

February 2018

# All complications should count Using our data to make hospitals safer

Stephen Duckett and Christine Jorm



## Grattan Institute Support

### Founding members



Australian Government



THE UNIVERSITY OF  
MELBOURNE

**BHP**

### Endowment Supporters

The Myer Foundation  
National Australia Bank  
Susan McKinnon Foundation

### Affiliate Partners

Google  
Medibank Private

### Senior Affiliates

Maddocks  
PwC  
McKinsey & Company  
The Scanlon Foundation  
Wesfarmers

### Affiliates

Ashurst  
Corrs  
Deloitte  
GE ANZ  
Jemena  
Urbis  
Westpac  
Woodside

## Grattan Institute Report No. 2018-01, February 2018

This report was written by Stephen Duckett, Grattan Institute Health Program Director, Christine Jorm, Honorary Fellow, and Lucille Danks and Greg Moran, Associates. Hugh Parsonage provided extensive assistance in the finalisation of this report.

We would like to thank numerous people from the health policy community including Ian Brownwood, Michael Daly, Irfan Dhalla, Jean-Frédéric Lévesque, Alison Verhoeven, John Wakefield, and Grattan Institute's Health Program Reference Group for their helpful comments.

The opinions in this report are those of the authors and do not necessarily represent the views of Grattan Institute's founding members, affiliates, individual board members, reference group members or reviewers. Any remaining errors or omissions are the responsibility of the authors.

Grattan Institute is an independent think tank focused on Australian public policy. Our work is independent, practical and rigorous. We aim to improve policy outcomes by engaging with both decision-makers and the community.

For further information on the Institute's programs, or to join our mailing list, please go to: <https://www.grattan.edu.au/>

This report may be cited as: Duckett, Stephen, Jorm, Christine, Danks, Lucille, and Moran, Greg (2018). *All complications should count: Using our data to make hospitals safer*. Grattan Institute.

ISBN: 978-0-6482307-1-7

All material published or otherwise created by Grattan Institute is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License

## Overview

One in every nine patients who go into hospital in Australia suffers a complication – about 900,000 patients each year. If they stay in overnight, the figure rises to one in four – about 725,000 patients each year. A patient's risk of developing a complication varies dramatically depending on which hospital they go to: in some cases, the additional risk of a complication at the worst-performing hospitals can be four times higher than at the best performers. If all hospitals lifted their safety performance to the level of the best 10 per cent of Australian hospitals, the complication rate across the nation would fall by more than a quarter.

This report exposes the flaws in Australian hospitals' safety and quality monitoring regime, and recommends reforms that could result in an extra 250,000 patients leaving hospital each year free of complications.

At the moment, a veil of secrecy hangs over which hospitals and clinicians have higher rates of complications and which are safety leaders. Hospital safety statistics are collected, but they are kept secret, not just from patients but from doctors and hospitals. This has to change. Patients have a right to know the data on complication rates in different hospitals and for different procedures, so they – and their GPs – can make better-informed decisions about how and where they are treated. Doctors and hospitals need to know how they are

performing compared to their peers, so that they can learn from the best-performing hospitals and clinicians.

At the moment, hospital safety policies focus on only a small subset of complications classified by government as being 'preventable'. Instead policy should be directed towards reducing all complications to the best rate achievable. This requires building up a comprehensive picture of patient outcomes, and understanding how some hospitals and clinical teams reduce all complications and achieve excellent outcomes.

Private health insurers should release the information they gather on private hospitals: reducing complication rates would mean quicker recoveries and lower premiums for their members. State and territory governments should release detailed data on the performance of both public and private hospitals. This data needs to show the whole gamut of hospital performance, from catastrophic but rare errors to less harmful but prevalent complications. It should highlight the areas where there is a big gap between the best and worst performers. Governments need to set ambitious goals for every hospital – public and private – to improve their safety and quality of care. And they need to ensure the data is published widely so that patients and taxpayers can see which hospitals are improving and which are not.

## Table of contents

Overview . . . . .	3
1 The harm done . . . . .	7
2 Australia needs to set more ambitious hospital safety targets . .	15
3 Data can inform hospitals about their specific strengths and weaknesses . . . . .	25
4 We need to share data openly . . . . .	33
5 Conclusion . . . . .	43
A The risks of transparency can be mitigated . . . . .	44
 <b>Methodological supplement</b>	 45
B Data . . . . .	47
C Conceptual framework . . . . .	58
D Results . . . . .	78
E Additional details regarding proof-of-concept app . . . . .	100

## List of Figures

1.1 Complications are common . . . . .	9
1.2 Procedural complications increase a patient's stay in hospital . . . . .	10
1.3 In Australia the rate of complications is not falling . . . . .	13
2.1 Many important complications are not included in HACs . . . . .	19
2.2 More than a quarter of all complications could be avoided if all hospitals were as safe as the top 10 per cent of hospitals . . . . .	22
2.3 Complication rates vary a lot between Australian hospitals . . . . .	23
3.1 Headline complication rates don't reveal whether a hospital is performing better than expected . . . . .	28
3.2 Hospital performance explains 8-10 per cent of the variation in patient outcomes . . . . .	28
3.3 Hospital safety varies significantly within states, and within sectors . . . . .	29
3.4 The greatest opportunity to make Australian hospitals safer is to move average hospitals closer to excellent hospitals . . . . .	30
3.5 The safety of a hospital's care can vary by specialty . . . . .	31
3.6 The safest choice can vary for patients of different ages . . . . .	32
4.1 Hospitals should regularly receive heat maps of their relative performance . . . . .	36
B.1 Most dropped observations related to empty readmission fields . . . . .	48
B.2 Several variables can be derived from a DRG . . . . .	54
C.1 Coding depth varies over time and by state . . . . .	65
C.2 Use of condition onset flag 1 varies over time and by state . . . . .	66
C.3 Hospital ranks are reasonably stable over time . . . . .	69
D.1 Excess risk varies substantially for multiday admissions . . . . .	82
D.2 Excess risk varies modestly for sameday admissions . . . . .	83

D.3	Most hospitals provide medical cardiology care . . . . .	84
D.4	Few hospitals complete bariatric surgery . . . . .	85
D.5	Obstetric patients face a uniformly high risk of a complication . . . . .	86
D.6	Some metrics are more useful for detecting differences in the safety of hospitals' medical cardiology care . . . . .	87
D.7	CHADx+ provides little information about the relative safety of knee replacements in younger patients . . . . .	88
D.8	Excess risk varies substantially within states . . . . .	89
D.9	Excess risk estimates explain the difference between hospitals' expected and observed complication rates . . . . .	90
D.10	Model predictions accord with observed rates across all subsamples . . . . .	91
D.11	All subsamples demonstrate satisfactory specificity and sensitivity . . . . .	92
D.12	Hospital performance is approximately normally distributed across our case studies . . . . .	95
D.13	Hospital performance is approximately normally distributed across our full sample . . . . .	96
D.14	Some indicators are much more stable than others . . . . .	99

## 1 The harm done

If a friend or family member is going to hospital, we usually wish them luck. Why? Because we know that sometimes, things go wrong.

Even in the best of hands, there is the risk that a patient's problem proves trickier to deal with than anticipated: treatments aren't effective, perhaps more radical surgery might be required, or they may be left with a long-term impairment after treatment. Less acceptably, errors are made and care doesn't go according to plan. Patients are harmed: they fall, receive a drug they are allergic to or, rarely, have the wrong limb operated on. Reviews of hospital charts suggest 'adverse events' occur in more than 10 per cent of hospital admissions in Australia (ranging across hospitals from 2.9 per cent to 16.6 per cent) and at least half are considered to have been preventable.<sup>1</sup>

Media reports focus on the most shocking cases: a cluster of baby deaths at Bacchus Marsh Hospital in Victoria,<sup>2</sup> or a sponge left inside a patient at Poplars Private Hospital in Sydney.<sup>3</sup> But most unsafe care is less dramatic: an otherwise healthy patient contracting an infection after their operation, for example.<sup>4</sup> Often no one is quite sure how or when a complication happened – but the patient still needs extra treatment or a longer stay in hospital.

Australians expect all hospitals to provide high-quality care. They may be surprised to learn that some hospitals achieve much better results for a procedure than others, and that the chance of something going wrong differs depending on which hospital treats you.

- 
1. Baines et al. (2015). The range is partly attributable to methodological differences in the studies.
  2. Duckett et al. (2016).
  3. Courts (2011).
  4. Vincent and Amalberti (2016).

Good quality care has many aspects (see Box 1 on the following page), and the perspectives of patients, clinicians and managers might differ as to how different aspects of quality ought to be given priority.<sup>5</sup> However, all agree that a key dimension of quality is safety; avoiding harm to patients from the care that is intended to help them.

This report proposes four reforms to reduce variation in complication rates and improve the safety of hospital care. The rest of this chapter argues that we should measure all complications, not just some. Chapter 2 shows that hospitals and doctors need to have the information necessary to know where they need to improve. Chapter 3 suggests that Australia should set a bold ambition: to reduce complication rates by more than a quarter of the current rate. Chapter 4 shows why there should be regular public reporting of the progress of each state and each public and private hospital toward this ambitious goal.

This chapter shows how a narrow view of safety has resulted in limited measurement of complications. Safety monitoring should move away from a focus on errors to a more panoramic view of patient outcomes.

### 1.1 One in nine patients suffer in-hospital harm

Over 2012-13 to 2014-15, one in every nine people admitted to hospital in Australia developed a complication.<sup>6</sup> For patients who were in

- 
5. Duckett and Ward (2008).
  6. That is, an additional diagnosis was made that was not present when the patient went into hospital. The diagnosis may be quite serious, or it may be of less consequence and may not even delay discharge from the hospital (although it will need to have been treated to be included in this data). A patient might have multiple complications across separate categories in the Classification of Hospital Acquired Diagnoses (CHADx+), see Jackson et al. (2009). All analyses in this report are based on patients classified as 'acute' in the National Hospital Morbidity

### Box 1: Safety is one component of a high-quality health system

The *quality* of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.<sup>a</sup>

The Australian Safety and Quality Framework for Health Care gives a vision for care that is: patient centred, organised for safety, and driven by information.<sup>b</sup>

The dominant quality assessment framework developed by the US Institute of Medicine lists six aims for the health care system:<sup>c</sup>

**Safe:** Avoiding harm to patients from the care that is intended to help them.

**Effective:** Providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those not likely to benefit (avoiding underuse and misuse, respectively).

a. Runciman et al. (2009).

b. See: <https://www.safetyandquality.gov.au/national-priorities/australian-safety-and-quality-framework-for-health-care/>.

c. See: <https://www.ahrq.gov/professionals/quality-patient-safety/talkingquality/create/sixdomains.html>.

**Patient-centred:** Providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions.

**Timely:** Reducing waits and sometimes harmful delays for both those who receive and those who give care.

**Efficient:** Avoiding waste, including waste of equipment, supplies, ideas, and energy.

**Equitable:** Providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status.

The framework makes some potential trade-offs explicit. It places safety below quality. But care that is not timely is unsafe, as is ineffective care. In practice, the terms safety and quality are often used interchangeably.

hospital overnight, the rate was even higher: more than one in four (Figure 1.1). About 900,000 patients each year have a complication.<sup>7</sup>

These complications are a real problem. Some cause patients discomfort, delay recoveries, and extend hospital stays.<sup>8</sup> The most serious cause permanent injury or death.

On average, patients who suffer a complication after a procedure<sup>9</sup> end up staying in hospital for five extra days. Figure 1.2 on the following page shows how this impact varies for two procedures, and for all overnight patients, other than women admitted to give birth. Patients may continue to suffer the consequences of these incidents after they are discharged.

The following sections describe the evolution of a scientific literature on hospital safety as clinicians and managers doggedly chased better results for patients.

## 1.2 The old way of thinking about safety normalised harm to patients

Over the past 40 years the health care system has moved from a focus on doctors' mistakes toward an understanding that poor outcomes are often the result of system rather than individual failures. The publication of a landmark US report, 'To err is human', in the year 2000 drew attention to the large amount of harm caused by health care and

---

Dataset; see the Methodological Supplement for more details of the data set and our analytical approach, including how we standardise for patient factors.

7. About 725,000 overnight patients suffer a complication.
8. While many complications result in longer hospital stays, it is also true that more complex patients have longer stays, increasing their risk of complications such as infections. Figure 1.2 focuses on procedural complications. Unlike medication errors, the risk of procedural complications does not increase one-for-one with the length of a patient's stay.
9. Defined as Major CHADx Class 1: Procedural complications. The CHADx+ classification system is described in Section 2.1.2.

**Figure 1.1: Complications are common**

Share of admissions involving at least one complication, per cent

30

25

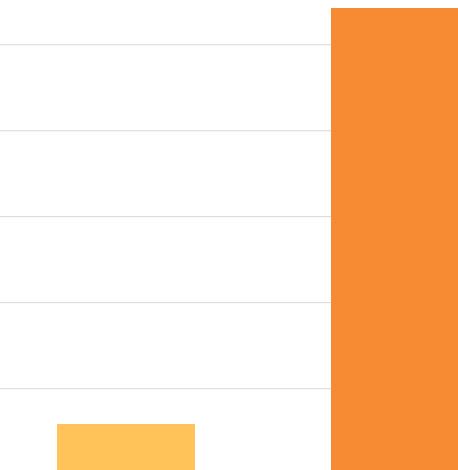
20

15

10

5

0



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

promoted a ‘systems approach’.<sup>10</sup> This led to an emphasis on ‘learning from mistakes’, and trying to ensure that the culture of health services was just and facilitated that learning.<sup>11</sup>

Yet safety was still defined as an absence of specific negative events. When such events happened, there was a search for who or what caused the errors. Such thinking inspired the development of health care incident reporting systems modelled on aviation safety reporting systems. The hope was that the process of reporting, investigating, and determining how an event could have been prevented, and then changing systems, would make hospitals safer.<sup>12</sup>

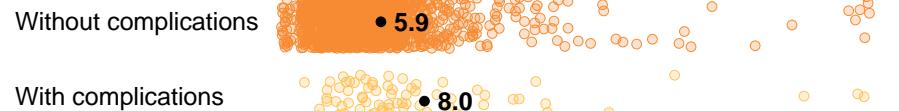
But this focus on preventability was misguided in several respects. Most fundamentally, ‘preventability’ was not a useful concept because what is ‘preventable’ is subjective, changes over time, and depends on the context of care (see Box 2).<sup>13</sup>

Disagreement among clinicians about what should be considered ‘preventable’ mired safety improvement efforts in an unproductive debate. The concept of preventability also established a culture of blame.<sup>14</sup> This limited learning because it engendered defensiveness among those involved in safety incidents, and quests for root causes distracted from lessons that could be applied more generally.<sup>15</sup> Detailed investigations of rare, dramatic health care events have identified few patterns in the

- 10. Institute of Medicine (2000).
- 11. Marx (2001); and Waring et al. (2015).
- 12. The limitations of incident reporting systems are discussed in detail in the Grattan Institute’s previous report, *Strengthening safety statistics*.
- 13. Vincent and Amalberti (2016).
- 14. A leading Australian safety expert and a reviewer of this paper noted: “Despite significant efforts over more than a decade, I and others have not been able to de-couple the notion of safety with ‘medical mistakes’.”
- 15. Cook and Nemeth (2010); Hayward and Hofer (2001); Shojania and Mheen (2015); and Vincent and Amalberti (2016).

**Figure 1.2: Procedural complications increase a patient’s stay in hospital**  
Average length of stay for patients with and without procedural complications, by admission type, number of days

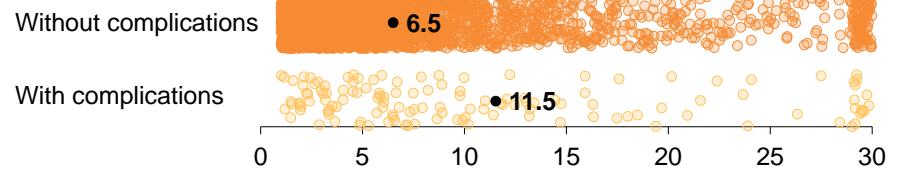
#### Knee replacement



#### Bariatric surgery



#### All multiday non-obstetric



Source: Grattan analysis of the 2012–15 National Hospital Morbidity Dataset.

#### Box 2: Preventing the ‘unpreventable’

On occasions, system changes that eradicate certain types of ‘unpreventable’ incidents have been identified. For example, central line-related bloodstream infections in intensive care units were not considered ‘preventable’ until a determined improvement and measurement effort in Michigan intensive care units proved otherwise.<sup>a</sup>

- a. Pronovost et al. (2010); and Dixon-Woods et al. (2012).

circumstances in which they arise, and so, unsurprisingly, have led to little progress in reducing their prevalence.<sup>16</sup>

But the most serious consequence of the focus on preventability was that it normalised harm to patients. Focusing safety improvement efforts on ‘errors’ that caused ‘preventable’ harm implied that other instances of harm to patients were acceptable, and less worthy of being tackled.

Some clinicians routinely achieve better outcomes for patients than most. For instance, elderly, obese patients with renal failure are at high risk of physiological deterioration, such as acquired fluid and electrolyte disturbances or pneumonia. These complications may occur without any errors being made by clinicians, and, in some patients, the complications may be inevitable. Yet some clinicians deliver care that beats the odds.

Focusing on only a subset of complications is a problem because it removes the impetus to innovate and learn from clinicians who manage, through clinical excellence, to defy the odds across a broad range of potential complications.

### **1.2.1 New definitions of hospital safety focus on opportunities to improve**

Safe care is now understood to mean more than ensuring no egregious errors occur – it means ensuring that every patient’s recovery is as swift and comfortable as possible.

The new emphasis is on reducing risk by improving system performance overall.<sup>17</sup> This requires studying frequent events, and normal (and excellent) performance. Safety monitoring needs to include the full spectrum of patient outcomes. The aim is to ensure that the number of

16. Wears (2016); and Kellogg et al. (2017).

17. Hollnagel (2014); Besnard and Hollnagel (2014); and Dekker (2014).

good outcomes is as high as possible and the number of events that could be harmful to patients is acceptably low.<sup>18</sup>

### **1.2.2 New definitions of hospital safety put patients at the centre**

Patients’ priorities and experiences are now considered to be central to the safety of hospital care.<sup>19</sup>

This means safety improvement policies should acknowledge and facilitate patients’ right to be informed about the likely outcomes of their care.<sup>20</sup> It is now possible for a hospital to know the rate of complications which occurred for a patient in a given demographic having a given procedure in the past year. Patients should know that information too, so they can appreciate the risks they face.<sup>21</sup>

‘Patient centredness’ also means patients should be engaged in key decisions about their care.<sup>22</sup> Too often, only lip service is paid to ‘patient centredness’.<sup>23</sup> While relatively few decision-making aids have

18. Hollnagel (2014). It is also recognised by safety academics that even egregious errors such as incorrect site surgery will recur (Pandit (2016)). The concept of a ‘never event’ is fundamentally flawed for a complex system like health care (I. Moppett and S. Moppett (2016)), although attention to the processes of care can reduce the likelihood of an undetected error causing harm.

19. Jorm et al. (2009); Hor et al. (2013); and Pronovost et al. (2017).

20. Rubin (2017).

21. Stacey et al. (2017).

22. Patient decision aids, where used, improve patients’ knowledge (Stacey et al. (2017) and Brown et al. (2015)). Patient decision aids have been developed for many conditions (see: <https://decisionaid.ohri.ca/index.html>), and the Australian Commission on Safety and Quality in Health Care has a program of work on patient decision aids (see: <https://www.safetyandquality.gov.au/our-work/shared-decision-making/patient-decision-aids/>). However, patient decision aids have a number of limitations (Agoritsas et al. (2015)), not least of which is their implementation and use in practice (Elwyn et al. (2008)).

23. Légaré and Thompson-Leduc (2014); and Stiggelbout et al. (2015).

been developed to engage patients in clinical decisions,<sup>24</sup> sometimes clinicians may not have enough information to know the facts about risks and benefits themselves. Consequently, they frequently overestimate benefits and underestimate risks.<sup>25</sup> This lack of – or inaccurate – information hinders proper informed consent.<sup>26</sup> As one consumer argued recently:

One of the reasons that clinicians struggle to form partnerships with patients and consumers is that there is inadequate information for proper informed consent. The numbers that are collected don't filter through to clinicians dealing with the patient, and certainly not to the patient themselves. I encourage people to ask doctors three questions: 'What are my options?', 'What are the treatment outcomes – both benefits and risks?', and 'How likely are those outcomes to happen to me?'. Doctors though say 'We just don't have that information'.

— Carey (2017)

The shift toward patient centredness furthers, rather than competes with, clinical objectives. When data about patients' likely outcomes is made public, clinicians evaluate the benefits and risks of care more accurately.<sup>27</sup> When patients are more engaged in shared decision-making, they gain a more accurate appreciation of the risks involved,<sup>28</sup> and become more comfortable with the decisions made and more satisfied with their care.<sup>29</sup> They also suffer fewer adverse events.<sup>30</sup>

- 
- 24. A good exception is the wealth of decision-making aids available for breast cancer treatments.
  - 25. Hoffmann and Mar (2017).
  - 26. Medical practitioners always need to consider individual patient circumstances and their own experience, but we argue that relevant information often does exist and should be made available to consumers and practitioners.
  - 27. Sacks et al. (2016).
  - 28. Stacey et al. (2017).
  - 29. Boss et al. (2016); Walsh et al. (2014); and Stacey et al. (2017).
  - 30. Schiffinger et al. (2016); Holzmueller et al. (2012); and Weingart et al. (2011).

### 1.3 Australia's safety improvement efforts are yet to reflect these changes to safety thinking

Despite these substantial changes in the way experts are thinking about safety improvement, in Australia preventability is still the key criterion used, patients don't receive the information they could about risks and outcomes, and the data sources available to support safety improvement are not fully utilised.

In our previous report, *Strengthening safety statistics*, sources of data on patient outcomes examined included: routine data, clinical quality registry data, death audit data, incident reporting and investigation data, patient-reported experience measures, and patient-reported outcome measures. Many instances were found where the data could and should be more accurate and relevant, and importantly more accessible and understandable.

Incremental safety improvement efforts by clinicians are obstructed where there is limited transparency of data. Such limited transparency was justified only when there was a real threat of misplaced blame.<sup>31</sup>

Given this, it's unsurprising that Australia appears to be making negligible measurable progress on reducing the incidence of complications in hospitals. Figure 1.3 on the following page shows that the prevalence and mix of complications in Australian hospitals didn't change significantly between 2012 and 2015.

This lack of measurable progress on this broad outcome measure might be a measurement problem that disguises an improvement trend: perhaps hospitals are becoming safer (as a result of a raft of State and National initiatives) – but hospitals are also treating sicker patients overall. However, there is little evidence of this latter trend. The lack of overall improvement is especially disappointing when the 'January effect' demonstrates that temporary organisational dysfunction results

---

31. Duckett et al. (2017).

in measurable worsening of complications.<sup>32</sup> It's not unreasonable to expect improvements in hospital function and patient safety to have a measurable impact, gradually reducing the rate of complications.

To reduce the harm caused to patients, Australia needs to modernise its hospital safety improvement strategy. The emphasis needs to move away from analysing in excruciating detail where things went wrong,<sup>33</sup> and towards understanding how some clinical teams achieve exceptional outcomes; from a focus on blame to a focus on incremental improvement.

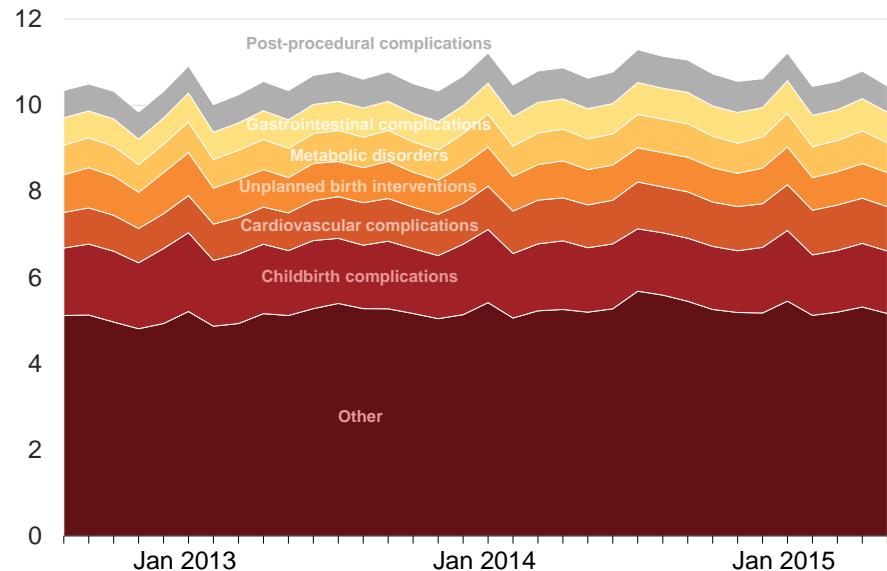
This report uses a powerful dataset to show how much scope there is for hospital safety to be improved. In the tradition of Florence Nightingale, we focus on the rates and patterns of complications – the epidemiology of patient outcomes.<sup>34</sup>

We take the position that it doesn't matter whether these complications have been declared to be 'preventable' – the fact that some hospitals consistently achieve much lower (risk-adjusted) rates than others is enough to demonstrate that there is scope for the incidence of all of these complications to be reduced. Accordingly, we call for relative, rather than absolute, safety improvement goals – a focus on *reducing* all complications to the best rate achievable, rather than just focussing on *preventing* some complications.

- 
32. Figure 1.3 illustrates this 'January effect' – an uptick associated with transitions in hospitals over summer and at the start of a new year. Patients have longer stays, higher mortality and suffer from more adverse events in the month after the mass changeover of junior medical staff and commencement of work by new, inexperienced interns: Young et al. (2011), Jen et al. (2009) and Wen et al. (2015). This period of significant organisational disruption happens in July in the US, August in the UK, and January in Australia, and hence is named differently.
  33. The limitations of the current reporting and investigation systems are discussed in *Strengthening safety statistics*.
  34. Neuhauser (1999).

**Figure 1.3: In Australia the rate of complications is not falling**

Share of admissions involving at least one complication, categorised by the most common CHADx+ categories, per cent



Note: Rates of individual CHADx+ categories have been scaled such that they sum to the share of admissions involving at least one of any complication.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

This report reveals the current variation in the safety of hospital care in Australia, and calls for a new national commitment to making all hospitals as safe as the safest 10 per cent. It's an ambitious goal: it would reduce the national complication rate by more than a quarter.

## 2 Australia needs to set more ambitious hospital safety targets

A lot of effort and expense is devoted to monitoring the safety and quality of care in Australian hospitals, but we are neither using the information as well as we should,<sup>35</sup> nor focussing on the right improvement targets. This chapter argues for a more ambitious approach.

Indicators that assess quality of care have proliferated.<sup>36</sup> Data and indicators are not the problem. Australia has enough data to gauge the scope for improvement in hospital safety, using the same data that the Commonwealth Government uses to allocate funding to states for increases in hospital admissions and states use to determine hospital funding.<sup>37</sup> The following sections describe the main measures of patient harm that are tracked for all hospital admissions, and compares them to reveal the lack of ambition in Australia's current safety monitoring policies.

### 2.1 Australia has three key measures of unsafe care

Australia has three major classifications of harm to patients: eight sentinel events, 16 priority 'Hospital Acquired Complications' (HACs), and the comprehensive Classification of Hospital Acquired Diagnoses (CHADx+).<sup>38</sup>

Table 2.1 shows the rates of the different types of complications. The rates equate to about 100 sentinel events recorded each year, about

**Table 2.1: Far more harm is caused to patients in hospital than is captured in sentinel events or 'Hospital Acquired Complications' statistics**

Share of admissions involving at least one complication

Complication type	Sameday admissions	Multiday admissions	All admissions
Sentinel events	—	—	0.0012%
Designated 'Hospital Acquired Complications' (HACs)	0.1%	5.2%	1.7%
All complications	3.0%	27.1%	10.7%

*Note: The shares of sameday and multiday admissions involving a sentinel event are not published and cannot be inferred from our dataset.*

*Source: Grattan analysis of the 2012–15 National Hospital Morbidity dataset and the Steering Committee for the Review of Government Service Provision (2017).*

35. As outlined in *Strengthening safety statistics*.

36. Copnell et al. (2009).

37. This is not to say the data is perfect. Indeed we have identified areas where it should be improved (Duckett et al. (2017)), but coding of the data set has been shown to be reliable (Henderson et al. (2006)) and is regularly audited by states.

38. The Annual Report on Government Services uses another classification set, but this is not used elsewhere (Steering Committee for the Review of Government Service Provision (2017)).

140,000 priority Hospital Acquired Complications, and almost 900,000 complications of any kind.

Since 2002, Australia has had an agreed-upon list of eight extremely serious complications that are publicly reported. This list was developed from one initially created as part of hospital accreditation in the United States in the late 1990s.<sup>39</sup> Since then the US list has been revised multiple times. It now has 29 items, which are now called ‘serious reportable events’.<sup>40</sup>

Australia’s eight sentinel events are all catastrophic (Box 3). Fortunately, they are also rare: only 99 sentinel events were recorded across Australia in 2014-15.<sup>41</sup> Most of the listed events are likely to involve preventable errors. But not all: maternal death is not always preventable, nor is intravascular gas embolism.<sup>42</sup>

### 2.1.1 Hospital Acquired Complications (HACs)

The HACs list, released in 2016, was developed as a subset of codes originally used in the Classification of Hospital Acquired Diagnoses (see next section). Table 2.2 on the following page shows the incidence of HACs.

The HACs list was developed with a focus on preventability to be used for top-down accountability and funding. Heads of government then agreed that the designated HACs met the following criteria:

- Clinical evidence is available to demonstrate that the HAC can be prevented with ‘best clinical practice’.

39. Kizer and Stegun (2005); Leape (2002); and Berman (1998).

40. See: [http://www.qualityforum.org/Topics/SREs>List\\_of\\_SREs.aspx](http://www.qualityforum.org/Topics/SREs>List_of_SREs.aspx).

41. Steering Committee for the Review of Government Service Provision (2017).

42. We understand that removal of some of the non-preventable events from this list may be imminent.

### Box 3: Australia’s eight sentinel events

1. Procedures involving the wrong patient or body part, resulting in death or major permanent loss of function
2. Suicide of a patient in an inpatient unit
3. Retained instruments or other material after surgery requiring re-operation or further surgical procedure
4. Intravascular gas embolism resulting in death or neurological damage
5. Haemolytic blood transfusion reaction resulting from ABO incompatibility
6. Medication error leading to the death of a patient reasonably believed to be due to incorrect administration of drugs
7. Maternal death associated with pregnancy, birth and the puerperium
8. Infant discharged to the wrong family

- Evidence shows that individual hospitals and Local Health Networks are able to prevent the HAC and that the causes of the condition are within the control of the hospital.
- The strength of external influences (e.g. patient factors) does not unduly affect the hospital's ability to avoid the HAC.
- There is sufficient evidence to inform / instruct health services on how to avoid the HAC.
- The development of the HAC measure has been subjected to valid construction.<sup>43</sup>

The Australian Commission on Safety and Quality in Health Care, which was responsible for the development of the HACs list, adopts a softer phrasing than the government position, stating that:

A hospital-acquired complication (HAC) refers to a complication for which clinical risk mitigation strategies may reduce (but not necessarily eliminate) the risk of that complication occurring.<sup>44</sup>

This softer phrasing is at odds with the use of the HACs list by the Commonwealth to adjust funding to states. The state penalty varies with the complexity of the patient but for low complexity patients it assumes that all HACs can be prevented and so the state is to be penalised for the full incremental cost of the HAC.<sup>45</sup>

### HACs ignore some patients

The HACs list excludes complications which occur in patients admitted for mental illness or who have drug or alcohol problems. Yet these

43. COAG (2017).

44. See: <https://www.safetyandquality.gov.au/our-work/indicators/hospital-acquired-complications/>.

45. IHPA (2017).

**Table 2.2: Incidence of Hospital Acquired Complications, 2012-15**

Share of admissions involving at least one complication

HAC1	Pressure Injury	0.04%
HAC2	Falls resulting in fracture or other intracranial injury	N/A
HAC3	Healthcare associated infection	0.76%
HAC4	Surgical complications requiring unplanned return to theatre	N/A
HAC5	Unplanned intensive care unit admission	N/A
HAC6	Respiratory complications	0.08%
HAC7	Venous thromboembolism	0.05%
HAC8	Renal failure	0.01%
HAC9	Gastrointestinal bleeding	0.09%
HAC10	Medication complications	0.15%
HAC11	Delirium	0.30%
HAC12	Persistent incontinence	0.05%
HAC13	Malnutrition	0.07%
HAC14	Cardiac complications	0.45%
HAC15	Third and fourth degree perineal laceration during delivery	0.08%
HAC16	Neonatal birth trauma	0.00%
Any HAC		1.72%

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

patients are not excluded from any other hospital-wide measure of the safety of care.

As illustrated in Table 2.3, these patients also suffer complications in the HAC list. There is no legitimate reason to exclude complications occurring in patients with a principal diagnosis of mental illness or alcohol and drug conditions from system-wide monitoring.

#### HACs ignore many complications

The complications chosen for inclusion in the HACs list are somewhat idiosyncratic. Figure 2.1 on the following page shows that there is little consistency in the types of conditions included, nor in the narrowness with which conditions are defined.

The complications excluded from the list include serious complications, such as lacerations during procedures, and common complications with effective clinical risk mitigation strategies,<sup>46</sup> such as constipation. The HACs classification also defines some complications narrowly, such as acute renal failure and pressure ulcers, while applying no threshold at all for delirium, a common complication especially in dying patients.<sup>47</sup>

#### Improvements in small areas don't show up in the big picture

Because HACs focus on a subset of complications, which don't occur with a high frequency, improvements in rates may not be seen when looking at complications overall.

There have been measurable improvements in the safety of hospital care in Australia in recent years. For example, cases of *Staphylococcus aureus* bacteraemia have declined from 0.96 to 0.73 cases per 10,000 days of patient care over the last four years, but because the number of cases involved is less than 1,500 in any year, this improvement is

46. That is, their prevalence can be reduced by appropriate drugs.

47. O'Regan et al. (2013); Hosie et al. (2013); and Grassi et al. (2015).

**Table 2.3: Mental health patients and drug and alcohol patients also experience complications**

Share of admissions involving at least one complication, by patient type

	Mental health patients	Drug and alcohol patients	Other patients	All patients
HACs	0.8%	0.4%	1.8%	1.7%
All complications	6.9%	4.7%	10.8%	10.7%

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

not picked up in looking at broad trends.<sup>48</sup> The same issue affects other HAC indicators.

#### 2.1.2 Classification of Hospital Acquired Diagnoses (CHADx+) or 'all complications'

The CHADx+, first developed in 2009,<sup>49</sup> aims to provide a comprehensive classification of all hospital-acquired complications present in the routine data (as recorded in the clinical record). It has been revised a number of times and now also includes incidents detected using procedure codes (CHAPx).<sup>50</sup>

The CHADx+ classification does not flag which types of complications should be the focus of prevention initiatives. It was designed to assist bottom-up improvement initiatives by local clinicians, by enabling hospitals and hospital departments to identify the complications relevant to their situation. CHADx+ is publicly available, but used in only a few hospitals.

---

48. Burgess et al. (2017).

49. Jackson et al. (2009).

50. We use CHADx+ version 1.4 in this report.

Table 2.4 on the next page shows the overall incidence of CHADx+ complications, and the incidence of each major CHADx+ class.<sup>51</sup> Information on the most common complications for elective procedures – categorised into CHADx+ classes and stratified by age, sex, and whether or not the patient stayed overnight – is provided for each specialty at: <https://grattan.shinyapps.io/dummyLiftingLidApp/>. This app provides a ‘proof-of-concept’ as to the type of reporting that existing data sources allow.<sup>52</sup>

## 2.2 Australia's current policies are unambitious

The characteristics of the key patient safety metrics underscore the modest ambition of Australia's current safety improvement policies.

### 2.2.1 Current policies imply that focusing on a subset of complications is sufficient

Australia's hospital reporting and pricing policies principally focus on sentinel events and HACs.<sup>53</sup> In recent years, the Australian Institute of Health and Welfare has expanded reporting to include the HACs and CHADx+ in its publications of patient care statistics.<sup>54</sup> Hospital managers will soon receive information about their prevalence of the 16 priority complications (HACs) in their hospitals. This is a useful advance, but hospitals should receive information about all complications.

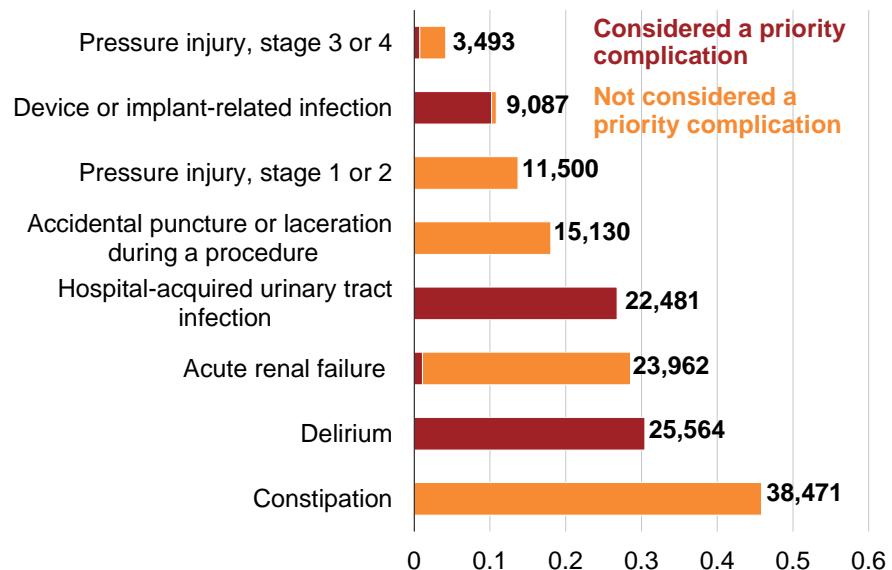
51. These estimates differ slightly from those published by the Australian Institute of Health and Welfare because this report uses a later, more comprehensive edition of the classification. For further details, see the Methodological Supplement to this report.

52. At the time of publication, the app was still being refined. As such, results produced by the app should be taken as indicative only and not relied upon.

53. However, the annual Report on Government Services also reports on a broader range of adverse events with a prevalence of around 7 to 8 per cent per annum, see Steering Committee for the Review of Government Service Provision (2017, Table 12A.35).

54. Burgess et al. (2017).

**Figure 2.1: Many important complications are not included in HACs**  
Share and annual number of admissions involving at least one complication, classified by inclusion in the HACs priority complication list, per cent



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

**Table 2.4: Incidence of CHADx+, 2012-15**

Share of admissions involving at least one complication

MCHADx1 Procedural complications	1.25%	MCHAPx1 Ventilatory support	1.12%
MCHADx2 Adverse drug events	0.48%	MCHAPx2 Haemorrhage/haematoma management	1.03%
MCHADx3 Accidental injuries	0.31%	MCHAPx3 Return to theatre or procedure room	0.20%
MCHADx4 Hospital-acquired infections	1.18%	MCHAPx4 Procedural complications relating to childbirth	1.54%
MCHADx5 Cardiovascular complications	1.77%	MCHAPx5 Nutrition support	0.16%
MCHADx6 Respiratory complications	0.66%	MCHAPx6 Fluid management	0.06%
MCHADx7 Gastrointestinal complications	1.26%		
MCHADx8 Skin conditions	0.60%	Any CHADx	9.18%
MCHADx9 Genitourinary complications	0.81%	Any CHAPx	3.84%
MCHADx10 Hospital-acquired psychiatric states	0.61%	Any CHADx+	10.63%
MCHADx11 Early pregnancy complications	0.01%		
MCHADx12 Labour and delivery complications	2.83%		
MCHADx13 Perinatal complications	0.10%		
MCHADx14 Haematological disorders	0.55%		
MCHADx15 Metabolic disorders	1.32%		
MCHADx16 Nervous system complications	0.15%		
MCHADx17 Other complications	1.57%		

*Notes: A limitation of our dataset precludes us from detecting adverse drug events (CHADx 2.01-2.13) and falls (CHADx 3.01-3.04). Please see this report's Methodological Supplement for a fuller discussion. CHAPx refers to complications identified using procedure codes.*

*Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.*

The prevailing narrow view of hospital safety is reflected in hospital pricing arrangements. Since July 2017 public hospital admissions in which a sentinel event occurred attract no Commonwealth subsidy under the Commonwealth-state growth funding arrangements. In addition, penalties based on a number of the 16 prescribed HACs will apply from 1 July 2018.<sup>55</sup>

This focus on a limited range of priority HACs may have been designed to serve a political end: start with an uncontested, clinically agreed, narrow target for change to minimise the potential opposition to this approach to clinical accountability.<sup>56</sup> It might have been hoped that a small target might minimise gaming through manipulating coding.<sup>57</sup> It might also be consistent with a planned approach to change management – start with a clear, highly specific indicator (or indicator set) to reduce the initial amount of change expected, increase the salience and specificity of any identified problem, and facilitate clinician engagement.<sup>58</sup> A narrow focus may be a sensible initial approach, but it risks entrenching narrowly based measures. Australia's medium-term goal should not be so limited.

Penalising unsafe care provides a financial incentive for safety improvement efforts. But basing safety policy on incomplete lists of complications is a problem because it detracts attention from other complications which may be more significant in some hospitals or specialties, and where improvement may be possible too. It may encourage hospitals to focus their improvement efforts too narrowly, in line with perceived financial incentives, to the detriment of broader safety issues which are not the focus of rewards or penalties.<sup>59</sup> Focusing on sentinel events and HACs is not enough, because these classifications

---

55. Duckett (2017).

56. Mehlman (2013).

57. This explanation was advanced during the consultation phase of this report.

58. This too was suggested during the consultation phase.

59. Gillam et al. (2012).

represent just a fraction of the complications that affect patients in Australian hospitals.

Focusing on a select list of complications reflects a view that 'health care improvement will come by improving one process at a time ... through bounded projects, rather than designing an integrated system of operations to eliminate or reduce all harms'.<sup>60</sup> This view denies the reality that 'patients are all at risk for dozens of harms, many of them not clearly confined or easily targeted by highly specific efforts'.<sup>61</sup>

## 2.2.2 Australia could dramatically reduce the incidence of complications if it aimed higher

Comparisons of risk-adjusted rates of all complications across Australian hospitals indicate that if all hospitals provided care as safe as the top 10 per cent of hospitals, the average rate of complications could be reduced by more than a quarter.<sup>62</sup> This would mean an extra 250,000 patients would leave Australian hospitals complication-free each year. Yet, as illustrated in Figure 2.2 on the following page, eliminating all HACs would only go 15 per cent of the way towards achieving this objective.

---

60. Pronovost et al. (2017).

61. Ibid.

62. In fact, setting a target needs to take into account the stability and distribution of the specific complications. This issue is discussed in the Methodological Supplement. The Methodological Supplement also shows our risk-adjustment process and the steps taken to ensure these figures reflect the true scope to reduce the incidence of complications, not differences in hospitals' coding practices. Systematic differences in the thoroughness of coding associated with hospital size and across states have been netted out of these figures directly. Our approach to calculating the scope for improvement relative to order statistics like the best decile is also robust to unobservable differences in the thoroughness of specific hospitals' coding, as it is unlikely that there are enough hospitals with outlier coding practices to make up the entire best decile, or quartile.

This finding has important implications for how ambitious Australia wants to be with its hospital safety improvement policies. While focusing on the subset of complications labelled ‘preventable’ may be politically expedient, it is an unnecessarily narrow medium-term target.

Some hospitals have achieved a much lower rate for many complications not on the designated HAC list. Many sustain this achievement year after year.<sup>63</sup> This good performance should be delivered to all hospital patients. It means we need to look at what good hospitals are achieving, rather than simply trying to explain what went wrong in poorly performing hospitals.

Focusing on *reducibility* of complications also has the advantage of avoiding the vexed issue of debating and defining what is *preventable*.

The focus on a narrow HACs list also means that some hospitals with low HACs rates may be misled into believing they have no scope to improve, when in fact they may have considerable scope to reduce complications not on the HACs list. As discussed in Chapter 3, hospitals should be encouraged to identify their own scope for improvement.

### 2.3 More ambitious targets would bring greater success

How Australia defines its safety improvement targets matters because these targets affect the amount of progress we can expect.

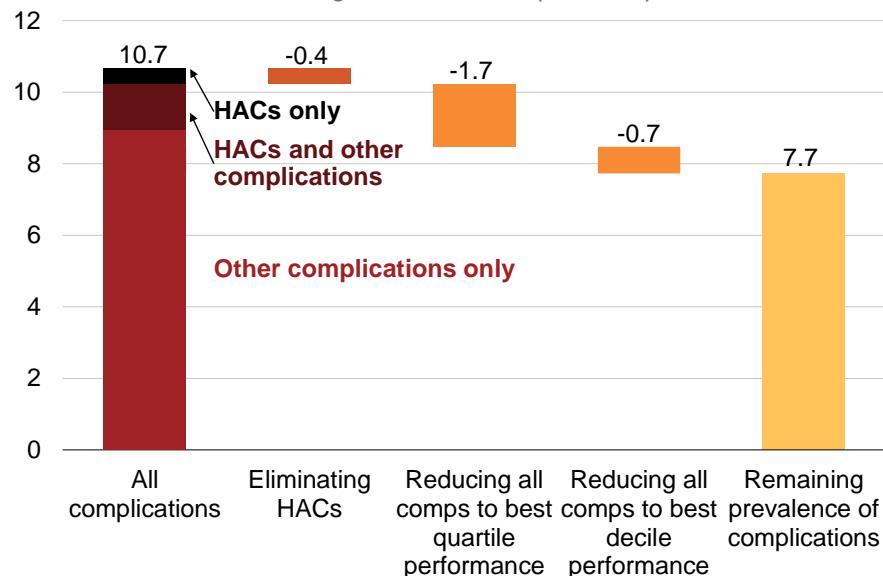
We can learn a lot from common events and ‘positive deviants’

Australia will be able to improve hospital safety more if we learn not just from rare, catastrophic errors but from all complications.

Many patients are considered high risk for complications due to the severity of their illness and their co-morbidities. Yet by excluding most

63. The stability of different safety metrics is examined in the Methodological Supplement.

Figure 2.2: More than a quarter of all complications could be avoided if all hospitals were as safe as the top 10 per cent of hospitals  
Share of admissions involving at least one complication, per cent



Note: Other complications defined as CHADx+ events. ‘HACs and other complications’ refers to admissions that involve a HAC but would still involve a CHADx+ event even if HACs were eliminated. Does not sum to 7.7 per cent because of rounding.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

complications from transparent reporting practices, we provide insufficient support to clinicians who are striving to achieve better outcomes for these patients.

The ability of medical teams to adapt to varying circumstances is astounding. Analysis of everyday successes and outstanding results can be more instructive for safety improvement efforts than analysis of failures.<sup>64</sup>

Complete complications data captures information about exceptionally good results, as well as the average outcomes and the causes for concern. Doctors and hospitals with exceptionally good performance ('positive deviants') should be identified and their methods studied, so others can understand how they succeeded. Their 'recipe' can then be trialled elsewhere and, if success continues, disseminated widely.<sup>65</sup>

#### Hospitals should identify their specific safety priorities

More ambitious safety improvement targets would also encourage hospitals to identify their own priorities. The biggest opportunities for safety improvement will not be the same across all hospitals. Some hospitals may be industry leaders at reducing some complications, while performing poorly in other regards.

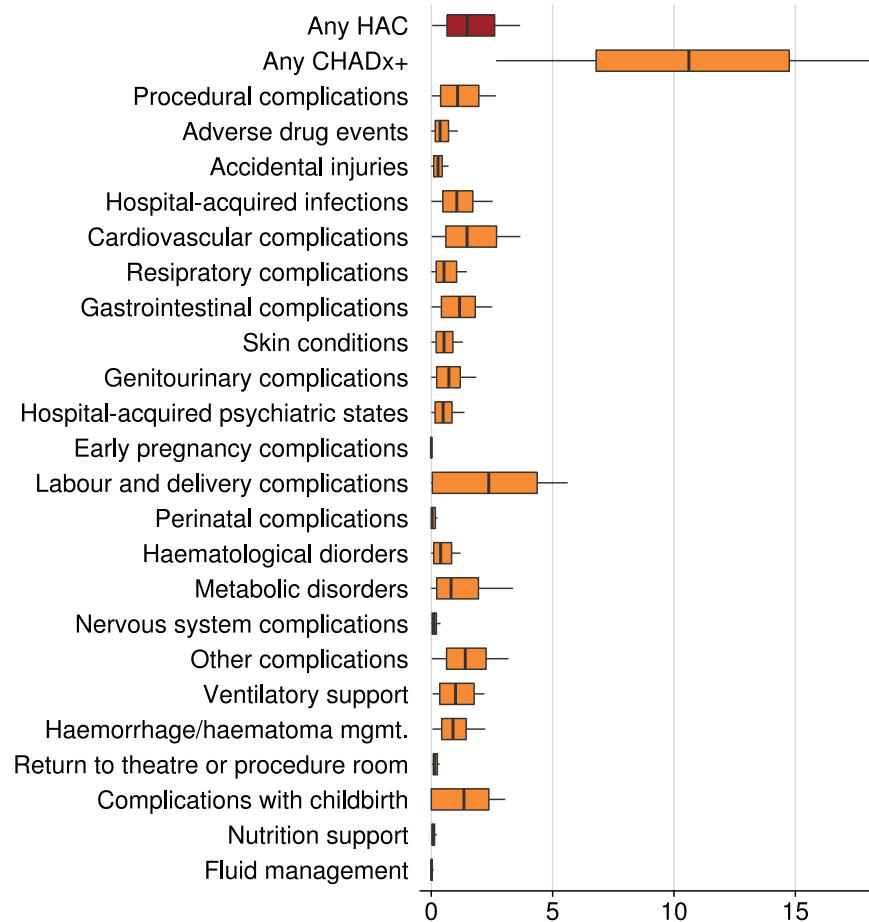
Figure 2.3 shows that there is substantial variation between hospitals in the raw incidence of complications overall, and in the complications declared to be the hospital sector's priority (HACs). However, there is also variation in the incidence of other categories of complications, such as nervous system complications and cardiovascular complications. A hospital's performance may also vary within a category of complications.<sup>66</sup>

64. Hollnagel (2014).

65. Lawton et al. (2014); and Baxter et al. (2015).

66. As revealed by analysis of the Minor CHADx+ classes.

**Figure 2.3: Complication rates vary a lot between Australian hospitals**  
Boxplots of the shares of admissions involving at least one complication at Australian hospitals, per cent



Notes: As is typical of boxplots, the central line marks the median, the shaded box extends across the interquartile range, and the whiskers of each plot extend out to the 10th and 90th quantiles. This figure is based on raw, not risk-adjusted, hospital complication rates.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

All of this variation matters, because it represents opportunities to reduce discomfort and harm to patients. As the following chapter shows, the key opportunities to improve the safety of care are different across hospitals. Accordingly, safety improvement priorities should be tailored for each hospital, rather than set uniformly across the sector.

More ambitious safety improvement targets would encourage each hospital to focus on the areas where they have the greatest scope to improve. More granular performance metrics would help them do so.

#### Common events are better suited to performance monitoring

When safety priorities are defined narrowly, we end up monitoring events that are rare. This data won't necessarily be useful to a hospital that is trying to evaluate and refine its safety improvement initiatives.

It's necessary to monitor rare events over a longer period in order to isolate the trend in the data from the statistical noise. By the time you have enough information to pin-point a hospital's relative performance, the circumstances surrounding the result may be a distant memory. It's possible to measure a hospital's performance with greater precision if the performance metric is based on a more common event. More common events – such as any complication in overnight patients – can also be validly used in smaller hospitals and smaller specialties.

Figure 2.3 on the preceding page also shows there is considerably more variation in the rate of all complications across hospitals than in HACs rates.<sup>67</sup> As a consequence, over a given time period rates for all complications identify hospitals' relative performance with greater precision than the incidence of HACs.

Where a given level of precision is required, it's also possible to report rates for all complications more frequently. Consequently, it's easier to

observe changes in each hospital's safety through its overall incidence of complications. This issue is discussed in more depth, including the implications for small hospitals, in the Methodological Supplement to this report.

#### Easy targets are lazy targets

Setting a target involves balancing the costs of getting to the target with the benefits that will be achieved. We know that it is easier to improve the performance of the worst hospitals than in the average hospital,<sup>68</sup> but that does not mean our ambition should be so limited.

Vastly more patients are treated in the middle-performing 80 per cent of hospitals than in the worst 10 per cent of hospitals. So improving the performance of that middle group will benefit vast numbers of patients. Hospitals in the middle band should not be allowed to wallow in complacency, when other hospitals, faced with the same budget constraints, are doing so much better.

Ambitious targets should therefore be set: the aim should be to get all hospitals up to the level of the best 10 per cent, rather than focussing only on improving care in the worst-performing hospitals.

The following chapter shows that, with granular performance metrics, Australia can move beyond artificially narrow lists and start identifying where hospitals' specific safety improvement opportunities lie.

---

67. Based on the interquartile ranges (*i.e.* the difference between the top and bottom 25 per cent for the two measures).

68. Ivers et al. (2012); and Mendelson et al. (2017).

### 3 Data can inform hospitals about their specific strengths and weaknesses

Chapter 1 argued that Australia should study the full range of patient outcomes, particularly the better-than-expected results, rather than only the rare catastrophes.

Chapter 2 showed that the narrow view of ‘harm’ to patients in Australian hospitals matters because it drastically underestimates the scope for improving hospital safety and may create incentives to ignore important opportunities.

A comprehensive view of complications is needed. The substantial variation found in the general quality of hospital care means that substantial data-driven quality improvement is within reach.

This chapter demonstrates how that can be done. It interrogates the routine data to answer three questions: What are specific hospitals’ strengths and weaknesses? Who excels at different aspects of care? And what does this mean for specific patients, and their healthcare decisions?

We’ve used anonymised routine data (that is, it does not identify particular hospitals or patients) to quantify how much variation there is in the safety of Australia’s hospitals. The Australian Institute of Health and Welfare dataset we use includes data from all admissions to Australia’s public and private hospitals between July 2012 and June 2015. Box 4 on the next page gives an example of how much information this dataset contains about individual patients’ hospital stays.

By comparing risk-adjusted complication rates across hospitals, we identify the extent to which safety could be improved overall, and what this means for specific institutions and their patients.

The fact that there are big differences between hospitals highlights another important policy issue – hospitals and their clinicians need to

have access to this information so they can analyse their performance and identify opportunities for improvement. We discuss this issue in Chapter 4.

Hospitals and states differ in their coding practices – an issue discussed in the Methodological Supplement to this report. In particular, the use of the indicator of whether a diagnosis occurred after admission – our measure of complications – varies across states. These differences make it difficult to validly compare hospital performance across states, and so the analysis in this report focuses on within-state variation.<sup>69</sup>

There are weaknesses in the existing data – as described in *Strengthening safety statistics*.<sup>70</sup> But this same dataset is used for determining funding flows from the Commonwealth Government to the states, and for determining the apparent safety-related penalties applied by the Commonwealth Government. This suggests that the dataset is good enough to be used more extensively for measuring safety of care.

Over time, coding can be expected to improve – especially as the routine data improves with use and as recommendations in our previous report on improving data quality are implemented.

In the longer term, the quality of recording and coding will also improve with the advent of an electronic health record and associated decision-support systems which will automate many of the existing manual data recording and coding tasks.<sup>71</sup>

---

69. All figures that relate to the scope to improve complication rates are net of differences in the mean complication rate across states.

70. Duckett et al. (2017).

71. Stanfill et al. (2010); Scheurwegen et al. (2017); and Berndorfer and Henriksson (2017).

#### Box 4: Belinda's story

Australia's routine hospital datasets contain rich information about patients' hospital stays. For example, *here is Belinda's story*:

Belinda, a 55-year-old woman from Dromana in Victoria, was recommended to have a knee replacement.<sup>a</sup> She is a smoker, but presented with no comorbidities.

The procedure went according to plan, except she bled heavily after the operation. Belinda's rehab went well until she contracted a bad strain of pneumonia that appeared to be making its way around the hospital.

What should have been a five-day admission extended to eight, and Belinda had to spend an extra week in bed at home recovering from pneumonia. This hindered her rehabilitation efforts after the knee replacement and delayed her return to work.

---

a. Not her real name, age or suburb.

**And here is what we see in the data set:**

**Principal diagnosis:**

M171 Primary gonarthrosis

**Additional diagnosis present on admission:**

F171 Harmful use of tobacco

**Principal procedure:**

4951800 Total arthroplasty of the knee, unilateral

**Additional diagnoses which occurred during the course of the admission:**

J189 Pneumonia

D649 Anaemia

D686 Thrombophilia (with presence of the lupus anticoagulant)

J440 Chronic obstructive pulmonary disease (with acute exacerbation)

R090 Asphyxia

### 3.1 Fair comparisons of complication rates across hospitals are now possible

Of course, complications are easier to avoid in patients who are in better health. This means fair comparisons can only be made across hospitals if the complication rates are adjusted for patients' 'risk profiles'.<sup>72</sup>

Risk adjustment has been employed in medical research since the 19<sup>th</sup> century.<sup>73</sup> However, there is continuing debate about what constitutes adequate risk adjustment, and whether data held in the routine data sets is adequate for the task.<sup>74</sup>

Australia's routine hospital data details patients' diagnoses, comorbidities and procedures, as well as their age, sex and socio-economic status.<sup>75</sup> In this report we use this information to conduct very extensive risk-adjustment. In our analyses, we adjust for the impact of patient characteristics such as age, sex, the severity of diagnoses and whether the patient has any comorbidities. Our model allows the risk associated with each of these factors to vary by Diagnosis Related Group.<sup>76</sup> For example, we account for whether a patient has other health issues such as a respiratory condition, and the contribution of these patient factors

72. Or are presented in terms of like groups – a process known as risk stratification. We use risk adjustment in this report.

73. Iezzoni (1996); and Iezzoni (1997).

74. Alexander et al. (2017).

75. Even better risk adjustment could be possible if more clinical data, such as pathology test results, were incorporated into the routine data set, as Grattan Institute has previously shown (Duckett et al. (2015)).

76. Whether factors related to health inequalities should be controlled for is debatable: these factors contribute to patients' risk profiles in ways that we hope clinicians can overcome. To control for these factors is to excuse different outcomes. We take a moderate approach and control for socio-economic status but not other factors such as indigeneity and cultural and linguistic diversity. See the Methodological Supplement for a full discussion of our approach.

to our estimates of their risk of a complication are different for patients with different primary diagnoses.<sup>77</sup>

After these adjustments have been made, the scope for better outcomes can be estimated as the difference between each hospital's rate of complications and the lowest risk-adjusted rate observed at any hospital.

Figure 3.1 on the following page illustrates the importance of this risk-adjustment process. A hospital can have a higher rate of complications than its peers, but still have a lower rate than expected given the risk profile of its patients. For example, hospital 558 has a higher complication rate than hospital 302. But hospital 558's rate is lower than expected given the sort of patients it treats, whereas hospital 302's rate is higher than expected given the sort of patients it treats.

### 3.2 Routine data can illuminate paths to safer care

Comparisons of risk-adjusted outcomes across hospitals can provide useful information about hospitals' relative strengths and weaknesses, the nature of the harm that particular patients are most susceptible to, and which hospitals are likely to be safest for patients with particular characteristics.

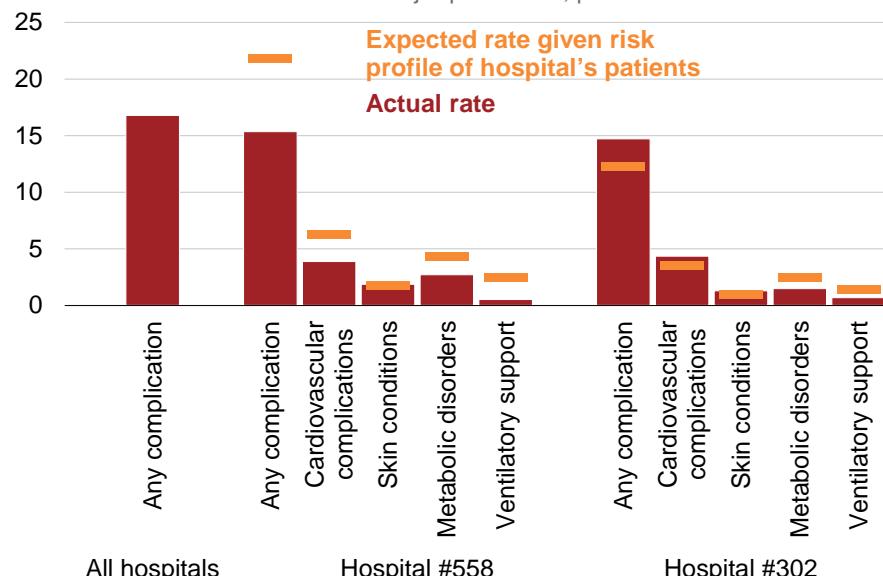
To illustrate these trends, we use three case studies: patients admitted for cardiology issues who do not have a procedure, patients admitted for knee replacement, and patients admitted for bariatric surgery. Only multiday admissions are included.

Figure 3.2 on the next page shows that hospital performance explains about 8-10 per cent of the variation in a patient's risk of a complication.

77. The risk adjustment underpinning the following results is substantially more detailed than the methodology proposed by IHPA (2017) for introducing risk-adjusted financial penalties for HACs and the level of risk-adjustment that is considered appropriate in many academic publications (for example, Zhang et al. (2013)). Please see this report's Methodological Supplement for further details.

**Figure 3.1: Headline complication rates don't reveal whether a hospital is performing better than expected**

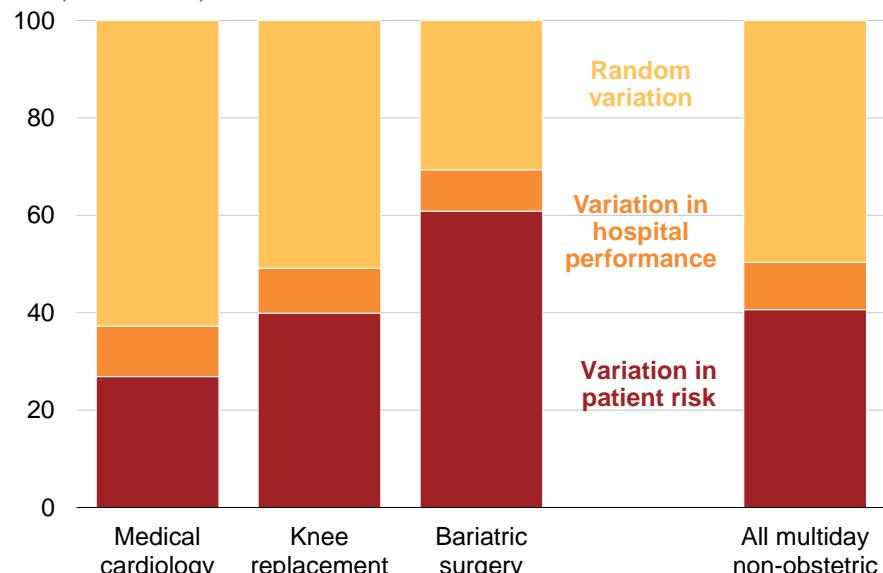
Share of admissions involving at least one complication (actual rate) relative to expected rate given the risk profile of hospital's patients, multiday cardiology admissions that do not involve a major procedure, per cent



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

**Figure 3.2: Hospital performance explains 8-10 per cent of the variation in patient outcomes**

Proportion of variation in complication rates explained by hospital performance and patient risk, per cent



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

In the following sections, we use these three case studies to show what can be learned from Australia's existing routine data.

Of course, the routine data is not the only source of information on hospital safety. Useful and complementary information about bariatric surgery and knee replacements are captured by clinical quality registries.<sup>78</sup> Our analyses of routine data illustrate what's currently possible to know about all hospital admissions, regardless of whether any registry data is available.

### 3.2.1 Data can illuminate where the biggest opportunities for safety improvement lie

Our analysis of cardiology admissions indicates that if all hospitals were to become as safe for cardiology patients as the safest 10 per cent, 12,000 fewer cardiology patients would experience complications during their admission. The size of this opportunity varies by institution.

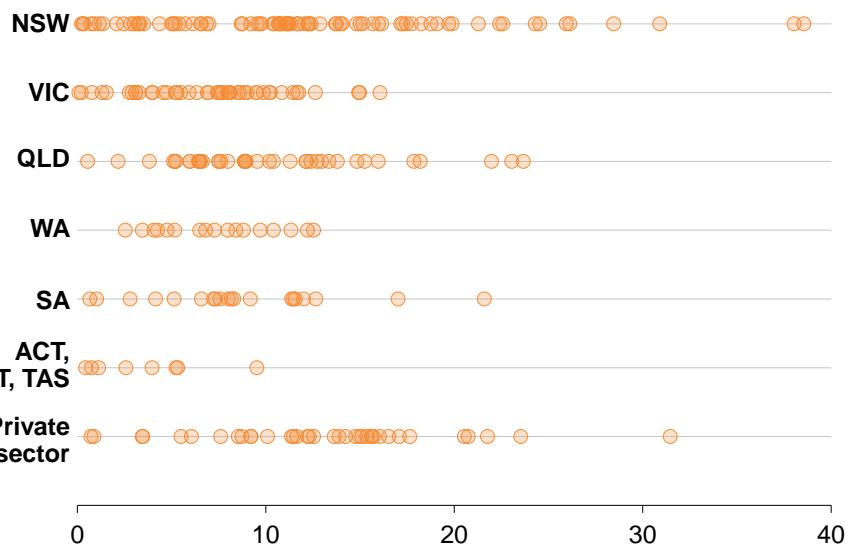
Figure 3.3 shows there are safer and less safe hospitals in every state, and in the private sector.<sup>79</sup> Each dot represents how much greater the risk of a complication is for patients at that particular hospital, relative to the risk they would face at the safest 10 per cent of hospitals in that state or group.

Some hospitals are substantially less safe for cardiology patients than others. Figure 3.3 shows that a patient with an average risk profile in New South Wales faces a 38 per cent higher risk of a complication if they attend the least safe hospital, relative to the risk they'd face at the safest New South Wales hospitals.

78. For more on how to improve the clinical quality registries, see *Strengthening safety statistics*.
79. It is not possible to draw comparisons across states from Figure 3.3. We have standardised for differences in the average complication rate by state, to take account of state-based differences in coding practices.

**Figure 3.3: Hospital safety varies significantly within states, and within sectors**

Excess risk of a complication relative to safest 10 per cent of hospitals in that state or group, multiday cardiology admissions that do not involve a major procedure, by hospital, per cent



Notes: Excludes small hospitals that were grouped together for the purposes of risk adjustment.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

However, this doesn't mean that efforts to improve the safety of cardiology care should be targeted exclusively at those outlier hospitals. Figure 3.4 shows that there is also a substantial difference between the excess risk at average-performing hospitals and the best decile. Most patients are treated in these average hospitals. If the general safety of hospital care could be improved to the best decile level, 12,000 more admissions could be complication-free every year. We should be helping every hospital learn from the best.

### 3.2.2 Data can show hospitals where improvement is needed

Hospitals would be better able to improve the safety of their care if they had two sets of information about the incidence of complications in their hospital.

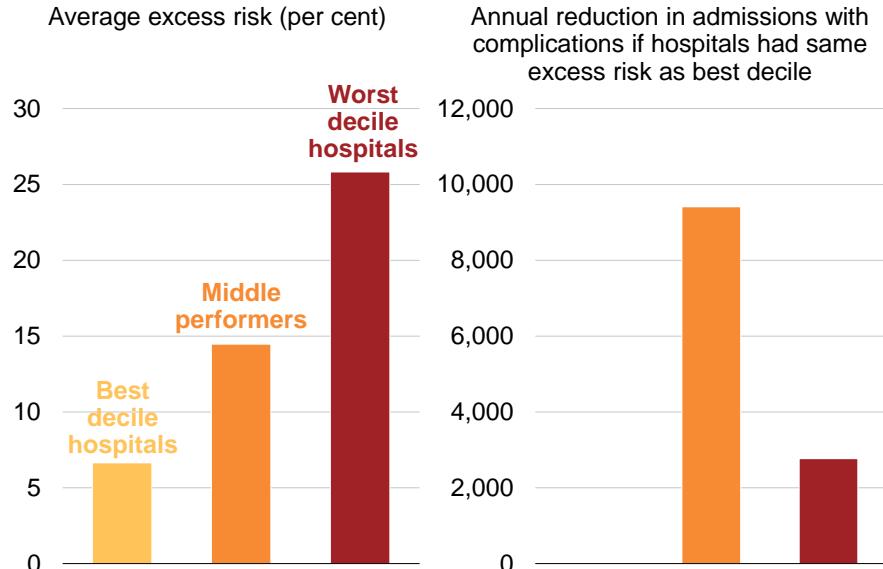
Firstly, hospitals need to know where they have the greatest scope to improve. Figure 3.5 on the following page shows that the safety of hospitals' care varies by specialty.

Hospital 803 is in the top 10 per cent of hospitals for medical cardiology and just outside the top 10 per cent for all multiday non-obstetric admissions, but its knee replacement patients face a risk of a complication more than 13 per cent higher than such patients at hospitals which excel in this speciality. Hospital 105, on the other hand, is a relatively poor performer: its patients generally face at least a 10 per cent higher chance of a complication than patients at the safest hospitals. However, its knee replacement patients are safer than those at hospital 803.

Hospitals do tend to be good in a number of areas or less good in a number of areas. But data can be used to identify "hot spots" so that the scarce resources devoted to improving care can then be allocated to them.<sup>80</sup> At the moment this hospital-specific information is retained

80. Accepting also that institution-wide strategies, such as improved handover, may influence the occurrence of a range of complications and in many types of patients.

**Figure 3.4: The greatest opportunity to make Australian hospitals safer is to move average hospitals closer to excellent hospitals**  
Excess risk of a complication relative to safest hospital for multiday cardiology admissions that do not involve a major procedure, by performance category



Notes: By definition, there is no reduction in admissions with complications at hospitals in the best decile of performance if they continue to perform at that level.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity dataset.

by health departments. If it were made available to the hospitals, they would be better equipped to improve the safety of their care.

The second set of information that could help hospitals improve the safety of their care is detailed data on the types of complications which are most common among patients of different types.<sup>15</sup>

Providing such detailed information about hospitals' relative performance across different specialties, and the complications that affect particular types of patients, would enable hospitals to target their safety improvement efforts where they have the greatest scope to improve.

### 3.2.3 Data can help patients make better decisions about their health care

Patients should know what outcomes to expect, and what complications they may face. Such information may affect their care choices and the focus of their pre-operative preparation and post-operative rehabilitation.

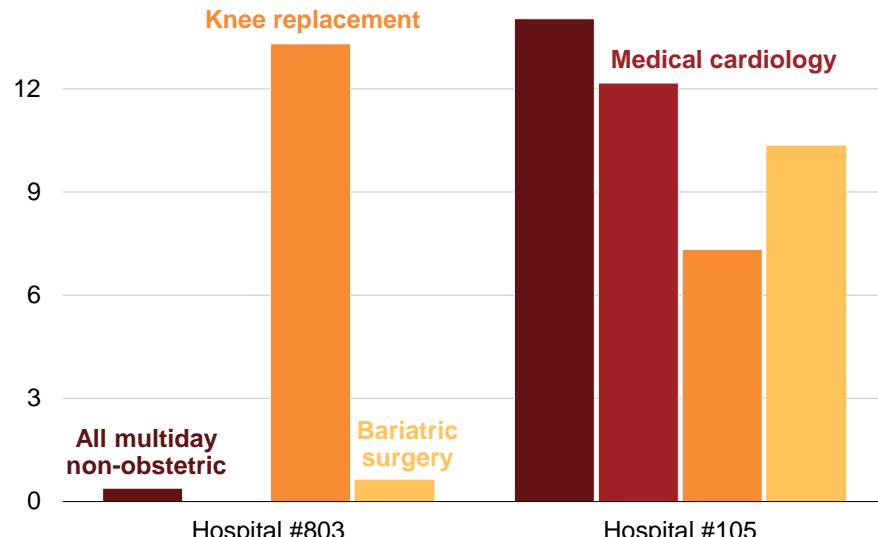
Existing data can sometimes provide answers to common questions such as: 'Which hospitals achieve the best outcomes for people like me?'

Just as different hospitals have different patterns of complications, so too some hospitals are better than others at treating patients in different age groups. Figure 3.6 on the next page shows estimates of excess risk for three hospitals that perform knee replacements.

It shows that Hospital A is a uniformly good hospital to go to for a knee replacement, and Hospital C is uniformly bad. At Hospital B, patients over 75 face twice the excess risk of a complication compared to similar patients at Hospital A, but their excess risk for other patients is similar to the performance in Hospital A.

**Figure 3.5: The safety of a hospital's care can vary by specialty**  
Excess risk of a complication relative to safest 10 per cent of hospitals, by admission type, multiday admissions only, per cent

15



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

More multidisciplinary care and specialised attention to rehabilitation may be reasons why Hospital A is able to reduce the occurrence of complications in the very old. The value of such data for identifying complications for the elderly is evident.

### 3.3 Current approaches to measuring safety are out of date

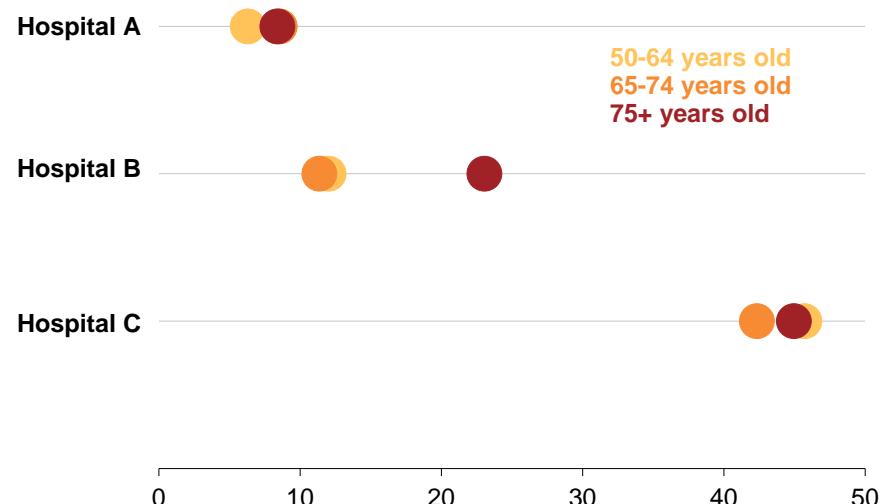
The insights that can be extracted from our existing data sources using approaches that have been available for ten years raises the question: Why are these analyses not performed regularly in Australia?

The answer may relate to implicit safety beliefs: that harm is uncommon and extreme, and that someone is to be blamed. This mind-set leads to clinicians being anxious about data being shared and analysed. Yet this chapter has demonstrated that complications are common, rather than rare and extreme events. This finding justifies a totally different approach to transparency.

This report particularly focuses on reporting comparative hospital performance, an area where Australia is a relative laggard given international recognition that ‘Robust comparison of performance with peers is fundamental to securing improvement’.<sup>81</sup>

The next chapter shows how increased transparency about the data could transform Australia’s approach to safety improvement.

**Figure 3.6: The safest choice can vary for patients of different ages**  
Excess risk of a complication relative to safest hospital for that age group, knee replacement admissions, per cent



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

---

81. OECD (2017).

## 4 We need to share data openly

Australia needs a comprehensive approach to variations in the outcomes of hospital care and the occurrence of complications. While once patient safety efforts were focused on serious incidents and deaths judged to be preventable, we now know that everyday instances of suboptimal care cause substantial harm.

This report's analysis of routine data shows how much improvement can be made. This final chapter details the changes required to close the gap we have identified between the safety of care that is given in Australian hospitals and that which could be given. It makes recommendations designed to ensure more hospitals deliver care that is as good as the top 10 per cent of hospitals.

The recommendations in this chapter complement those in Grattan Institute's previous report, *Strengthening safety statistics*. That report looked at improving the data sources we use to monitor safety, including the need to ensure the quality of coding remains strong so that routine data can be used confidently for both payment and safety monitoring.<sup>82</sup>

Hospitals' own internal efforts to improve safety of care requires attributable and specific information.<sup>83</sup> Grattan's next report will look at strengthening the external motivators of hospital activity, such as accreditation processes.

All the recommendations in this report are about better using the existing data which is collected from hospitals – publishing it, and providing it to those best placed to use it. The recommendations do not require collection of any additional data, and so implementation costs are

---

82. Duckett et al. (2017).

83. Levesque and Sutherland (2017).

small. The benefits of implementation, however, are large in terms of improving the safety of patient care.

The recommendations are also about redirecting existing effort – away from dramatic complications that happen rarely, toward more 'ordinary' complications that happen often. Because so many more people would benefit, this would be a better use of the time that clinicians already spend on addressing safety issues in their hospitals.

### 4.1 Governance mechanisms need to assure safety

Accreditation, governance and pricing systems play an important role in ensuring minimum safety standards are met in Australian hospitals. They affect the priorities of and constraints on hospital management teams. Safety performance information should be central to these systems.<sup>84</sup>

Hospital management is an extremely difficult task. Hospitals are big and complex organisations, expected to deliver care that is accessible and cost-effective as well as of high quality. This 'triple aim' means that hospitals' efforts to improve the quality and safety of care must be managed alongside efforts to improve access and to constrain costs.<sup>85</sup> Time and resources invested into quality improvement is not available to improve waiting times.

The trade-offs between these objectives are not pre-determined – they are the consequence of policy. Accreditation standards define

---

84. Ibid.

85. The term 'triple aim' was coined by the United States Institute for Healthcare Improvement and referred to improving the experience of care, improving the health of populations, and reducing per capita costs of health care; see Berwick et al. (2008).

minimum safety and quality processes, the hospital pricing regime defines the financial constraints within which this must be delivered, and health departments set access targets. Governance activities – through boards and health departments – routinely monitor progress on all these measures.

When these systems work well, they highlight meaningful safety requirements, effectively support management to deliver them, and minimise the tensions between quality and cost objectives. But when they don't, it can be difficult for management to improve safety. Tensions with financial and access objectives can obscure the business cases for safety initiatives.<sup>86</sup>

The role of a regulator should be to eliminate unwarranted variation in performance.<sup>87</sup> For the hospital system – both public and private – the regulator is the state government, through its health department. States should address the wide variation in rates of complications highlighted in this report. They should rise to the challenge of helping hospitals and clinicians to drive down rates of all complications, not just focus on the smaller subset of complications which have been labelled as 'preventable'.

The reforms needed to ensure Australia's hospital accreditation, governance and pricing systems all support reliably safer care are beyond the scope of this report, and will receive fuller attention in our next report. But governance changes need to be accompanied by quality improvement work – which requires providing data to clinicians and patients, setting targets, and supporting the subsequent improvement endeavours.

86. A future Grattan Institute report will deal more explicitly with financial incentives and business cases.

87. Hollnagel (2014).

## 4.2 Measure and report on all complications – not a subset

Clinicians and managers are hungry for more data.<sup>88</sup> Currently 92 public and private health service organisations across Australia and New Zealand are members of a private benchmarking group called The Health Roundtable.<sup>89</sup> They pay their fees and submit their data because participation provides comparative information for improvement that they cannot obtain otherwise.

Comparative data for improvement should systematically identify all opportunities to reduce harm. Narrow 'indicator' sets are inconsistent with new theoretical thinking on how to make systems safer. They also distort priorities because, as shown in Figure 3.5 on page 31, within hospitals some clinical specialties and care teams do better than others.

There is another issue with indicator sets, as illustrated by the ossification of the Australian sentinel event lists, discussed in Section 2.1. It was borrowed from a US list that has since developed considerably. Unfortunately, the Australian sentinel events list did not evolve and a revision is only now awaiting approval, and there is clearly the risk that HACs will meet the same fate.<sup>90</sup>

Measuring all complications ensures a dynamic approach to improving safety. As the complication profile of patients and hospitals changes, new targets for improvement can be set.

### 4.2.1 Reporting to hospitals

The data currently kept within state health departments should be made available to all relevant parties: hospitals, clinicians, patients and

88. Jorm (2017); and Duckett et al. (2016).

89. See: <https://www.healthroundtable.org/>.

90. Although this risk is mitigated by revision of the HACs list being on the work program of the Australian Commission on Safety and Quality in Health Care.

the wider public. This information should be provided to hospitals in a form that can be analysed and then used to help clinicians improve the safety of care. It should also be provided to the public, so people can hold hospital managers accountable.

State health departments have the capacity to risk-adjust and report comparative data. They should give hospitals regular, specific updates of how their performance compares to their peers'. Figure 4.1 on the next page provides an example of how this could be done. It shows the quintile of a given hospital's risk-adjusted performance overall, by each of the Major CHADx and CHAPx classes, and by each of the Minor CHADx+ classes.

Providing Australian hospitals with 'heat maps' like this would not be difficult. The Methodological Supplement to this report outlines the analytic approaches underpinning Figure 4.1, and indicates how large a hospital needs to be for the metric to be meaningful.

The newest edition of the Australian Commission on Safety and Quality in Health Care's National Standards includes a specific requirement under Clinical Governance for hospitals to analyse how clinical practice varies, and to communicate this to their workforce. This is a step forward, but the exercise will be futile unless states provide hospitals with comparative information.

The National Standard on Clinical Governance requires hospitals to focus on variation.

Under Standard 1.28, Variation in clinical practice and health outcomes, the health service organisation has systems to:

- a. Monitor variation in practice against expected health outcomes
- b. Provide feedback to clinicians on variation in practice and health outcomes

- c. Review performance against external measures
- d. Support clinicians to take part in clinical review of their practice
- e. Use information on unwarranted clinical variation to inform improvements in safety and quality systems
- f. Record the risks identified from unwarranted clinical variation in the risk management system

#### 4.2.2 Reporting to clinicians

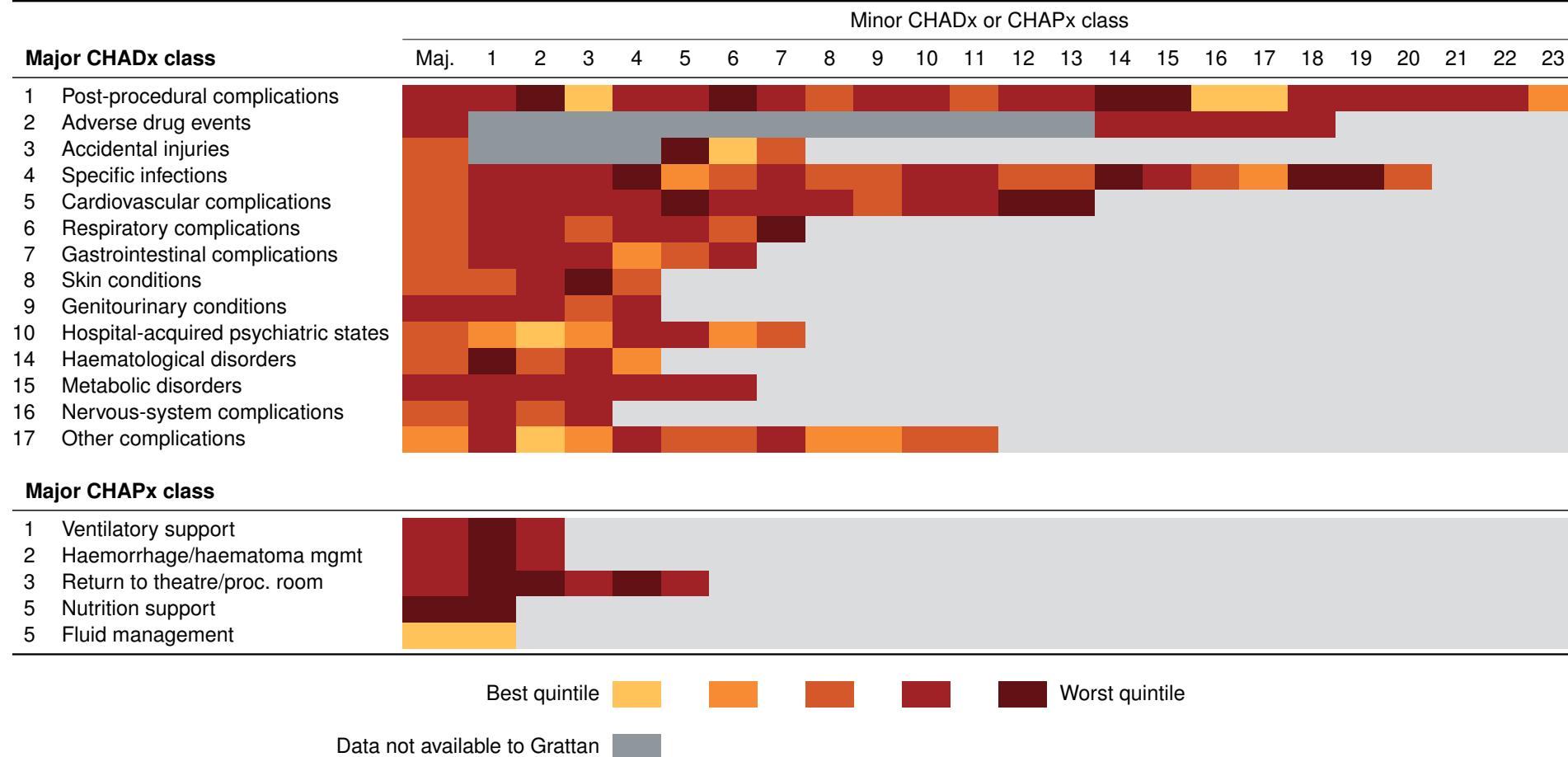
Australia has generally relied heavily on grassroots innovation for safety improvement. Individual clinicians and hospitals have developed their own internal monitoring strategies, and a plethora of small initiatives are underway in most hospitals.<sup>91</sup>

Clinical and managerial staff strive for improvement, but are not supported adequately by data that helps them achieve it.<sup>92</sup> Unfortunately, time-limited projects unsupported by data do not create sustainable change. Few quality improvement initiatives (including internationally) have been rigorously evaluated.<sup>93</sup> Without continuous provision of robust data, there is 'action without knowledge' – that is, activities that may add to staff workload without improving patient care.<sup>94</sup>

Our analysis in *Strengthening safety statistics* of all available safety and quality data sources revealed limitations can be fixed.<sup>95</sup> Inadequate access to data is making it harder for clinicians to improve safety than it should be.

- 
- 91. Indeed, Victorian clinicians and managers recently complained about an exhausting excess: Jorm (2017).
  - 92. Leggat and Balding (2017).
  - 93. Nicolay et al. (2011); E. Jones et al. (2016); Dixon-Woods and Martin (2016); Dixon-Woods and Pronovost (2016); and Walsh et al. (2014).
  - 94. Pronovost et al. (2017); Kreindler (2016); and Höög et al. (2016).
  - 95. Duckett et al. (2017).

Figure 4.1: Hospitals should regularly receive heat maps of their relative performance



Notes: Minor CHADx+ classes are particular conditions that are classified within the major CHADx or CHAPx category.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

More data should be placed in the hands of clinical teams. Considerable information about their practice is collected; they should be given timely access to it. With the advent of electronic health records in hospitals, this could be near real time (see Box 5).

Data at the level of detail in the heat map in Figure 4.1 on the previous page should be provided to clinical teams to help target efforts to specific groups of clinicians and to specific conditions. Reporting should also describe excess complications in each specialty, and it should provide detail on the outcomes (including length of stay, and readmissions) of high-volume conditions and procedures.

In the UK's 'Getting it right the first time' program, for example, specialist clinicians in hospitals are presented with all the available data about their patients, including activity and cost as well as outcomes. This enhances peer and management review of the performance of individual specialists.<sup>96</sup> In Australia, the medical colleges would be ideally placed to lead such reviews.

A focus on all complications would cast the net for identifying improvements much wider – specialty groups in larger hospitals would be able to focus on their own priorities, unconstrained by whether they are on the HAC list or not.

New South Wales already provides clinicians with access to a dataset which enables them to compare their performance with other similar clinical teams.<sup>97</sup> All states should follow suit. But each state does not need to reinvent this wheel. The National Benchmarking Portal established by the Independent Hospital Pricing Authority<sup>98</sup> currently focuses on cost-benchmarking. It should be expanded to enable easy comparison of complication rates.

---

96. Timmins (2017).

97. See: <http://www.health.nsw.gov.au/wohp/Documents/mc2-abm-adamato.pdf>.

98. See: <https://www.ihsa.gov.au/what-we-do/data-collection/national-benchmarking-portal>.

#### Box 5: Developing real-time safety data

This report uses a large national dataset to demonstrate the opportunity for improvement. However, as we argued in *Strengthening safety statistics*, to be useful for improvement, comparative information needs to be timely.<sup>a</sup> Information about safety reaches its 'use-by' date surprisingly quickly. Real-time data is the new 'gold standard'. Information about recent complications is interesting, but information about unfolding events is riveting, because clinicians can act in response.<sup>b</sup> An electronic medical record that provides decision support, and with data collected during care rather than coded *after* discharge, is a basic 21<sup>st</sup> century tool. Yet, implementation of these systems in Australia is still embryonic.<sup>c</sup>

- a. Duckett et al. (2017).
- b. During the consultation phase on this report one commentator highlighted this, saying: 'We have been giving really extensive retrospective data to our hospitals for five years now, outlining many opportunities for improvement. Yet, as soon as it was real time, an impetus for change was generated that all the retrospective data given to date couldn't generate. I don't quite understand why, but it is very real.' A second commentator echoed this: 'My early experience with the digital hospital is that we are able to build, in real time, lead indicators which trigger clinician action and intervention in the face of variation, flagging signs of deterioration, poor glycaemic control, etc.'
- c. Sullivan et al. (2016).

#### 4.2.3 Reporting to the public

Expectations about public reporting are changing both by governments – who are actively looking at improving reporting<sup>99</sup> – and from patients. States should report hospital outcomes data relevant to patients’ care decisions in a way that is readily accessible to patients and GPs. State reporting about public and private hospitals should be quite detailed – reporting all complications classified into Minor CHADx+ classes. An interactive website should be created to enable more specific reporting relevant to patients, such as by age and sex.<sup>100</sup>

New South Wales has made a good start, and its example should be followed by other states and extended to include private hospitals. Private patients pay extra for (among other things) the opportunity to choose their treating doctor and hospital. Despite some private hospital chains starting to publish performance information themselves (see Box 6), at present patient choice is poorly informed, because existing data that robustly compares hospitals and doctors is not made available.

We have provided an example of an interactive website and app at: <https://grattan.shinyapps.io/dummyLiftingLidApp/>. However, we are not permitted to release hospital-specific information and so the website simply provides Australia-wide information. States should develop similar websites which would allow patients to compare complication rates between hospitals, taking account of the patient’s age and sex, and ideally any other conditions they have – such as diabetes – which influence complication rates. This information – as we show in our prototype app – is available now and should be made available to prospective patients.

99. A recent decision of Health Ministers supported work on national consistency in reporting, see: <https://www.coaghealthcouncil.gov.au/Portals/0/COAG%20Health%20Council%20Commuque%20-%204%20August%202017.pdf>.

100. Greenhalgh et al. (2017).

#### Box 6: Reporting by private hospitals

A number of private hospital groups, such as Healthscope<sup>a</sup> and Ramsay,<sup>b</sup> publicly share their safety and quality measures. Intending Healthscope and Ramsay patients are able to see a wide range of results for the group, together with details of improvement initiatives underway. A comparison with ‘public hospitals’ is also provided, but this information is not risk adjusted – probably because the private chains do not have the information to do so – and so patients are not able to make comparisons between particular private hospitals or between particular public and private hospitals.<sup>c</sup>

- a. See: <http://www.healthscopehospitals.com.au/quality/my-healthscope/>.
- b. See: <http://www.ramsayhealth.com/Sustainability/Patient-Safety-and-Quality/Latest-Results/>.
- c. An alternative approach suggested in the consultation phase on this report is to provide unadjusted data for particular groups (e.g. stratified by age), so that patients have a better understanding of the risks of admissions.

Private health insurers have an important role here too. The larger insurers (Medibank and BUPA), and through a cooperative venture (the Australian Health Services Alliance) many smaller insurers, have sufficient information to produce robust measures of hospital performance. Indeed, Medibank already provides scorecard information to hospitals which includes data on rates of readmissions, complications, and admissions to intensive care units.<sup>101</sup>

Private health insurers should share that same information with their members, to help members make fully informed choices about where they have their elective procedures.

Australians should have access to comprehensive, risk-adjusted, institution-level data for every institution.

Australia's hospital system lags behind international peers in making this transition to greater transparency. New York State, for instance, commenced public reporting of cardiac surgeons' clinical outcomes 25 years ago. The US and the UK now provide detailed, publicly reported information on the safety performance of hospitals and some clinicians.<sup>102</sup>

There are two compelling reasons for Australia to follow suit. Most fundamentally, it's important in principle that information about variation in the quality of hospitals' care is shared openly with patients and taxpayers. Secondly, transparent reporting of performance data is an important tool for improving safety.<sup>103</sup>

---

101. Rankin et al. (2017).

102. For example, Leapfrog Group (2017), NHS (2016), Behrendt and Groene (2016) and Findlay (2016a). The UK's hospital-specific atlas of variation in outcomes demonstrates the specificity with which hospital performance data can be reported back to hospitals (and also publicly shared), see NHS (2017).

103. There are some risks associated with transparency; the nature, magnitude and mitigation strategies for these risks is discussed in Appendix A.

Review of comparative performance of professionals and hospitals has long been seen as 'secret squirrel' business, with review of safety incidents to be conducted by consenting adults in private and shielded from scrutiny by legal privilege.<sup>104</sup> In this approach, a doctor's or a hospital's right to privacy or commercial interests has been given greater weight than a patient's or the public's right to know about relative performance.

However, the public's expectations about what they ought to know are changing across all areas. Comparison websites are common in almost every area of consumer choice. Health care is not immune from this transition. The culture change from the old-style, secrecy approach will be immense. It will also require work to ensure that data is presented as clearly as possible so that people can evaluate health risks and benefits.<sup>105</sup> The change management task for states, hospitals and insurers will also be significant, but it will be worthwhile for patients and for the system as a whole.

### The public has a right to know

The Australian Charter of Healthcare Rights includes the right to receive safe and high-quality care, the right to be informed about services, treatment options and costs in a clear and open way, and the right to participate in decisions about care.<sup>106</sup> It has been argued that practitioners may have a legal obligation – as part of their consent

---

104. Through national or state schemes; see, for example: <http://www.health.gov.au/internet/main/publishing.nsf/Content/qps-info>.

105. The importance of this and methods of supporting both patients and clinicians are discussed in the Grattan Institute's previous report: Duckett et al. (2017).

106. See: <https://www.safetyandquality.gov.au/national-priorities/charter-of-healthcare-rights/>. One of the four key themes of the recently released Australian Digital Health Strategy is 'Support me in making the right healthcare choices, and provide me with options', see: <https://www.digitalhealth.gov.au/australias-national-digital-health-strategy>. However, the strategic priorities in the strategy will do little to achieve this goal.

procedures – to disclose information they have about the hospital’s rate of complications.<sup>107</sup> Our changing understanding of the nature of most harm caused to patients in hospitals has clear implications for public reporting obligations.

Information about the incidence of extreme harm in specific hospitals is of little relevance to elective patients’ choice of hospital or taxpayers’ appraisal of their local hospital – the rarity of these incidents means they are unlikely to reflect hospitals’ normal performance. However, information on the general quality of hospitals’ care derived from their relative rates of complications is directly relevant to patients’ choices and taxpayers’ satisfaction.

Historically, consumers have made limited use of publicly reported comparative performance measures for hospitals.<sup>108</sup> However, this does not mean that they won’t use such information in future, or that they are not entitled to it.<sup>109</sup> Citizens also require information to fulfil their democratic role of holding government to account.<sup>110</sup>

The patient’s GP has an important role here too. If GPs have better public information – especially if it is also delivered to their desktop

---

107. McWhirter (2017).

108. Greenhalgh et al. (2017). N Marshall et al. (2000) found that patients were distrustful of public benchmarking data, and early studies found that leading hospitals don’t gain market share after public benchmarking. Vladeck et al. (1988), Weller et al. (2010), Chassin (2002), Hibbard et al. (2005) and Jha et al. (2010). But Mennemeyer et al. (1997) did observe a market response to safety scandals.

109. Recent evidence from the US suggests patient engagement with publicly reported performance data is increasing (Carman et al. (2016) and Findlay (2016b)). Other studies are detecting market responses to public benchmarking Merle et al. (2009), Pope (2009), Chen et al. (2012) and Blake and Clarke (2017). Mukamel et al. (2004) found that public benchmarking diminishes the influence of experience and price on a patient’s choice of surgeon. This indicates that patients value information on surgeons’ past performance, where such information is available and user-friendly.

110. Levay (2016).

– their advice to patients about referrals could be more informed and based on the total experience of a specialist or hospital rather than just what the individual GP might have observed from previous referrals.

#### Public reporting of hospital safety creates additional incentives for safety improvement

When data on hospital performance is fed back to hospitals in a timely fashion, greater safety improvements are achieved.<sup>111</sup> Data makes it easier to identify safety improvement opportunities, and causes management teams and boards to devote more attention to safety.<sup>112</sup>

Some studies have found that public reporting improves the business case for safety improvements by influencing patients’ choice of hospital.<sup>113</sup>

But the biggest benefit from public reporting appears to be hospitals acting to protect their reputation.<sup>114</sup> Public reporting appears to prompt hospitals to redouble their safety efforts and rethink their hiring strategies.<sup>115</sup> And sharing data publicly also creates an incentive for state governments to act. Some hospitals may need more central support.

#### 4.2.4 States need to pay attention to issues that are beyond the scope of individual hospitals

Many safety challenges are beyond the capacity of a single institution. In the absence of support, making individual hospitals solely responsible for improving safety could be as pointless as blaming individuals for system defects.<sup>116</sup> The size, staffing and scope of operations of a

---

111. Hibbard et al. (2005).

112. Tsai et al. (2015).

113. Chandra et al. (2016).

114. Frølich et al. (2007); and Totten et al. (2012).

115. Mukamel et al. (2002); Mukamel et al. (2014); Chassin (2002); and Fung et al. (2008).

116. Dixon-Woods and Pronovost (2016).

hospital may all affect its ability to reduce complication rates among patients. None of these issues are fully within the control of the hospital; all require the involvement of the state.<sup>117</sup>

State governments also need to build quality improvement capacity across the health system;<sup>118</sup> foster the sharing of knowledge about what is working well; and address broader, whole-system safety issues. Section 4.2.5 discusses four concepts that are central to supporting improvement efforts. States also need to reinforce the priority of quality improvement. It is too easy for states to slip into sending an implicit signal to hospital managers that all that matters are budgets and waiting times.

There are a range of data sources including clinical quality registry data, death audit data, incident reporting and investigation data, patient-reported experience measures, and patient-reported outcome measures that can assist in providing guidance for safety and quality improvement.<sup>119</sup> All should be employed when appropriate in the quest to reduce all complications.

The Commonwealth Government should support smaller states and the territories, by providing them with the data to allow them to develop their own approach to national benchmarking (for example, by identifying a relevant cohort of hospitals against which to benchmark), and by publishing national benchmarking data (for example, on what is the best safety performance in a specialty nationally, rather than locally). Smaller states should also engage with the routine work of the quality improvement bodies of the larger states, for example, the NSW Agency

---

117. For instance a shortage of hospital physiotherapists may be a state-wide or regional problem; solutions may include changes in the relevant employment award.

118. Mery et al. (2017).

119. Duckett et al. (2017). For an example of an attempt to create a balanced framework that also includes culture and compliance measures, see Wakefield and Jorm (2009).

for Clinical Innovation, the Queensland Clinical Excellence Division, and Safer Care Victoria.

#### 4.2.5 Supporting quality improvement

##### Increase clinician engagement

A pre-condition for serious safety improvement is close working relationships between policy makers, regulators and clinicians. Australia needs to get better at this. Managers need to give more attention to clinician engagement. They need to ensure clinicians have sufficient time, resources and skills to pursue safety improvements. The message should be that safety improvement is not ‘managers’ business’ but everyone’s business.<sup>120</sup>

##### Foster collaboration

Cooperative structures such as ‘clinical networks’ or ‘clinical collaboratives’ create peer accountability and can drive improvement, although reviews have evaluated their effectiveness as ‘promising’ rather than proven.<sup>121</sup> Nevertheless, some form of systematic clinical engagement is needed to facilitate sharing of good practice, so that the experience, protocols and procedures used by ‘positive deviants’ becomes part of the routine in all hospitals. Gone are the days when it was thought that ‘each organisation should solve their patient safety problems alone’. The more organisations share experiences, data and evidence, the better.

---

120. Duckett et al. (2016); Jorm (2017); and Alderwick et al. (2017).

121. Scott and Phelps (2009); and Nelson et al. (2007).

### Design systems that combine accountability and learning

Any thinking that accountability and learning are incompatible needs to end. There should be no dichotomy between information for performance ('for managers') and information for improvement ('for clinicians'). Both require reliable data. (Scotland provides one good model for combining stakeholder engagement, learning and accountability for best results.<sup>122</sup>)

### Fund programs to reduce complications

Getting to grips with the full gamut of reducible complications will require new research. A developmental program should include evidence synthesis, expert review, and research commissioning. It should study how 'positive deviants' deliver safer care, and test the success of applying those lessons in other institutions. Given the savings which could be generated by reducing complication rates, the Medical Research Future Fund could be directed to develop a program of this kind. It could be coordinated by the Australian Commission on Safety and Quality in Health Care.

---

122. Damschroder et al. (2014); and Schang and Morton (2017).

## 5 Conclusion

It is unknown how much progress Australia has made on improving patient safety outcomes over the past 20 years. Internationally, progress is considered to have been disappointing.<sup>123</sup> For Australia, neither longitudinal adverse event studies nor information about the prevalence of the full range of patient complications has been freely available.

The current incomplete reporting practice, together with a media culture that amplifies sensational events, misleads the public to believe that harm in hospitals is both infrequent and catastrophic. This obstructs their ability to make informed decisions about their own care.

The lack of sustained attention to reducing all complications is a tragedy, given that some hospitals do so much better than the average. Australia needs new safety thinking that doesn't normalise harm to hospital patients. It is time to adopt an epidemiological focus – examining broad patterns – and to learn from hospitals and clinicians who achieve low rates of harm. The policy question is not whether we move to addressing all complications, but when. The sooner we set more ambitious targets the better.

Priority should be given to seizing opportunities at the local specialty level. Hospitals and clinicians should be given all the information required to facilitate change.

This report has identified substantial scope to reduce patient harm in Australia, and made recommendations to make our hospitals safer. Policy makers should now act.

---

123. Vincent and Amalberti (2016). Some studies, for instance in Holland, have demonstrated improvement but there is controversy associated with the validity of longitudinal adverse events studies and their focus on death and error and a call for better, more granular safety metrics. Shojania and Mheen (2015), Baines et al. (2015) and Shojania and Dixon-Woods (2016).

### Recommendations:

- All states and territories establish goals for reducing the overall rate of complications in public and private hospitals.
- All states and territories give hospitals and clinicians the ability to interrogate the state hospitals data (without individual patients being able to be identified), so they can see how their performance measures up against the best-performing hospitals and clinicians. All hospitals develop strategies to identify opportunities to improve.
- All states and territories publish reports on excess complications, by specialty and institution (including private hospitals).
- Major private health insurers provide their members with comparative information on complication rates.

## Appendix A: The risks of transparency can be mitigated

Australia has developed a culture of anxiety around the public reporting of hospital activities. It has manifested as a reluctance to embrace transparency. This is both understandable and unfortunate. Problems caused by poor use of performance measures are eagerly seized upon, but rarely weighed against the benefits of careful use of the data.<sup>124</sup> Measurement and reporting are powerful drivers of improvement, but they need to be used sensitively.

Transparency comes with risks: security around sensitive data needs to be watertight, the gaming of metrics needs to be minimised and managed, and the data needs to be of high quality.

Security risks can be mitigated by: ensuring the data is not able to be used to identify individual patients; providing authorised access to trustworthy parties only; and using secure distribution channels.

Gaming risks can be mitigated by imposing penalties for erroneous coding, as identified by auditors and metric designers.<sup>125</sup> Evidence suggests the incidence of doctors ‘cherry-picking’ healthier patients, and the seriousness of potential adverse consequences, appear to be exaggerated.<sup>126</sup>

Data quality risks are also real. But when data is published, used and audited, the quality of the data improves. For instance, the accuracy of hospitals’ records of the types of care they were providing improved

after the introduction of activity-based hospital funding. Similarly, the first stage in any investigation of safety data is to check that the patient record (and the coded data) accurately reflects what happened to the patients.<sup>127</sup>

Whether data is deemed to be of adequate quality depends on the purpose for which it is being used.<sup>128</sup> For instance, data used for clinical quality improvement, such as surgical death audit data, needs to be far more detailed than data used for institutional performance monitoring.

---

124. Mannion and Braithwaite (2012).

125. For example, metrics based on complications rates that are reducible are less susceptible to gaming because it’s unclear at the time of coding what the advantageous rate of occurrence of the complication might be.

126. Tweddell et al. (2017) show that increased risk aversion in patient selection is often beneficial to the patient, with surgeons now less willing to operate on high-risk patients when they are insufficiently experienced or specialised, or when death is a highly likely outcome (emergency departments).

---

127. Duckett et al. (2007).

128. Damschroder et al. (2014).

## Methodological supplement

The Grattan Institute report *All complications should count: Using our data to make hospitals safer* sets out to measure the extent of safety problems in Australian hospitals and to develop appropriate policy responses.

Our approach was predicated on the assumption that complications of care can be measured using the routine data reported on every patient discharged from hospital. There is a voluminous literature supporting this approach and we have reviewed different sources of measurement in our previous report *Strengthening safety statistics*.

We were particularly interested in identifying persistent variations in the prevalence of complications of care. Variation in complication rates among hospitals indicates that some hospitals are able to achieve better outcomes than others and therefore suggests that these rates are *reducible*. In order to make such comparisons across hospitals, careful adjustment for differences in the patients treated by each hospital was required.

In this document, we present the data and methodology underpinning this analysis. We start by presenting our data sources and summary statistics. We then outline the conceptual framework unpinning our analysis, and present our findings regarding which patients are most at risk of experiencing complications of care.

## Appendix B: Data

Our analysis is completed on three years of data from the National Hospital Morbidity Dataset (NHMD) provided to Grattan Institute by the states and territories through the Australian Institute of Health and Welfare (AIHW). The years provided were 2012-13, 2013-14 and 2014-15.

This dataset contains anonymised information about the demographics and hospitalisations of all patients who attended a public or private hospital in Australia over this period. Separations for which care-type was reported as *Newborn* (without qualified days), and records for *Hospital boarders* and *Posthumous organ procurement* were excluded. We were not provided with hospitals' names, or permitted to release analysis which would identify specific hospitals.

In this chapter, we summarise how we cleaned this dataset, created variables relating to the incidence of complications, and derived all other imputed variables.

### B.1 Data cleaning and sample selection

The raw data Grattan Institute received from AIHW contained 29 million observations, which each included data on an initial admission and up to one linked readmission.<sup>129</sup> After separating the linked admissions into separate observations, removing observations where there was insufficient information for a Diagnosis Related Group (DRG) to be

129. Readmissions were linked to initial admissions where they related to patients readmitted to the same public hospital within 90 days, excluding same-day readmissions for dialysis and chemotherapy.

**Table B.1: Key subsamples of the 2012-15 NHMD**

Original sample	29,216,399
Expanded original sample	58,432,798
Observations to be excluded from all analysis	
Empty readmission fields	26,115,179
Duplicates in all variables	4,795,317
Flawed records: MDC=8 or 9, or no principal diagnosis	44,944
Outlier rates of COF 1	643,979
Non-acute admissions	1,657,421
Sample used in general analysis	
	25,175,958

*Notes: Although some observations could be excluded on the basis of more than one criterion, the sample sizes listed here are derived by applying the criteria sequentially. The observations missing DRGs were empty readmission fields that were initially appended to diagnoses with a readmission within 90 days to the same public hospital.*

assigned and removing duplicates, flawed records and outliers, 25 million observations remained.<sup>130</sup>

Figure B.1 on the following page provides a visual summary of how the original sample was expanded to allocate each set of readmission information its own observation, and then reduced. Almost all the observations discarded for lacking a DRG were observations created in the process of separating the linked readmission fields into their own observations, as this produced empty observations for the 90 per cent of initial observations which did not have linked readmissions.

130. We defined outliers as hospitals which were more than three standard deviations from the mean in their rate of allocating Condition Onset Flag (COF) 1, the variable that indicates whether a diagnosis was acquired in hospital.

A further 150,000 observations were excluded from the data when completing regression analysis because some observations were missing data on independent variables of interest.

Full details regarding the number of observations affected by our data cleaning decisions are provided in Table B.1. We also conducted some of our analysis on subsets of the data. Definitions and sample sizes of these subsamples are given in Table B.2.

## B.2 Identifying and defining complications

A significant component of the analysis contained in *All complications should count: Using our data to make hospitals safer* is concerned with the rate of complications in Australian hospitals, and the circumstances surrounding them. In this section, we review the various definitions of complications used in *All complications should count: Using our data to make hospitals safer*, and how their incidence is derived using the NHMD.

### B.2.1 The Classification of Hospital Acquired Diagnoses

The Classification of Hospital Acquired Diagnoses (CHADx) aims to be a comprehensive classification of all complications that can be incurred by patients during their hospitalisation. Developed and maintained in Australia, this classification is continually being extended and refined.

The CHADx+ algorithm has two functions.<sup>131</sup> Firstly, it defines a set of events that constitute hospital-acquired complications. The CHADx+ algorithm “cleans” the raw incidence of hospital-acquired diagnoses by:

- unflagging diagnoses as being hospital acquired if it is implausible that this is the case;

**Figure B.1: Most dropped observations related to empty readmission fields**

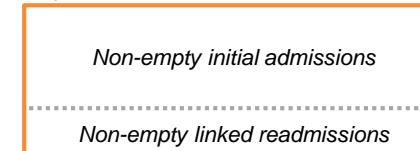
Original data format



Expanded data format



Expanded data format, after removing empty observations



Source: Grattan analysis.

---

131. Michel et al. (2009).

- counting diagnoses as being hospital acquired if it is implausible that they were present on admission; and
- drawing on combinations of diagnoses, external cause codes, external place codes and external activity codes to identify a single complication.

Secondly, the CHADx+ algorithm classifies complications into minor and major CHADx+ classes. This is useful for investigating the composition of aggregate complication rates, and makes clear where multiple complications may be double-counted, or sequelae of earlier complications.

Grattan Institute is grateful to the Victorian Department of Health and Human Services for providing us with access to the most recent edition of the Classification of Hospital Acquired Diagnoses, CHADx+ version 1.4 (CHADx+). This version of the classification differs from previous editions of the classification in a number of ways:

- Most significantly, all “Plus” versions of the classification have been extended to include procedures that indicate that a complication must have occurred, like blood transfusions, as well as diagnoses that constitute complications.
- There have also been a number of changes to how complications are categorised. Most significantly, infections previously categorised under the affected body system are now grouped together in Major CHADx Class 4: Specific Infections.
- Some changes have also been made to complications’ definitions. Most significantly, a COF of 1 (acquired in hospital) is no longer required to identify complications in Major CHADx Class 12: Labour, delivery and postpartum complications. In other instances, definitions of complications have expanded, for example CHADx 15.02: ‘Electrolyte disorders / fluid management’ has been expanded from a definition that excluded dehydration.

**Table B.2: Details of subsamples used for analysis**

**Collectively exhaustive subsamples:**

Obstetric admissions	1,608,104
<i>Admissions with Major Diagnostic Category (MDC) of “O”</i>	
Non-obstetric sameday admissions	16,388,968
<i>Sameday admissions MDC not equal to “O”</i>	
Non-obstetric multiday admissions	7,028,984
<i>Multiday admissions MDC not equal to “O”</i>	

**Case studies:**

Medical multiday cardiology admissions	561,101
<i>Multiday admissions with adjacent DRGs of F40-F59. These include admissions for circulatory disorders, unstable angina, syncope and collapse, chest pain, arrhythmia and other circulatory disorders.</i>	
Multiday knee replacements	139,697
<i>Multiday admissions with adjacent DRGs of I04 and I32.</i>	
Multiday bariatric surgery	36,641
<i>Multiday admissions with procedure codes relating to bariatric surgery. These include gastric reduction, gastric banding, gastroplasty, gastrectomy, gastric bypasses, biliopancreatic diversion, duodenal-jejunal bypass, ileal interposition, other procedures for obesity and their revisions.</i>	

<sup>†</sup>Specific bariatric procedure codes are: 3051100-6, 3051108-10, 3051200-3, 3051401, 3144100-1, 9094000, 9094100, 9094300, 9094301, 9094302, 9095000-1, 9094200-2.

*Note: These subsample sizes exclude observations that were missing data on independent variables to be used in regression analysis.*

Table B.3 on the next page compares the rate of complications reported in AIHW's 2014-15 Admitted Patient Care report with the numbers observed across the 2014-15 admissions in our data.<sup>132</sup> These figures are broadly aligned. However, the overall rate of complications we observe is higher than that reported by AIHW, and differs notably in a few cases.

There are three reasons for these discrepancies. Firstly, AIHW appears to have used CHADx v.5, whereas we have used CHADx+ v.1.4, so our comparison is affected by the differences listed above. Another key difference between these classifications is that our overall complication rate (CHADx+) includes complications detected through procedures (CHAPx) in addition to the complications detected from diagnoses (CHADx) included in CHADx v.5.

Secondly, our data is a subset of AIHW's: we have applied more restrictive cleaning requirements. The exclusion of newborn admissions implied by our focus on acute admissions substantially reduces the observed rate of Major CHADx Class 13: Perinatal complications.<sup>133</sup> Our exclusion of observations without sufficient information to assign a DRG and other flawed records is expected to contribute to our higher overall rate of complications.

Finally, we note that we do not observe COFs for external cause, place or activity codes, or which particular diagnoses they relate to, in our dataset. Moreover, none of the International Classification of Disease – Version 8 codes which can serve as external cause, place or activity codes appear in our diagnosis fields, even though these codes can also serve as diagnoses. This precludes us from detecting adverse

---

132. AIHW (2016).

133. The figures published in AIHW's Admitted Patient Care statistics relate to all episodes, and only exclude admissions where the COF is missing. The exclusions listed in Table B.1, notably the focus on acute admissions, result in a sample that is around 500,000 observations smaller.

drug events (CHADx 2.01-2.13), and all falls (CHADx 3.01-3.04). As a consequence, our estimates of the average prevalence of complications are expected to be downwardly biased by about 0.1 percentage points in absolute terms.<sup>134</sup>

## B.2.2 The “Priority Complications”, known as Hospital Acquired Complications

Over 2012-2016, the Australian Commission on Safety and Quality in Health Care (ACSQHC) and the Independent Hospital Pricing Authority (IHPA) developed a list of 16 national priority complications, known as Hospital Acquired Complications (HACs).<sup>135</sup> Our analysis identifies which, if any, HACs occurred during each admission using the HACs specification version 1.1, as published by ACSQHC.<sup>136</sup>

Unfortunately, indicators of unplanned theatre or intensive care admissions are not currently collected in the NHMD.<sup>137</sup> Consequently, we could not identify HAC 4, or HAC 5. As discussed earlier, our data also doesn't allow us to identify falls. This prevents us from detecting instances of HAC 2. Our exclusion of newborn admissions from our sample also means that HAC 16 is not applicable. We estimate that these shortcomings have caused us to underestimate the prevalence of HACs by 14 per cent, or 0.28 per cent in absolute terms.<sup>138</sup>

---

134. This has been estimated by comparing our current numbers to those that come from applying the CHADx+ v.1.4 algorithm to the 2014-15 National Hospital Cost Dataset provided to Grattan Institute by the Independent Hospital Pricing Authority, where it is possible to identify the incidence of CHADx 2.01-2.13 and CHADx 3.01-3.04.

135. ACSQHC (2017).

136. ACSQHC (2016).

137. Ibid.

138. We assume we would have observed the same rate of HAC 5 as IHPA, and would have also observed HAC 6 at the average prevalence of an additional HAC. By this method, IHPA's inability to estimate HAC 6 means they probably also underestimate the prevalence of HACs, but by about 7 per cent.

Table B.4 on the following page demonstrates that, for public hospital patients in 2014-15, our estimated rates of HACs line up closely with the rates estimated by IHPA. This comparison is aided by a common set of definitions: both IHPA and Grattan have specified the HACs according to version 1.1. It is unsurprising that IHPA finds a slightly higher prevalence, as they have applied more restrictive data cleaning criteria.

### B.2.3 Reconciling definitions of complications

When describing the various types of harm that occur to patients in hospitals, *All complications should count: Using our data to make hospitals safer* refers to CHADx+, HACs and sentinel events separately, and to any of these events as “complications”. CHADx+ is intended to be a comprehensive classification of complications, which should mean that HACs and sentinel events refer to subsets of CHADx+. Unfortunately, it is not quite this neat in practice.

CHADx+ and HACs are defined using routine data, so can both be identified within the NHMD and reconciled against each other. Table B.5 on page 53 shows that 0.04 per cent of admissions, or 2.4 per cent of admissions with at least one HAC, are found to have a HAC event but not a CHADx+ event. The main cause of this discrepancy is a more expansive definition of hypoglycaemia in the HACs classification.<sup>139</sup>

In theory, sentinel events should also be a subset of the comprehensive CHADx+. However, sentinel events are manually recorded, so include some complications – like infants being discharged to the wrong family

139. Diagnoses E1064, E1164, E1364 and E1464 constitute hypoglycaemia in HAC 10.3, when accompanied by a COF 1, but they do not contribute to the prevalence of the analogous CHADx 15.04.

**Table B.3: Incidence of CHADx+, 2014-15**  
Share of admissions involving at least one complication

	<b>Estimates</b>		
	<b>Grattan</b>	<b>AIHW</b>	
MCHADx1	Procedural complications	1.28%	1.13%
MCHADx2	Adverse drug events	0.50%	0.70%
MCHADx3	Accidental injuries	0.31%	0.29%
MCHADx4	Hospital-acquired infections	1.19%	0.26%
MCHADx5	Cardiovascular complications	1.94%	1.38%
MCHADx6	Respiratory complications	0.73%	0.63%
MCHADx7	Gastrointestinal complications	1.34%	1.15%
MCHADx8	Skin conditions	0.64%	0.51%
MCHADx9	Genitourinary complications	0.85%	0.79%
MCHADx10	Hospital-acquired psychiatric states	0.65%	0.46%
MCHADx11	Early pregnancy complications	0.01%	0.01%
MCHADx12	Labour and delivery complications	2.71%	1.52%
MCHADx13	Perinatal complications	0.10%	0.76%
MCHADx14	Haematological disorders	0.53%	0.34%
MCHADx15	Metabolic disorders	1.37%	1.00%
MCHADx16	Nervous system complications	0.17%	0.12%
MCHADx17	Other complications	1.64%	1.30%
Any CHADx		9.36%	8.3%
Any CHAPx		3.80%	N/A
Any CHADx+		10.80%	8.3%

*Notes: The names of the MCHADx1, MCHADx4, MCHADx12 have been revised between the CHADx v.5 seemingly used by AIHW and CHADx+ v1.4 used by Grattan. In the version used by AIHW, these categories were titled Post-procedural complications, Specific infections and Labour, delivery and postpartum complications.*

*Source: AIHW (2016) and Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.*

– that cannot be identified retrospectively from routine data. Consequently, it is unclear how many sentinel events are additional to the complications identified in the NHMD by the CHADx+ algorithm.

These messy definitions make it challenging to arrive at a clean figure of the total number “complications”. In *All complications should count: Using our data to make hospitals safer*, the total number of complications is defined as the total number of CHADx+ events and any HACs events which are not also flagged as CHADx+ events (see Table B.5 on the following page). This figure excludes sentinel events that are not flagged as CHADx+ events because, although they should be included, it’s unclear how to do so and there are too few of them (approximately 100 per year) for their omission to be of consequence.

We avoid these technicalities in the report’s main text, and simply refer to HACs as a subset of the comprehensive CHADx+ classification of complications.

#### B.2.4 Comparing CHADx+ and HACs

In Figure 2.1 of *All complications should count: Using our data to make hospitals safer*, we illustrate that there are significant and idiosyncratic differences between the complications included in CHADx+, and those included in HACs. Table B.6 sets out which minor CHADx+ classes and HACs were compared in this graphic.

#### B.3 Other imputed variables

To facilitate the analysis of the CHADx+ and HACs variables, we generated a number of other patient and hospital characteristics from the variables contained within the NHMD. This section provides definitions and analytical notes on these variables.

**Table B.4: Incidence of HACs in public hospitals, 2014-15**  
Share of admissions involving at least one complication

		Estimate	
		Grattan	IHPA
HAC1	Pressure Injury	0.06%	0.06%
HAC2	Falls resulting in fracture or other intracranial injury	N/A	0.03%
HAC3	Healthcare associated infection	1.03%	1.12%
HAC4	Surgical complications requiring unplanned return to theatre	N/A	0.21%
HAC5	Unplanned intensive care unit admission	N/A	N/A
HAC6	Respiratory complications	0.12%	0.19%
HAC7	Venous thromboembolism	0.06%	0.06%
HAC8	Renal failure	0.02%	0.01%
HAC9	Gastrointestinal bleeding	0.12%	0.12%
HAC10	Medication complications	0.24%	0.26%
HAC11	Delirium	0.42%	0.43%
HAC12	Persistent incontinence	0.06%	0.07%
HAC13	Malnutrition	0.10%	0.10%
HAC14	Cardiac complications	0.65%	0.64%
HAC15	Third and fourth degree perineal laceration during delivery	0.11%	0.15%
HAC16	Neonatal birth trauma	0.00%	0.01%
Any HAC		2.37%	2.68%
Sample size		5,443,561	3,779,338

Source: IHPA (2017) and Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

### B.3.1 Variables derived from Diagnosis Related Groups

Key information about patients' diagnoses and procedures is routinely summarised into Diagnosis Related Groups (DRGs). The DRGs included in the NHMD provided to Grattan Institute by AIHW were assigned according to the version 6 of the Australian Refined DRG classification.

Each component of a DRG codifies a different piece of information about a patient's principal diagnosis or procedure:

- The first letter indicates the body system primarily affected, and is referred to as the Major Diagnostic Category (MDC).
- Together with the two numbers which follow, the MDC identifies a patient's principal diagnosis or procedure and is referred to as their Adjacent DRG.
- The two numbers of a DRG also indicate the DRG's type: 01-39 are used in surgical DRGs; 40-59 are used in medical procedures; and 60-99 are used in medical DRGs.
- The DRG's suffix indicates resource consumption, with complexity decreasing from A to D. The suffix Z is used if there is no split in an Adjacent DRG.

We derive each of these variables from the DRGs contained within our dataset. Figure B.2 on the next page provides a visual summary of the components of a DRG each variable draws upon.

### B.3.2 Index of socioeconomic advantage and disadvantage

From the data collected through each Australian census, the Australian Bureau of Statistics releases sets of socio-economic indices for areas (SEIFA). We used the 2013 edition of the Index of Relative Socio-economic Advantage and Disadvantage from the SEIFA to

**Table B.5: Derivation of the total number of complications**  
Admissions involving at least one complication

	Number per year	Share
CHADx+	891,957	10.63%
HACs, not otherwise counted	3,407	0.04%
Total number of complications	895,364	10.67%

*Note: Figures are calculated over 2012-15.*

*Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.*

**Table B.6: Definitions employed in Figure 2.1 of *All complications should count: Using our data to make hospitals safer***

	CHADx+	HACs
Acute renal failure	9.01	8
Accidental puncture or laceration during a procedure	1.05	N/A
Device or implant-related infection	4.11	3.5, 3.7
Pressure injury, stage 1 or 2	8.01	N/A
Pressure injury, stage 3 or 4	8.02	1.1, 1.2
Constipation	7.03	N/A
Delirium	10.03	11
Hospital-acquired urinary tract infection	4.16	3.1

*Notes: We define these minor complication classes relative to v.1.4 and v.1.1 of the CHADx+ and HAC classifications, respectively.*

estimate the socio-economic status of each patient, on the basis of the Statistical Local Area 2 they resided in when they were hospitalised.<sup>140</sup> The 2011 census data used to construct this index is relevant to our data on 2012-2015 hospital admissions.

For simplicity, we refer to this index as SEIFA from here on.

### B.3.3 Patient age

The information about patient age contained in our dataset is grouped into five year categories, with the exception of the open-ended category of 75 or older. For regression analysis, we relabel these categorical variables with their central value: age category 0 to 4 is recoded as 2, and so forth.<sup>141</sup> Summary statistics for age presented in Appendix C relate to this transformed variable.

We also complete some analysis of knee replacement patients by age category. Table B.7 defines the categories used, and the number of knee replacement patients in each category.

### B.3.4 Comorbidities

We identify patients' comorbidities from the diagnoses assigned to them in the NHMD using the Multipurpose Australian Comorbidity Scoring System (MACSS).<sup>142</sup> Analogous to the better-known Charlson and Elixhauser comorbidity indices, the MACSS index groups specific diagnoses into particular comorbidities like hypertension, and categorises these comorbidities into body system categories.

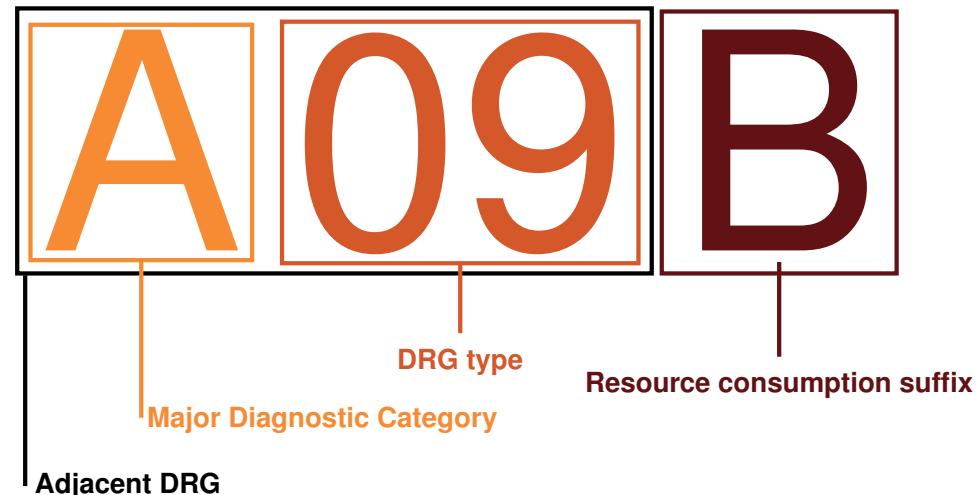
We chose to use this Australian-derived comorbidity index because studies have found that it outperforms the Charlson and Elixhauser

140. ABS (2013).

141. When treated as a count variable, age category "75 and older" was coded as 80.

142. Toson et al. (2016).

Figure B.2: Several variables can be derived from a DRG



Source: Grattan analysis.

Table B.7: Age categories used for knee replacement analysis

Age range	Number of knee replacements, 2012-15
0 – 49	3,132
50 – 64	43,043
65 – 74	55,177
75+	38,402

indices in predicting mortality, readmissions and length of stay for Australian hospital records.<sup>143</sup>

The MACSS body system categories used in our analysis have been modestly refined from those suggested by the index's researchers, in order to better capture characteristics of the data that we saw to be relevant to our analysis of complications. The original body system categories can be found in Toson et al. (2016), our modifications to these categories are described in Table B.8 and our refined body system categories are listed in Table B.9 on the next page.

### B.3.5 Hospital size and scope

We define hospital size as the number of admissions per hospital over the three years for which we have data. We also calculate the number of each hospital's operations over the three years in relation to each of our three case studies: bariatric surgery; medical cardiology; and knee replacements.

To measure the scope of hospitals' services we use an Information Theory Index. These have long been used to capture how similar the concentration of a hospital's activity is to the concentration of the system overall.<sup>144</sup> Our index is constructed using patients' MDCs, so reflects the degree to which hospitals have specialised in particular MDCs:

$$\text{scope}_h = \sum_{k=1}^N p_{k,h} \ln\left(\frac{p_{k,h}}{\varphi_k}\right)$$

Where:

- $p_{k,h}$  = the proportion of hospital  $h$ 's patients that are assigned MDC  $k$
- $\varphi_k$  = the proportion of all patients that are assigned MDC  $k$

143. Toson et al. (2016); and Holman et al. (2005).

144. Kobel and Theurl (2013).

**Table B.8: Refinements made to MACSS' Body System Categories (BSCs)**

Description	Changes required
Diabetes categorised separately from other Endocrine, Metabolic and Immune Diseases.	Indices 16, 17 and 22 removed from BSC 3, and categorised separately.
Drug and alcohol use categorised separately from other Mental Disorders.	Indices 28 and 29 removed from BSC 5, and categorised separately.
Incontinence reclassified as Diseases of the Genitourinary System, rather than Mental Disorders.	Index 36 removed from BSC 5 and reclassified as BSC 14.
Eye diseases categorised separately from Diseases of the Nervous System or Sense Organs.	Indices 44, 45 and 46 removed from BSC 6 and categorised separately.
Chronic renal disease classified separately from Diseases of the Genitourinary System.	Index 74 removed from BSC 10.
Body system category labels Congenital Abnormalities, Injuries and Poisonings, Factors Influencing Health Status and Contact with Health Services, and Symptoms, Signs and Ill-Defined Conditions combined into an "Other" category.	BSCs 14, 15, 16 and 17 combined.

This index increases with the concentration of a hospital's services.

### B.3.6 Coding quality

As discussed in our earlier report, *Strengthening safety statistics*, the quality of hospital coding is known to vary by institution. This variation complicates analysis of hospital safety in two ways.

Firstly, variation in the quality of hospital coding makes it difficult to distinguish a low rate of complications from a low rate of recording whether a condition has been acquired in hospital. Through this channel, shortcomings in coding could cause the safety of some hospitals to be overestimated.

Poor coding quality can also result in fewer comorbidities being recorded or complex DRGs being assigned less frequently. Where this is the case, the risk profiles of patients in some hospitals may be systematically underestimated, causing the hospital's risk-adjusted rate of complications to be overestimated.

To capture these aspects of coding quality, we use two metrics:

#### Prevalence of hospital-acquired conditions among diagnoses:

$$\text{COF 1 Prevalence}_{s,t} = \sum_{i \in I_{s,t}} \frac{\text{Number of COF 1's}_i}{\text{Number of diagnoses}_i}$$

Where:

COF 1 denotes condition onset flag equal to 1, and indicates that a condition has been acquired in hospital

$I_{s,t}$  is the set of all of admissions for state  $s$  in month  $t$

**Coding depth**, which summarises the severity of DRG codes used by a state in a given time period:

Table B.9: MACSS Refined Body System Categories

- |    |  |
|----|--|
| 1  | Infectious diseases  |
| 2  | Neoplasms  |
| 3  | Endocrine, metabolic or immune diseases                      |
| 4  | Diabetes   |
| 5  | Blood diseases   |
| 6  | Mental disorders   |
| 7  | Drug or alcohol use  |
| 8  | Diseases of the nervous system or sense organs               |
| 9  | Eye disease  |
| 10 | Diseases of the circulatory system                           |
| 11 | Diseases of the respiratory system                           |
| 12 | Diseases of the digestive system                             |
| 13 | Chronic renal disease  |
| 14 | Diseases of the genitourinary system                         |
| 15 | Pregnancy, childbirth and puerperium                         |
| 16 | Chronic skin ulcer   |
| 17 | Diseases of the musculoskeletal system and connective tissue |
| 18 | Other  |

$$\text{DRG severity } z\text{-score}_{i,s,t} = \frac{\text{DRG suffix}_{i,s,t} - E[\text{DRG suffix} | \text{ADRG}]}{\sqrt{\text{Var}[\text{DRG suffix} | \text{ADRG}]}}$$

$$\text{Coding depth}_{s,t} = \frac{1}{N_{s,t}} \sum_{i \in I_{s,t}} \text{DRG severity } z\text{-score}_i$$

Where:

- DRG suffix refers to the DRG's resource consumption suffix. For this purpose, the suffixes A, B, C, and D are converted to 1, 2, 3, and 4, respectively.<sup>145</sup> This direct mapping means that low DRG severity  $z$ -scores indicate high severity, as the suffix A is the most severe.<sup>146</sup>
- The expectation and variance of DRG suffixes are calculated across all admissions that share the same adjacent DRG.
- DRG severity  $z$ -scores are calculated for each admission,  $i$ .
- Coding depth summarises the DRG severity  $z$ -scores across admissions within a given state,  $s$ , and month,  $t$ .

We use these metrics to monitor and control for differences in coding quality across states and time.

---

145. If a DRG has the suffix Z, it is not allocated a DRG severity  $z$ -score. This is because the Z suffix indicates there is only one potential severity level for that adjacent DRG. Consequently, that DRG cannot be used to infer the average complexity of DRGs at a particular hospital.

146. It's possible that the DRG severity  $z$ -score will also capture some differences in states' case-mix. However, we expect these differences to be minimal and, regardless, standardise hospital performance estimates across states.

## Appendix C: Conceptual framework

The Grattan Institute report *All complications should count: Using our data to make hospitals safer* finds that one in nine patients in Australian hospitals are harmed during the course of their stay. This rate is as high as one in four for patients who stay in hospital overnight. This does not need to be the case: as we show in the report, some hospitals are succeeding at delivering far safer care. In this chapter, we outline our methodology for identifying how much scope there is for hospitals to reduce their complication rate, given the risk profile of their patients.

### C.1 Conceptual framework

The central premise of our approach is that a patient's hospital outcomes are jointly determined by the patient's condition upon admission, the quality of the care that they receive while admitted and random chance:

$$G(\Pr(\text{adverse outcome})) = \text{patient's condition} + \text{hospital characteristics} + \\ \text{quality of hospital's care} + \text{random chance}$$

When patient outcomes are understood in this way, it follows that the systematic differences in patients' outcomes across hospitals that persist after differences in patients' risk profiles have been accounted for reflect differences in hospitals' quality of care.<sup>147</sup> Hospital characteristics are also included when they are perceived to reflect otherwise unobserved aspects of patients' conditions. The objective of our research is to estimate the magnitude of these differences.

147. The unspecified function  $G(\cdot)$  simply indicates that we expect this additive relationship to be non-linear. We specify and justify our choice of  $G(\cdot)$  in Appendix C.5.1.

This methodology can be applied to any adverse patient outcome, and is known in the health economics literature as *outcomes research*.<sup>148</sup> Our approach is modelled on the recommendations made in Ash et al. (2012), and is complemented by subsequent research that builds on issues raised in that White Paper.

In the following sections, we outline our: choice of dependent variable; approach to risk adjustment; precautions regarding small hospitals, varying coding quality and regression to the mean. We conclude by formally stating our full model specification.

### C.2 Choice of dependent variable

The most common outcomes studied to infer the safety of hospital care are 30-day mortality or readmission.<sup>149</sup> However, coding innovations in Australia over the last 10 years allow us to apply the outcomes research methodology to the prevalence of any hospital-acquired complication.

Complications are a superior choice of adverse patient outcome for our purpose, because they can occur to patients who have strong recovery prospects as well as patients who have ongoing health problems or are close to death. The frequency with which complications occur also improves the statistical power of our analysis.<sup>150</sup> Complications of care – as measured using the HAC classification – is also the contemporary focus of policy about adjusting Commonwealth-state payments. For these reasons, we use hospitals' complication rates to investigate the safety of hospitals' care.

148. Alemayehu et al. (2017).

149. Ash et al. (2012).

150. Krumholz et al. (2006).

The majority of our analysis is completed on the incidence of any complication, as defined by the CHADx+ classification. However, we also fit one model on the incidence of HACs, for the purpose of assessing how these classifications differ in their usefulness for distinguishing the safety of hospital care.

We observe every instance of each type of complication in our data. However, we choose to focus our analysis on whether patients experience at least one complication. Collapsing our count data on complications into a binary variable reduces the amount of variation in outcomes we observe, which may limit the scope for improvement in outcomes that we can identify. We have decided to take this conservative approach because we cannot determine whether secondary complications have been caused by earlier complications or further shortcomings in care.

### C.3 Risk adjustment

The key challenge involved in outcomes research is adequately controlling for each patient's risk of experiencing a complication during their hospitalisation. This is important because, in its absence, a high incidence of complications that is associated with above-average patient complexity could erroneously be attributed to the quality of a hospital's care.

In this section, we discuss the appropriate treatment of alternative outcome measures in risk adjustment, the observable patient and hospital characteristics that we adjust for in our models, and the implications of unobservable patient risk factors on the validity of our estimates.

#### C.3.1 Alternative outcome measures

Our methodology involves estimating hospital performance from the difference between a hospital's observed rate of complications and

the rate which is to be expected, given the variables used for risk adjustment.

To ensure that this expected rate is not set too high, all risk factors that are determined before a patient is admitted to hospital should be included in the risk adjustment model. To ensure that the expected rate is not set too low, intermediate and alternative outcome measures should be deliberately excluded.<sup>151</sup>

The distinction between hospitals' circumstances and outcomes is, in most of these cases, clear cut: patients' demographic characteristics are beyond hospitals' control, but patient outcomes – like length of stay and mortality and readmission rates – are influenced by the safety of hospitals' care. None of these variables are included in our risk adjustment model.<sup>152</sup>

However, when it comes to patient characteristics that relate to health disadvantage, the distinction between circumstances and outcomes of care is less clean cut. Existing disadvantage does mean that patients in these demographics are more likely to experience adverse outcomes, which makes it less likely that hospitals treating these patients will achieve best-in-class results.<sup>153</sup> However, unlike equalising outcomes

---

151. Intermediate outcomes are defined as those that may contribute to the outcome of interest but are determined in hospital, like length of stay and whether a patient was admitted to an Intensive Care Unit. If alternative outcomes were included in the risk adjustment model, hospitals' differences in this metric would be excused. That is, hospital performance as measured by the over-risk-adjusted complication rate would only capture differences in performance that are not correlated with differences in the alternative outcome rate (Ash et al. (2012, p. 19)).

152. The number of procedures a patient has could also be considered an intermediate outcome, if their complications necessitate further procedures. This is not always the case for most complications, but most procedures increase patients' risk of a complication. For this reason, we consider it to be more important, on balance, to include patients' number of procedures in our risk adjustment models. This decision may cause our estimates of hospital performance to be overly conservative.

153. Iezzoni (2012); and Krumholz et al. (2006).

for patients who differ in terms of age, comorbidities and other clinical risk factors, achieving equitable outcomes for disadvantaged Australians should not be beyond the ambition of our health system.

One way of clarifying the appropriate treatment of health disadvantage in outcomes research models of this type is to define the time horizon of reference. In the short term, a high burden of health disadvantage is a reality that should be expected to result in worse outcomes at a hospital level, so this risk should be adjusted for. Over the medium term, the outcomes of these patients' hospitalisations should be improved, rather than excused.

For the most part, our analysis takes a short-term perspective and controls for patients' socio-economic status as a circumstance of care. As we do not have data on indigeneity or other patient-characteristics like cultural and linguistic diversity that are correlated with health disadvantage, we do not control for these factors. Accordingly, our estimates of hospital performance should be interpreted as taking a longer-term view in these regards, as they do not excuse sub-par outcomes that may be attributable to these forms of health disadvantage.

### C.3.2 Observable patient-level risk factors

The biggest determinant of a patient's probability of acquiring a complication is how sick they are on admission. Observable aspects of this risk are often conceptually grouped by:<sup>154</sup>

- patient demographics, like age and sex;
- clinical factors, like diagnoses and comorbidities;
- health behaviours, like exercise habits and diet; and
- attitudes and perceptions, like religion and care preferences.

---

154. Iezzoni (2012).

The routine data we rely on for our analysis is a good source of information on patient demographics and clinical factors. We can identify patients' primary diagnosis, their comorbidities and basic demographic information, such as patients' age, gender and whether they have diabetes. Unfortunately, we do not observe health behaviours or patients' attitudes and preferences.

We are also able to capture some aspects of patients' clinical risk and perhaps health behaviours indirectly, through patients' socio-economic status. The social determinants of health mean that public hospitals and hospitals located in low socioeconomic areas are likely to be treating patients with poorer overall health than those hospitals in areas with high socioeconomic status or private hospitals.<sup>155</sup> We control for this by including socioeconomic status in our risk adjustment model.

Table C.1 on the following page provides summary statistics on all of our patient-level risk factors.

#### Treatment of principal diagnosis in risk adjustment

A patient's diagnosis is likely to affect the impact of every patient-level risk factor on their overall risk of a complication. Accordingly, we think it would be appropriate to allow the risk loading associated with each observable characteristic to depend on a patient's DRG. However, this would result in an unworkably large model specification.

We pursue the same overall model specification by fitting the risk adjustment model using a two-stage estimation approach:

1. First, we estimate each patient's risk of a complication using models that are estimated across only groups of patients who share the same DRG or similar DRGs. Only patient-level risk factors are

---

155. Marmot (2015).

included in this model:

$$G(\Pr(\text{adverse outcome})) = \text{patient's condition} + \text{random chance}$$

2. From these models, we obtain estimates of each patient's risk of experiencing a complication. We collate these estimates from the various models into a single patient risk variable,  $\hat{p}$ .

$$G(\Pr(\text{adverse outcome})) = \hat{p} + \text{hospital characteristics} + \\ \text{quality of hospital's care} + \\ \text{random chance}$$

3. Second, we estimate the overall model outlined in Appendix C.1 using our first-stage risk estimates,  $\hat{p}$ , as the "patient condition" component of the risk adjustment model. This single model is fitted across the full sample of patients being considered, regardless of their DRG.

This two-step estimation approach was applied to each of the six samples we were interested in modelling. On some of these samples, it was feasible to achieve our goal of completing the patient-level risk adjustment separately for each DRG. For example, there are only five types of DRGs that relate to knee replacements, so DRG-specific risk-adjustment could be achieved using just five first-stage models.

For others, the huge variety of DRGs within the sample required us to compromise. For example, there are 658 DRGs among the sample of patients with non-obstetric multiday admissions. We ran 21 first-stage models on this sample, which allowed us to reduce the number of DRGs in each first-stage model to 31 on average. We ensured that the DRGs which were grouped together into the same first-stage model were as alike as possible.

**Table C.1: Summary statistics, patient-level risk variables**

	Mean	Std. dev.
Age (central year of 5 year category)	52.09	22.82
Sex (0 = male, 1 = female)	0.54	0.50
SEIFA	997.79	77.85
Number of procedures	2.08	2.03
Emergency (0 = non-emergency, 1 = emergency)	0.36	0.48
<i>Comorbidities:</i>		
Infectious diseases	0.053	0.224
Neoplasms	0.135	0.342
Endocrine, metabolic or immune diseases	0.080	0.272
Diabetes	0.104	0.305
Blood diseases	0.045	0.208
Mental disorders	0.054	0.225
Drug or alcohol use	0.028	0.165
Diseases of the nervous system or sense organs	0.025	0.157
Eye disease	0.038	0.190
Diseases of the circulatory system	0.123	0.328
Diseases of the respiratory system	0.056	0.229
Diseases of the digestive system	0.131	0.337
Chronic renal disease	0.047	0.211
Diseases of the genitourinary system	0.081	0.273
Pregnancy, childbirth and puerperium	0.058	0.234
Chronic skin ulcer	0.004	0.063
Diseases of the musculoskeletal system and connective tissue	0.051	0.219
Other comorbidities	0.146	0.353

*Note:* Summary statistics calculated on full pre-regression sample, excluding 88 admissions for which sex of patient was not identified as either male or female.

*Source:* Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

In total, we fitted 89 first-stage models. Table C.2 summarises the number of first-stage models run over each of the six samples, and the average number of DRGs within each of them. Where there was more than one DRG in a sample's first-stage models, we also note the principle on which we decided to group DRGs together.

The greater granularity of risk adjustment across our three case studies allows us to have greater confidence in the sufficiency of our risk adjustment efforts across these subsamples, which is why all detailed inference is conducted on these case studies. Estimates relating to all hospital admissions have been derived from aggregating model estimates across the three collectively exhaustive subsamples.

A potential critique of this approach is that the relative contributions of patient characteristics to overall patient risk factors are determined without reference to hospital characteristics. As a consequence, the estimates of the coefficients of patient characteristics in the patient risk models could be biased if they are correlated with the hospital characteristics excluded from this first-stage model. This imperfection is not of concern however because the coefficients of each of the patient risk factors are not of interest in this analytical exercise, and this two-stage approach will not bias the second stage estimates of hospital performance.<sup>156</sup>

### C.3.3 Observable hospital-level risk factors

We also include hospital characteristics in our risk adjustment model where we believe they may reflect otherwise observable differences in patients' risk.

156. This two-stage approach will not bias the overall contribution of patient risk to the risk adjustment model, because the coefficient of the combined patient risk term can adjust to accommodate the inclusion of hospital-level risk factors and hospital performance indicators in the second stage.

**Table C.2: Granularity of patient-level risk adjustment**

Characteristic on which subsamples were defined	Number of subsamples	Average number of DRGs in each subsample	
<i>Collectively exhaustive subsamples:</i>			
Obstetric	DRG & sameday or multiday	20	1
Non-obstetric multiday	MDC & DRG type	21	31
Non-obstetric sameday	MDC & DRG type	21	32
<i>Case studies:</i>			
Multiday bariatric surgery	Procedure type	7	1 <sup>†</sup>
Multiday cardiology	By adjacent DRG	15	2
Multiday knee-replacement	By DRG	5	1

<sup>†</sup> Excluding the 2% of bariatric patients who are assigned very uncommon DRGs. If the 72 very uncommon DRGs are included, this figure jumps to 11.

For example, patients with particularly complex conditions are more likely to seek out the care of large or specialised hospitals.<sup>157</sup> This means that a hospital's size or degree of specialisation may indicate the complexity of their patients. For this reason, we include the scale and scope (*i.e.* breadth of services / extent of specialisation) of hospitals' operations in our risk adjustment model. Each hospital's state and variables that reflect the quality of hospitals' coding are also included, for reasons elaborated on in Appendix C.4.1. Table C.3 presents summary statistics on these hospital-level risk factors.

We note that our decision to include these hospital characteristics in our risk adjustment model is conservative. This methodology attributes all differences in outcomes across hospitals of different scales and scopes to differences in their patients' complexity, and all differences across states to differences in the way they code hospital activity. This is equivalent to assuming there is no difference in the average safety of care in large versus small hospitals, or across states. Consequently, our estimates of the total scope to improve the safety of hospital care will be conservative.

### C.3.4 Unobservable risk factors

Of course, not all factors that affect patients' likelihood of experiencing a complication are observable in routine data. Consequently, we are unlikely to be able to predict a particular patient's likelihood of experiencing a complication with great accuracy.

The impossibility of such *perfect* risk adjustment is sometimes raised as a cause for concern regarding the validity of outcomes research. However, it is important to recognise that risk adjustment is *sufficient* for comparing patient outcomes across hospitals if the patient risk factors that have not been adjusted for are equally distributed

---

157. Rajaram et al. (2015).

Table C.3: Summary statistics, hospital-level risk variables

	Mean	Std. dev.
Hospital volume per annum	71,646	75,492
Hospital scope	0.260	0.401
COF1 prevalence	0.030	0.011
DRG severity <i>z</i> -score	<0.001	0.032
<i>States and Territories:</i>		
New South Wales	0.302	
Victoria	0.262	
Queensland	0.218	
Western Australia	0.096	
South Australia	0.076	
Tasmania	0.022	
Australian Capital Territory	0.014	
Northern Territory	0.011	

*Note:* Summary statistics calculated on full pre-regression sample, excluding 88 admissions for which sex of patient was not identified as either male or female.

*Source:* Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

across hospitals. To mitigate against this risk we focus our analysis on subsets of admissions which are alike.

We complete the majority of our analysis through case studies of admissions for bariatric surgery, medical cardiology and knee replacements because these admissions are quite homogeneous: bariatric surgery and knee replacements are essentially only elective procedures, medical cardiology only emergency. Relative homogeneity within these patient groups more generally also provides a measure of protection against any unacknowledged sources of unobserved variation in the average risk of hospitals' patients.

#### C.4 Other analytical challenges

Beyond adequate risk adjustment, there are two additional challenges involved in arriving at robust estimates of the safety of hospitals' care. The first challenge is accounting for the disjoint between what happens in hospitals and what gets recorded. The second challenge is isolating hospital performance from statistical noise, particularly in the case of small hospitals. In this section, we discuss the strategies we've employed to overcome these challenges.

##### C.4.1 Accounting for differences in coding quality

Each hospital is responsible for converting information collected about patients' hospitalisations into coded records. Coding audits are conducted to ensure that coding practices are consistent across hospitals. However, differences in how routine data has been used to determine hospitals' funding entitlements has meant that the incentives to invest in high-quality coding has varied across states, and over time.

Each Australian state has a different history of using hospital records to determine hospitals' funding entitlements. Victoria introduced casemix funding in 1993-94, and South Australia adopted a similar model in 1994-95. Western Australia and Tasmania introduced forms

of casemix funding in 1996-97, and Queensland started phasing it in from 1997-98, leaving New South Wales to be a slow adopter.<sup>158</sup> In 2011, it was agreed that growth funding from the Commonwealth would transition to an Activity-Based Funding approach, coming into effect from 2014-15.<sup>159</sup>

This varied history of incentives for detailed coding manifests as differences in coding depth across states. Figure C.1 on the following page shows that more DRGs with suffixes that indicate "catastrophic" and "severe" complications or comorbidities are more common in Victoria, the ACT and the NT than the other states, and less severe DRGs are less common. For the larger states, this is more likely to reflect different incentives to code patients' conditions thoroughly, than differences in health across these geographies.

In addition to depending on the general thoroughness of coding practices, our analysis relies on the accuracy of diagnoses' condition onset flags. Condition onset flags (COFs) identify whether diagnoses were present on admission or hospital-acquired, and so are required to distinguish complications from comorbidities.

The accuracy of COFs is also expected to vary by states. Victoria pioneered the use of COFs in 1992, Queensland adopted the practice in 2006 and all other states introduced the COF in 2008. State health departments have also used this data for different purposes over this time, which we expect would have affected the emphasis placed on thorough and accurate COF coding. Given this, it is unsurprising that we also observe substantial variation in the prevalence of COFs equal to 1 – indicating that a complication was hospital-acquired – across states and over time.

---

158. Duckett (1998); and Eager (2010).

159. COAG (2011).

As we detailed in Appendix B of *Strengthening safety statistics* there is also evidence that coding depth, use of the COFs and general coding accuracy also varies across hospitals within states.

We employ a number of strategies to account for differences in coding quality across states and over time, and within states.

#### Accounting for among-state differences in coding depth

Our first strategy for controlling for between-state differences in coding depth is to include in our risk adjustment model the state-by-time period coding depth and COF 1 prevalence variables displayed in Figure C.1 and Figure C.2. By doing so, we account for the fact that a higher rate of complications detected will be higher in months where a hospital's states has been recording a greater proportion of the complications that occur.

To capture any remaining variation in coding quality across states, we also control for differences in the average rate of complications observed across states. While this precaution also excuses any systematic differences in the safety of states' hospitals, identifying differences in states' performances was not the objective of our analysis.

As a further precaution, we de-emphasise among-state differences in our report.

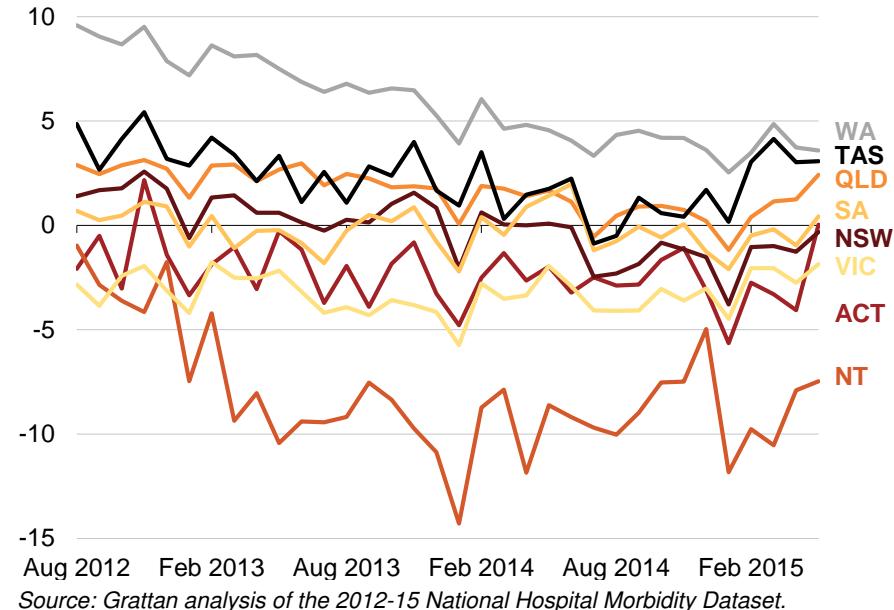
#### Accounting for within-state differences in coding depth

In addition to controlling for among-state variation, we take two precautions to account for within-state differences in coding depth.

Firstly, we control for differences in coding quality that are correlated with hospital size by including hospital size in our risk adjustment model. Of course, there are many other characteristics that vary by hospital size – principally, resources and patient risk. We are not concerned about capturing these characteristics in our attempts to

**Figure C.1: Coding depth varies over time and by state**

Average DRG severity  $z$ -score, by state and month, per cent



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

control for differences in coding quality because, as discussed in Appendix C.3, we also want to risk adjust for these factors.

It's likely that there are also legitimate differences in the safety of large and small hospitals.<sup>160</sup> Insofar as this is the case, excusing differences in safety as if they were attributable to differences in patient risk or coding practices will have made our risk-adjusted complication rates overly conservative.

As a final protection against unobserved within-state differences in coding depth, we identify hospitals' scope for improvement relative to the top decile and quartile of hospitals. Order statistics like deciles and quartiles provide a degree of protection from poor coding quality because they are determined entirely by the observation occupying a particular rank in the ordered sample. This means that, unless all hospitals that make up the best decile are extreme outliers in their coding practices, some hospitals are achieving at the benchmark complication rate against which hospitals' scope for improvement is being calculated.

#### C.4.2 Ensuring estimates of hospital performance are representative of usual outcomes

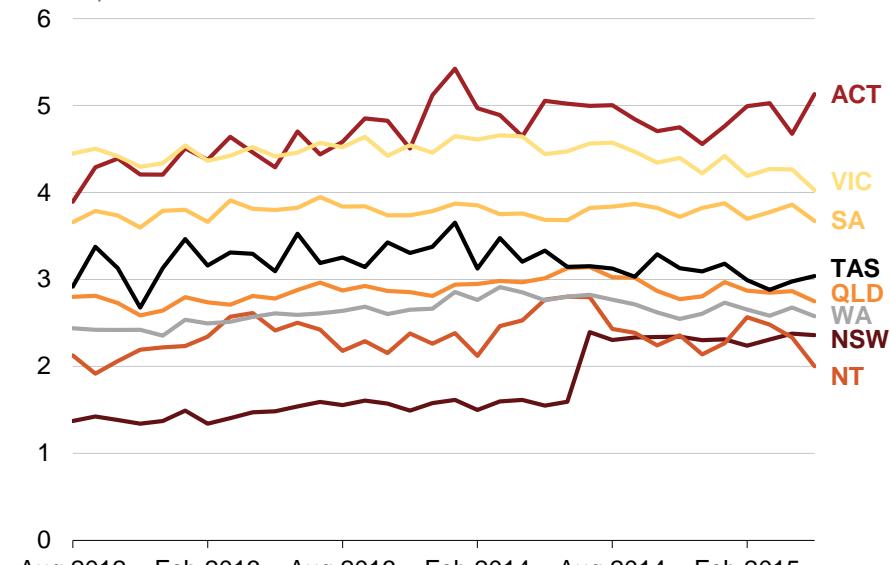
Even where the incidence of complications is recorded accurately, it can be unclear whether the rate of complications incurred over a particular window of time is representative of a hospital's performance generally. This is especially of concern for small hospitals, where each patient outcome has a larger influence on a hospital's complication rate.

When risk-adjusted complication rates are used to estimate the scope for safety improvement, as they are in our analysis, it is particularly important that short-term fluctuations in complication rates are averaged out. This is because performance benchmarks that are set during moments of favourable conditions and good fortune may be

---

160. Pieper et al. (2013).

**Figure C.2: Use of condition onset flag 1 varies over time and by state**  
Average prevalence of condition onset flag 1 per diagnosis, by state and month, per cent



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

near impossible to sustain under regular conditions, even for the safest hospitals. This phenomenon is known as *mean reversion*: extreme outcomes tend to be followed by more moderate ones, reverting towards the mean over time.

To ensure that our estimates are robust to mean reversion, we employ four strategies.

#### Strategy 1: Measure hospital performance over a long time window

Our analysis is based on estimates of hospital performance which are derived from hospitals' outcomes over a three-year period. In Appendix D.4, we illustrate that short-term waves of fortune are expected to even out over a far briefer period: perhaps one month. This accords with our expectations, as the most obvious examples of favourable conditions – a quiet period, a below-average period of infection risk, a cohort of patients that happens to be more complex than their characteristics suggest – would be expected to vary over quite short time periods.

#### Strategy 2: Offset the volatility of small hospitals' performance estimates

Our second defence is to account for the higher volatility of small hospitals' outcomes in our risk-adjustment methodology. Under Bayesian approaches, statisticians make assumptions about what they expect the true value of parameters to be. Rather than simply concluding that parameters are equal to the average value they're observed to take in the data, Bayesian analysis combines these empirical observations with prior expectations. By incorporating an external source of information, Bayesian analysis can arrive at estimates that are more robust to outliers in the data.

We use an empirical Bayes estimation methodology to protect against undue volatility in estimates of small hospitals performances. We start

with the assumption that every hospital probably has the average risk-adjusted rate of complications.<sup>161</sup> Our estimation methodology then calculates the risk-adjusted rate of complications for each hospital observed in our data, and "shrinks" it towards this prior expectation in proportion to hospitals' size. This means that more extreme data is required from small hospitals in order for our model to conclude that their performance is above or below average.

A more formal description of the shrinkage effect we achieve by using an empirical Bayes estimator is provided in Appendix C.5.2.

#### Strategy 3: Impose a minimum sample size for hospitals

A further precaution we employ to protect against outlier performance scores is to impose a minimum sample size for hospitals. Table C.4 presents the cut-off hospital volumes employed in each subsample, and the number of observations affected. Rather than excluding these observations entirely, we grouped all hospitals smaller than these thresholds, and treated each of these groups of hospitals as a single institution.

#### Strategy 4: Benchmark performance relative to a large group of hospitals

Finally, we minimise the sensitivity of our performance benchmark to any remaining mean reversion through general robustness to outlier performances. As argued in Appendix C.4.1, defining the benchmark rate of complications with order statistics reduces the chance that hospitals' scope to improve is calculated relative to an outlier complication rate.

---

161. It's from this assumption that our approach got its name: *empirical Bayes*. This approach is empirical because, unlike most Bayesian approaches, our assumption regarding the true values of our estimated parameters is derived entirely from our data. Consequently, this methodology sits half way between the frequentist and Bayesian traditions.

Just as this approach minimises the risk of setting aspirations relative to what can be appears to be achieved when coding quality is poor, it also minimises the risk that aspirations are defined relative to an outlier result that, instead of being sustainable, will revert to the mean.

## Diagnostics

Figure C.3 on the next page presents the proportion of hospitals in each quintile of cardiology performance by their quintile of performance in the previous year. These rankings are reasonably stable: about 85 per cent of hospitals are in the same or neighbouring performance quintile from year to year.

This finding provides assurance that the hospitals with the most extreme performance estimates are not particularly prone to mean reversion. We expect that estimates employed in *All complications should count: Using our data to make hospitals safer* are even less so because they are calculated over three years rather than the single years used for this analysis.

## C.5 Model specification

As introduced in Appendix C.1, the objective of our analysis is to identify the rate of complications at each hospital in excess of the rate expected, given the risk profile of their patients. In this section, we describe the model specification we use to estimate each hospital's excess complication rate.

Further to the general model specification described in Appendix C.1 and the choice of risk adjustment terms given in Appendix C.2, there are three distinctive characteristics of our model specification. Firstly, we use a logit generalised linear model to accommodate the binary nature of our dependent variable: whether patients experience at least one complication. Secondly, we use a random effects specification to estimate the excess risk associated with each hospital. Finally, we

**Table C.4: Categorisation and grouping of small hospitals**

Cut-off, number of admissions	Number of hospitals	Number of admissions
<i>Collectively exhaustive subsamples</i>		
<b>Obstetric</b>		
<50	259	3,895
50-99	38	2,708
100-199	35	5,006
<b>Non-obstetric, sameday</b>		
<50	34	649
50-99	21	1,481
100-149	26	3,194
150-199	14	2,398
<b>Non-obstetric, multiday</b>		
<100	60	2,380
100-149	16	1,962
150-199	21	3,704
<i>Case studies</i>		
<b>Medical cardiology</b>		
<100	252	10,495
100-149	51	6,144
150-199	45	7,727
<b>Knee replacement</b>		
<100	40	1,851
100-149	15	1,892
150-199	14	2,499
<b>Bariatric surgery</b>		
<50	108	1,105
50-149	9	825
150-199	5	864

include variables containing the mean value of patient characteristics for each hospital in our risk adjustment model.

In this section, we discuss the justifications for and implications of these decisions. We conclude by formally stating our full model specification.

### C.5.1 A Logit Generalised Linear Model functional form

As discussed in Appendix C.2, we treat the incidence of complications as a binary variable to avoid challenges associated with dependence between the first and subsequent complications. This has implications for the appropriate specification of our outcomes research model.

The multiple linear regression approach used in outcomes research requires the error component of the model to be normally distributed. Binary dependent variables cannot satisfy this requirement because the residual error term that results from modelling a binary variable with a linear combination of predictors is not normally distributed.

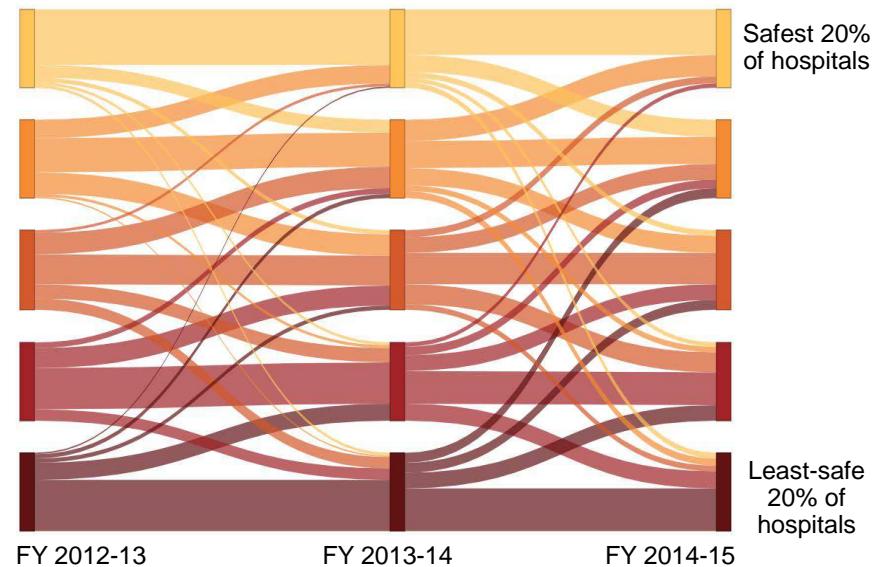
The consequences of proceeding with a linear model specification regardless would include: predictions of the model not being constrained within zero and one, unreliable model predictions at both extremes and heteroscedasticity, which would reduce the efficiency and result in incorrect estimates of the coefficients' standard errors.<sup>162</sup> These problems would impede our ability to accurately identify differences in hospitals' performances, especially for particularly poor and high performing hospitals.

To avoid these problems, we adopt a more general model specification: the logit Generalised Linear Model (GLM). The logit GLM assumes that the log-odds of the dependent variable can be described by a linear combination of predictors:

$$\log \left( \frac{p}{1 - p} \right) = \beta_1 X_1 + \dots + \beta_K X_K$$

<sup>162</sup> Angrist and Pischke (2009).

**Figure C.3: Hospital ranks are reasonably stable over time**  
Hospital performance quintile by year, multiday cardiology admissions



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

Where:

$p$  is the probability of a patient experiencing at least one combination;<sup>163</sup>

$X_1, \dots, X_K$  are the risk factors and hospital performance indicators that linearly explain the log-odds of a patient experiencing a complication

The GLM specification is important for accommodating the binary nature of our dependent variable. It also facilitates an intuitive interpretation of how patients' risk factors interact: the marginal risk associated with any risk factor, such as age, is assumed to depend on all of the other characteristics of a patient, like whether they have diabetes.<sup>164</sup>

The interdependent interpretation of GLMs' components means that the excess risk associated with a hospital's care can only be quantified relative to a particular patient's risk profile.<sup>165</sup> Given this, there are three ways that we can calculate the marginal impact of a hospital's safety on its patients' safety:

1. We can calculate the additional risk of a complication associated with a particular hospital for each patient, accounting for the values that other components of the model, like age and gender, take in each case. Such estimates of hospital risk will be different for each individual.

163.  $p$  is a latent variable. We observe whether a complication occurred, which is linked to the latent probability variable through the inverse Binomial distribution.

164. Specifically, the first order derivative of the probability of a complication,  $p$ , relative to any given covariate,  $X_k$ , is given by the following expression:

$$\frac{dp}{dX_k} = \frac{\exp(\beta_1 X_1 + \dots + \beta_n X_n)}{(1 + \exp(\beta_1 X_1 + \dots + \beta_n X_n))^2} * \beta_k.$$

165. Wooldridge (2010).

2. We can calculate the additional risk of a complication associated with a particular hospital for the average patient in each particular hospital.
3. We can calculate the additional risk of a complication associated with a particular hospital as if the average patient across all hospitals was admitted to that hospital.

We use the second approach when estimating the number of complications that could be avoided, but the third approach when drawing comparisons across hospitals. This is to ensure that we are not suggesting that fictional complications could be avoided, but comparisons between hospitals are made on a like for like basis.

### C.5.2 A random effects specification of hospital performance

As stated in Appendix C.1, the general model specification used for outcomes research is given by:

$$G(\Pr(\text{adverse outcome})) = \text{patient's condition} + \text{hospital characteristics} + \text{quality of hospital's care} + \text{random chance}$$

In practice, there are three ways that the hospital component of this model can be estimated.

#### Pooled effects approach

A *pooled effects* approach – also known as *indirect standardisation* – is the most intuitive to understand. This approach involves estimating the likelihood of a complication on the basis of the risk adjustment terms only:

$$G(\Pr(\text{adverse outcome})) = \text{patient's condition} + \text{hospital characteristics} + \text{random chance}$$

This formula is then used to predict the rate of complications that should be expected for each hospital, given the risk profile of each of their patients. The safety of each hospital is then summarised by the ratio of their observed rate of complications, relative to their expected rate:<sup>166</sup>

$$\text{Hospital performance}_h = \frac{\text{Observed rate}_h}{\text{Expected rate}_h}$$

Where, for a given hospital,  $h$ :

$$\begin{aligned}\text{Expected rate}_h &= \sum_{i \in h} \Pr(\text{adverse outcome})_{ih} \\ &= \sum_{i \in h} G^{-1}(\text{patient's condition}_{ih} + \text{hospital characteristics}_{ih})\end{aligned}$$

The pooled effects approach is the most common approach used in public policy applications of outcomes research.<sup>167</sup> However, the methodology has been consistently recognised in the statistical and health economics literature as being second-best for two reasons.<sup>168</sup>

Firstly, when hospital performance is correlated with patient characteristics, estimates of hospitals' performance will be biased.<sup>169</sup> Secondly, the unacknowledged correlation between the outcomes of patients admitted to the same hospital will result in systematically underestimated standard errors. This error in the coefficient estimates can be corrected at a slight cost to the model's efficiency by using

166. Iezzoni (2012).

167. Ibid.

168. Krumholz et al. (2006); and Ash et al. (2012).

169. This is also the case with the random effects approach we employ. In both instances, this problem can be rectified by including Mundlak means in the model specification, as discussed in Appendix C.5.3. That is, the exclusion of hospital indicators from the model specification will result in Omitted Variable Bias where there is correlation between the independent variables and hospital indicators: Cameron and Trivedi (2005).

cluster robust standard errors. However, reasonable standard errors for the ratio of observed to expected complication rates are difficult to obtain.<sup>170</sup>

Regardless, we use this methodology when estimating hospital performance on each of the 160 minor CHADx+ classes in the heat map in *All complications should count: Using our data to make hospitals safer*, because it is far less computationally intensive than the alternative approaches.

#### Fixed effects approach

An alternative to pooled estimation is to include terms in the risk adjustment model directly – that is, to undertake *direct standardisation*. The simplest way to do this is through a *fixed effects* specification which means that a variable is included in the model for each hospital:

$$\begin{aligned}G(\Pr(\text{adverse outcome})) &= \text{patient's condition} + \text{hospital characteristics} \\ &\quad + \delta_1 H_1 + \dots + \delta_H H_H + \text{random chance}\end{aligned}$$

Where  $H_1, \dots, H_H$  are indicator variables for each hospital, 1 to  $H$ ; equal to 1 for patients that attend that hospital, and 0 otherwise.

In such model specifications, the coefficient associated with each hospital,  $\delta_h$ , can be interpreted in a similar way to the observed-to-expected ratio used in pooled effects specifications: they are estimates of the additional risk of a complication associated with that hospital.

The fixed effects approach produces unbiased estimates of hospital performance and facilitates easy computation of the standard errors of each hospital's performance metric.<sup>171</sup> However, it has a higher mean

170. These could be obtained within a frequentist approach using bootstrapping. However, this would be extremely computationally intensive.

171. Cameron and Trivedi (2005).

squared error, lower efficiency and provides less information about the impact of hospital choice on patient outcomes than can be achieved with a random effects specification.<sup>172</sup> The fixed effects approach can also suffer from the incidental parameter problem, which can cause estimates of hospital performance to be inconsistent.

### Random effects approach

Our analysis employs an alternative *direct standardisation* approach that treats hospital indicator variables as *random*, rather than *fixed*. At a high level, this random effects approach can be specified in the same terms and interpreted in the same way as the preceding fixed effects equation.

However, in the context of measuring hospital performance, the random effects approach offers three distinct advantages.

Firstly, a random effects approach is, statistically, more efficient than a fixed effects approach.<sup>173</sup> This is because random effects models use the assumption that hospital's performances belong to a common distribution to reduce the number of parameters in the model – which means there are more observations with which to estimate each of the models' parameters.<sup>174</sup> Consequently, hospital performance can be estimated more precisely using a random effects approach.

---

172. Ash et al. (2012); and Bell and K. Jones (2015).

173. Ash et al. (2012).

174. Both this attribute of the random effects approach and fixed effects' *incidental parameter problem*, a source of potential bias in the estimates of fixed effects, are attributable to the substantial difference in the degrees of freedom under the two approaches when there is a large number of groups. We don't emphasise the *incidental parameter problem* because the findings of Moran and Solomon (2014) suggest this is not a big problem when analysing hospital performance because hospitals' sample sizes are so large.

A second advantage is that the default methodology for estimating random effects is more robust to outliers than the alternative approaches.<sup>175</sup> Under pooled and fixed effects approaches, the only information that determines hospitals' performance estimates is the relative prevalence of complications in that institution, relative to the expectations set by risk adjustment. The random effects approach differs in that because the distance of each performance estimate from the average is also taken into account.

Constrained by the assumption that hospital performance is normally distributed, the random effects estimates of hospital performance are "shrunk" towards the mean performance in proportion to each hospital's size.<sup>176</sup> By requiring more evidence of outlier performance from smaller institutions, the statistical tendency of smaller groups to be more prone to outlier results is counteracted.

This innovation makes the random effects estimates of hospital performance more accurate than those from fixed effects models, on average.<sup>177</sup> True outlier performances will likely be estimated with a greater error under this specification than under a fixed effects specification. However, this error will be in the conservative direction.

The third – and, in our context, most significant – advantage of a random effects approach is that it allows us to explicitly acknowledge

---

175. Shahian et al. (2005). As this attribute of the random effects approach is attributable to the empirical Bayes estimation methodology with which they are routinely estimated, it could also be achieved under a fixed effect specification. See, for instance, Jacob and Lefgren (2005).

176. Random effects differ in this way from fixed effects because they are derived from the structure of the residual variance after the fixed component has been estimated. The shrinkage property comes from using Empirical Bayes estimation to derive these estimates. This the default estimation methodology of random effects for many statistical programs, including the Stata package GLLAMM employed in our analysis.

177. James and Stein (1961).

our hypothesised data generating process in our model, and estimate particular characteristics of this process.

We hypothesise that the complication risks patients face are not independent. Rather, we expect these risks to be correlated within hospitals, because patients in the same hospital are exposed to many of the same processes and staff.

We're interested in these sources of correlation for two reasons. Firstly, the proportion of total variation in outcomes that is explained by each of these sources of correlation is of policy relevance. The variance decomposition analysis summarised in Appendix D.1.1 would not be possible using a fixed effects specification.

Secondly, if these sources of correlation were not acknowledged in the covariance structure, estimates of the standard error of hospital performance metrics would be downwardly biased unless corrected for using clustered standard errors,<sup>178</sup> which would be less efficient – and as a consequence, less precise.<sup>179</sup>

For these reasons, as well as the advantage over the pooled approach shared with the fixed effects approach, we opt to use a random effects model approach. This results in a model specification which can be written as:

$$G(\Pr(\text{adverse outcome})) = \text{patient's condition} + \text{hospital characteristics} + \alpha_h$$

Where:

$\alpha_h$  is the hospital-level random effects that are normally distributed:  
 $\alpha_h \sim N(0, \sigma_H^2)$

178. For example, Moran and Solomon (2014) use a fixed effects specification with cluster-robust standard errors.

179. Cameron and Trivedi (2005).

Another way of thinking of this model is as a hierarchical equi-correlation model. This means that we assume the outcomes achieved for patients in the same hospital,  $h$ , are all correlated to each other by the same amount. The outcomes of patients treated in different hospitals are assumed to be independent.

Of course, the random effects approach also has disadvantages relative to the fixed effects approach. We summarise these in Box 7 on the following page and discuss how we mitigate the most restrictive of these – potential bias in the estimates of hospital performance associated with correlation between hospital performance and patient risk – in the next section.

### C.5.3 The inclusion of *Mundlak means* in the risk adjustment model

The advantages and disadvantages of random effects specifications relative to fixed effects specifications are summarised in Box 7. The first of each of these listed are critical for our analysis: we want to be able to estimate the proportion of variation attributable to hospital performance but also require our estimates to be correct on average, even if hospital performance is correlated with patient risk characteristics.

We can achieve both these objectives without requiring hospital performance to be uncorrelated with patient characteristics by including hospital-level *Mundlak means* in our risk adjustment model.<sup>180</sup> A random effects model with Mundlak means is considered “part-way” between a random effects and fixed effects approach but out-performs both approaches in finite samples.<sup>181</sup>

Conceptually, the inclusion of Mundlak means constitutes controlling for potential externalities from individual patients' risk profiles in our

180. Bafumi and Gelman (2006); and Bell and K. Jones (2015).

181. Ash et al. (2012); and Dieleman and Templin (2016).

**Box 7: Summarising the advantages of random effects relative to fixed effects**

**Advantages:**

1. **More information:** the random effects approach allows nested hierarchical variance structures to be estimated directly, which then allows the proportion of variation in patient outcomes attributable to hospital performance to be estimated.<sup>a</sup>
2. **More accurate:** the random effects approach reduces the average error in hospital performance estimates by shrinking estimates towards the mean.<sup>b</sup>
3. **More efficient:** by requiring the estimation of fewer parameters, the random effects approach allow all elements of the model, including hospital performance estimates, to be estimated with greater precision. The standard errors estimated by random effects models are also correct, which means post-estimation corrections that reduce efficiency are not required.

a. Ash et al. (2012).

b. As flagged earlier, empirical Bayes estimation can also be applied to fixed effects estimation methodologies.

c. The assumption of normality is justified in this context because the Bayesian central limit theorem establishes that, as the number of units in a random effects cluster increase, the posterior density tends toward multivariate normality. Our average hospital size of 231,760 admissions and minimum size of 100 is sufficient for this asymptotic property to hold approximately. Regardless, the evidence on whether violating this assumption results in significant bias is mixed.

d. Kalbfleisch and Wolfe (2013).

**Disadvantages:**

1. **Stronger assumptions:** Random effects specifications require two more assumptions than fixed effects models:
  - a) that hospital performance is uncorrelated with patient risk factors; and
  - b) that hospital performance is normally distributed.If the first of these assumptions is violated, estimates of hospital performance will be biased.<sup>c</sup>
2. **Less extreme estimates:** The flip side of the greater accuracy achieved using the shrinkage of empirical Bayes is that some improbable extreme performances will be estimated less accurately.<sup>d</sup>
3. **More computationally intensive:** Random effects models are more computationally intensive, and can be substantially so in non-linear models.

risk adjustment model, in addition to controlling for individual patients' risk profiles. Such externalities may arise when the complexity of one patient's condition affects the resources available for assisting other patients.

In practice, it means that the hospital-level average of each covariate is included in the risk adjustment model as well as the raw patient-level variable. Including separate variables for within-hospital and between-hospital variation in risk factors ensures that the impact of both of these factors on hospitals' outcomes are properly accounted for.

Controlling for these differences is not an overly conservative risk-adjustment choice. It's difficult to believe that the hospitals with the sickest patients are systematically worse than other hospitals. In fact, it improves our model's ability to account for hospital-level differences in patient risk which, as discussed in Appendix C.3, is the primary objective of our risk adjustment model.

#### C.5.4 Overall model specification

In the preceding sections, we have described the overarching logic of outcomes research models, the basis on which variables have been included (or excluded) from the risk adjustment component of our model, and the reasons why we have employed a logit model specification with hospital-level random effects. In this section, we combine these characteristics and formally state the models we have estimated.

We've chosen to use a  $\text{logit}(\cdot)$  link function and applied to the dependent variable  $\text{CHADx+}$ , which is equal to 1 if a patient experienced at least one complication and 0 otherwise.  $\text{logit } \text{CHADx+}$  is equivalent to

$$\log \frac{p_{ih}}{1 - p_{ih}},$$

where  $p_{ih}$  is a given patient's (unobserved) probability of experiencing a complication. Consequently, our model

can be interpreted as linear model of patients' *log-odds* of experiencing a complication. All of the results that feature in *All complications should count: Using our data to make hospitals safer* have been transformed so that they are expressed in terms of a patient's *probability* of experiencing a complication.

Our formal outcomes research model specification is:

$$\text{logit}(\text{CHADx+}_{ih}) = \mathbf{x}'_{ih}\mathbf{B} + \bar{\mathbf{x}}'_h\hat{\mathbf{B}}_{MM} + \mathbf{x}'_h\mathbf{B}_H + \alpha_h$$

We estimate this model in two stages:

#### Stage 1:

$$\text{logit} (\text{CHADx+}_{ih}) = \mathbf{x}'_{ih}\mathbf{B}$$

From which we obtain:

$$\hat{p}_{ih} = \frac{\exp(\mathbf{x}'_{ih}\hat{\mathbf{B}})}{1 + \exp(\mathbf{x}'_{ih}\hat{\mathbf{B}})}$$

#### Stage 2:

$$\text{logit}(\text{CHADx+}_{ih}) = \hat{p}_{ih}\beta + \hat{p}_h\beta_{MM} + \mathbf{x}'_h\mathbf{B}_H + \alpha_h$$

Where:

$i$  refers to any of the  $N$  admissions,

$h$  refers to any of the  $H$  hospitals.

**Bold** is used to indicate where a term is a vector containing multiple variables;

$\mathbf{x}_{ih}$  is the vector of patient-level independent variables employed for risk adjustment, which are listed in Table C.1 on page 61. These variables are combined into a single patient-level risk estimate,  $\hat{p}_{ih}$ , in our first-stage models. They are then only included indirectly, through  $\hat{p}_{ih}$ , in our second stage model;

$\bar{\mathbf{x}}'_{ih}$  is the vector of hospital-level (“Mundlak”) means of the patient-level dependent variables employed for risk adjustment. As the independent variables are included in the second-stage model through  $\hat{p}_{ih}$ , the hospital-level average of this term,  $\hat{p}_h$ , is the appropriate transformation of  $\mathbf{x}_h$  for the second-stage model;

$\mathbf{x}'_h$  is the vector of hospital-level independent variables employed for risk adjustment, which are listed in Table C.3 on page 63. They are included directly in our second stage model.

$\mathbf{B}$  is the vector of coefficients associated with  $\mathbf{x}_{ih}$  and  $\mathbf{B}_{MM}$  is the vector of coefficients associated with the Mundlak means of the patient-level regressors. These vectors become scalars in the second stage model, when the vectors of covariates and Mundlak means are collapsed into single variables.  $\mathbf{B}_H$  is the vector of coefficients associated with the hospital-level regressors;

$\alpha_h$  refers to the hospital membership random effects term, which takes the same value for every patient within each hospital. Across all hospitals, this variable is assumed to be normally distributed with variance  $\sigma_H^2 : \alpha_h \sim N(0, \sigma_H^2)$ .

13 models of this specification have been estimated, one on each of the following subsamples:

- Obstetric admissions
- Non-obstetric multiday admissions
- Non-obstetric sameday admissions

- Multiday bariatric admissions
- Multiday knee replacement admissions
  - Also run separately by patient age group
- Multiday medical cardiology admissions
  - Also run separately by financial year

We also estimated one model using the incidence of HACs, rather than CHADx+ as the dependent variable. This was completed using the multiday medical cardiology subsample.

### C.5.5 Estimation methodology

The estimation of random effects models proceeds in three stages. First, the correlation matrix implied by the random effects specification is integrated out of the model’s likelihood function. Second, the fixed component of the model is estimated by maximising the remaining likelihood function. That is, the estimates of the fixed component of the model’s parameters are optimised taking the random component’s correlation structure as a constraint. Finally, the estimates of the random effects terms are recovered from the model’s residuals.

We use numerical integration for the first stage of our model’s estimation. This is because nonlinearities in the functional form of random effects models precludes analytical optimisation by either least squares or maximum likelihood estimation. We choose to use adaptive quadrature over other approaches like Maximum or Penalised Quasi-likelihood

approaches because it's the most stable and accurate of these approaches.<sup>182</sup>

We then use the Newton Raphson algorithm to estimate the parameters of the fixed component of the model from the remaining likelihood function.

Finally, we estimate the values of the random effects terms using empirical Bayes estimation. This involves conditioning the likelihood function for the hospitals' random effects terms, which is derived from the model's residual, on the prior distribution of the random effects. The approach is called *empirical* Bayes because the prior distribution of the random effects is defined by the empirical moments of the data and the assumption of normality. The posterior means that constitute the random effects estimate for each hospital are then obtained using non-adaptive quadrature.

We prefer this approach to estimating the random effects terms to ordinary frequentist or Bayesian estimation because it achieves a lower mean squared error than either approach and mitigates against the 'bouncing beta problem' that can afflict models with a large number of random effects terms.<sup>183</sup> It achieves these properties by shrinking hospitals' random effects estimates towards the prior in inverse-proportion to hospitals' sizes.

Finally, we note the standard errors of the random effects terms that we have used to make inferences about hospital performance are comparative, rather than diagnostic, standard errors. These errors

---

182. The superior performance of adaptive quadrature is attributable to the fact that the approach is *Gaussian* (the spacing of integration points are determined using the roots of a polynomial rather than naïve equal spacing) and uses *importance-based* sampling (it samples more intensively in areas of the likelihood function where there is more data): Rabe-Hesketh et al. (2002) and Haan and Uhlendorff (2006).

183. Rabe-Hesketh and Skrondal (2008).

are calculated separately for each hospital because they relate to the sampling error of each individual random effect, and this depends on each hospital's size and the proportion of its variation in outcomes that can't be explained by the fixed component of the model.<sup>184</sup>

We estimate these error terms from the random effects' posterior distributions because, assuming the prediction errors are normally distributed,<sup>185</sup> these standard deviations are accurate conditional and unconditional measures of the mean squared error of prediction even in nonlinear contexts.<sup>186</sup> The limitation of this approach is that the sampling variation of the model's parameters is not accounted for. Accordingly, these estimates of the random effects' terms comparative standard errors are not exact.<sup>187</sup>

We implement this estimation methodology using Rabe-Hesketh and Skrondal GLLAMM package in version 15 of Stata.

---

184. See Rabe-Hesketh and Skrondal (*ibid.*) for a full discussion of the difference between diagnostic and comparative standard errors for empirical Bayes estimates of random effects, and the various alternative approaches to estimating comparative standard errors.

185. The normality assumption is justified in binary response models because the Bayesian central limit theorem establishes that, as the number of units in a cluster increase, the posterior density tends toward multivariate normality. As this is an asymptotic property, we ensure a minimum hospital sample size of 100 admissions.

186. Rabe-Hesketh and Skrondal (2008).

187. This could be resolved through bootstrap estimation of each hospital's comparative standard error, but the computational intensity of this approach was prohibitive in our context.

## Appendix D: Results

In the preceding chapters, we have set out our data sources, explained the derivation of additional variables, and laid out the model specification we used to estimate the excess risk of a complication at each hospital. In this chapter, we take all these variables – including the hospital performance metrics – as given, and summarise our results.

Table D.1 on the next page presents summary statistics on our dependent variable, CHADx+. Appendix D.1 summarises the results of our hospital performance models, Appendix D.2 presents our diagnostic analysis of these models, Appendix D.3 presents other supporting analysis and Appendix D.4 presents a stability analysis.

### D.1 Hospital performance metrics

*All complications should count: Using our data to make hospitals safer* uses 13 models of hospital performance to investigate variation in safety across hospitals: one for each of the six subsamples defined in Table B.2, and the seven required to compare hospital performance across age categories for knee replacement patients, and compare performance across years using the medical cardiology sample. In this section, we collate the key findings of these models.

#### D.1.1 Variance decomposition

For each model, we estimate the proportion of variation explained by the regressors, hospital performance, and state performance or coding differences, and the proportion of variation that remains unexplained.

Estimates of the standard deviation of our models' hospital effects ( $\sigma_H$ ) are obtained directly from the output of each model. By construction, the variance of the residual term of a logit model is  $\pi^2/3$ . To estimate the proportion of variation in outcomes explained by the regressors, we

use the pseudo- $R^2$  proposed for binary dependent variable models by McKelvey and Zavoina (1975).

$$R_{\text{MZ}}^2 = \frac{\frac{1}{N} \sum_i (x_i \hat{\beta} - \bar{x} \hat{\beta})^2}{\pi^2/3 + \sigma_H^2 + \frac{1}{N} \sum_i (x_i \hat{\beta} - \bar{x} \hat{\beta})^2}$$

We follow Zhang et al. (2013) in using this pseudo- $R^2$  in combination with the estimated variances of our random effects terms to define variance partition coefficients:

$$\text{VPC}_H | \sigma_H^2, R_{\text{MZ}}^2 = \frac{\sigma_H^2}{\frac{\pi^2}{3} + \sigma_H^2 + \frac{1}{N} \sum_i (x_i \hat{\beta} - \bar{x} \hat{\beta})^2}$$

$$\text{VPC}_\varepsilon | \sigma_H^2, R_{\text{MZ}}^2 = \frac{\pi^2/3}{\frac{\pi^2}{3} + \sigma_H^2 + \frac{1}{N} \sum_i (x_i \hat{\beta} - \bar{x} \hat{\beta})^2}$$

These figures can be interpreted as the proportion of variation in outcomes that can be explained by observable hospital and patient characteristics, hospital performance, and the proportion of variation in outcomes that cannot be explained by these factors. Our estimates of these figures for each model are presented in Table D.2.

Estimates relating to the whole sample are calculated as a weighted average of the estimates from the obstetric, non-obstetric multiday and non-obstetric sameday samples, weighted by sample size. These figures are also summarised in Table D.2.

**Table D.1: Prevalence of complications, by complication type and sample**

	<b>All admissions</b>		<b>Public hospital admissions</b>		<b>Private hospital admissions</b>		<b>Case studies</b>	
	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
<i>Any length of stay</i>								
Number	25,175,958		15,648,510		9,527,448		37,691	
CHADx+	10.63%	30.82%	13.19%	33.83%	6.43%	24.53%	15.49%	36.18%
CHADx	9.18%	28.87%	11.47%	31.86%	5.41%	22.63%	14.76%	35.47%
CHAPx	3.84%	19.21%	4.76%	21.30%	2.31%	15.04%	2.20%	14.68%
HACs	1.72%	12.99%	2.29%	14.95%	0.78%	8.79%	1.85%	13.49%
<i>Multiday admissions</i>								
Number	8,037,258		6,043,377		1,993,881		562,725	
CHADx+	27.01%	44.40%	27.72%	44.76%	24.87%	43.23%	16.79%	37.38%
CHADx	24.25%	42.86%	24.77%	43.17%	22.67%	41.87%	14.96%	35.67%
CHAPx	9.31%	29.05%	9.65%	29.53%	8.26%	27.52%	2.80%	16.50%
HACs	5.17%	22.15%	5.70%	23.18%	3.58%	18.57%	4.58%	20.91%
<i>Sameday admissions</i>								
Number	17,138,612		9,605,094		7,533,518		139,754	
CHADx+	2.95%	16.91%	4.04%	19.70%	1.55%	12.35%	33.96%	47.36%
CHADx	2.11%	14.37%	3.10%	17.33%	0.85%	9.17%	27.97%	44.89%
CHAPx	1.27%	11.20%	1.68%	12.87%	0.74%	8.58%	8.61%	28.05%
HACs	0.10%	3.09%	0.14%	3.74%	0.04%	1.98%	5.27%	22.34%

*Notes:* These subsample sizes include observations that were missing data on independent variables to be used in regression analysis. Differences between the coding practices of public and private hospitals mean that direct comparisons of public and private sector complication rates are not meaningful.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

### D.1.2 Estimating each hospital's excess risk of a complication

The second component of our model of relevance to *All complications should count: Using our data to make hospitals safer* is the estimates of hospital performance: that is, the random effects series,  $\alpha_h$ , which takes a particular value for each hospital. Except where surgeons operate across a large number of hospitals, any “surgeon effect” will swept up in these terms.

As described in Appendix D.1.1, our logit model specification makes these random effects terms difficult to interpret. When estimated, they are expressed in terms of patients’ log odds of a complication. To express these terms as probabilities, we calculate the probability that a given patient experiences a complication with and without accounting for the performance of their hospital, average across all patients and take the difference:<sup>188</sup>

$$\Pr(\text{CHADx}_{+ih} | X_{ih}) = \frac{\exp(\mathbf{X}'_{ih}\mathbf{B})}{(1 + \exp(\mathbf{X}'_{ih}\mathbf{B}))}$$

$$\Pr(\text{CHADx}_{+ih} | X_{ih}, \alpha_h) = \frac{\exp(\mathbf{X}'_{ih}\mathbf{B} + \alpha_h)}{(1 + \exp(\mathbf{X}'_{ih}\mathbf{B} + \alpha_h))}$$

$$\text{Risk}_h = \frac{1}{N} \left( \sum_{i \in N} \Pr(\text{CHADx}_{+ih} | X_{ih}, \alpha_h) - \sum_{i \in N} \Pr(\text{CHADx}_{+ih} | X_{ih}) \right)$$

Where  $\mathbf{X}'_{ih}\mathbf{B}$  is shorthand for  $\mathbf{x}'_{ih}\mathbf{B} + \mathbf{x}'_h\mathbf{B}_{MM} + \mathbf{x}'_h\mathbf{B}_H$ , which is defined in Appendix C.5.4.

We have specified our model such that  $\alpha_h$  has a mean of 0 across all hospitals. It follows that patients in the top performing half of hospitals

188. In every instance where we draw comparisons across hospitals,  $\text{Risk}_h$  is calculated by averaging  $\Pr(\text{CHADx}_{+ih} | X_{ih})$  and  $\Pr(\text{CHADx}_{+ih} | X_{ih}, \alpha_h)$  across all patients, even those who did not attend that particular hospital,  $h$ . This is so that all hospitals are being assessed relative to the same cohort of patients.

Table D.2: Variance partition coefficients by model

	$R^2_{MZ}$	VPC <sub>H</sub>	VPC <sub><math>\varepsilon</math></sub>
<b>Mutually exclusive, collectively exhaustive subsamples</b>			
Obstetric	72%	2%	26%
Non-obstetric multiday	41%	10%	50%
Non-obstetric sameday	33%	10%	57%
<i>Full sample</i>	38%	9%	53%
<b>Case studies</b>			
Multiday bariatric surgery	61%	8%	31%
Multiday cardiology	27%	10%	63%
2012-13	28%	12%	60%
2013-14	27%	11%	62%
2014-15	26%	9%	65%
Multiday knee replacement	40%	9%	51%
0 – 49 years	41%	6%	53%
50 – 64 years	38%	9%	53%
64 – 74 years	39%	9%	51%
75+ years	42%	9%	49%

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

will have a lower probability of a complication than the risk adjustment component of our model predicts, and patients in the bottom performing half of hospitals will have a higher probability. Consequently,  $\text{Risk}_h$  will be positive for half the hospitals in our data, and negative for half the hospitals.

We further calculate the excess risk of a complication that a patient faces because they did not attend the best performing hospital as follows:

$$\begin{aligned}\text{Excess risk}_h &= \text{Risk}_h - \min_{h \in H} \text{Risk}_h \\ &= \Pr(\text{CHADx+}_{ih} | X_{ih}, \alpha_h) - \Pr(\text{CHADx+}_{ih} | X_{ih}, \min_{h \in H} (\alpha_h))\end{aligned}$$

Accordingly,  $\text{Excess risk}_h = 0$  for the safest hospital, and is positive for all other hospitals.

We calculate approximate 95% confidence intervals around these excess risk estimates as follows:

$$\begin{aligned}\text{Excess risk}_{h,LB} &= \Pr(\text{CHADx+}_{ih} | X_{ih}, \alpha_h - tcrit_{LOS=0.025} * se(\alpha_h)) \\ &\quad - \Pr(\text{CHADx+}_{ih} | X_{ih}, \min_{h \in H} (\alpha_h))\end{aligned}$$

$$\begin{aligned}\text{Excess risk}_{h,UB} &= \Pr(\text{CHADx+}_{ih} | X_{ih}, \alpha_h + tcrit_{LOS=0.025} * se(\alpha_h)) \\ &\quad - \Pr(\text{CHADx+}_{ih} | X_{ih}, \min_{h \in H} (\alpha_h))\end{aligned}$$

$$95\% \text{ confidence interval}_h = (\text{Excess risk}_{h,LB}, \text{Excess risk}_{h,UB})$$

These confidence estimates are approximate for two reasons. Firstly, as we note in Appendix C.5.5, the comparative standard errors of the random effects estimates obtained from Stata's GLLAMM do not account for the estimation error around each of the parameters in the fixed component of our model.

Secondly, our approach to calculating confidence intervals treats our estimate of  $\min_{h \in H} (\text{Risk}_h)$  as a fixed benchmark off. Of course,  $\min_{h \in H} (\text{Risk}_h)$  is also estimated with uncertainty. There are no simple remedies to these problems.<sup>189</sup> Instead, we recommend that our estimates of the confidence intervals surrounding our random effects estimates are approximate.

Secondly, we're interested in is constructed around these estimates of excess risk using the standard errors estimated for each random effect by Stata's GLLAMM package,  $se(\alpha_h)$ :

### Inter-hospital comparisons

*All complications should count: Using our data to make hospitals safer* uses inter-hospital comparisons to estimate the scope for complications to be reduced in aggregate, and to showcase that there is such scope for improvement within every state.

Here, we present caterpillar plots of all the excess risk estimates underpinning these calculations, by admissions sample. Each dot on a caterpillar plot represents a specific hospital's scope to improve, and the line extending above and below each point represents the 95 per cent confidence interval of this estimate.

These confidence intervals are appropriate for testing single hypotheses, such as whether a hospital's complication rate is different from a particular peers'. They are not appropriate for drawing inference about whether a given hospital's performance is different from a set of their peers' performances.<sup>190</sup>

---

189. These problems could be resolved within our frequentist approach using bootstrapping, or by switching to a Bayesian estimation approach.

190. We emphasise this point because caterpillar plots are often misused in this way: Moran and Solomon (2014). A Bonferroni correction would need to be applied to the confidence intervals we have presented in order to apply the same level of

Figure D.1 and Figure D.2 on the following page present excess risk estimates for each hospital's multiday and same day non-obstetric patients. The most significant difference between these graphs is that the excess risk for multiday patients at any hospital dwarfs that faced by same day patients.

This difference is in line with our expectations, as complications are nine times more common among multiday admissions. It also indicates that, when multiday and same day admissions are considered together, estimates of the scope to reduce complications will be substantially lower, but the relative safety of a hospital's same day care will have little bearing on how its safety compares to its peers overall.

Figure D.3 on page 84 and Figure D.4 on page 85 present excess risk estimates for medical cardiology admissions and multiday bariatric surgery. The substantial difference in the number of hospitals providing these services affects the number of outliers we observe across the samples. Evidently, the precision with which meaningful feedback on relative performances can be provided depends on the generality of the admissions compared.

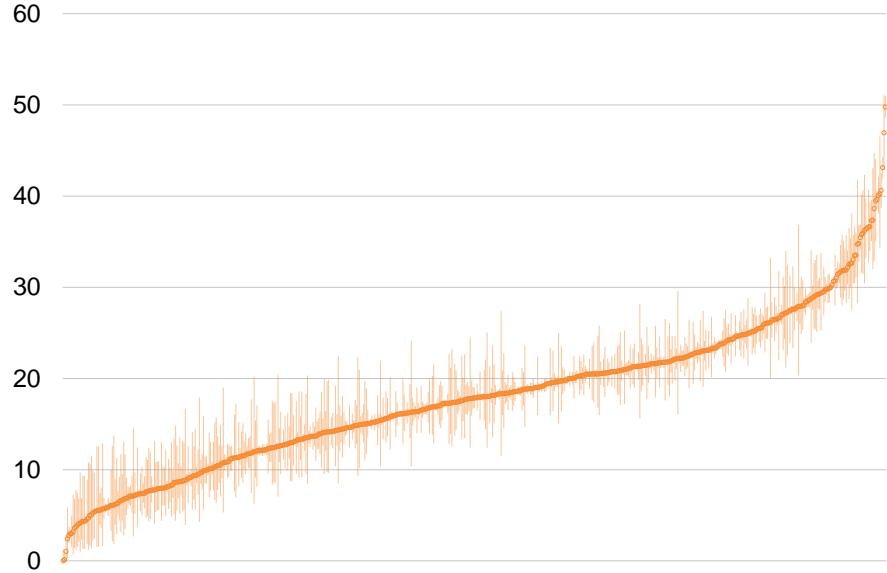
The large number of medical cardiology admissions makes this sample particularly informative about the relative safety of hospitals' care. The medical cardiology sample also has the advantage of being relatively homogeneous, which made it feasible to conduct risk adjustment separately for each DRG. Together, these characteristics make the medical cardiology excess risk estimates the most robust of our samples.

However, even in the bariatric surgery sample, hospital performance can be estimated with sufficient precision to identify which hospitals are above average performers, and which are below average performers.

---

confidence to multiple hypothesis tests, as the joint probability of multiple events with 95 per cent probability is less than 95 per cent: Dunn (1959).

**Figure D.1: Excess risk varies substantially for multiday admissions**  
Excess risk by hospital, non-obstetric multiday admissions



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

Figure D.5 on page 86, Figure D.6 on page 87 and Figure D.7 on page 88 present the excess risk estimates for obstetric patients, medical cardiology patients (by category of complication) and knee replacement patients (by age group). The key takeaway from these estimates is that the amount of variation in the outcome measure of interest is a key determinant of how much information can be derived from performance comparisons.

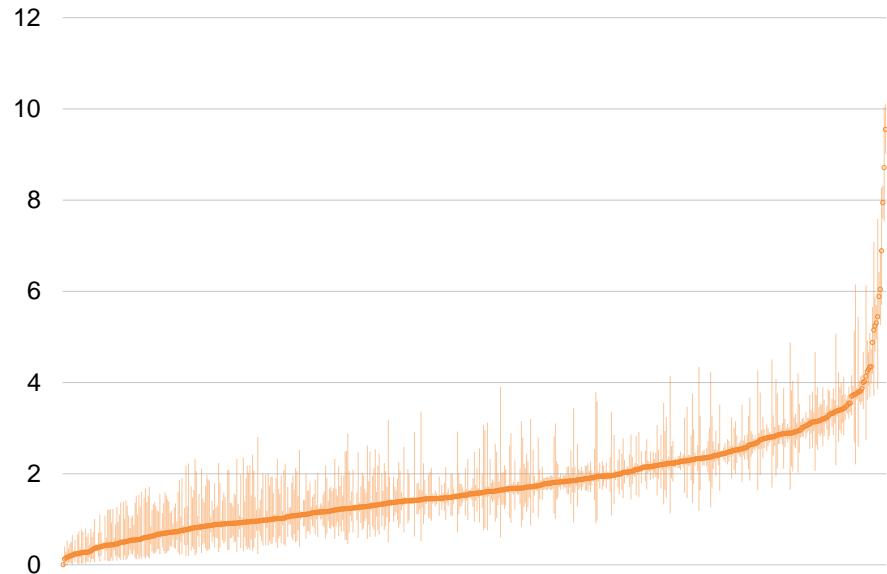
For example, 46 per cent of obstetric patients experiences a CHADx complication, but the variance of the risk-adjusted rate of complications across hospitals is less than 1 per cent. This is because some of the obstetric diagnoses included in the CHADx classification are common but extremely difficult to avoid. Such events do not make particularly informative metrics.

The HACs classification of complications was created to address such concerns. However, the collection of complications deemed to have good clinical preventability has not produced a more informative performance metric. Figure D.6 illustrates that the CHADx+ classification identifies hospitals' scope to reduce complication rates with far greater precision.

The usefulness of particular metrics for identifying the scope for hospitals to improve the safety of their care also depends on the group of patients that are of interest. Figure D.7 shows that the risk-adjusted prevalence of CHADx+ provides useful information about the safety of knee replacements for older patients – even providing enough information to identify hospitals' different strengths. However, this indicator is not particularly informative about which hospitals may be better knee replacements for young people.

Finally, the variation in hospitals' excess risk observed across these samples also exists within states. Figure D.8 demonstrates that excess risk for multiday non-obstetric admissions varies substantially across hospitals within all states, and within the private sector. Figure 3.3

**Figure D.2: Excess risk varies modestly for sameday admissions**  
Excess risk by hospital, non-obstetric sameday admissions



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

in *All complications should count: Using our data to make hospitals safer* presents the same analysis of medical cardiology admissions. No average differences in the safety across states are observable in these figures because we have controlled for differences in coding depth and any remaining differences across states.<sup>191</sup>

#### Intra-hospital comparisons

We also investigated differences in the safety of care within hospitals. For this analysis, hospitals' excess risk was estimated relative to their average patient, rather than the average patient across all hospitals. The comparison of these excess risk estimates for particular hospitals across each sample of admissions is presented as Figure 3.5 of *All complications should count: Using our data to make hospitals safer*.

*All complications should count: Using our data to make hospitals safer* also uses these estimates to demonstrate that risk adjustment is required in order to infer whether a hospital's complication rate is above or below what should be expected. Reproduced as Figure D.9, Figure 3.1 of *All complications should count: Using our data to make hospitals safer* was calculated by reporting the rate of complications expected per hospital given patient risk, and then the rate observed after hospital's excess risk was accounted for.

We note that the rates of specific major CHADx+ classes presented in this figure were obtained by extrapolating the amount that each hospital was expected to exceed (or underrun) the overall complication rate relative to the rates of major CHADx+ classes. These figures could be made more robust by replicating our analysis at the major CHADx+ class level.

---

191. Any apparent differences in the average excess risk across states is attributable to different numbers of private hospitals within each state, and differences in the number of outliers that needed to be excluded.

**Figure D.3: Most hospitals provide medical cardiology care**  
Excess risk by hospital, multiday cardiology admissions

50

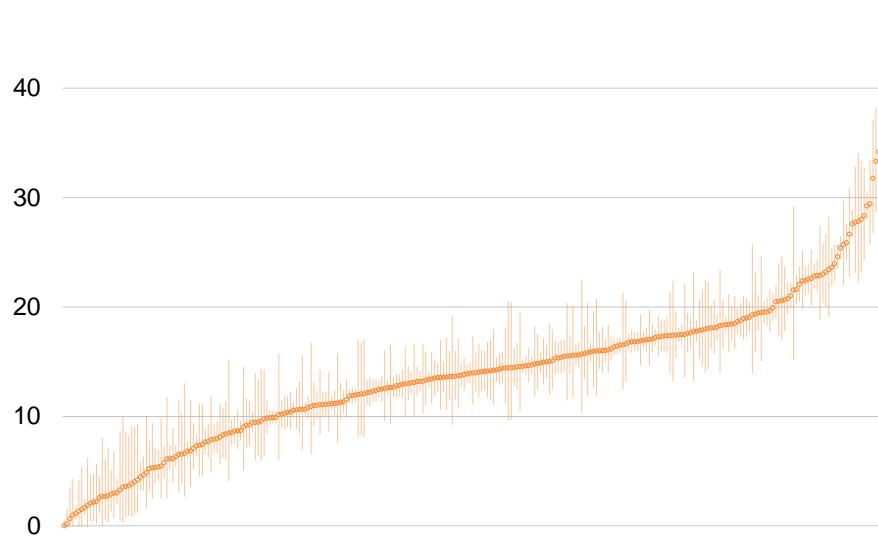
40

30

20

0

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.



### D.1.3 Overall scope for improvement

In *All complications should count: Using our data to make hospitals safer*, we summarise the excess risk observed across hospitals by calculating the total scope for improvement. These metrics are calculated relative to a particular quantile of performance,  $q$ , as follows:

$$\text{Scope for improvement}_q = \frac{1}{N_H} \sum_{h \in H'} (\text{Excess Risk}_h - \text{Benchmark}_q)$$

Where:

$q$  the quantile that defines the target rate of complications, and  
 $\text{Benchmark}_q$  is the corresponding complication rate;

$\text{Excess risk}_h$  current excess risk of a complication at hospital  $h$ , evaluated relative to the average patient profile at each hospital;

$H$  is the set of all hospitals in the sample

$N_H$  is the total number of hospitals in  $H$ .

$H'$  is the set of hospitals such that  $\text{Excess risk}_h - \text{Benchmark}_q > 0$ .

Our estimates of the scope for improvement by sample are listed in Table D.3.

For the cardiology case study, we estimate the scope to improve both CHADx+ and HACs. We find that the scope to reduce these events implied by the difference between the average and top decile hospitals' rates is similar. This implies that the reducibility of these events is similar, even if the base prevalence of the indicators differs substantially.

**Figure D.4: Few hospitals complete bariatric surgery**

Excess risk by hospital, multiday bariatric admissions

50

40

30

20

10

0

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.



## D.2 Model diagnostics

To be comfortable with our hospital performance estimates, we have carefully assessed each of our six key models' fit and the reasonableness of their assumptions.<sup>192</sup> We discuss these aspects of our models in turn.

### D.2.1 Goodness of fit

The reliability of our estimates hinges on the adequacy of our risk adjustment, and the fit of the model overall.

Our McKelvey and Zavoina pseudo- $R^2$  estimates show that the risk adjustment component of our model explains between around 30 and 70 per cent of the variation in outcomes (Table D.4 on page 93). The maximum amount of variation in outcomes that could be explained by patient risk with perfect information is unknown, but we note that our pseudo- $R^2$ 's are high relative to those achieved elsewhere in the outcomes research literature.<sup>193</sup>

Our model also appears to fit our data well overall. Figure D.10 shows that the proportion of patients who experienced a complication accords closely with the probability of a complication that our model assigned them. Each bubble on these graphs represents the patients that were allocated a particular decile of risk (x axis). This compares closely to the proportion of these patients who actually experienced a complication, except where there were very few patients in the risk decile – as indicated by a small bubble size.

**Figure D.5: Obstetric patients face a uniformly high risk of a complication**

Excess risk by hospital, obstetric admissions

60

50

40

30

20

10

0

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

192. Full diagnostics for our analysis of cardiology admissions in 2012-13, 2013-14 and 2014-15 separately and our separate analysis of knee replacement patients by each age group are not presented here. However, these models fit similarly to the overarching cardiology and knee replacement models, and are not drawn upon extensively.

193. Zhang et al. (2013).

The classification accuracy rates presented in Table D.4 indicate that our models accurately predict whether a patient will experience a complication in 73 to 86 per cent of cases, depending on the subsample.<sup>194</sup> This accuracy rate applies to the group of patients who experienced complications as well as those who didn't.

We achieve this by setting the cut-off probability above which we classify admissions as being expected to incur a complication equal to the probability that minimises Youden's index. This index weights the model's likelihood of correctly identifying true positives (sensitivity) and correctly identifying true negatives (specificity) equally:<sup>195</sup>

$$\text{Youden's index(pr)} = (1 - \text{sensitivity(pr)})^2 + (1 - \text{specificity(pr)})^2$$

It is important to define the cut-off probability in this way because we care about the accuracy with which we can predict the rare outcome – experiencing a complication.<sup>196</sup> If we only cared about the overall accuracy of our model, we could surpass these accuracy rates by simply setting the cut-off probability at 100 per cent, and assuming patients wouldn't experience a complication regardless of their estimated probability. Given that only 10 per cent of patients incur complications, we'd still be right 90 per cent of the time but would learn nothing about the fit of our model.

For comprehensiveness, we also present the Receiver Operator Curves (ROC) for each model (Figure D.11 on page 92). The curves are formed by lining up observations by the rank of their predicted probabilities, and extending the line upwards if the admission didn't involve a complication, and towards the right if it did.<sup>197</sup> These curves

194. While Figure D.10's calibration graphs illustrate these accuracy rates calculated across deciles, these accuracy rates are calculated at the observation level.

195. Youden (1950).

196. Böhning et al. (2008).

197. Cameron and Trivedi (2005).

**Figure D.6: Some metrics are more useful for detecting differences in the safety of hospitals' medical cardiology care**

Excess risk by hospital across cardiology admissions, for CHADx+ and HACs

50

40

30

20

10

0

CHADx+

HACs

Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

present a fuller picture of each model's sensitivity and specificity than classification accuracy alone.

All of our ROC curves bow out towards the top-left corner to a satisfactory degree. The extent to which they do so is summarised by the AUC statistic, and is generally considered adequate if the AUC is approximately 80 or greater. The AUC stands for the Area Under the ROC Curve, but is more easily understood when thought of as the average rank of admissions which did involve a complication, given that the observations are ordered by their predicted probability of a complication.

Each of the pseudo- $R^2$ , classification accuracy and AUC together mean that we can explain between 30 and 70 per cent of the variation in outcomes by patients' characteristics, we can accurately predict a patient's outcome about three-quarters or more of the time, and when our predictions are wrong, tend to be wrong by much.

## D.2.2 Validity of model assumptions

In addition to assuring each model's in-sample fit, we have done our best to ensure the underlying assumptions of our model specification are valid. The key assumptions are summarised in Box 8 on page 94. The assumptions pertaining to the random effects specification discussed in Appendix C.5 are restated here as assumptions 3 and 5.

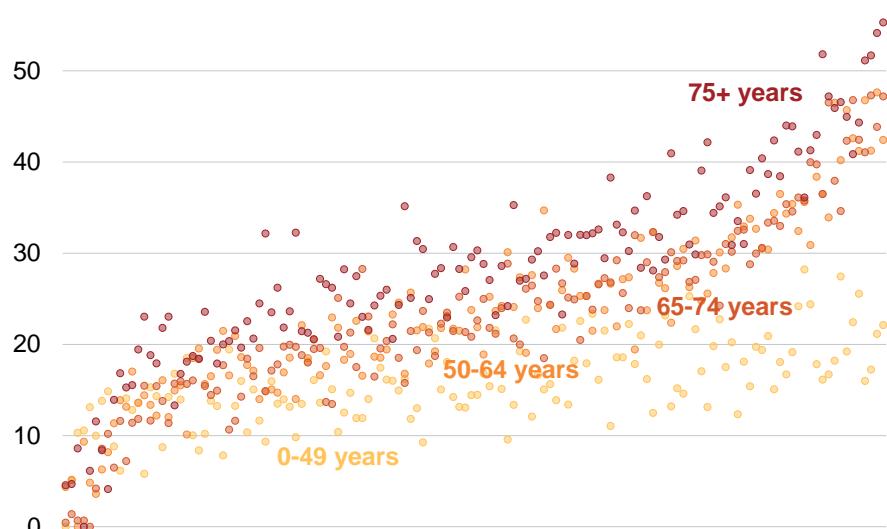
### No multicollinearity

We confirmed that our independent variables were not collinear by estimating the pairwise correlations between our independent variables. Estimates of the correlations between patient-level covariates were all below 20 per cent.

Our two coding quality variables, coding depth and condition onset flag prevalence, were 67 per cent correlated. While high, this correlation is

**Figure D.7: CHADx+ provides little information about the relative safety of knee replacements in younger patients**

Excess risk by hospital and age group, knee replacement admissions  
60



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

still below the threshold of 80 per cent usually used to define collinearity. This is not of great consequence because we do not aim to distinguish their contribution to patients' likelihood of having a complication recorded. For the purpose of controlling for overall coding quality, the 33 per cent difference between these variables is sufficient to justify their inclusion.

### No endogeneity

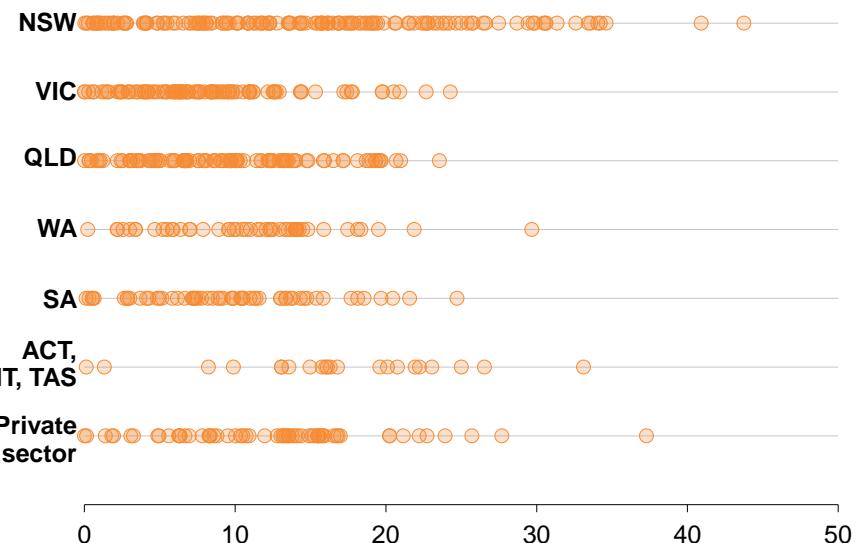
As discussed in Appendix C.5.3, we protect against the most likely source of endogeneity: correlation between the random effects terms and covariates, proactively by including Mundlak means in our model specification.

As generally the case in outcomes research, there are other potential sources of endogeneity that could affect our model. Length of stay is excluded because it's an intermediate outcome variable. However, its exclusion could cause endogeneity via omitted variable bias. Our data also contains little information about patients' health behaviours, religions and care preferences. The exclusion of these types of risk factors could also some degree of endogeneity. However, without a valid instrument, we can't test for or address these potential sources of endogeneity. This is an inherent limitation of our model.

### Distributional assumptions hold

We assess whether our residual term is over-dispersed by comparing it to its assumed distribution. With a sample size of  $N$  and with  $P$  parameters, the residuals of logit generalised linear models are assumed to be  $\chi^2$  distributed with  $N - P$  degrees of freedom.<sup>198</sup> Consequently, a logit model can be said to be over-dispersed if the model's total error is greater than its degrees of freedom, and problematically so if total error

**Figure D.8: Excess risk varies substantially within states**  
Excess risk by hospital across all multiday non-obstetric admissions



Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.

---

198. Collett (2002).

exceeds the model's degrees of freedom by more than a factor of 5.<sup>199</sup>  
We estimate the dispersion of each of our models,  $\phi$ , as follows:

$$\hat{\phi} = \frac{\text{Pearson } \chi^2 \text{ statistic}}{N - P} = \frac{\sum_{d=1}^{10} \frac{(O_d - E_d)^2}{E_d}}{N - P}$$

Table D.5 on page 94 presents our estimated dispersion parameters by subsample. It shows that all of the models exceed the assumed rate of dispersion of one, but not by much except for the non-obstetric multiday sample. Such low dispersion is a positive indication of models' fit, as over-dispersion is very much the norm in practice. It is likely to be attributable to the realistic way our multilevel model structure accounts for independence between outcomes.<sup>200</sup>

The over-dispersion of our model on the non-obstetric multiday subsample is likely to be attributable to the substantial heterogeneity of the patients included in this sample, which limits the accuracy with which it's possible for us to adjust for patients' risk profiles. We draw this conclusion because other common causes of over-dispersion in logit models, such as an incorrect choice of link function, influential outliers or an exceedingly rare dependent variable, are more true of the subsamples which do not exhibit this shortcoming.<sup>201</sup>

The over-dispersion may bias the standard error estimates of this model, so we refrain from drawing conclusions from this sample that are not also supported in our case studies. However, we note that this model still performed favourably on measures of goodness of fit.

Figures D.12 and D.13 on page 95 and on page 96 demonstrate that our additional assumption regarding the normality of the random effects

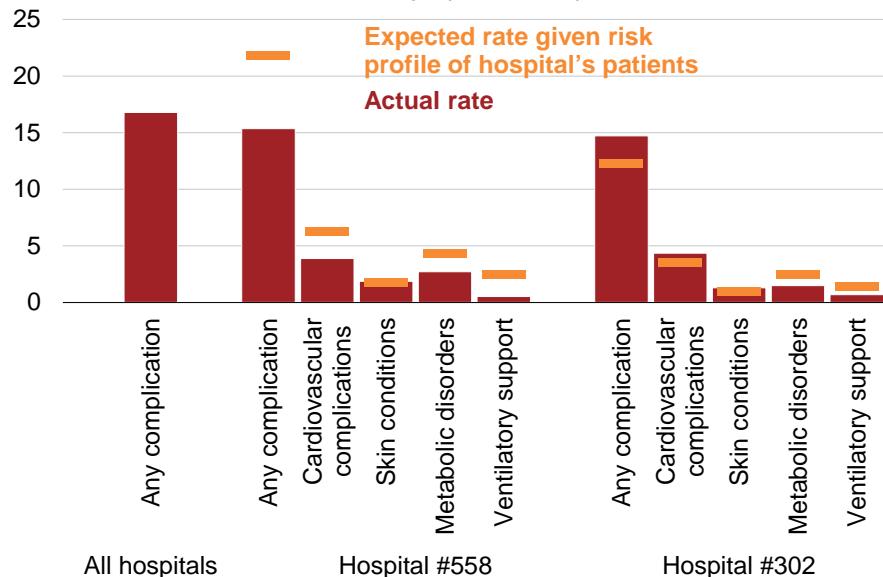
199. Carruthers et al. (2008).

200. Rabe-Hesketh and Skrondal (2008).

201. Collett (2002).

**Figure D.9: Excess risk estimates explain the difference between hospitals' expected and observed complication rates**

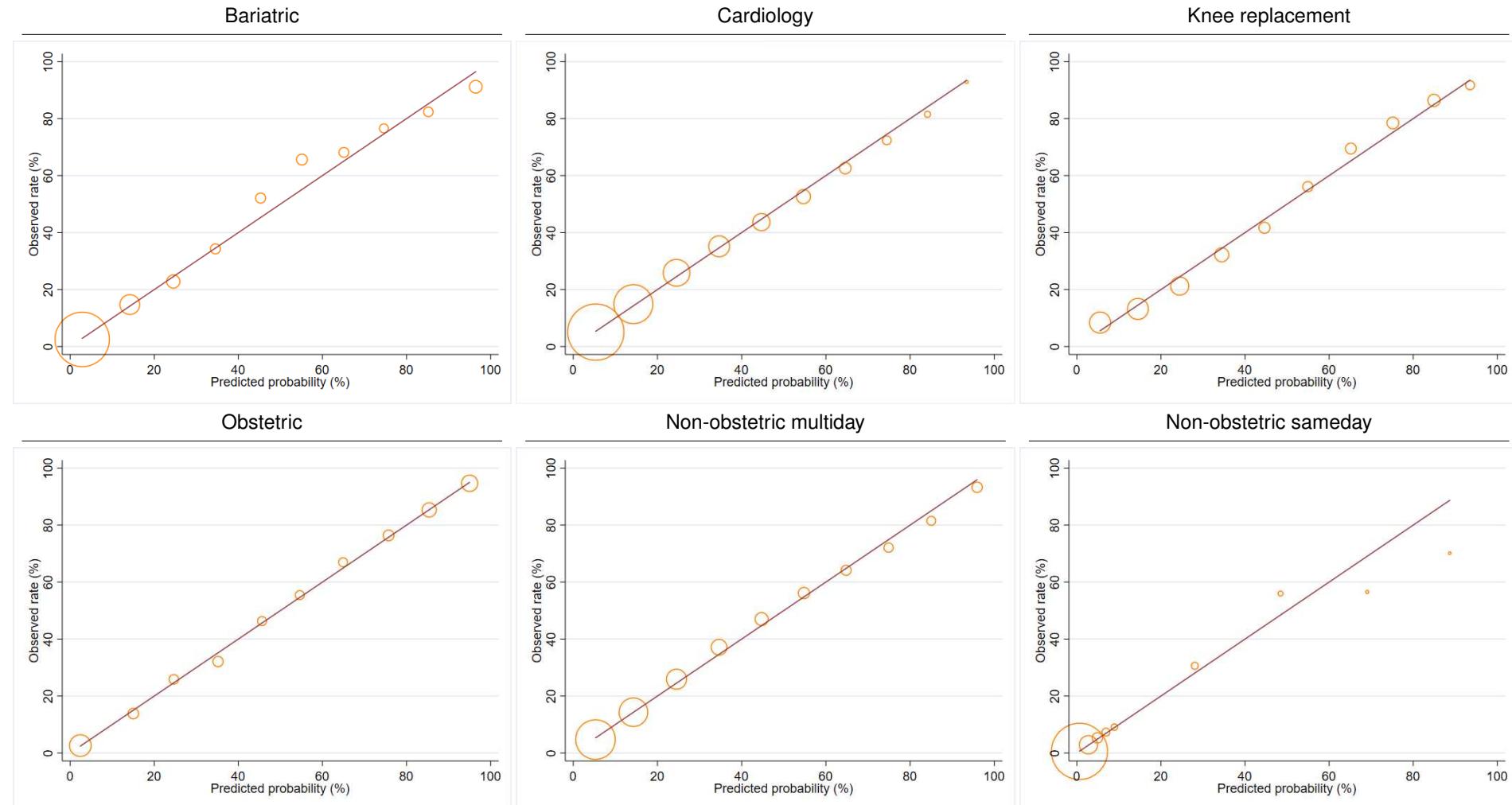
Share of admissions involving at least one complication (actual rate) relative to expected rate given the risk profile of hospital's patients, multiday cardiology admissions that do not involve a major procedure, per cent



Source: Grattan analysis of the 2012–15 National Hospital Morbidity Dataset.

**Figure D.10: Model predictions accord with observed rates across all subsamples**

Calibration plots of model predictions and observed outcomes by subsample, deciles of predicted probabilities scaled by the number of observations

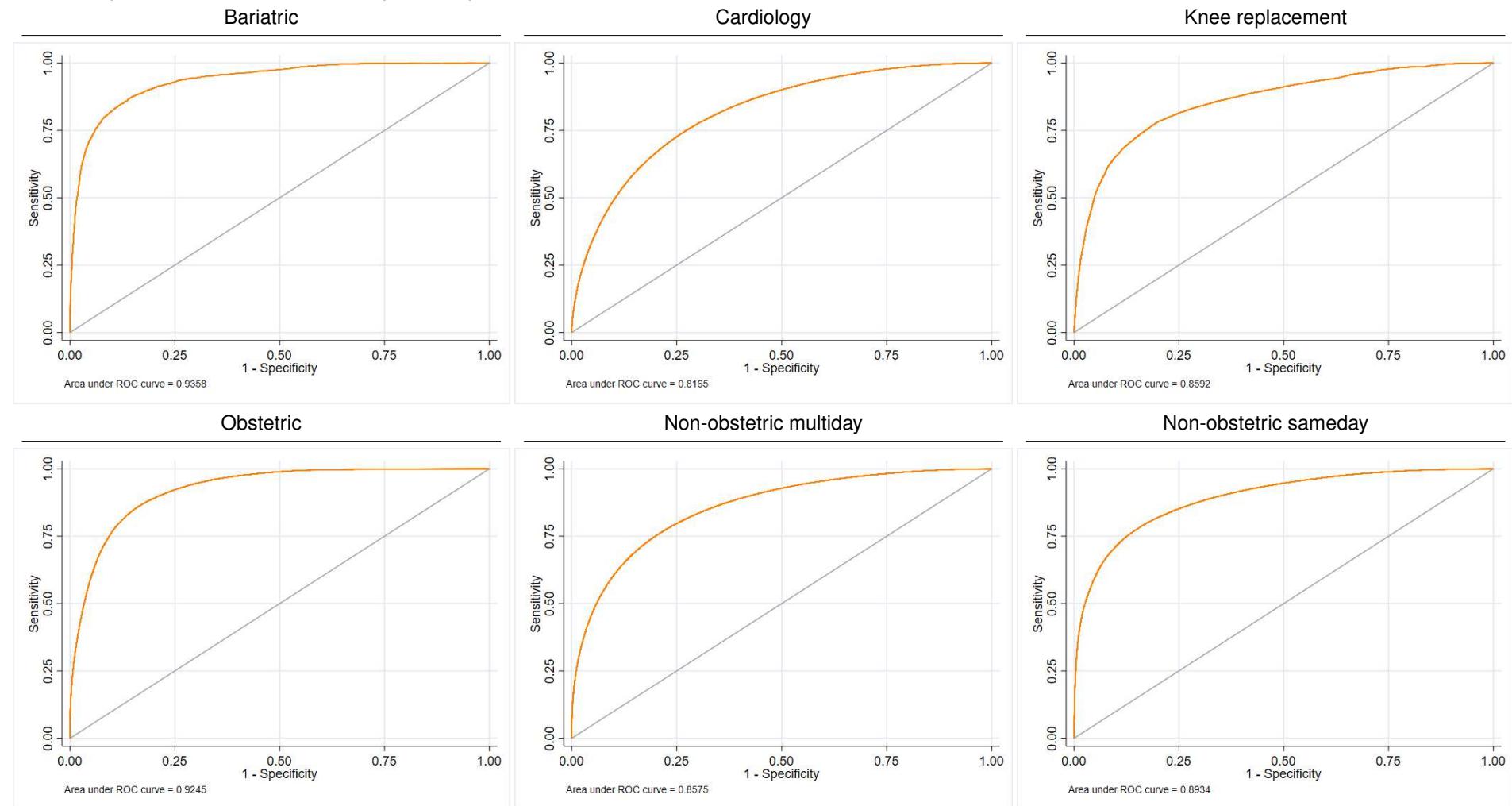


*Note: Calibration plot for the sameday non-obstetric subsample focuses on the first decile because almost all observations fell within that range.*

*Source: Grattan analysis of National Hospital Morbidity Dataset.*

**Figure D.11: All subsamples demonstrate satisfactory specificity and sensitivity**

Receiver Operator Characteristic curves by subsample



Source: Grattan analysis of National Hospital Morbidity Dataset.

terms is also reasonably well supported. These estimates of the distribution of hospital performance by subsample are not perfectly normal, but we note model estimates have been found to be robust to modest misspecification of random effects' distributions.<sup>202</sup>

### D.3 Other supporting analysis

Not all of the analysis contained in *All complications should count: Using our data to make hospitals safer* is based on the outcomes research models. In this section, we describe the methodologies for supporting pieces of analysis.

#### D.3.1 Complications and length of stay

One metric with which to measure the impact of a complication on patients' wellbeing is number of extra days in hospital that is associated with it, on average. However, the impact of complications on length of stay is difficult to quantify because length of stay generally also affects the likelihood of most complications occurring.

Conscious of this obstacle, we chose to focus our inquiry into complications' impact on length of stay on post-procedural complications. Post-procedural complications are not endogenous with length of stay in the same way as adverse drug events because the risk of these complications increases with the number of procedures a patient undergoes, rather than the days in hospital. This makes patients who undergo a single, standardised operation, like knee replacements patients, an appropriate focus for this type of analysis.

Figure 1.2 of *All complications should count: Using our data to make hospitals safer* presents the differences in average length of stay for patients who do and do not experience procedural complications, by

202. Neuhaus and McCulloch (2011).

Table D.3: Scope for improvement in complication rates

	Average rate of complications	Max rate among the top quartile	Max rate among the top decile
Full sample	10.67%	8.47%	7.73%
Obstetric	46.17%	43.47%	41.22%
Nonobstetric, multiday	22.26%	16.94%	15.40%
Nonobstetric, sameday	2.12%	1.34%	1.10%
Multiday bariatric surgery	15.49%	12.56%	11.52%
Multiday knee replacement	33.96%	26.21%	23.36%
Multiday cardiology			
– CHADx+	16.79%	12.31%	10.35%
– HACs	4.58%	3.21%	2.69%

Note: As fewer complications is better, the "top" decile and quartile of hospitals are those with risk-adjusted complication rates in the lowest decile or quartile.

Table D.4: Goodness of fit statistics

	R <sup>2</sup> <sub>MZ</sub>	Classification accuracy	AUC
Cardiology	27%	73.32%	81.65%
Knee replacement	40%	79.29%	85.92%
Bariatric	61%	85.56%	93.58%
Non-obstetric multiday	41%	77.41%	85.75%
Non-obstetric sameday	33%	81.88%	89.34%
Obstetric	72%	84.76%	92.45%

sample of admissions. Table D.6 in Box 9 on page 97 presents the conclusions of our formal hypothesis tests that length of stay differs among patients on the basis of whether they experienced a post-procedural complication, for each sample.

### D.3.2 Hospital performance aggregated to minor CHADx+

Making up the Classification of Hospital Acquired Diagnoses are 17 major classes of complications, and 159 complications. *All complications should count: Using our data to make hospitals safer* recommends that risk-adjusted data on each complication and each major CHADx+ class is provided to hospitals, as well as risk adjusted data on their overall prevalence of any complication.

In Chapter 4 of *All complications should count: Using our data to make hospitals safer*, we provide an example of how each individual hospital's data could be reported to them as a risk-adjusted heat map. The first column of the chart is colour coded by the hospital's quintile of performance by each major CHADx+ class, and the subsequent columns indicate the hospital's quintile of performance in each of the specific complications that make up each major CHADx+ class. This analysis was completed on multiday non-obstetric multiday admissions.

To obtain these risk-adjusted estimates of hospitals' relative performance on specific complications, we estimate additional outcomes research models: one for each complication and for each of the 17 major classes of complications.

Without sufficient time or computing resources to fit each of these models using a random effects, or even fixed effects, model specification, we used a pooled effects model specification for this segment of the analysis. This is similar to the approach employed in IHPA's risk adjustment model for Hospital Acquired Complications.<sup>203</sup>

---

203. IHPA (2017).

### Box 8: Model assumptions

#### No multicollinearity:

1. The independent variables are not collinear;

#### No endogeneity:

2. The residual term is not correlated with the independent variables;
3. The random effects terms are not correlated with the residual term;

#### Distributional assumptions hold:

4. The residual term is not over-dispersed;
5. The random effects estimates are approximately normally distributed.

Table D.5: Estimated dispersion by model

	Estimated dispersion parameter	Is this model over-dispersed?
Non-obstetric MD	14.22	Yes
Non-obstetric SD	1.26	No
Obstetric	1.22	No
Cardiology	1.01	No
Knee replacement	1.29	No
Bariatric	2.10	No

To expedite this exercise, we use the patient risk term previously estimated in relation to the prevalence of any complication. We combine this risk factor with the hospital-level risk factors in a logit risk adjustment model:

$$\log \left( \frac{p_i}{1 - p_i} \right) = \beta_1 X_{i,1} + \dots + \beta_K X_{i,K} + \varepsilon_i$$

Where:

$p_i$  is the probability that patient  $i$  experiences a given complication;

$X_{i,1}, \dots, X_{i,K}$  are the risk factors included in the model, evaluated relative to patient  $i$ ;

And the residual term is normally distributed and homoscedastic:  $\varepsilon_i \sim N(0, \sigma^2)$ .

From this model, we predict the expected log-odds ratio,  $r_i$ , of each complication for the average patient of each hospital. We convert these to the probability that the average patient from each hospital,  $h$ , experiences a complication using the transformation:

$$\hat{p}_h = \frac{\exp(\hat{r}_h)}{1 + \exp(\hat{r}_h)}$$

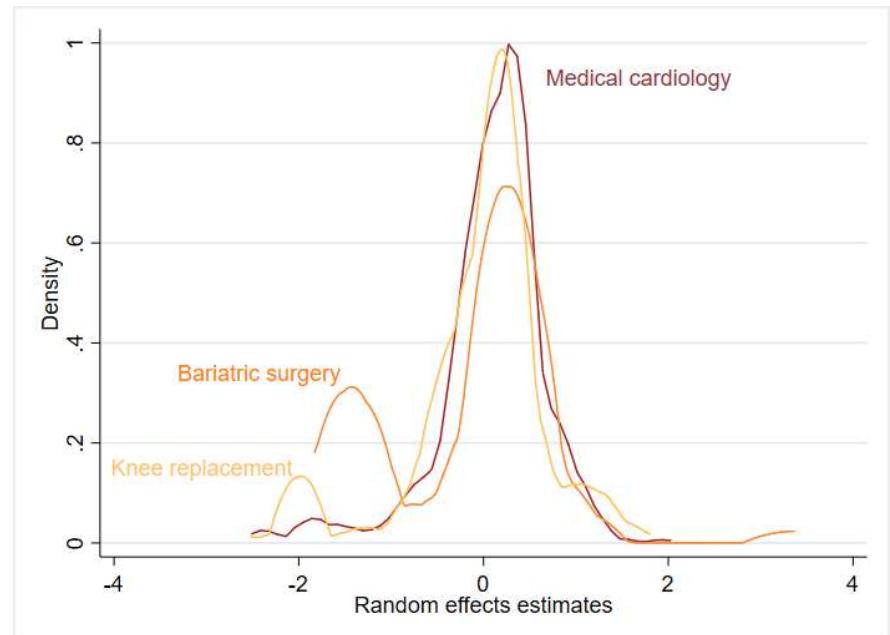
The probability that a hospital's average patient experiences a given complication can be interpreted as the expected rate of that complication for that hospital. Accordingly, we compute the observed to expected ratio of complications used to measure hospital performance from pooled specifications of outcomes research models using  $\hat{p}_i$ :

$$\text{Hospital performance}_{c,h} = \frac{\sum_{i \in h} \text{CHADx}_c / n_h}{\hat{p}_h} = \frac{\text{Observed rate}_{c,h}}{\text{Expected rate}_{c,h}}$$

Where:

**Figure D.12: Hospital performance is approximately normally distributed across our case studies**

Kernel densities of hospital random effects terms by subsample



Notes: Densities have been estimated using an Epanechnikov kernel with bandwidths ranging from 0.1 to 0.2 on the raw random effects terms, depending on the series' volatility.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity dataset.

$c$  indicates the complication of interest

$h$  identifies any particular hospital

$n_h$  is the number of admissions within hospital  $h$

Using this methodology, risk adjusted estimates of hospitals' performance in each complication and major class of complications were obtained. The exemplar for our proposed reporting scheme was then colour coded on the basis of which quintile of performance hospital 1 was classified in for each complication and major CHADx+ class.

#### D.4 Stability analysis

In Appendix C.4.1, we discussed the importance of ensuring performance metrics are representative of hospitals' usual performance, rather than being prone to mean reversion. We have taken several precautions in our analysis to ensure this is the case.

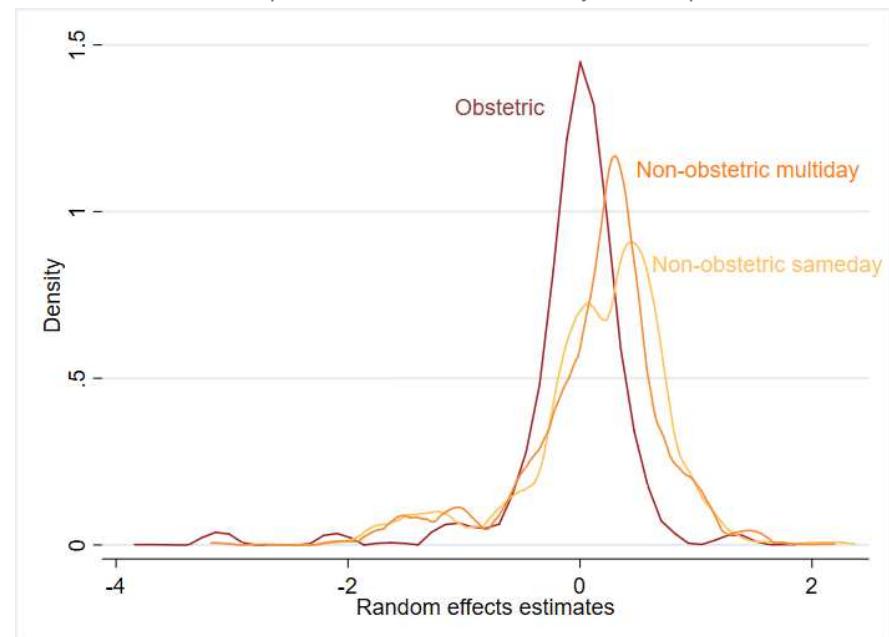
However, *All complications should count: Using our data to make hospitals safer* also recommends that hospital performance is reported on regularly and granularly. This raises the more general question of which indicators are sufficiently stable to be useful measures of hospitals' performances, and under what circumstances.

There are four factors that affect the stability of a performance metric. Fundamentally, performance metrics are more stable when random chance plays a smaller role in determining its incidence, and when the underlying event is more common.<sup>204</sup> Metrics are also more stable when they're measured across bigger institutions, or over longer periods of time. Whether a metric is stable *enough* depends on the level of detail and reliability required of the estimate.

---

204. That is, stability increases alongside a metric's coefficient of dispersion.

**Figure D.13: Hospital performance is approximately normally distributed across our full sample**  
Kernel densities of hospital random effects terms by subsample



Notes: Densities have been estimated using an Epanechnikov kernel with bandwidths ranging from 0.1 to 0.2 on the raw random effects terms, depending on the series' volatility.

Source: Grattan analysis of the 2012-15 National Hospital Morbidity dataset.

For the purpose of exploring these relationships, we investigate whether particular performance metrics are stable enough to reliably identify a hospital's decile of performance.<sup>205</sup> To do this, we first calculate the average amount that a hospitals' metric varies across consecutive periods, and then scale it by the average width of that metric's deciles:

$$\text{Stability}_M = \frac{1}{H(T-1)} \sum_{h=1}^H \left( \sum_{t=2}^T (M_{h,t} - M_{h,t-1}) \right)$$

$$\text{Average decile width}_M = \frac{1}{9} \sum_{d=10}^{100} (Q_{M,d} - Q_{M,d-1})$$

$$\text{Scaled stability}_M = \frac{\text{Stability}_M}{\text{Average decile width}_M}$$

Where:

$M$  is any given metric, which takes the value  $M_{h,t}$  for hospital  $h$  in period  $t$ ;

$Q_{M,d}$  refers to the  $d^{\text{th}}$  quantile of metric  $M$

$h$  is any of the  $H$  hospitals

$t$  is any of the  $T$  time periods

$d$  refers to any number in the series: 10, 20, ..., 80, 90, 100.

We identify the minimum sample size for which scaled stability $_M \leq 1$  by splitting our data into seven categories by hospital size, estimating scaled stability $_M$  for each of these subsamples and fitting a line of best with the following functional form fit to these seven data points:

205. This is an arbitrary threshold and could readily be tailored to a particular policy objective.

### Box 9: Conclusions of the hypothesis test

Table D.6: Procedural complications increase length of stay

	Marginal effect	P-value
Non-obstetric multiday	4.93	0.000
Multiday knee replacement	2.10	0.000
Multiday bariatric surgery	4.34	0.000

These hypothesis tests relating to the marginal effect of procedural complications on length of stay are derived from models on each subsample that are specified as follows:

$$\text{length of stay} = \beta_0 + \beta_1 \text{MCHADx1} + \beta_2 \text{Age} + \beta_3 \text{Sex} + \varepsilon$$

The marginal effect of incurring at least one procedural complication is given by  $\beta_1$  in the model specification.

Reported p-value's relate to the hypothesis test of whether  $H_0 : \beta_1 = 0$  or  $H_1 : \beta_1 \neq 0$ .

$$\text{Scaled stability}_M = b * \ln(\text{hospital size}) + c$$

This simple model specification explains about 85 per cent of variation in the scaled stability of each of the indicators assessed.<sup>206</sup> We infer the minimum hospital size for which we can expect a metric's stability to be less than or equal to with the width of a decile as follows:

$$(\text{Hospital size} | \text{Scaled stability}_M = 1) = \exp\left(\frac{1 - \hat{c}}{\hat{b}}\right)$$

Where  $b$  and  $c$  are the parameters of the model, and are estimated from the data to take the values  $\hat{b}$  and  $\hat{c}$ .

We find that overall complication rates are actually quite stable within hospitals. For example, a single performance decile spans a four per cent range of complication rates, on average, and complication rates at hospitals with at least 600 admissions a year tend to vary across years by less than this amount.<sup>207</sup> This implies that, when hospitals are compared on the basis of their overall complication rates, it's likely that they'll be ranked in the same performance decile in each consecutive year – even if they are a relatively small hospital.<sup>208</sup>

Figure D.14 illustrates how this stability changes when performance is measured over a shorter period, or when only a particular group of complications are considered. As expected, performance metrics are

206. Assessing three variables (CHADx+, MCHADx5, HACs) across four time periods (monthly, quarterly, six-monthly, annually), we estimated this relationship across 12 subsamples.

207. Of course, extreme deciles are wider and moderate performance deciles are narrower than this average range.

208. More precisely, they're likely to be *within a decile* of their previous position, which includes the scenario where a hospital moves from the upper bounds of one decile to the lower bounds of the above decile.

less stable when measured over shorter periods of time. Still, overall complication rates are stable enough to identify a hospital's precise decile of performance for hospitals with 3000 or more admissions per year.

Importantly, we also observe that HACs and cardiology complications are substantially less stable measures than the incidence of complications overall. This is in line with our expectations: affecting approximately 2 per cent of admissions, these subsets of complication affect only a fifth of patients that experience at any complication. Consequently, their incidence will be more volatile.

Figure D.14 on the next page usefully clarifies two characteristics of this relationship. Firstly, which outcome is measured has a greater impact on the stability of the outcome metric than the length of time over which it is measured. Secondly, the prevalence of a particular event affects – but does not determine – a metric's stability. We note that HACs and cardiology complications are equally prevalent, yet hospitals' rates of cardiology complications are more stable.<sup>209</sup>

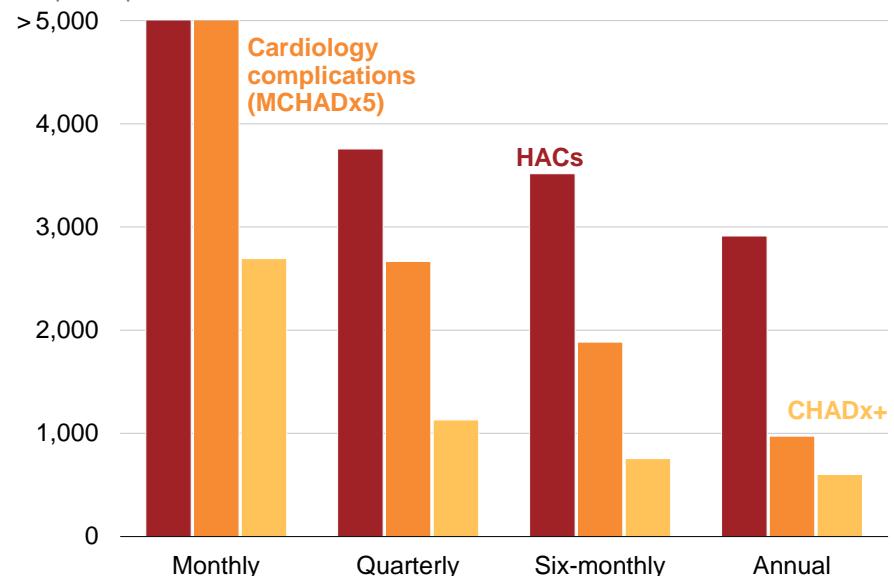
In total, this analysis provides useful assurance that there is enough information amongst the noise to distinguish hospitals' performances with reasonable precision, even over short time periods or using relatively rare events.<sup>210</sup> While the stability of metrics is not self-evident, it is readily assessable – especially relative to decision criteria. Like analysis should be repeated to determine whether particular metrics are stable *enough* for a given policy objective, for the hospital size and measurement window of interest.

209. This does not imply one measure is necessarily superior, just that the stability of particular metrics cannot be inferred from their prevalence.

210. We emphasise that this analysis uses raw prevalence rates. Extending this analysis to risk adjusted complication rates would be a worthwhile endeavour.

**Figure D.14: Some indicators are much more stable than others**

Minimum hospital size required to be able to reliably identify the decile of a hospital's performance



*Notes: We consider an indicator to be able to reliably identify the decile of a hospital's performance over a particular sample if the average change in hospitals' rates between periods is less than the average range of a performance decile.*

*Source: Grattan analysis of the 2012-15 National Hospital Morbidity Dataset.*

## Appendix E: Additional details regarding proof-of-concept app

In *All complications should count* we refer to an app that Grattan Institute has developed as a proof-of-concept. The app presents the average complication rates for elective surgical procedures by age, sex, specialty and the length of the patient's stay, using admissions data from 2012 to 2015. Results produced by the app should be taken as indicative only.

No personal information is disclosed by this app. Only subgroups with more than 20 admissions in each of the years were used, only counts and averages are used for each subgroup, and the original data was perturbed before the tables were prepared and uploaded.

## Bibliography

- ABS (2013). *Socio-Economic Indexes for Areas (SEIFA), Australia, 2011*. Cat. 2033.0.55.001. Australian Bureau of Statistics.  
<http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001>.
- ACSQHC (2016). *Hospital-acquired complications*. Version 1.1. Originally available at <https://safetyandquality.gov.au/wp-content/uploads/2016/08/Hospital-acquired-complications-specifications-V1.1-Oct-2016.xlsx>. Australian Commission on Safety and Quality of Hospital Care.
- (2017). *Hospital-acquired complications*. Australian Commission on Safety and Quality of Hospital Care.  
<https://www.safetyandquality.gov.au/our-work/indicators/hospital-acquired-complications/>.
- Agoritsas et al. (2015). Agoritsas, T. et al. "Decision aids that really promote shared decision making: the pace quickens". *BMJ* 350.feb10 14, g7624–g7624. doi: 10.1136/bmj.g7624.  
<https://doi.org/10.1136/bmj.g7624>.
- AIHW (2016). *Admitted patient care 2014–15: Australian hospital statistics*. Australian Institute of Health and Welfare.  
<http://www.aihw.gov.au/publication-detail/?id=60129554702&tab=3>.
- Alderwick et al. (2017). Alderwick, H., Charles, A., Jones, B. and Warburton, W. *Making the case for quality improvement: lessons for NHS boards and leaders*. Kings Fund and The Health Foundation.  
<https://www.kingsfund.org.uk/publications/making-case-quality-improvement>.
- Alemayehu et al. (2017). Alemayehu, D., Cappelleri, J., Emir, B. and Zou, K. *Statistical Topics in Health Economics and Outcomes Research*. Ed. by B. Emir. Chapman and Hall. ISBN: 9781498781879.
- Alexander et al. (2017). Alexander, M., Evans, S. M., Wolfe, R., Ball, D. L. and Burbury, K. "Risks of using medical record and administrative data for prognostic models". *The Medical journal of Australia* 207.3, p. 126. ISSN: 1326-5377. PMID: 28764629.
- Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: an empiricist's companion*. Princeton University Press. ISBN: 9780691120355.
- Ash et al. (2012). Ash, A., Fienberg, S., Louis, T., Normand, S.-L., Stukel, T. and Utts, J. *Statistical issues in assessing hospital performance*. Committee of Presidents of Statistical Societies.  
<https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf>.
- Bafumi, J. and Gelman, A. (2006). *Fitting Multilevel Models When Predictors and Group Effects Correlate*. [http://www.stat.columbia.edu/~gelman/research/unpublished/Bafumi\\_Gelman\\_Midwest06.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/Bafumi_Gelman_Midwest06.pdf).
- Baines et al. (2015). Baines, R., Langelaan, M., Bruijne, M. de, Spreeuwenberg, P. and Wagner, C. "How effective are patient safety initiatives? A retrospective patient record review study of changes to patient safety over time". *BMJ quality & safety* 24.9, pp. 561–71. ISSN: 2044-5423. doi: 10.1136/bmjqqs-2014-003702. PMID: 26150548.
- Baxter et al. (2015). Baxter, R., Taylor, N., Kellar, I. and Lawton, R. "What methods are used to apply positive deviance within healthcare organisations? A systematic review". *BMJ Quality & Safety*. ISSN: 2044-5415. doi: 10.1136/bmjqqs-2015-004386. eprint: <http://qualitysafety.bmjjournals.com/content/early/2015/11/20/bmjqqs-2015-004386.full.pdf>.  
<http://qualitysafety.bmjjournals.com/content/early/2015/11/20/bmjqqs-2015-004386>.

- Behrendt, K. and Groene, O. (2016). "Mechanisms and effects of public reporting of surgeon outcomes: A systematic review of the literature". *Health policy (Amsterdam, Netherlands)* 120.10, pp. 1151–1161. ISSN: 1872-6054. DOI: 10.1016/j.healthpol.2016.08.003. eprint: 27638232.
- Bell, A. and Jones, K. (2015). "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data". *Political Science Research and Methods* 3.1. DOI: 10.1017/psrm.2014.7.
- Berman, S. (1998). "Identifying and Addressing Sentinel Events: An Interview with Richard Croteau". *The Joint Commission Journal on Quality Improvement* 24.8, pp. 426–434. DOI: 10.1016/s1070-3241(16)30393-5. [https://doi.org/10.1016/s1070-3241\(16\)30393-5](https://doi.org/10.1016/s1070-3241(16)30393-5).
- Berndorfer, S. and Henriksson, A. (2017). "Automated Diagnosis Coding with Combined Text Representations". *Studies in health technology and informatics* 235, pp. 201–205. ISSN: 0926-9630. eprint: 28423783.
- Berwick et al. (2008). Berwick, D., Nolan, T. and Whittington, J. "The Triple Aim: Care, Health, And Cost". *Health Affairs* 27.3, pp. 759–769. DOI: 10.1377/hlthaff.27.3.759. <https://doi.org/10.1377/hlthaff.27.3.759>.
- Besnard, D. and Hollnagel, E. (2014). "I want to believe: some myths about the management of industrial safety". *Cognition, Technology & Work* 16.1, pp. 13–23.
- Blake, R. and Clarke, H. (2017). "Hospital Compare and Hospital Choice". *Medical Care Research and Review*. DOI: 10.1177/1077558717699311. <https://doi.org/10.1177/1077558717699311>.
- Böhning et al. (2008). Böhning, D., Böhning, W. and Holling, H. "Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies". *Statistical methods in medical research* 17.6, pp. 543–54. ISSN: 0962-2802. DOI: 10.1177/0962280207081867.
- Boss et al. (2016). Boss, E., Mehta, N., Nagarajan, N., Links, A., Benke, J., Berger, Z., Espinel, A., Meier, J. and Lipstein, E. "Shared Decision Making and Choice for Elective Surgical Care: A Systematic Review". *Otolaryngology—head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 154.3, pp. 405–20. ISSN: 1097-6817. DOI: 10.1177/0194599815620558. eprint: 26645531.
- Brown et al. (2015). Brown, J., Joyce, K., Stacey, D. and Thomson, R. G. "Patients or Volunteers? The Impact of Motivation for Trial Participation on the Efficacy of Patient Decision Aids: A Secondary Analysis of a Cochrane Systematic Review". *Medical Decision Making* 35.4, pp. 419–435. DOI: 10.1177/0272989x15579172. <https://doi.org/10.1177/0272989x15579172>.
- Burgess et al. (2017). Burgess, K., Christian, L., McIntyre, J. and Mole, T. *Admitted patient care 2015-16: Australian hospital statistics*. Australian Institute of Health and Welfare. <https://www.aihw.gov.au/reports/hospitals/ahs-2015-16-admitted-patient-care/contents/table-of-contents>.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeometrics: Methods and Applications*. ISBN: 978-0-511-12581-2. <http://www.centroportici.unina.it/centro/Cameron&Trivedi.pdf>.
- Carey, K. (2017). "Speech". *SPHERE 2017 International Symposium*.
- Carman et al. (2016). Carman, K. et al. "Understanding An Informed Publics Views On The Role Of Evidence In Making Health Care Decisions". *Health Affairs* 35.4, pp. 566–574. DOI: 10.1377/hlthaff.2015.1112. <https://doi.org/10.1377/hlthaff.2015.1112>.
- Carruthers et al. (2008). Carruthers, A. et al. "A Validated Grading Scale for Marionette Lines". *Dermatologic Surgery* 34, S167–S172. DOI: 10.1111/j.1524-4725.2008.34366.x. <https://doi.org/10.1111/j.1524-4725.2008.34366.x>.

- Chandra et al. (2016). Chandra, A., Finkelstein, A., Sacarny, A. and Syverson, C. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector". *American Economic Review* 106.8, pp. 2110–2144. DOI: 10.1257/aer.20151080.  
<https://doi.org/10.1257/aer.20151080>.
- Chassin, M. (2002). "Achieving And Sustaining Improved Quality: Lessons From New York State And Cardiac Surgery". *Health Affairs* 21.4, pp. 40–51. DOI: 10.1377/hlthaff.21.4.40.  
<https://doi.org/10.1377/hlthaff.21.4.40>.
- Chen et al. (2012). Chen, L. M., Orav, E. J. and Epstein, A. M. "Public reporting on risk-adjusted mortality after percutaneous coronary interventions in New York State: forecasting ability and impact on market share and physicians' decisions to discontinue practice". *Circulation. Cardiovascular quality and outcomes* 5.1, pp. 70–5. ISSN: 1941-7705. DOI: 10.1161/CIRCOUTCOMES.111.962761. PMID: 22235066.
- COAG (2011). *National Health Reform Agreement*. Council of Australian Governments. [http://www.federalfinancialrelations.gov.au/content/npa/health\\_reform/national-agreement.pdf](http://www.federalfinancialrelations.gov.au/content/npa/health_reform/national-agreement.pdf).
- \_\_\_\_\_. (2017). *Schedule I – Addendum to the National Health Reform Agreement: Revised Public Hospital Arrangements*. Council of Australian Governments.  
[http://www.federalfinancialrelations.gov.au/content/npa/health/other/Addendum\\_to\\_the\\_National\\_Health\\_Reform.pdf](http://www.federalfinancialrelations.gov.au/content/npa/health/other/Addendum_to_the_National_Health_Reform.pdf).
- Collett, D. (2002). *Modelling Binary Data, Second Edition*. Chapman and Hall. ISBN: 9781584883241.
- Cook, R. I. and Nemeth, C. P. (2010). "Those found responsible have been sacked": some observations on the usefulness of error". *Cognition, Technology & Work* 12.2, pp. 87–93. ISSN: 1435-5566. DOI: 10.1007/s10111-010-0149-0.  
<https://doi.org/10.1007/s10111-010-0149-0>.
- Copnell et al. (2009). Copnell, B., Hagger, V., Wilson, S. G., Evans, S. M., Sprivulis, P. C. and Cameron, P. A. "Measuring the quality of hospital care: an inventory of indicators". *Internal medicine journal* 39.6, pp. 352–60. ISSN: 1445-5994. DOI: 10.1111/j.1445-5994.2009.01961.x. eprint: 19323697.
- Courts, L. H. (2011). "Woman to sue after sponge left in her body". *Sydney Morning Herald*. <http://www.smh.com.au/nsw/woman-to-sue-after-sponge-left-in-her-body-20110731-1i6fv.html>.
- Damschroder et al. (2014). Damschroder, L. J., Robinson, C. H., Francis, J., Bentley, D. R., Krein, S. L., Rosland, A.-M., Hofer, T. P. and Kerr, E. A. "Effects of performance measure implementation on clinical manager and provider motivation". *Journal of general internal medicine* 29 Suppl 4, pp. 877–84. ISSN: 1525-1497. DOI: 10.1007/s11606-014-3020-9. eprint: 2523454.
- Dekker, S. W. (2014). "The bureaucratization of safety". *Safety Science* 70. Supplement C, pp. 348–357. ISSN: 0925-7535. DOI: <https://doi.org/10.1016/j.ssci.2014.07.015>. <http://www.sciencedirect.com/science/article/pii/S092575351400174X>.
- Dieleman, J. and Templin, T. (2016). "Correction: Random-Effects, Fixed-Effects and the within-between Specification for Clustered Data in Observational Health Studies: A Simulation Study". *PLOS ONE* 11.5. DOI: 10.1371/journal.pone.0156508.  
<https://doi.org/10.1371/journal.pone.0156508>.
- Dixon-Woods et al. (2012). Dixon-Woods, M., McNicol, S. and Martin, G. "Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature". *BMJ quality & safety* 21.10, pp. 876–84. ISSN: 2044-5423. DOI: 10.1136/bmjqqs-2011-000760. eprint: 22543475.
- Dixon-Woods, M. and Martin, G. (2016). "Does quality improvement improve quality?" *Future Healthcare Journal* 3.3. DOI: 10.7861/futurehosp.3-3-191.  
<http://futurehospital.rcpjournals.org/content/3/3/191.full>.

- Dixon-Woods, M. and Pronovost, P. (2016). "Patient safety and the problem of many hands". *BMJ quality & safety* 25.7, pp. 485–8. ISSN: 2044-5423. DOI: 10.1136/bmjqqs-2016-005232. eprint: 26912578.
- Duckett, S. (1998). "Casemix funding for acute hospital inpatient services in Australia". *Medical journal of Australia* 169.
- \_\_\_\_\_. (2017). "Cutting funding for hospital complications is unlikely to change patient care – here's why". *The Conversation*. <https://theconversation.com/cutting-funding-for-hospital-complications-is-unlikely-to-change-patient-care-heres-why-88945>.
- Duckett et al. (2007). Duckett, S., Coory, M. and Sketcher-Baker, K. "Identifying variations in quality of care in Queensland hospitals". *The Medical journal of Australia* 187.10, pp. 571–5. ISSN: 0025-729X. eprint: 18021046.
- Duckett, S. and Ward, M. (2008). "Developing 'robust performance benchmarks' for the next Australian Health Care Agreement: the need for a new framework". *Australia and New Zealand Health Policy* 5.1, p. 1. DOI: 10.1186/1743-8462-5-1. <https://doi.org/10.1186/1743-8462-5-1>.
- Duckett et al. (2015). Duckett, S., Breadon, P., Romanes, D., Fennessy, P. and Nolan, J. *Questionable care: Avoiding ineffective treatment*. 2015-7. Grattan Institute. <https://grattan.edu.au/report/questionable-care-avoiding-ineffective-treatment/>.
- Duckett et al. (2016). Duckett, S., Cuddihy, M. and Newnham, H. *Targeting zero: Supporting the Victorian hospital system to eliminate avoidable harm and strengthen quality of care*. Report of the Review of Hospital Safety and Quality Assurance in Victoria. Victorian Government. <https://www2.health.vic.gov.au/hospitals-and-health-services/quality-safety-service/hospital-safety-and-quality-review>.
- Duckett et al. (2017). Duckett, S., Jorm, C. and Danks, L. *Strengthening safety statistics: how to make hospital safety data more useful*. 2017-11. Grattan Institute. <https://grattan.edu.au/report/strengthening-safety-statistics/>.
- Dunn, O. J. (1959). "Estimation of the Medians for Dependent Variables". *The Annals of Mathematical Statistics* 30.1, pp. 192–197. DOI: 10.1214/aoms/1177706374. <https://doi.org/10.1214/aoms/1177706374>.
- Eager, K. (2010). *What is activity-based funding?* <https://ahsri.uow.edu.au/content/groups/public/@web/@chsd/documents/doc/uow082633.pdf>.
- Elwyn et al. (2008). Elwyn, G., Légaré, F., Weijden, T. van der, Edwards, A. and May, C. "Arduous implementation: Does the Normalisation Process Model explain why it's so difficult to embed decision support technologies for patients in routine clinical practice". *Implementation Science* 3.1. DOI: 10.1186/1748-5908-3-57. <https://doi.org/10.1186/1748-5908-3-57>.
- Findlay, S. (2016a). "Consumers' Interest In Provider Ratings Grows, And Improved Report Cards And Other Steps Could Accelerate Their Use". *Health affairs (Project Hope)* 35.4, pp. 688–96. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2015.1654. eprint: 27044970.
- Findlay, S. (2016b). "Consumers Interest In Provider Ratings Grows, And Improved Report Cards And Other Steps Could Accelerate Their Use". *Health Affairs* 35.4, pp. 688–696. DOI: 10.1377/hlthaff.2015.1654. <https://doi.org/10.1377/hlthaff.2015.1654>.
- Frølich et al. (2007). Frølich, A., Talavera, J. A., Broadhead, P. and Dudley, R. A. "A behavioral model of clinician responses to incentives to improve quality". *Health policy (Amsterdam, Netherlands)* 80.1, pp. 179–93. ISSN: 0168-8510. DOI: 10.1016/j.healthpol.2006.03.001. eprint: 16624440.
- Fung et al. (2008). Fung, C., Lim, Y.-W., Mattke, S., Damberg, C. and Shekelle, P. "Systematic review: the evidence that publishing patient care performance data improves quality of care". *Annals of internal medicine* 148.2, pp. 111–23. ISSN: 1539-3704.

- Gillam et al. (2012). Gillam, S., Siriwardena, N. and Steel, N. "Pay-for-Performance in the United Kingdom: Impact of the Quality and Outcomes Framework—A Systematic Review". *The Annals of Family Medicine* 10.5, pp. 461–468. DOI: 10.1370/afm.1377. <https://doi.org/10.1370/afm.1377>.
- Grassi et al. (2015). Grassi, L., Caraceni, A., Mitchell, A. J., Nanni, M. G., Berardi, M. A., Caruso, R. and Riba, M. "Management of delirium in palliative care: a review". *Current psychiatry reports* 17.3, p. 550. ISSN: 1535-1645. DOI: 10.1007/s11920-015-0550-8. eprint: 25663153.
- Greenhalgh et al. (2017). Greenhalgh, J., Dalkin, S., Gibbons, E., Wright, J., Valderas, J. M., Meads, D. and Black, N. "How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis". *Journal of Health Services Research & Policy* 23.1, pp. 57–65. DOI: 10.1177/1355819617740925. <https://doi.org/10.1177/1355819617740925>.
- Haan, P. and Uhlendorff, A. (2006). "Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood". *STATA journal* 6.2, pp. 229–245.
- Hayward, R. A. and Hofer, T. P. (2001). "Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer". *JAMA* 286.4, pp. 415–20. ISSN: 0098-7484. eprint: 11466119.
- Henderson et al. (2006). Henderson, T., Shepheard, J. and Sundararajan, V. "Quality of Diagnosis and Procedure Coding in ICD-10 Administrative Data". *Medical Care* 44.11, pp. 1011–1019. DOI: 10.1097/01.mlr.0000228018.48783.34. <https://doi.org/10.1097/01.mlr.0000228018.48783.34>.
- Hibbard et al. (2005). Hibbard, J., Stockard, J. and Tusler, M. "Hospital Performance Reports: Impact On Quality, Market Share, And Reputation". *Health Affairs* 24.4, pp. 1150–1160. DOI: 10.1377/hlthaff.24.4.1150. <https://doi.org/10.1377/hlthaff.24.4.1150>.
- Hoffmann, T. C. and Mar, C. D. (2017). "Clinicians' Expectations of the Benefits and Harms of Treatments, Screening, and Tests: A Systematic Review". *JAMA internal medicine* 177.3, pp. 407–419. ISSN: 2168-6114. DOI: 10.1001/jamainternmed.2016.8254. eprint: 28097303.
- Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management*. Ashgate. ISBN: 978-1472423054.
- Holman et al. (2005). Holman, D., Preen, D., Baynham, N., Finn, J. and Semmens, J. "A multipurpose comorbidity scoring system performed better than the Charlson index". *Journal of clinical epidemiology* 58.10, pp. 1006–14. ISSN: 0895-4356. DOI: 10.1016/j.jclinepi.2005.01.020. eprint: 16168346.
- Holzmueller et al. (2012). Holzmueller, C. G., Wu, A. W. and Pronovost, P. J. "A framework for encouraging patient engagement in medical decision making". *Journal of patient safety* 8.4, pp. 161–4. ISSN: 1549-8425. DOI: 10.1097/PTS.0b013e318267c56e. eprint: 22892584.
- Höög et al. (2016). Höög, E., Lysholm, J., Garvare, R., Weinehall, L. and Nyström, M. E. "Quality improvement in large healthcare organizations: Searching for system-wide and coherent monitoring and follow-up strategies". *Journal of Health Organization and Management* 30.1. PMID: 26964854, pp. 133–153. DOI: 10.1108/JHOM-10-2013-0209. <https://doi.org/10.1108/JHOM-10-2013-0209>.
- Hor et al. (2013). Hor, S.-y., Godbold, N., Collier, A. and Iedema, R. "Finding the patient in patient safety". *Health* 17.6, pp. 567–83. ISSN: 1461-7196. DOI: 10.1177/1363459312472082. eprint: 23349385.
- Hosie et al. (2013). Hosie, A., Davidson, P. M., Agar, M., Sanderson, C. R. and Phillips, J. "Delirium prevalence, incidence, and implications for screening in specialist palliative care inpatient settings: a systematic review". *Palliative medicine* 27.6, pp. 486–98. ISSN: 1477-030X. DOI: 10.1177/0269216312457214. eprint: 22988044.

- Iezzoni, L. (1996). "100 apples divided by 15 red herrings: a cautionary tale from the mid-19th century on comparing hospital mortality rates". *Annals of internal medicine* 124.12, pp. 1079–85. ISSN: 0003-4819. eprint: 8633823.
- \_\_\_\_\_. (1997). "The risks of risk adjustment". *JAMA* 278.19, pp. 1600–7. ISSN: 0098-7484. eprint: 9370507.
- \_\_\_\_\_. (2012). *Risk Adjustment for Measuring Health Care Outcomes*. 4th. Health Administration Press. ISBN: 978-1567934373.
- IHPA (2017). *Risk adjustment model for Hospital Acquired Complications – Technical Specifications*. Version 1.0. Independent Hospital Pricing Authority. [https://www.ihsa.gov.au/sites/g/files/net636/f/risk\\_adjustment\\_model\\_for\\_hospital\\_acquired\\_complications\\_-technical\\_specifications\\_v1.0\\_july\\_2017\\_pdf.pdf](https://www.ihsa.gov.au/sites/g/files/net636/f/risk_adjustment_model_for_hospital_acquired_complications_-technical_specifications_v1.0_july_2017_pdf.pdf).
- Institute of Medicine (2000). *To Err Is Human: Building a Safer Health System*. Ed. by L. Kohn, J. Corrigan and M. Donaldson. Washington, DC: The National Academies Press. ISBN: 978-0-309-26174-6. DOI: 10.17226/9728. <https://www.nap.edu/catalog/9728/to-err-is-human-building-a-safer-health-system>.
- Ivers et al. (2012). Ivers, N. et al. "Audit and feedback: effects on professional practice and healthcare outcomes". *Cochrane Database of Systematic Reviews*. DOI: 10.1002/14651858.cd000259.pub3. <https://doi.org/10.1002/14651858.cd000259.pub3>.
- Jackson et al. (2009). Jackson, T., Michel, J., Roberts, R., Jorm, C. and Wakefield, J. "A classification of hospital-acquired diagnoses for use with routine hospital data". *The Medical journal of Australia* 191.10, pp. 544–8. ISSN: 0025-729X. eprint: 19912086.
- Jacob, B. and Lefgren, L. (2005). *Principals as Agents: Subjective Performance Measurement in Education*. 11463. National Bureau of Economic Research. DOI: 10.3386/w11463. <http://www.nber.org/papers/w11463>.
- James, W. and Stein, C. (1961). "Estimation with Quadratic Loss". *Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. Berkeley: University of California Press.
- Jen et al. (2009). Jen, M. H., Bottle, A., Majeed, A., Bell, D. and Aylin, P. "Early in-hospital mortality following trainee doctors' first day at work". *PLoS one* 4.9, e7103. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0007103. eprint: 19774078.
- Jha et al. (2010). Jha, A., Prasopa-Plaizier, N., Larizgoitia, I. and Bates, D. W. "Patient safety research: an overview of the global evidence". *Quality and Safety in Health Care* 19.1, pp. 42–47. DOI: 10.1136/qshc.2008.029165. <https://doi.org/10.1136/qshc.2008.029165>.
- E. Jones et al. (2016). Jones, E., Lees, N., Martin, G. and Dixon-Woods, M. "How Well Is Quality Improvement Described in the Perioperative Care Literature? A Systematic Review". *The Joint Commission Journal on Quality and Patient Safety* 42.5, 196–AP10. DOI: 10.1016/s1553-7250(16)42025-8. [https://doi.org/10.1016/s1553-7250\(16\)42025-8](https://doi.org/10.1016/s1553-7250(16)42025-8).
- Jorm, C. (2017). *Clinical engagement - scoping paper*. Victorian Department of Health. <https://www2.health.vic.gov.au/about/publications/policiesandguidelines/clinical-engagement-scoping-paper>.
- Jorm et al. (2009). Jorm, C., Dunbar, N., Sudano, L. and Travaglia, J. "Should patient safety be more patient centred?" *Australian health review : a publication of the Australian Hospital Association* 33.3, pp. 390–9. ISSN: 0156-5788. eprint: 20128754.
- Kalbfleisch, J. and Wolfe, R. (2013). "On Monitoring Outcomes of Medical Providers". *Statistics in Biosciences* 5.2, pp. 286–302. DOI: 10.1007/s12561-013-9093-x. <https://doi.org/10.1007/s12561-013-9093-x>.
- Kellogg et al. (2017). Kellogg, K., Hettinger, Z., Shah, M., Wears, R. L., Sellers, C. R., Squires, M. and Fairbanks, R. J. "Our current approach to root cause analysis: is it contributing to our failure to improve patient safety?" *BMJ quality & safety* 26.5, pp. 381–387. ISSN: 2044-5423. DOI: 10.1136/bmjqqs-2016-005991. eprint: 27940638.

- Kizer, K. and Stegun, M. (2005). "Serious Reportable Adverse Events in Health Care". *Advances in Patient Safety: From Research to Implementation*. Vol. 4. <https://www.ncbi.nlm.nih.gov/books/NBK20598/>.
- Kobel, C. and Theurl, E. (2013). *Hospital specialisation within a DRG-Framework: The Austrian case*. 06. University of Innsbruck. <https://eeecon.uibk.ac.at/wopec2/repec/inn/wpaper/2013-06.pdf>.
- Kreindler, S. (2016). "What if implementation is not the problem? Exploring the missing links between knowledge and action". *The International journal of health planning and management* 31.2, pp. 208–26. ISSN: 1099-1751. DOI: 10.1002/hpm.2277. eprint: 25424863.
- Krumholz et al. (2006). Krumholz, H. et al. "Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group". *Circulation* 113.3, pp. 456–62. ISSN: 1524-4539. DOI: 10.1161/CIRCULATIONAHA.105.170769. eprint: 16365198.
- Lawton et al. (2014). Lawton, R., Taylor, N., Clay-Williams, R. and Braithwaite, J. "Positive deviance: a different approach to achieving patient safety". *BMJ quality & safety* 23.11, pp. 880–3. ISSN: 2044-5423. DOI: 10.1136/bmjqqs-2014-003115. eprint: 25049424.
- Leape, L. (2002). "Reporting of Adverse Events". *New England Journal of Medicine* 347.20, pp. 1633–1638. DOI: 10.1056/nejmnejmhpr011493. <https://doi.org/10.1056/nejmnejmhpr011493>.
- Leapfrog Group (2017). *Compare hospitals*. <http://www.leapfroggroup.org/compare-hospitals>.
- Légaré, F. and Thompson-Leduc, P. (2014). "Twelve myths about shared decision making". *Patient education and counseling* 96.3, pp. 281–6. ISSN: 1873-5134. DOI: 10.1016/j.pec.2014.06.014. eprint: 25034637.
- Leggat, S. and Balding, C. (2017). "A qualitative study on the implementation of quality systems in Australian hospitals". *Health services management research* 30.3, pp. 179–186. ISSN: 1758-1044. DOI: 10.1177/0951484817715594. eprint: 28695775.
- Levay, C. (2016). *Health Care Transparency in Organizational Perspective*. Ed. by E. Ferlie, K. Montgomery and A. R. Pedersen. DOI: 10.1093/oxfordhb/9780198705109.013.12. <https://doi.org/10.1093/oxfordhb/9780198705109.013.12>.
- Levesque, J.-F. and Sutherland, K. (2017). "What role does performance information play in securing improvement in healthcare? a conceptual framework for levers of change". *BMJ Open* 7.8. ISSN: 2044-6055. DOI: 10.1136/bmjopen-2016-014825. eprint: <http://bmjopen.bmjjournals.com/content/7/8/e014825.full.pdf>. <http://bmjopen.bmjjournals.com/content/7/8/e014825>.
- Mannion, R. and Braithwaite, J. (2012). "Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service". *Internal medicine journal* 42.5, pp. 569–74. ISSN: 1445-5994. DOI: 10.1111/j.1445-5994.2012.02766.x. eprint: 22616961.
- Marmot, M. (2015). *The Health Gap: The Challenge of an Unequal World*. Bloomsbury. ISBN: 9781408857991.
- Marx, D. (2001). *Patient Safety and the "Just Culture" A Primer for Health Care Executives*. Columbia University. [http://www.chphso.org/sites/main/files/file-attachments/marx\\_primer.pdf](http://www.chphso.org/sites/main/files/file-attachments/marx_primer.pdf).
- McKelvey, R. and Zavoina, W. (1975). "A statistical model for the analysis of ordinal level dependent variables". *The Journal of Mathematical Sociology* 4.1, pp. 103–120. DOI: 10.1080/0022250x.1975.9989847. <https://doi.org/10.1080/0022250x.1975.9989847>.
- McWhirter, R. (2017). "Performance data and informed consent: a duty to disclose?" *The Medical Journal of Australia* 207.3, pp. 100–101. DOI: 10.5694/mja16.01195. <https://doi.org/10.5694/mja16.01195>.
- Mehlman, M. (2013). "Professional Power and the Standard of Care in Medicine". *Arizona State Law Journal* 1165.44. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2205485](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205485).

- Mendelson et al. (2017). Mendelson, A., Kondo, K., Damberg, C., Low, A., Motúapuaka, M., Freeman, M., O'Neil, M., Relevo, R. and Kansagara, D. "The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care". *Annals of Internal Medicine* 166.5, p. 341. DOI: 10.7326/m16-1881.  
<https://doi.org/10.7326/m16-1881>.
- Mennemeyer et al. (1997). Mennemeyer, S. T., Morrisey, M. A. and Howard, L. Z. "Death and reputation: how consumers acted upon HCFA mortality information". *Inquiry : a journal of medical care organization, provision and financing* 34.2, pp. 117–28. ISSN: 0046-9580. PMID: 9256817.
- Merle et al. (2009). Merle, V. et al. "Influence of infection control report cards on patients' choice of hospital: pilot survey". *Journal of Hospital Infection* 71.3, pp. 263–268. DOI: 10.1016/j.jhin.2008.11.025.  
<https://doi.org/10.1016/j.jhin.2008.11.025>.
- Mery et al. (2017). Mery, G., Dobrow, M., Baker, G. R., Im, J. and Brown, A. "Evaluating investment in quality improvement capacity building: a systematic review". *BMJ open* 7.2, e012431. ISSN: 2044-6055. DOI: 10.1136/bmjopen-2016-012431. eprint: 28219957.
- Michel et al. (2009). Michel, J., Nghiem, H. D. and Jackson, T. "Using ICD-10-AM codes to characterise hospital-acquired complications". *The HIM journal* 38.3, pp. 18–25. ISSN: 1833-3575.
- Moppett, I. and Moppett, S. (2016). "Surgical caseload and the risk of surgical Never Events in England". *Anaesthesia* 71.1, pp. 17–30. ISSN: 1365-2044. DOI: 10.1111/anae.13290. eprint: 26597516.
- Moran, J. and Solomon, P. (2014). "Fixed Effects Modelling for Provider Mortality Outcomes: Analysis of the Australia and New Zealand Intensive Care Society (ANZICS) Adult Patient Data-Base". *PLoS ONE* 9.7, e102297. DOI: 10.1371/journal.pone.0102297.  
<https://doi.org/10.1371/journal.pone.0102297>.
- Mukamel et al. (2002). Mukamel, D., Zwanziger, J. and Bamezai, A. "Hospital competition, resource allocation and quality of care". *BMC Health Services Research* 2.1. DOI: 10.1186/1472-6963-2-10.  
<https://doi.org/10.1186/1472-6963-2-10>.
- Mukamel et al. (2004). Mukamel, D., Weimer, D., Zwanziger, J., Gorthy, S.-F. H. and Mushlin, A. "Quality Report Cards, Selection of Cardiac Surgeons, and Racial Disparities: A Study of the Publication of the New York State Cardiac Surgery Reports". *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 41.4, pp. 435–446. DOI: 10.5034/inquiryjnl\_41.4.435.  
[https://doi.org/10.5034/inquiryjnl\\_41.4.435](https://doi.org/10.5034/inquiryjnl_41.4.435).
- Mukamel et al. (2014). Mukamel, D., Haeder, S. and Weimer, D. "Top-Down and Bottom-Up Approaches to Health Care Quality: The Impacts of Regulation and Report Cards". *Annual Review of Public Health* 35.1, pp. 477–497. DOI: 10.1146/annurev-publhealth-082313-115826.  
<https://doi.org/10.1146/annurev-publhealth-082313-115826>.
- N Marshall et al. (2000). N Marshall, M., G Shekelle, P., Leatherman, S. and H Brook, R. "The Public Release of Performance Data: What Do We Expect to Gain? A Review of the Evidence". *JAMA : the journal of the American Medical Association* 283, pp. 1866–74.
- Nelson et al. (2007). Nelson, E., Batalden, P. and Godfrey, M. *Quality By Design: A Clinical Microsystems Approach*. John Wiley & Sons. ISBN: 978-0-7879-7898-3.
- Neuhaus, J. and McCulloch, C. (2011). "Generalized linear models". *Wiley Interdisciplinary Reviews: Computational Statistics* 3.5, pp. 407–413. DOI: 10.1002/wics.175. <https://doi.org/10.1002/wics.175>.
- Neuhauer, D. (1999). *Florence Nightingale: Measuring Hospital Care Outcomes*.
- NHS (2016). *Organisation patient safety incident reports: 1 October 2015 to 31 March 2016*. UK National Health Service.  
<https://improvement.nhs.uk/resources/organisation-patient-safety-incident-reports-28-september-2016/>.

- NHS (2017). *NHS Atlas series*. National Health Service.  
<https://www.england.nhs.uk/rightcare/products/atlas/>.
- Nicolay et al. (2011). Nicolay, C., Purkayastha, S., Greenhalgh, A., Benn, J., Chaturvedi, S., Phillips, N. and Darzi, A. "Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare". *British Journal of Surgery* 99.3, pp. 324–335. DOI: 10.1002/bjs.7803.  
<https://doi.org/10.1002/bjs.7803>.
- OECD (2017). *Health at a Glance 2017*. Organisation for Economic Co-operation and Development. <https://www.oecd.org/health/health-systems/health-at-a-glance-19991312.htm>.
- O'Regan et al. (2013). O'Regan, N., Fitzgerald, J., Timmons, S., O'Connell, H. and Meagher, D. "Delirium: a key challenge for perioperative care". eng. *International journal of surgery (London, England)* 11.2, pp. 136–44. ISSN: 1743-9159. DOI: 10.1016/j.ijsu.2012.12.015. eprint: 23277227.
- Pandit, J. (2016). "Deaths by horsekick in the Prussian army - and other 'Never Events' in large organisations". *Anaesthesia* 71.1, pp. 7–11. ISSN: 1365-2044. DOI: 10.1111/anae.13261. PMID: 26597473.
- Pieper et al. (2013). Pieper, D., Mathes, T., Neugebauer, E. and Eikermann, M. "State of evidence on the relationship between high-volume hospitals and outcomes in surgery: a systematic review of systematic reviews". *Journal of the American College of Surgeons* 216.5, 1015–1025.e18. ISSN: 1879-1190. DOI: 10.1016/j.jamcollsurg.2012.12.049. PMID: 23528183.
- Pope, D. (2009). "Reacting to rankings: Evidence from "America's Best Hospitals"". *Journal of Health Economics* 28.6, pp. 1154–1165. DOI: 10.1016/j.jhealeco.2009.08.006.  
<https://doi.org/10.1016/j.jhealeco.2009.08.006>.
- Pronovost et al. (2010). Pronovost, P. et al. "Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study". *BMJ (Clinical research ed.)* 340, p. c309. ISSN: 1756-1833. eprint: 20133365.
- Pronovost et al. (2017). Pronovost, P., Mathews, S., Chute, C. and Rosen, A. "Creating a purpose-driven learning and improving health system: The Johns Hopkins Medicine quality and safety experience". *Learning Health Systems* 1.1.
- Rabe-Hesketh et al. (2002). Rabe-Hesketh, S., Skrondal, A. and Pickles, A. "Reliable estimation of generalized linear mixed models using adaptive quadrature". *STATA journal* 2.1, pp. 1–21.
- Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata, Second Edition*. Stata Press. ISBN: 978-1-59718-040-5.
- Rajaram et al. (2015). Rajaram, R., Chung, J. W., Kinnier, C. V., Barnard, C., Mohanty, S., Pavely, E. S., McHugh, M. C. and Bilimoria, K. Y. "Hospital Characteristics Associated With Penalties in the Centers for Medicare & Medicaid Services Hospital-Acquired Condition Reduction Program". *JAMA* 314.4, p. 375. DOI: 10.1001/jama.2015.8609. <https://doi.org/10.1001/jama.2015.8609>.
- Rankin et al. (2017). Rankin, D., Bailey, N., McPherson, C. and Sleeman, L. "Outcome Measurement in the Private Hospital Setting". *ACHSM/ACHS Asia-Pacific health leadership congress*. Sydney.
- Rubin, J. (2017). "Patient empowerment and the Learning Health System". *Learning Health Systems* 1.3. ISSN: 2379-6146. DOI: 10.1002/lrh2.10030. <http://dx.doi.org/10.1002/lrh2.10030>.
- Runciman et al. (2009). Runciman, W., Hibbert, P., Thomson, R., Schaaf, T. V. D., Sherman, H. and Lewalle, P. "Towards an International Classification for Patient Safety: key concepts and terms". *International Journal for Quality in Health Care* 21.1, pp. 18–26. DOI: 10.1093/intqhc/mzn057.  
<https://doi.org/10.1093/intqhc/mzn057>.
- Sacks et al. (2016). Sacks, G., Dawes, A., Ettner, S., Brook, R., Fox, C., Russell, M., Ko, C. and Maggard-Gibbons, M. "Impact of a Risk Calculator on Risk Perception and Surgical Decision Making: A Randomized Trial". *Annals of surgery* 264.6, pp. 889–895. ISSN: 1528-1140. DOI: 10.1097/SLA.0000000000001750.

- Schang, L. and Morton, A. (2017). "Complementary logics of target-setting: hierarchist and experimentalist governance in the Scottish National Health Service". *Health economics, policy, and law* 12.1, pp. 21–41. ISSN: 1744-134X. DOI: 10.1017/S1744133116000232. eprint: 27322065.
- Scheurwegs et al. (2017). Scheurwegs, E., Luyckx, K., Luyten, L., Goethals, B. and Daelemans, W. "Assigning clinical codes with data-driven concept representation on Dutch clinical free text". *Journal of Biomedical Informatics* 69, pp. 118–127. DOI: 10.1016/j.jbi.2017.04.007. <https://doi.org/10.1016/j.jbi.2017.04.007>.
- Schiffinger et al. (2016). Schiffinger, M., Latzke, M. and Steyrer, J. "Two sides of the safety coin?: How patient engagement and safety climate jointly affect error occurrence in hospital units". *Health Care Manage Rev* 41.4, pp. 356–367.
- Scott, I. and Phelps, G. (2009). "Measurement for improvement: getting one to follow the other". *Internal medicine journal* 39.6, pp. 347–351. DOI: 10.1111/j.1445-5994.2009.01966.x.
- Shahian et al. (2005). Shahian, D., Torchiana, D., Shemin, R., Rawn, J. and Normand, S.-L. "Massachusetts Cardiac Surgery Report Card: Implications of Statistical Methodology". *The Annals of Thoracic Surgery* 80.6, pp. 2106–2113. DOI: 10.1016/j.athoracsur.2005.06.078. <https://doi.org/10.1016/j.athoracsur.2005.06.078>.
- Shojania, K. and Mheen, P. M.-v. de (2015). "Temporal trends in patient safety in the Netherlands: reductions in preventable adverse events or the end of adverse events as a useful metric?" *BMJ Quality & Safety* 24.9, pp. 541–544. ISSN: 2044-5415. DOI: 10.1136/bmjqqs-2015-004461. eprint: <http://qualitysafety.bmjjournals.com/content/24/9/541.full.pdf>. <http://qualitysafety.bmjjournals.com/content/24/9/541>.
- Shojania, K. and Dixon-Woods, M. (2016). "Estimating deaths due to medical error: the ongoing controversy and why it matters: Table 1". *BMJ Quality & Safety*, bmjqs-2016-006144. DOI: 10.1136/bmjqs-2016-006144. <https://doi.org/10.1136/bmjqs-2016-006144>.
- Stacey et al. (2017). Stacey, D. et al. "Decision aids for people facing health treatment or screening decisions". *Cochrane Database of Systematic Reviews*. DOI: 10.1002/14651858.cd001431.pub5. <https://doi.org/10.1002/14651858.cd001431.pub5>.
- Stanfill et al. (2010). Stanfill, M., Williams, M., Fenton, S., Jenders, R. and Hersh, W. "A systematic literature review of automated clinical coding and classification systems". *Journal of the American Medical Informatics Association* 17.6, pp. 646–651. DOI: 10.1136/jamia.2009.001024. <https://doi.org/10.1136/jamia.2009.001024>.
- Steering Committee for the Review of Government Service Provision (2017). "Public hospitals". *Report on Government Services 2017*. Ed. by Productivity Commission. <http://www.pc.gov.au/research/ongoing/report-on-government-services/2017/health/public-hospitals>.
- Stigglebout et al. (2015). Stigglebout, A., Pieterse, A. and Haes, H. de. "Shared decision making: Concepts, evidence, and practice". *Patient education and counseling* 98.10, pp. 1172–9. ISSN: 1873-5134. DOI: 10.1016/j.pec.2015.06.022. eprint: 26215573.
- Sullivan et al. (2016). Sullivan, C., Staib, A., Ayre, S., Daly, M., Collins, R., Draheim, M. and Ashby, R. "Pioneering digital disruption: Australia's first integrated digital tertiary hospital". *The Medical Journal of Australia* 205.9, pp. 386–389. DOI: 10.5694/mja16.00476. <https://doi.org/10.5694/mja16.00476>.
- Timmins, N. (2017). *Tackling variations in clinical care: assessing the Getting It Right First Time (GIRFT) programme*. Kings Fund and The Health Foundation. <https://www.kingsfund.org.uk/publications/tackling-variations-clinical-care>.

- Toson et al. (2016). Toson, B., Harvey, L. and Close, J. "New ICD-10 version of the Multipurpose Australian Comorbidity Scoring System outperformed Charlson and Elixhauser comorbidities in an older population". *Journal of clinical epidemiology* 79, pp. 62–69. ISSN: 1878-5921. DOI: 10.1016/j.jclinepi.2016.04.004.
- Totten et al. (2012). Totten, A., Wagner, J., Tiwari, A., O'Haire, C., Griffin, J. and Walker, M. "Closing the quality gap: revisiting the state of the science (vol. 5: public reporting as a quality improvement strategy)". *Evidence report/technology assessment* 208.5, pp. 1–645. ISSN: 1530-4396. eprint: 24422977.
- Tsai et al. (2015). Tsai, T., Jha, A., Gawande, A., Huckman, R., Bloom, N. and Sadun, R. "Hospital Board And Management Practices Are Strongly Related To Hospital Performance On Clinical Quality Metrics". *Health Affairs* 34.8, pp. 1304–1311. DOI: 10.1377/hlthaff.2014.1282. <https://doi.org/10.1377/hlthaff.2014.1282>.
- Tweddell et al. (2017). Tweddell, J., Jacobs, J. and Austin, E. "Are there negative consequences of public reporting? The hype and the reality". *The Journal of thoracic and cardiovascular surgery* 153.4, pp. 908–911. ISSN: 1097-685X. DOI: 10.1016/j.jtcvs.2017.01.034. PMID: 28359373.
- Vincent, C. and Amalberti, R. (2016). *Safer Healthcare: Strategies for the Real World*. Springer. ISBN: 978-3-319-25559-0. <https://www.springer.com/gp/book/9783319255576>.
- Vladeck et al. (1988). Vladeck, B., Goodwin, E., Myers, L. and Sinisi, M. "Consumers and hospital use: the HCFA "death list"". *Health Affairs* 7.1, pp. 122–125. DOI: 10.1377/hlthaff.7.1.122. <https://doi.org/10.1377/hlthaff.7.1.122>.
- Wakefield, J. and Jorm, C. (2009). "Patient safety – a balanced measurement framework". *Australian Health Review* 33.3, p. 382. DOI: 10.1071/ah090382. <https://doi.org/10.1071/ah090382>.
- Walsh et al. (2014). Walsh, T., Barr, P. J., Thompson, R., Ozanne, E., O'Neill, C. and Elwyn, G. "Undetermined impact of patient decision support interventions on healthcare costs and savings: systematic review". *BMJ* 348, g188.
- Waring et al. (2015). Waring, J., Allen, D., Braithwaite, J. and Sandall, J. "Healthcare quality and safety: a review of policy, practice and research". *Sociology of Health & Illness* 38.2, pp. 198–215. DOI: 10.1111/1467-9566.12391. <https://doi.org/10.1111/1467-9566.12391>.
- Wears, R. (2016). "Why do we love to hate ourselves?" *BMJ Quality & Safety*. Comment on 'Premature closure? Not so fast'. <http://qualitysafety.bmjjournals.com/content/early/2016/03/15/bmjqqs-2016-005267.responses#why-do-we-love-to-hate-ourselves>.
- Weingart et al. (2011). Weingart, S. et al. "Hospitalized patients' participation and its impact on quality of care and patient safety". *International Journal for Quality in Health Care* 23.3, pp. 269–277. DOI: 10.1093/intqhc/mzr002.
- Weller et al. (2010). Weller, W. E., Kabra, G., Cozzens, K. S. and Hannan, E. L. "Appending limited clinical data to an administrative database for assessing institutional quality outliers for AMI care: Does it matter?" *International Journal of Cardiology* 140.1, pp. 95–101. DOI: 10.1016/j.ijcard.2008.11.003. <https://doi.org/10.1016/j.ijcard.2008.11.003>.
- Wen et al. (2015). Wen, T., Attenello, F. J., Wu, B., Ng, A., Cen, S. Y. and Mack, W. J. "The July effect: an analysis of never events in the nationwide inpatient sample". *Journal of hospital medicine* 10.7, pp. 432–8. ISSN: 1553-5606. DOI: 10.1002/jhm.2352. eprint: 25828426.
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd. MIT Press. ISBN: 9780262232586.
- Youden, W. (1950). "Index for rating diagnostic tests". *Cancer* 3.1, pp. 32–5. ISSN: 0008-543X.

Young et al. (2011). Young, J., Ranji, S., Wachter, R., Lee, C., Niehaus, B. and Auerbach, A. ““July effect”: impact of the academic year-end changeover on patient outcomes: a systematic review”. *Annals of internal medicine* 155.5, pp. 309–15. ISSN: 1539-3704. DOI: 10.7326/0003-4819-155-5-201109060-00354.

Zhang et al. (2013). Zhang, X., Hauck, K. and Zhao, X. “Patient safety in hospitals - a Bayesian analysis of unobservable hospital and specialty level risk factors”. *Health economics* 22.9, pp. 1158–74. ISSN: 1099-1050. DOI: 10.1002/hec.2972.