general approach to benchmarking cost performance across states and hospitals. It discusses the relative merits of random and fixed-effects estimation and introduces some notation.
- the second explores the impact that adjusting for legitimate factors has on state and hospital effects
- the third presents the output of the main benchmarking models

4.1 General approach to benchmarking cost performance

Much of the empirical hospital benchmarking literature has been based on provider-level data.[54] While useful, analysis of hospital-level data has well-established limitations.[55] Where possible, patient-level analysis is preferable as it accounts much more comprehensively for differences in patients, allowing for fairer benchmarking. Parametric analysis of patient-level data also maximises the ability to investigate cost drivers that may affect large groups of patients, be they the scope of services a hospital provides, or policy choices at the state level. The value of this method – i.e. assessing healthcare performance by explicitly modelling hierarchical effects using patient-level data – has long been acknowledged.[56] As more patient-level data has become available, this approach has become increasingly prevalent in the literature.[57] Given that Grattan has access to patient cost information, we believe that the best empirical strategy to estimating unexplained cost variation is to build on this growing body of research.

Overview of main models

Our data are clustered at a number of levels. To keep the analysis tractable, we focus on the two highest levels of aggregation: hospitals and states. Our ultimate aim is to estimate the systematic cost variation that exists at these levels that cannot be explained by legitimate and unavoidable cost drivers. This is reflected in our modelling approach. We estimate 2 different multi-level models:

Patient and State $$c_{is} = \boldsymbol{X_i'}\alpha + \boldsymbol{Z_h'}\delta + \psi_s + \varepsilon_{is} \quad (3a)$$

Patient and hospital $$c_{ih} = \boldsymbol{X_i'}\alpha + \boldsymbol{Z_h'}\delta + \gamma_h + \varepsilon_{ih} \quad (3b)$$

Where

- c_{ihs} is the cost of patient i (in hospital h, and State *s*).

[54] Within this literature, the main methodological debate centres on the relative merits of Data Envelopment Analysis and Stochastic Frontier Analysis. For an excellent summary see Hollingsworth and Peacock (2008).
[55] Newhouse (1994)
[56] Duncan, *et al.* (1998); Rice and Leyland (1996)
[57] Kristensen, *et al.* (2010); Kessler and McClellan (2001); Bradford, *et al.* (2001) Daidone and Street (2013); Hvenegaard, *et al.* (2009); Olsen and Street (2008); Dormont and Milcent (2004) Laudicella, *et al.* (2010) Gutacker, *et al.* (2011); Gaughan, *et al.* (2012)

- **X** is a set of patient factors controlling for unavoidable and legitimate cost differences related to patient needs

- **Z** is a set of provider factors, representing hospital scale and scope.

Variation that remains after controlling for differences in **X** and **Z** is described as 'unexplained'. We decompose this variation into 3 parts:

- ψ_s represents the ability of state *s* to control costs. One of these 'state effects' is estimated for each state and territory, and represents the average quantum of cost variation (after legitimate sources have been purged from the data) that can systematically be associated with a patient being treated in each state.[58] When estimated with random effects (our preferred approach, as discussed below) then $\psi_s \sim N(0, \sigma_\psi^2)$. When using fixed effects, we estimate a dummy for each state.

- γ_h represents the ability of hospital *h* to control costs. One 'hospital effect' is estimated for each hospital *h (*with a total of 149 in our sample) and represents the average quantum of cost variation (after legitimate sources have been purged from the data) that can systematically be associated with a patient being treated in each hospital. When estimated with random effects then we assume that $\gamma_h \sim N(0, \sigma_\gamma^2)$. When using fixed effects, we estimate a dummy for each hospital.

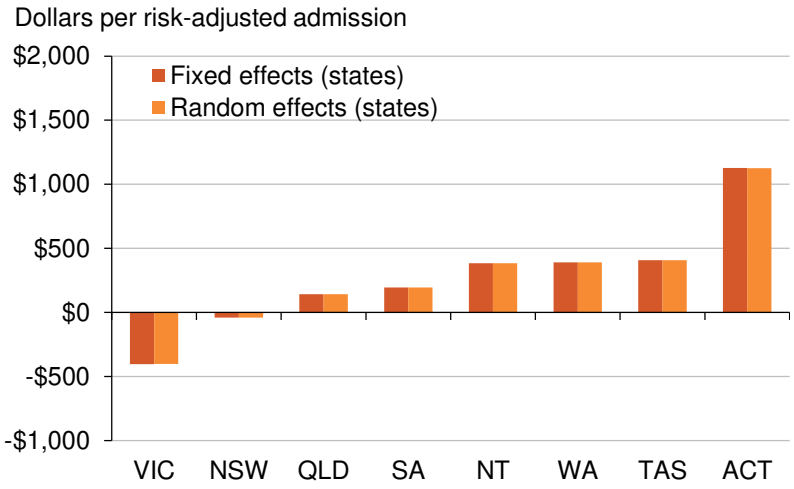
- $\varepsilon_{i[hs]}$, the error term that embodies remaining stochastic variation. We assume this is normally distributed.[59] $\varepsilon_{i[hs]} \sim N(0, \sigma_\varepsilon^2)$

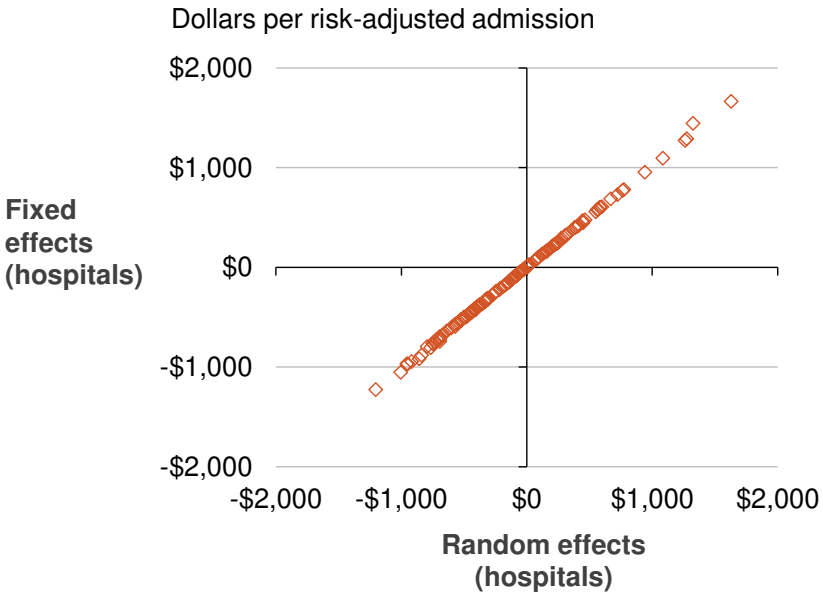
Random vs fixed effects

In the first instance, our preference is to estimate the state and hospital effects (ψ_s and γ_h) using random rather than fixed effects. There are three reasons for this.60 First, the random effects approach means that for hospitals with relatively few patients our point estimate of performance will shrink toward the mean. This has positive practical implications as it focusses our attention on providers where we have enough information to draw meaningful conclusions.61 Second, there aren't large differences in the estimates. In a majority of cases, estimates are extremely similar using either fixed or random effects (see Figure 4 for a comparison of state and hospital effects). When estimates differ it is generally a consequence of hospital size, and the resulting 'shrinkage towards the mean' of the random effects estimate, which as mentioned above, we believe to be desirable. Third, the random effects approach eases some of the practical constraints on model estimation. With around 4 million episodes (and 12 million when we estimate models using data from all three years), this proved useful.

[58] We use "states' to refer to state and territories from here on out.
[59] Even though the distribution of costs is not normal, as noted by Lumley*, et al.* (2002), in large health datasets a linear regression with normally distributed errors produces unbiased estimates.
[60] For a fuller discussion, see Gutacker*, et al.* (2011)
[61] Ibid., p.5

Figure 4 – comparison of fixed and random effects (2010-11)





Sample= 1,965,055
[i.e. 50% of 2010-11 sample, which was done to ensure the fixed effects models would estimate]

4.2 Impact of various legitimate and unavoidable costs on state and hospital estimates

To assess the impact of our controls (and to build towards the benchmarking models 3a and 3b) we estimate cost differences among states (4a) and hospitals (4b), *without adjusting for legitimate drivers of cost variation*. As is the case for all our models, the dependent variable is total cost (for patient *i*, in state *s* and hospital *h*):

$$c_{is} = \psi_s + \varepsilon_{is} \quad (4a)$$

$$c_{is} = \gamma_h + \varepsilon_{is} \quad (4b)$$

We then add the variables described in Table 2 individually, until we reach the main benchmarking models (equations 3a and 3b respectively). In other words, controls are added one by one, and the resulting state and hospital effects are logged. These estimates are plotted in Figure 5 and Figure 6.

Moving down the charts, as each new source of legitimate cost variation is added to the analysis, we track the impact on the performance benchmarks. A shift to the left indicates that our estimate of performance has improved (i.e. unexplained cost/separation has decreased). A shift to the right indicates that measured performance has declined.

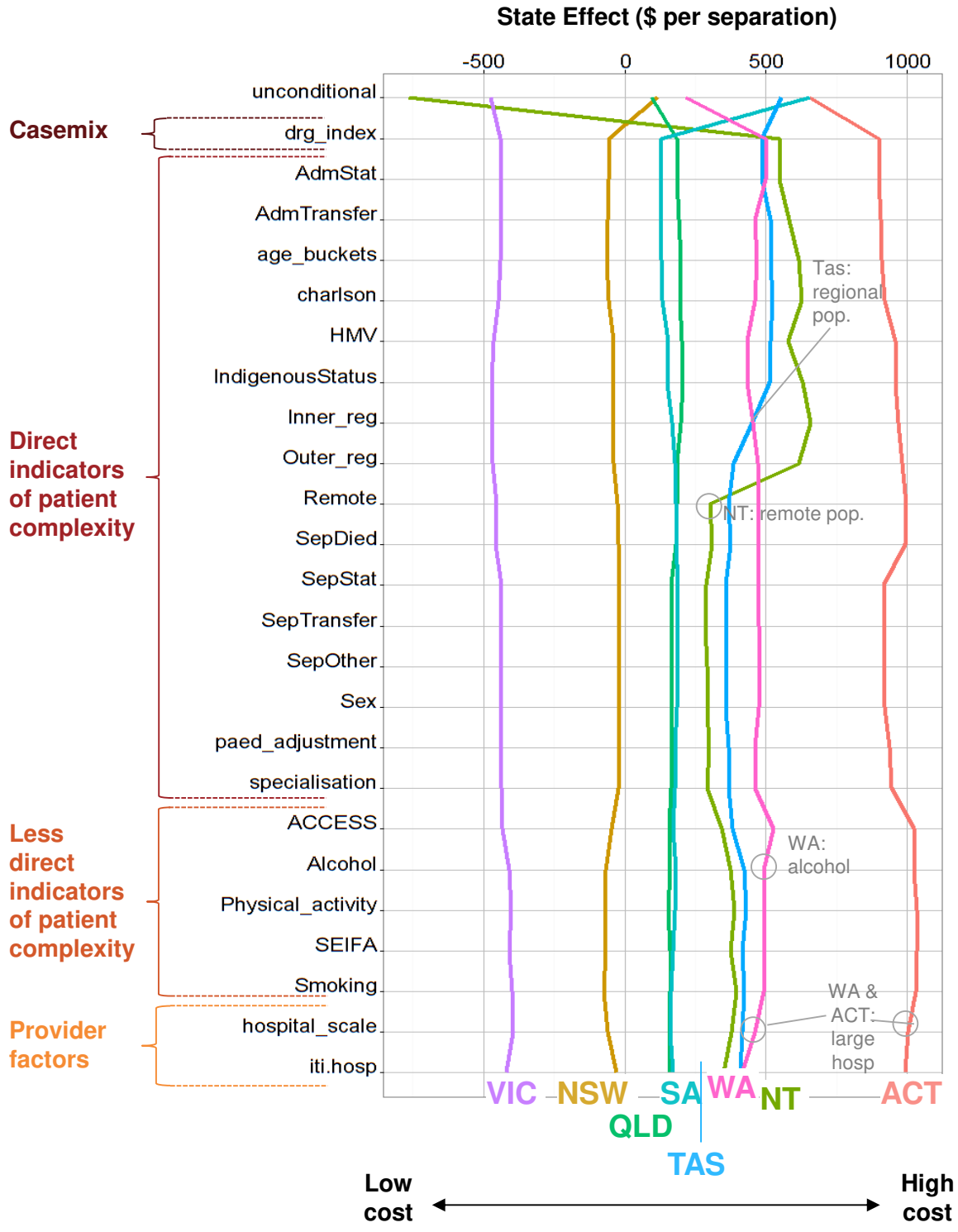
The impact of controls for state benchmarking

As expected, adjusting for casemix (i.e. adding the control for DRG_index) has a noticeable impact on our assessment of state performance, especially in the NT where there is a relatively high proportion of relatively simple separations (e.g. dialysis).

After the DRG adjustment, there are relatively few notable movements, and state estimates are – as expected – robust to these extra controls. Of note:

- NT's benchmark shifts once remoteness is taken into account
- Tasmania's measured performance increases once regionality (both inner and outer) is factored in.
- There are marginal effects relating to alcohol use in WA and Tasmania
- States which have above-average size hospitals (in the NHCDC) benefit from the scale adjustment (e.g. WA).

Figure 5 - how do State estimates change when unavoidable and legitimate costs are added?



Note that these estimates come from a 10% sample of the data, n=393,010

Impact on controls on hospital benchmarking

In the hospital figure (Figure 6), we removed the first step (i.e. accounting for DRG differences) from the figure, as this had a very large effect and overwhelmed the rest of the (perhaps more contentious) adjustments.

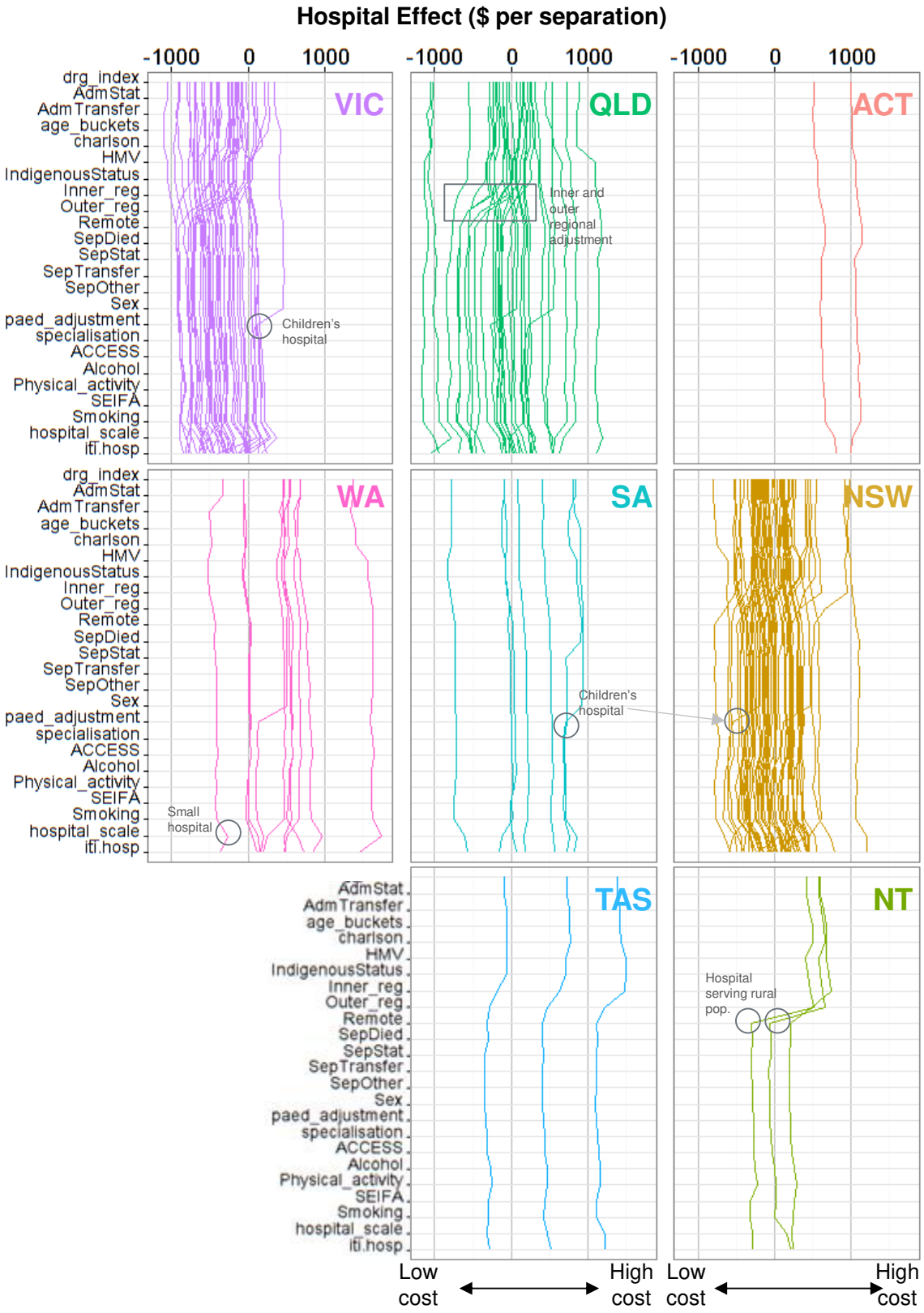
Arguably the most striking of these adjustments is the control for paediatric specialisation. Introducing this variable markedly improves the assessed performance of a number of children's hospitals. This variable is clearly important when assessing hospitals that provide any significant amount of paediatric care.

Another important set of factors are the regional/remote markers. These have a clear impact on our performance estimates for a number of hospitals. In the case of inner and outer regional patients, the adjustment has a noticeable influence on our assessment of a number of hospitals in QLD (as well as VIC, NSW and Tas). In the case of remote, this adjustment is particularly prominent in the NT.

Again, the provider variables have an effect. In our main analysis, adjustments for legitimate and unavoidable cost drivers serve to improve the estimated performance of larger hospitals (which we highlight in the WA panel of Figure 6).

In general, however, both figures illustrate that with a few notable exceptions, controlling for various patient and provider factors has a relatively limited bearing on our estimates of cost performance. Although many of these controls do not have a large impact on our estimates (e.g. smoking) our approach is to leave them in the model, as this maximises the chances that we isolate potentially-avoidable variation. In other words, by leaving in all controls that we believe represent unavoidable cost drivers, we minimise the risk that we construe a provider as having avoidably high costs when this isn't the case.

Figure 6 – how do hospital estimates change when unavoidable and legitimate costs are added?



Note that these estimates come from a 10% sample of the data, n=393,010

4.3 Results of main models

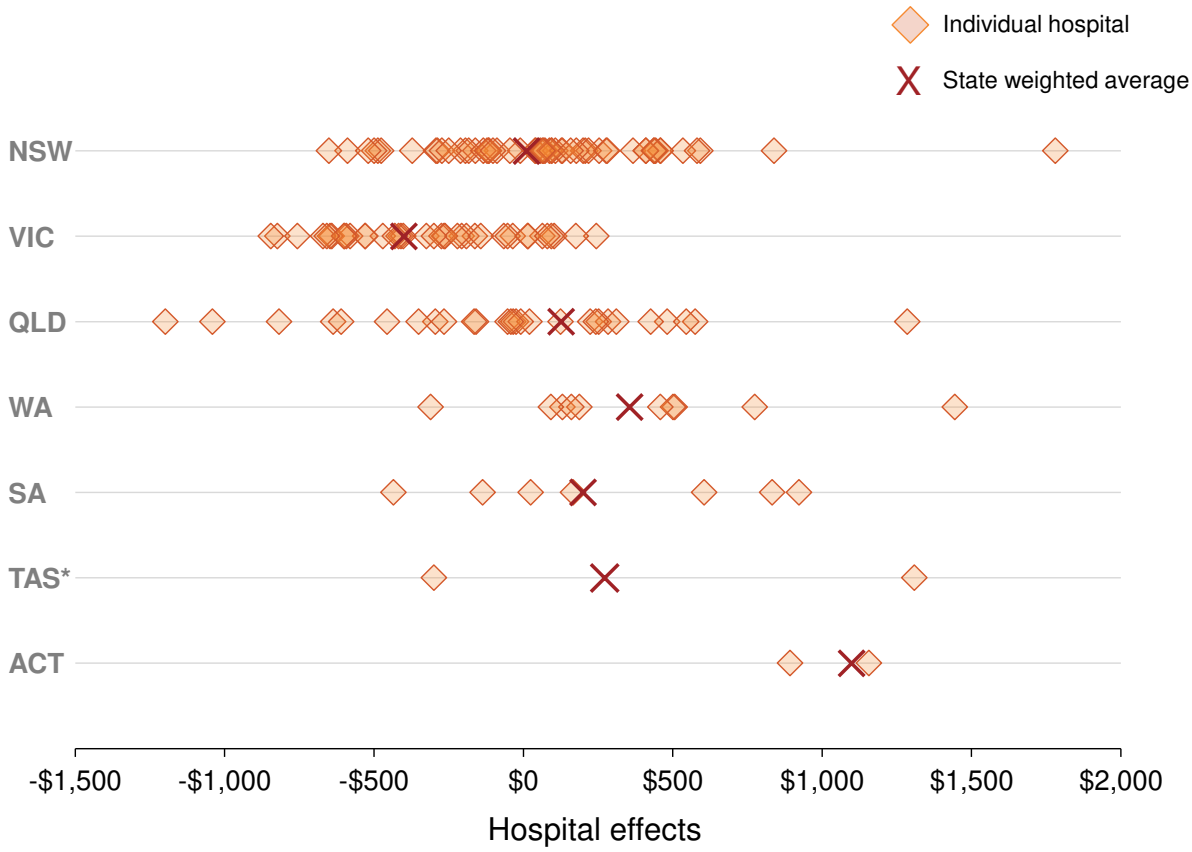
Table 7 presents the coefficients of models 3a and 3b (where state and hospital effects are estimated with random effects).

Table 7 – Results of main analysis. Risk adjustment coefficients for equations 3a and 3b.

	States (3a)		Hospitals (3b)	
Direct indicators of patient complexity (patient-level)	Estimate	(s.e.)	Est	(s.e.)
Admission mode:				
Adm_transfer = 1 if transferred from another hospital	$938	($15)	$933	($15)
Adm_stat = 1 if 'change of episode type'	-$1,356	($58)	-$1,397	($58)
Age (base category is <1 years)				
1-16	$938	($15)	-$461	($20)
17-26	-$467	($20)	-$353	($20)
27-36	-$336	($20)	-$283	($20)
37-46	-$255	($20)	-$363	($20)
47-56	-$349	($20)	-$329	($20)
57-66	-$314	($19)	-$275	($19)
67-76	-$251	($19)	-$176	($19)
77-86	-$160	($19)	-$157	($20)
87-96	-$152	($19)	-$362	($24)
97-116	-$362	($23)	-$887	($74)
Charlson (weighted Charlson score)	$60	($2)	$59	($2)
DRG_index	$3,601	($2)	$3,593	($2)
Hours of mechanical ventilation (HMV)	$94	($0)	$95	($0)
Indigenous status (= 1 if either Aboriginal or Torres Strait Islander)	-$111	($14)	-$69	($14)
Paediatric specialisation	$508	($17)	$503	($24)
Rurality				
Inner_reg = 1 if living in 'inner regional Australia'	$177	($9)	$280	($13)
Outer_reg = 1 if living in 'outer regional Australia'	$232	($13)	$411	($19)
Remote = 1 if living in 'remote Australia' or 'very remote Australia'	$504	($25)	$661	($29)
Separation mode				
Sep_transfer = 1 if transferred to healthcare provide (e.g. another hospital)	$179	($13)	$233	($13)
Sep_stat = 1 if statistical discharge - type change	$2,753	($25)	$2,718	($25)
Sep_died = 1 if patient died	-$2,426	($32)	-$2,405	($32)
Sep_other = 1 for other seps apart from discharge to regular accommodation	-$543	($30)	-$526	($30)
Sex (= 1 if female, 0 if male)	$83	($5)	$83	($6)
Specialisation (=1 if ICD code on NHS specialisation list)	$1,943	($15)	$1,932	($15)
Less direct indicators of patient complexity (neighbourhood level)				
Access to primary care (Full Time GP's per 1,000 people)	$2	($0)	$1	($0)
Alcohol over-use rate (per 100 people)	$54	($3)	$34	($4)
Physical inactivity rate (per 100 people)	$10	($1)	-$2	($1)
Index of relative advantage/disadvantage	$0	($0)	-$1	($0)
Smoking rate (per 100 people)	-$6	($2)	$12	($2)
Provider factors				
Scale (of hospital; seps per year)	$4	($0)	$7	($2)
Scope (Information Theory Index of hospital)	$240	($8)	$189	($86)
Model information				
intercept	-$223		$130	
n	3,930,109		3,930,109	
residual	$5,326		$5,315	
AIC	7.85969e+07		7.8581e+07,	

Figure 7 presents plots of state and hospital effects. Note that in *Controlling costly care* we are primarily focussed on within-state variation, and present hospital effects with the state effect stripped out. To be as consistent as possible, the state effects we subtract are not those specified by estimating model 3a, but calculated by a weighted average of the hospital effects in each state.[62] These averages are presented in Figure 7.

Figure 7 – estimates of hospital effects by state



*Only range shown

[62] Weighted by separations. Note that these two methods yield extremely similar state effects.

5. Robustness

5.1 How stable are performance estimates over time?

Over time we would expect some variation in the estimates of state and hospital performance due to actual changes in clinical practice as well as measurement error. In addition to these two sources of temporal variation, performance estimates may vary due to changes in the NHCDC methodology, which is continually improving, particularly with respect to comparability across jurisdictions.

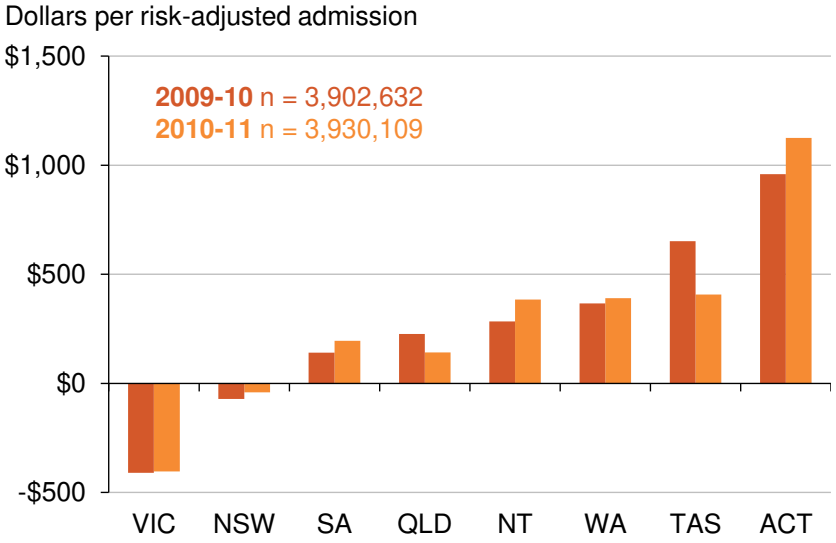
We estimate how volatile performance estimates are by comparing our main results (2010-11) to the previous year's data (2009-10). Figure 8 illustrates how state estimates changed over this period by estimating model 3a on the two different years of data.

$$c_{is} = \boldsymbol{X_i'}\alpha + \boldsymbol{Z_h'}\delta + \psi_s + \varepsilon_{is} \qquad (3a)$$

Note: terms are defined in section 4.1 (starting on page 23).

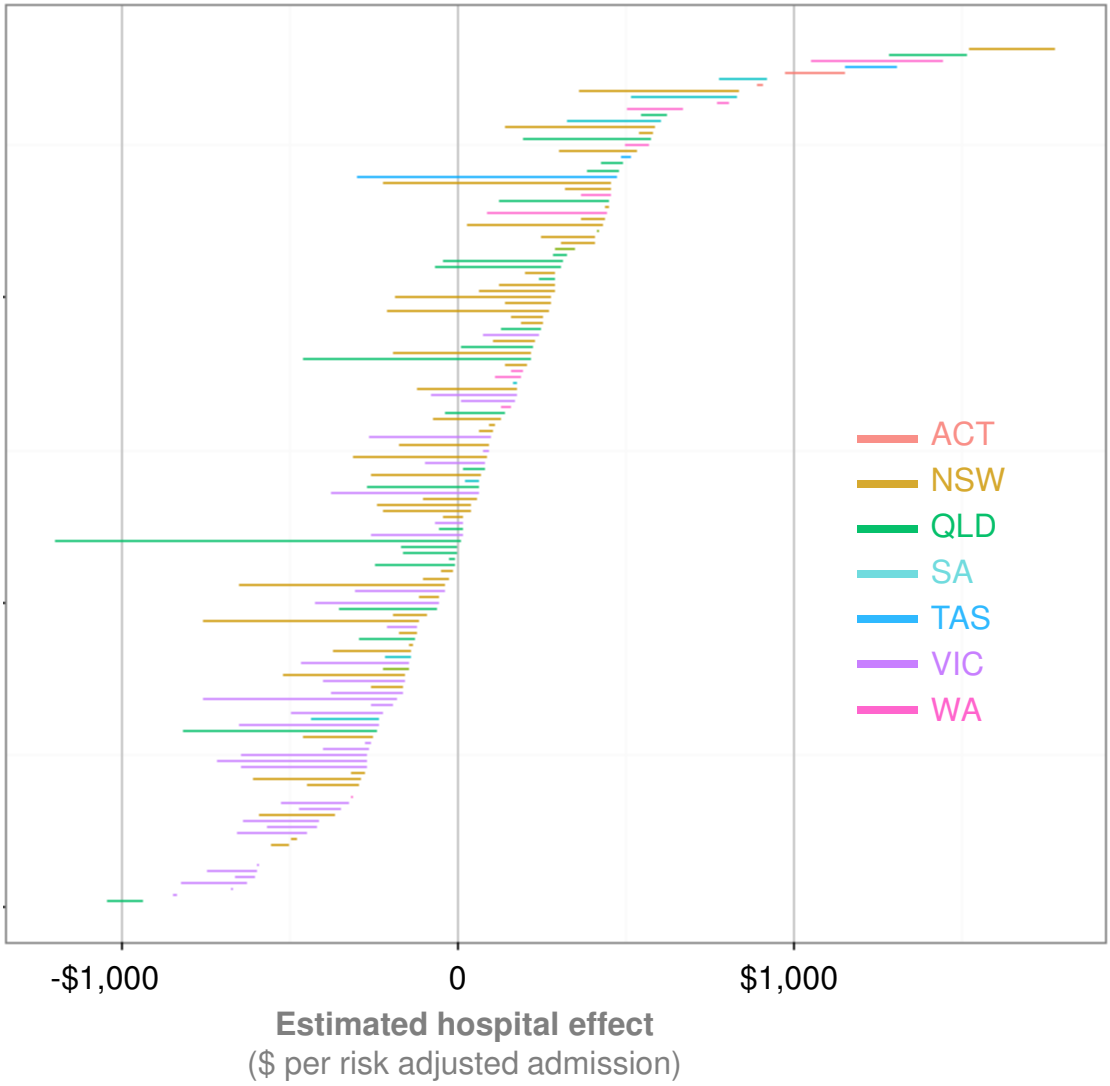
Given the potential sources of variation in the data, these results suggest that state effects are relatively stable.

Figure 8 – comparing State and hospital estimates across different years



We then undertake a similar exercise for our hospital estimates, comparing estimated performance in 2010-11 with that in 2009-10. The results are presented in Figure 9. Each hospital is represented by a line whose length illustrates the range of our performance estimates across 2009-10 and 2010-11. The correlation between estimates in the two years is 0.83. The median deviation between estimates from 2009-10 and 2010-11 is $160 (mean = $203). In other words, the median length of the lines in Figure 9 is 160, which represents 4% of the average cost per separation in 2010-11.

Figure 9 – range of hospital effects in 2009-10 and 2010-11.
Each represents one hospital



The hospitals in which estimates are the most volatile tend to be focussed (with a high proportions of separations in one area, for example obstetrics) or rural/regional. While understanding the volatility in this subset of hospitals is an important area for future research, in summary we believe that estimates across all hospitals show sufficient robustness as to be meaningful measures of performance.

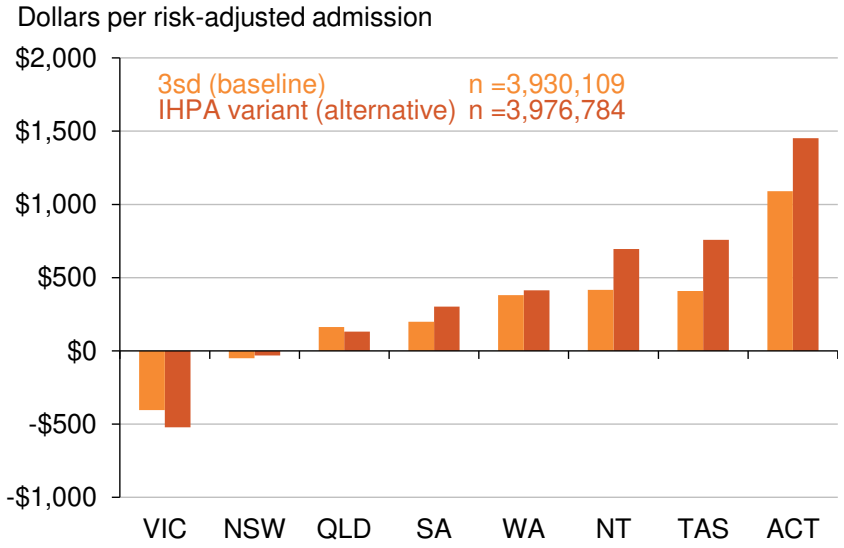
5.2 Changing the definition of an 'outlier'

One methodological choice which has the potential to influence our estimates is whether and how we defined some hospital admissions as 'cost outliers'. The rationale behind trimming outliers from the analysis is that some patients have very large costs, generally for reasons beyond hospitals' control. Whether or not a hospital treats such a patient may substantially affect its estimated performance, despite there being no difference in the hospital's processes,

staff quality, procurement practices, or other determinants of a hospital's ability to control costs. Stripping out outlier patients therefore provides a fairer way of assessing how well a hospital is performing.

Unfortunately, there is no widely agreed upon rationale for identifying 'outliers'. Given the absence of an orthodox method, we compare two procedures (outlined in detail in section 2). The first procedure identifies outliers as admissions where the cost is more than 3 standard deviations away from the mean cost within that DRG. This is our favoured assumption. The second is a variant of the approach IHPA's uses in setting the National Efficient Price, and is a much stricter definition (i.e., fewer admissions are defined as being outliers). Figure 10 compares our state effect estimates (model 3a) with different outlier assumptions defining the size of the dataset. While the estimates for the bigger states are robust to the difference in methodology, the stricter definition of outlier negatively impacts on the evaluation of smaller states.

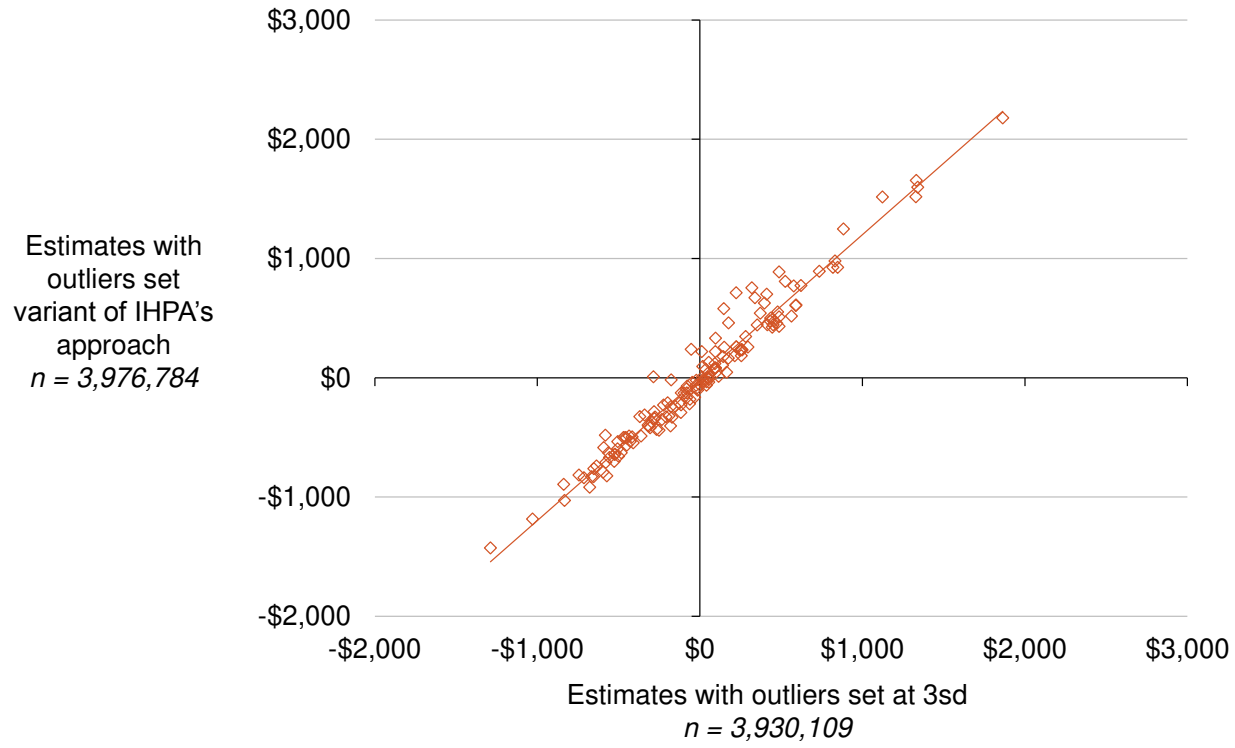
Figure 10 – comparing state estimates with 2 different definitions of 'outliers'
Model 3a estimated on two different datasets



We then replicate this analysis for hospital effects (Figure 11). The correlation between the two sets of estimates is high (0.98). Analogous to the state analysis, the main difference between the two estimates is the assessment of smaller hospitals, which fare worse under the stricter definition of outliers.

Given the potential to unfairly assess small providers' performance more harshly because of a chance encounter with a very high-resource patient, in *Controlling costly care* we present results using the '3 standard deviation' definition of outliers.

Figure 11 – comparing hospital estimates with 2 different definitions of 'outlier'
Model 3b estimated on two different datasets



5.3 Do we need to model separate service groups separately?

It's reasonable to assume that variations in costs can be modelled more precisely if separate types of care can have their own specific risk-adjustments for legitimate drivers of cost variation. It's plausible, for example, that the effect of age on cost will be different for psychiatry and orthopaedics.

The sheer quantity of data in the NHCDC allows for the possibility of estimating the cost impact of legitimate cost drivers in each service area, and using these estimates as adjustments in our benchmarking.

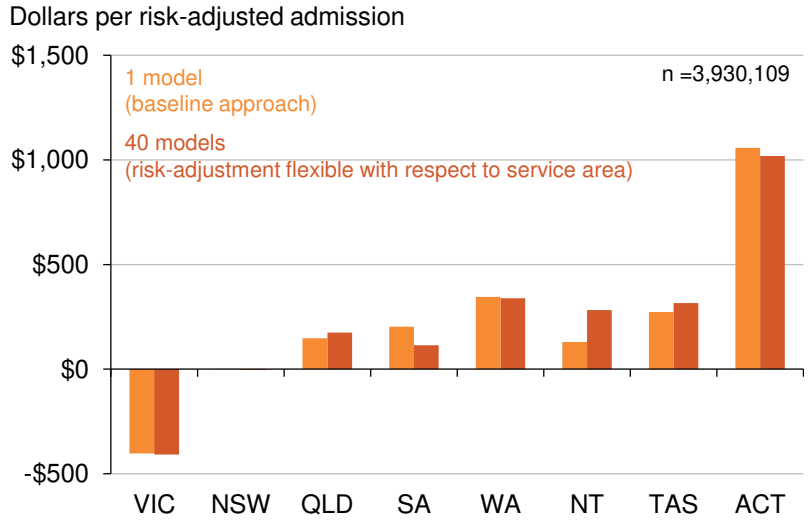
In this section, we explore the extent to which adopting this more complex and flexible modelling approach impacts our analysis. In short, we compare two approaches:

- Using all the 2010-11 data to estimate models 3a (states) and 3b (hospitals) [i.e. the approach that generated the main results in this document and *Controlling costly care*]
- Essentially splitting the data into 40 subsets on the basis of service groups, and generating (up to) 40 performance estimates for each state using model 3a, and (up to) 40 performance

estimates for hospitals using model 3b. These estimates are then recombined via a weighted average, to form an overall performance estimate for any given state or hospital.[63]

Figure 12 suggests that at the state level, allowing flexibility in controls makes very little difference – especially to the bigger states (NSW, QLD, WA, and Vic).

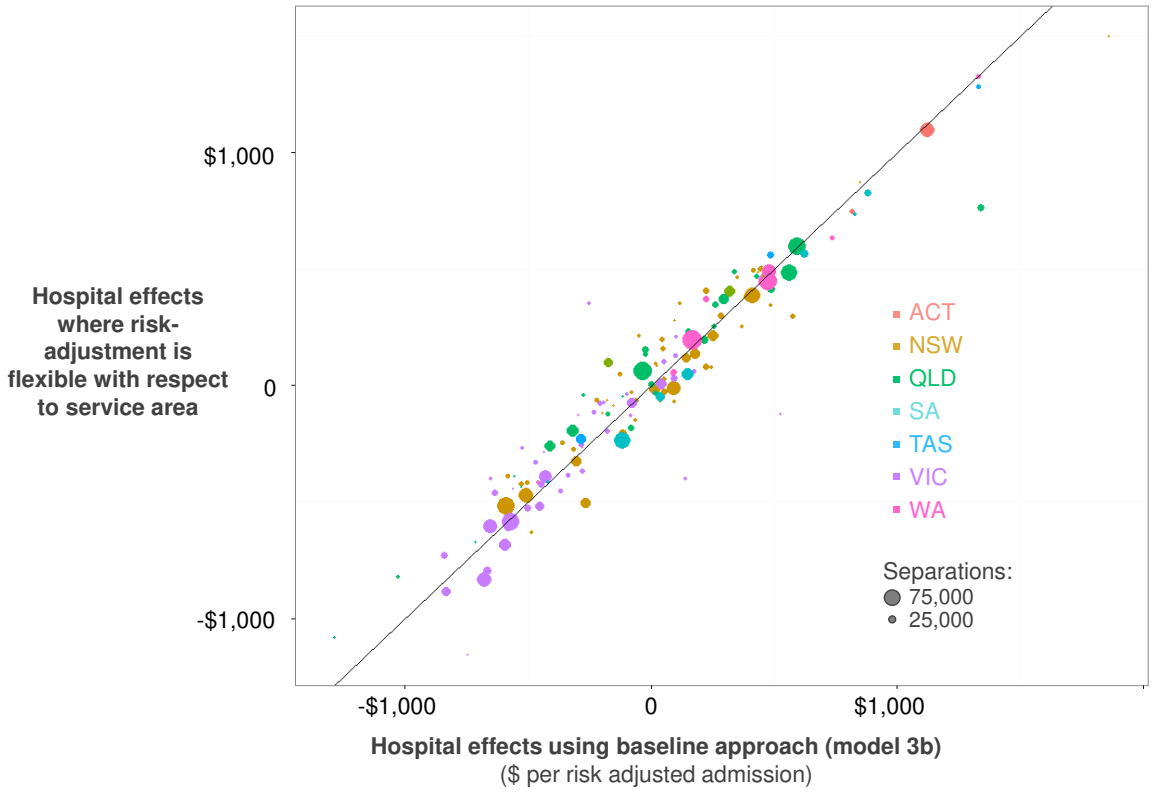
Figure 12 – comparison of state effects generated from 1 model, vs combined model of 40 different service groups



At the hospital level, there's more of a difference, but the results presented in *Controlling costly care* and earlier in this paper are still relatively robust to this change of approach. It's also worth noting from the chart that deviations tend to be in small hospitals.

[63] Full results of coefficients for specific service areas are available on request.

Figure 13 – comparison of hospital effects generated from 1 model, vs combined model of different service groups



5.4 Differences in coding across states

Given the importance of casemix adjustment to our benchmarking, we were conscious that interstate comparisons might be undermined by potential differences in coding practices across states. This might manifest itself with some states systematically coding fewer 'high resource' DRGs than others. Some states might, for example, code fewer patients on the borderline between I03B (hip replacement *without* catastrophic complications and comorbidities) and I03A (hip replacement *with* catastrophic complications and comorbidities) to the lower resource group.[64]

The first task was to assess whether there were any material differences in distribution of 'A' codes relative to 'B', 'C' and 'D'. Our method was adapted from Coory and Cornes (2005),[65] and first involved stripping out separations in DRGs where no reasonable analysis of inter-state coding differences was possible. Four categories of DRGs were removed. DRGs:

1. that had no 'adjacent' DRGs, e.g. those with a Z suffix [146 DRGs]

[64] Within a DRG category such as 'I03' (where this three character combination is sometimes called the DRG stem), there can be a number of 'adjacent' groups. These are partitioned according to disease severity. DRGs ending in "A' are the most complex patients, while "D' are the simplest and require the least resources.
[65] See "Interstate Comparison of Public Hospital Outputs using DRGs: Are they Fair', *Australian and New Zealand Journal of Public Health* (2005) vol 29, no. 2

2. where the partition was defined by sameday status [33 DRGs]

3. relating to the care of neonates [25 DRGs]

4. labelled as 'unallocated', e.g. 801A [3 DRGs]

There was some overlap in these categories. Ultimately, 189 of 705 DRGs were removed. The main analysis was conducted on the remaining 514 DRGs. In addition, a secondary analysis was performed on a subset of 6 DRGs described in Coory and Cornes as 'Major Medical Conditions' (MMC). [66] These are conditions where non-admission is very unlikely. The rationale behind the analysis was to check the extent to which results may be driven by different states having different admissions thresholds, i.e. potentially having tighter controls on admission for low-resource patients.

For each DRG stem we calculated the average percentage of patients coded as being in the highest resource category (e.g. the proportion of hip replacement separations coded as 'I03A').[67] This national average was then compared to the percentage of 'A' codes in each state. A new variable was created DRG_RECODE, which tracked what the DRG allocation *would have been* if every state had the national average rate of high-resource DRGs. In instances where a state had a lower rate of 'A' codes, the 'B' coded separations would be changed to 'A's (starting at the most expensive 'B' separations) until the state's proportion of A's equalled the national average. The reverse procedure was conducted if a state's proportion of A's was above the national average.[68]

The results suggest that there may well be differences across states in the proportion of patients assigned to the high-resource DRGs. Table 8 summarises this result by presenting the deviation from the national average percentage of 'A' codes (controlling for DRG). Results are relatively consistent across the main analysis of 514 DRGs and the 'Major Medical Conditions'.

[66] The DRG codes are: E61A, E61B, F62A, F62B, T60A, T60B (representing pulmonary embolism; heart failure and shock; septicaemia).

[67] In order to check that states with above-average levels of "A' cases aren't facing populations with more morbidity, we compared the 'proportion of 'A' cases' to a number of possible confounders: age; rurality; access to primary care; index of relative disadvantage; average Charlson score. None of these explained differences in upcoding. High upcoding states like Victoria didn't have disadvantages in these areas, while low upcoding states like WA, QLD and Tas were not blessed with healthier populations (with respect to the aforementioned confounders).

[68] Note that if all B's were converted to A's and the state's percentage was still below the average *and* the DRG stem had a 'C' category, then these C separations were recoded in DRG_RECODE. Note also that if a state had fewer than 5 'A' separations then no changes were made on the basis that we had too little information to make any kind of judgment about whether the lack of A's was due to coding practices, or just a lack of patients. This threshold is arbitrary, but makes no substantive differences to the results.

Table 8 –Comparing states to the national average: proportion of patients coded as being in the highest resource category within DRG stems (2010-11)

	Main Analysis (514 DRGs)	Major Medical Conditions (3 DRGs)
TAS	-4.1%	-8.6%
WA	-2.1%	-4.3%
ACT	-1.9%	-3.4%
QLD	-1.0%	-2.7%
NSW	-0.1%	-2.8%
SA	2.5%	1.4%
VIC	2.9%	9.4%

Impact of coding differences in dollar terms

To estimate the financial implications of coding differences, the casemix-adjusted cost was calculated for each state using a) actual DRG codes to adjust for casemix, and b) DRG_RECODE, i.e. what the DRG assignment would have been had the State coded the average proportion of separations to the highest-resource category (Table 2).

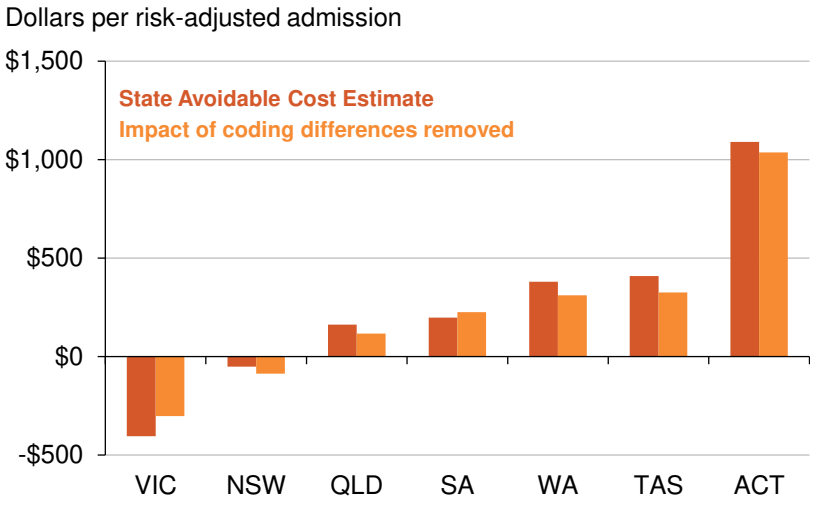
Table 9 – what is the impact of equalising the coding distribution in each State?

	Casemix-adjusted cost	Casemix-adjusted cost NO CODING DIFFERENCES	Value per separation	Total Value* ($Millions)
TAS	$5,014	$4,931	**-$83**	**-$10,370,045**
WA	$4,817	$4,748	**-$69**	**-$34,556,689**
ACT	$5,575	$5,521	**-$54**	**-$6,454,144**
QLD	$4,472	$4,427	**-$45**	**-$51,719,724**
NSW	$4,198	$4,163	**-$35**	**-$61,646,368**
SA	$4,532	$4,560	**$28**	**$10,030,397**
NT	$5,060	$5,101	**$41**	**$5,608,756**
VIC	$3,681	$3,783	**$102**	**$147,412,085**

Note: *scaled up to all ABF-funded activity in NHCDC public hospitals

Figure 14 (over the page) puts the per-patient impact of coding differences into the broader context of our state effects. While coding differences are important and do explain some of the variation between states (roughly 6% of the interstate differences disappear once coding differences are accounted for), this is a marginal factor relative to other differences across the states.

Figure 14 – Putting the impact of coding differences into the context of state effects



6. Conclusion

This document explains the data, methodology, and robustness of the benchmarking analysis used in *Controlling costly care.* The project itself was greatly aided by IHPA and the state and territory governments generously providing the National Hospital Cost Data Collection (NHCDC). This gave us more detailed and comprehensive data than has been available for previously published research on Australian hospital costs.

Naturally, these data have limitations, especially when making comparisons across jurisdictions. It appears, for example, that there are material differences in the coding practices across Australia that may impact efforts to quantify cost differences across states [section 5.4]. In *Controlling costly care* we have tried to minimise these issues by focussing our analysis on variation within states. We believe that the data quality is more than good enough for this purpose.

In analysing the NHCDC, we build on a growing body of literature that uses multi-level modelling and patient-level data to benchmark hospital performance. Within this framework, there are still methodological choices to be made. This document illustrates that the hospital effects reported in *Controlling costly care* are robust to a number of these choices: the definition of outliers [section 5.2]; allowing our risk-adjustments to differ in different areas of care [section 5.3]; and using fixed rather than random effects [section 4.1]. Our estimates of hospital effects are the primary data source used in calculating 'avoidable costs' and so the robustness demonstrated here illustrates that the key results of *Controlling costly care* would not be substantially altered by these modelling changes.

To the extent that changes in method would alter the results presented in *Controlling costly care*, they would tend to increase the estimated levels of avoidable cost. This is because our general approach was to make conservative methodological choices about defining variation between hospitals as 'avoidable'. In the case of outliers, for example, our main analysis used the 'three standard deviation' definition, which filtered out considerably more separations with very low and very high costs than IHPA's definition. Less conservative choices would lead to higher, but still comparable, estimates of avoidable cost.

We also show that, with a couple of exceptions, our estimates of hospital effects were fairly stable from 2009-10 to 2010-11 (especially in the context of the average cost of an admission). Even if the data and benchmarking methodology were perfect, some temporal variation would be expected. The relative consistency of estimates in different years is therefore a good indication that our benchmarking tells us something meaningful about how hospitals control their costs.

Since 2010-11 (the last year of data we were able to access), the NHCDC's users and contributors have reported an improvement in the costing processes and software that underpin the database.[69] We're confident that these data, in conjunction with the sorts of methods reported here, can provide practical and reliable insights to help hospitals and states continually improve the way they provide acute care.

[69] PWC (2013a)

References

ABS (2011) *SEIFA 2011*, from http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa2011?opendocument&navpos=260

AIHW (2013) *Australian hospital statistics*, from http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129543146

Australian Government Department of Health and Ageing (2007) *Hospital Reference Manual Round 11 (2006-07)* from http://www.health.gov.au/internet/main/publishing.nsf/Content/90F1A1681D35BCA4CA2570BC0015B1FC/$File/Round%2011.pdf

Bestawros A, Filion KB, Haider S, Pilote L and MJ., E. (2005) 'Coronary artery bypass graft surgery: do women cost more?', *Canadian Journal of Cardiology*, 21(13), p 1195-1200

Bradford, W. D., Kleit, A. N., Krousel-Wood, M. A. and Re, R. N. (2001) 'Stochastic Frontier Estimation of Cost Models Within the Hospital', *The Review of Economics and Statistics*, 83(2), p 302-309

Breadon, P. and Duckett, S. (2013) *Access All Areas*, Grattan Institute from http://grattan.edu.au/static/files/assets/31e5ace5/196-Access-All-Areas.pdf

Cookson, R. and Laudicella, M. (2011) 'Do the poor cost much more? The relationship between small area income deprivation and length of stay for elective hip replacement in the English NHS from 2001 to 2008', *Social Science and Medicine*, 72(2), p 173-184

Daidone, S. and Street, A. (2013) 'How much should be paid for specialised treatment', *Social Science and Medicine*, 84, p 110-118

Daley, J., McGannon, C. and Savage, J. (2013) *Budget pressures on Australian governments*, Grattan Institute from http://grattan.edu.au/static/files/assets/ff6f7fe2/187_budget_pressures_report.pdf

Deb, P. and Burgess, J. (2003) *A Quasi-experimental Comparison of Econometric Models for Health Care Expenditures*, Hunter College Department of Economics Working Papers

Dormont, B. and Milcent, C. (2004) 'The sources of hospital cost variability', *Health Economics*, 13, p 927-939

Duncan, C., Jones, K. and Moon, G. (1998) 'Context, composition and heterogeneity: using multilevel models in health research', *Social Science and Medicine*, 46(1), p 97-117

Gaughan, J., Mason, A., Street, A. and Ward, P. (2012) *English hospitals can improve their use of resources: an analysis of costs and length of stay for ten treatments*, CHE Research Paper 78, Centre for Health Economics

Gutacker, N., Bojke, C., Daidone, S., Devlin, N., Parkin, D. and Street, A. (2011) *Truly Inefficient or Providing Better Quality of Care? Analysing the Relationship Between Risk-Adjusted Hospital Costs and Patients' Health Outcomes*, Centre for Health Economics, Research Paper 68

Gutacker, N., Bojke, C., Daidone, S., Devlin, N. J., Parkin, D. and Street, A. (2011) 'Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes', *Health Economics*, 22(8), p 931-947

Hauck, K., Rice, N. and Smith, P. (2003) 'The influence of health care organisations on indicators of health system performance', *Journal of Health Services Research and Policy*, 8, p 68-74

Hollingsworth, B. and Peacock, S. (2008) *Efficiency Measurement in Health and Health Care*, Routledge

Hossain, F. (2009) 'Comparing Cost and Quality of Health Care Across the Country', *The New York Times,* 8 September 2009, accessed 31 October 2013, from http://www.nytimes.com/interactive/2009/09/08/us/cost-scatterplot.html

Hvenegaard, A., Arendt, J. N., Street, A. and Gyrd-Hansen, D. (2011) 'Exploring the relationship between costs and quality: does the joint evaluation of costs and quality alter the ranking of Danish hospital departments?', *The European Journal of Health Economics*, 12(6), p 541–551

Hvenegaard, A., Street, A., Sorensen, T. H. and Gyrd-Hansen, D. (2009) 'Comparing hospital costs: What is gained by accounting for more than case-mix index?', *Social Science and Medicine* 29, p 640-647

Independent Hospital Pricing Authority (2012) *National Efficient Price Determination 2012-2013*, accessed 24 September 2013, from http://www.ihpa.gov.au/internet/ihpa/publishing.nsf/Content/nep-determination-2012-13.htm/$File/NEPDetermination2012-2013.pdf

Independent Hospital Pricing Authority (2013) *National Pricing Model Technical Specifications 2013-2014* from http://www.ihpa.gov.au/internet/ihpa/publishing.nsf/Content/CA25794400122452CA257B15007D2F51/$File/National-Pricing-Model-Technical-specifications-2013-14.pdf

Jackson, T. (2000) 'Cost estimates for hospital inpatient care in Australia: evaluation of alternative sources', *Australian and New Zealand Journal of Public Health*, 24, p 234-241

Kessler, D. and McClellan, M. (2001) *The Effects of Hospital Ownership on Medical Productivity*, NBER Working Paper No. 8537 from http://www.nber.org/papers/w8537

Kobel, C. and Theurl, E. (2013) *Hospital specialisation within a DRG-Framework: The Austrian case*, Working Papers in Economics and Statistics, University of Innsbruck from http://eeecon.uibk.ac.at/wopec2/repec/inn/wpaper/2013-06.pdf

Kristensen, T., Laudicella, M., Ejersted, C. and Street, A. (2010) 'Cost variation in diabetes care delivered in English hospitals', *Diabetic Medicine*, 27, p 949-957

Laudicella, M., Olsen, K. R. and Street, A. (2010) 'Examining cost variation across hospital departments--a two-stage multi-level approach using patient-level data', *Social Science and Medicine*, 71, p 1872-1881

Lumley, T., Diehr, P., Emerson, S. and Chen, L. (2002) 'The importance of the normality assumption in large public health data sets', *Annual Review of Public. Health*, 2002(23), p 151-169

Macario, A., Vitez, T., Dunn, B., McDonald, T. and Brown, B. (1997) 'Hospital costs and severity of illness in three types of elective surgery', *Anesthesiology*, 86(1), p 92-100

Malyon, R., Zhao, Y. and Oates, B. (2013) 'Differences in the cost of admitted patient care for Indigenous people and people from remote locations', *Australian Health Review*, 37, p 26-31

McKay, N. L. and Deily, M. E. (2008) 'Cost inefficiency and hospital health outcomes', *Health Economics* 17(7), p 833–848

Newhouse, J. P. (1994) 'Frontier estimation: How useful a tool for health economics?', *Journal of Health Economics*, 13, p 317-322

NHS (2012) *PbR and the Market Forces Factor (MFF) in 2013-14*, Department of Health from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/214906/PbR-and-the-MFF-in-2013-14.pdf

NHS (2013) 'Specialised Services National Definitions Set', from http://www.specialisedservices.nhs.uk/info/specialised-services-national-definitions

Olsen, K. R. and Street, A. (2008) 'The analysis of efficiency among a small number of organisations: how inferences can be improved by exploiting patient-level data', *Health Economics*, 17, p 671-681

Paul, C. J. M. (2002) 'Productive structure and efficiency of public hospitals', in *Efficiency in the Public Sector*, K. J. Fox, Ed. Kluwer Academic Publishers

Peltola, M. (2012) 'Patient classification and hospital costs of care for stroke in 10 European countries', *Health Economics*, 21 (Supplement 2), p 129-140

PHIDU (2013) 'A Social Atlas of Health', from http://www.publichealth.gov.au/data/

Productivity Commission (2010) *Public and Private Hospitals: Multivariate Analysis*, from http://www.pc.gov.au/__data/assets/pdf_file/0008/97964/supplement.pdf

PWC (2013a) *Independent Review of the Round 16 National Hospital Cost Data Collection*, from http://www.ihpa.gov.au/internet/ihpa/publishing.nsf/Content/E941955AE78F18CACA257C160010C5F4/$File/NHCDC-Indp-Fin-Review-R16.pdf

PWC (2013b) *Protecting prosperity: Why we need to talk about tax*, from http://www.pwc.com.au/tax/assets/Protecting-prosperity-22Jul13.pdf

Quan, H., Li, B., Couris, C., Fushimi, K., Graham, P., Hider, P., Januel, J. and V., S. (2011) 'Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries', *Am J Epidemiol.*, 173(6), p 676-682

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J. C., Saunders, L., Beck, C., Feasby, T. E. and Ghali, W. A. (2005) 'Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data', *Medical Care*, 43(11), p 1130-1139

Rice, N. and Leyland, A. (1996) 'Multilevel models: applications to health data', *Journal of Health Services Research and Policy*, 1(3), p 154-164

Schleifer, A. (1985) 'A Theory of Yardstick Competition', *The RAND Journal of Economics*, 16(3),

Schreyögg, J. and Stargardt, T. (2010) 'The trade-off between costs and outcomes: the case of acute myocardial infarction', *Health Services Research* 45(6p1), p 1585–1601

SCRCSSP (1997) *Data Envelopment Analysis, A Technique for Measuring the Efficiency of Government Service Delivery*, AGPS

Sundararajan, V., Quan, H., Halfon, P., Fushimi, K., Luthi, J. C., Burnand, B. and Ghali, W. A. (2007) 'Cross-national comparative performance of three versions of the ICD-10 Charlson index.', *Medical Care*, 45(12), p 1210-1215

Wang, J. and Mahmood, A. (2000) 'Efficiency of the NSW Public Acute Hospitals: An Application of the Data Envelopment Analysis', in *Economics and Health: 2000 – Proceedings of the Twenty-second Australian Conference of Health Economists*, J. Bridges, Ed. University of New South Wales School of Health Services Management

Warner, D. O. (2006) 'Preoperative abstinence from cigarettes', *Anesthesiology*, 104(2), p 356-367

Webster, R., Kennedy, S. and Johnson, L. (1998) *Comparing Techniques for Measuring the Efficiency and Productivity of Australian Private Hospitals*, ABS Working Paper No. 98/3

Wilke, M. H. and Grube, R. (2010) 'Pharmaco-economic evaluation of antibiotic therapy strategies in DRG-based healthcare systems – a new approach', *European Journal of Medical Research*, 15, p 564-570

Yong, K. and Harris, A. (1999) *Efficiency of Hospitals in Victoria under Casemix Funding: A Stochastic Frontier Approach*, Centre for Health Program Evaluation, Working Paper 92

Appendix A: scale and scope

Figure 15 – exploring whether full-service, small hospitals are often found to have high avoidable costs