



Australian Government
Productivity Commission

SHIFTING THE DIAL



5 YEAR PRODUCTIVITY REVIEW

SUPPORTING PAPER NO. 3

COMPARATIVE
PERFORMANCE INDICATORS

3 AUGUST 2017

© Commonwealth of Australia 2017

ISBN 978-1-74037-628-0 (PDF)



Except for the Commonwealth Coat of Arms and content supplied by third parties, this copyright work is licensed under a Creative Commons Attribution 1.0 Australia licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/1.0/au>. In essence, you are free to copy, communicate and adapt the work, as long as you attribute the work to the Productivity Commission (but not in any way that suggests the Commission endorses you or your use) and abide by the other licence terms.

Use of the Commonwealth Coat of Arms

For terms of use of the Coat of Arms visit the 'It's an Honour' website: <http://www.itsanhonour.gov.au>

Third party copyright

Wherever a third party holds copyright in this material, the copyright remains with that party. Their permission may be required to use the material, please contact them directly.

Attribution

This work should be attributed as follows, *Source: Productivity Commission, Comparative Performance Indicators, Shifting the Dial: 5 year Productivity Review, Supporting Paper No. 3.*

If you have adapted, modified or transformed this work in anyway, please use the following, *Source: based on Productivity Commission data, Comparative Performance Indicators, Shifting the Dial: 5 year Productivity Review, Supporting Paper No. 3.*

An appropriate reference for this publication is:

Productivity Commission 2017, *Comparative Performance Indicators, Shifting the Dial: 5 year Productivity Review, Supporting Paper No. 3*, Canberra.

Publications enquiries

Media and Publications, phone: (03) 9653 2244 or email: maps@pc.gov.au

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au).

Contents

Key points	2
Comparative performance indicators	3
1 Measuring provider-level performance for government-funded services	3
2 A basis for developing comparative performance indicators	12
3 Data availability and accessibility	22
4 Links to financial incentives	26
References	32

Key points

- Although many government-funded service providers report their performance against internally-established benchmarks, the use of provider-level comparative performance reporting as a tool to inform and better incentivise performance is in its relative infancy.
- Comparisons between like providers can enhance competition, help consumers make more informed choices and, by identifying better performers, help providers to recognise best-practice methods and aid governments in dealing with poor performance.
- There has been some effort to develop comparative performance platforms in major service delivery areas, for example, the *My Hospitals* and *My School* websites. Experience here and overseas suggests the following lessons:
 - while the jurisdiction that collects the performance data depends on the role of different levels of government in different sectors, having a single national authority publish the data enhances accessibility and discoverability.
 - more granular data (down to specific facilities or individual professionals) is better, but only where performance can be reasonably attributed to that level.
 - the potential benefits of performance data are difficult to know *ex ante*, so should be made public by default, with other arrangements (including ‘trusted-user’ provisions or private feedback) where there are strong privacy or confidentiality concerns.
 - as the client base for services varies greatly between providers, risk-adjustment (or value-added) treatment of performance data is needed for valid comparisons.
 - to measure performance comprehensively, it is necessary to include qualitative factors among the reported indicators (not just quantitative factors), such as self-appraised consumer outcomes.
 - the fewest possible easily-understood performance indicators would have the most value, but care is needed to ensure indicators actually relate to government objectives and that they do not distort provider behaviour.
 - publishing provider prices (costs) alongside other performance indicators would also add value where prices are opaque and can differ substantially within the sector.
 - utilising data from existing sources (such as administrative data) helps to reduce costs, although new data sources (such as improved consumer surveys) may be needed where there are clear and important gaps in information.
 - while performance-based financial incentives (such as performance pay for teachers, activity-based funding for hospitals or social impact bonds) can be used, they have had mixed results in the past, generally depending on their context and design.

Comparative performance indicators

1 Measuring provider-level performance for government-funded services

Historically, much of the analysis of Australia's productivity performance has been focused on macroeconomic measures of productivity, reported as part of the national accounts. These accounts help to provide a summary of economic events and a stocktake of an economy's production and wealth (and the component factors). They are important for understanding the economy before formulating macroeconomic policy — including, for example, the Reserve Bank's interest rate decisions or broad fiscal policy settings.

However, in some areas of the economy, productivity is difficult to measure. In particular, multi-factor productivity (MFP) estimates are not available for industries in the 'non-market' sector — health care and social assistance, education and training and public administration and safety. Government provision, consumer subsidies and sometime mandated consumption (as in school education) mean that there are no market determined prices and quantities for most of the services in these industries. This means there is a computational barrier to estimating MFP in these industries and hence across the whole economy (see SP 2).

To ensure the national accounts provide a more complete picture of how the entire economy is performing (not just part of the economy), the Australian Bureau of Statistics (ABS) is currently developing measures of non-market MFP for the national accounts. This work is beyond the timeline for this inquiry, but will help to provide a more accurate picture of Australia's productivity performance, and guidance on how government-dominated industries are performing over time and whether taxpayers are getting improved value for their money. More importantly for government, these macro-level measures of an industry's productivity can act as a 'canary in a cage', detecting manifest failures in the performance of government services.

However, such aggregate data are not very useful for consumers, providers, funders or policymakers concerned about individual services. The information they seek on performance requires more granular data on the costs, outputs and outcomes of individual providers and the systems they operate in. More specifically, additional granular data on the nature of resources used by firms (inputs), their rates of change, the resulting goods and services produced (outputs) and their value to consumers (outcomes) can provide a basis for:

- improving information to support consumer choice and provide feedback to providers

-
- reviewing and assessing government programs
 - designing government policy
 - managing government contracts and external service providers
 - identifying influences on productivity, which along with performance data feed into improving policy and program design.

The value of disaggregated, provider-level productivity analysis has already been widely recognised for industries outside the non-market sector, with a range of recent international studies using firm level data to provide distributional analyses of firm productivity in the market economy (see Andrews, Criscuolo and Gal 2015; Conway 2016; McGowan, Andrews and Millot 2017). Although none of this work has yet been widely replicated in Australia,¹ the analysis has nonetheless provided useful insights on market dynamics in other economies, including the characteristics of the most and least productive firms and the level of competition within industries.

Information on provider-level productivity in the non-market sector would allow similar insights about the drivers of firm-level productivity in the sector. Importantly, it would also enhance government accountability and transparency, as most of the non-market sector is funded (at least in part) by taxpayers, and its outputs (the services to be delivered, and the price, quality and/or quantity of those services) are substantially regulated by governments. That is, the public has a right to know details of where their money is going, whether it is achieving meaningful results, and if they are getting the best possible value (DoF 2016).

This principle applies not only to services that are directly provided or (effectively) guaranteed by government (such as public hospitals, emergency services, public transport, and water and sewerage services). It also applies to services that are delivered and funded *indirectly* by governments, such as through the use of grants, subsidies or contracts to private sector organisations — examples include subsidised private education and health care, contracted employment services, and grants for creative and performing arts activities. Although none of these services are conducted directly by government, they are still funded by taxpayers. Some of these taxpayer-funded services are also in the market sector, rather than non-market sector. As such, indicators of government-funded performance are important for both the market and non-market sectors.

In addition, detailed information on the productivity of different service providers can help guide the development of productivity-enhancing reforms. Many of the most effective policy levers available to governments are microeconomic, relying on an understanding of the structure and dynamics of markets (such as regulations, individual tax and transfer policies and specific government programs) (Atkinson 2005; Schreyer 2010). As such, more detailed indicators of performance can help to identify patterns in underperformance

¹ The ABS and the Department of Industry, Innovation and Science have developed a firm level database (BLADE) that will allow such analysis to be undertaken in the future (see chapter 5 in the main report).

and high performance, providing insights for policies that encourage the latter and discourage the former.

Taxpayers and policy makers are also interested in performance metrics beyond strict estimates of productivity, so they can better assess whether governments are achieving their broader policy goals. In part, this reflects the difficulty in capturing improvements in service quality in the measured productivity of government-sponsored activity (given the absence of or small role played by price signals). To form a more complete picture of government performance, productivity indicators should be considered alongside other measures, such as whether quality, equity and access standards have been met, consumer perceptions of quality, and financial indicators.

Absolute and comparative performance indicators

Many indicators of the performance of government-funded service providers are already collected and reported on a regular basis — primarily to improve their accountability and transparency. Much of that performance reporting focuses on changes over time in a particular entity's performance and the degree to which they meet internally-established goals. For instance, at the Commonwealth level, the *Public Governance, Performance and Accountability Act 2013* aims to improve the line of sight between what was intended and what was delivered by Commonwealth government-owned entities, requiring Annual Performance Statements from each entity in their annual reports (DoF 2016). The States and Territories generally operate similar systems. For example, Western Australia requires each agency to report annually against a range of Key Performance Indicators (KPIs) under their Outcome Based Management system (WA DTF 2004).

Similar levels of reporting are often also required from contracted private providers of government-funded services. For example, Transport for NSW (TfNSW) publishes a Guideline for Construction Contractor Performance Reporting to 'enhance contractor performance via a continuous improvement process and recognition of good performance' (TfNSW 2014).

While such reporting frameworks are valuable in assessing some aspects of an entity's *absolute* performance, they do not provide consumers or taxpayers with any measure of its performance relative to its peers. For example, consider the different management imperatives that would be triggered by information indicating that a hospital's unplanned readmission rates had been improving (suggesting a good outcome), and information that demonstrated that its readmission rate was still more than twice as high as other comparable hospitals (an indicator that performance needs further improvement).

The value of performance comparisons

The analysis of comparative measures of performance between providers of government-funded services plays an important role in informing and incentivising

performance. Beyond the benefits of better policy development and enhanced taxpayer accountability and transparency, comparative performance measures can help to overcome a lack of adequate market information, which in turn drives improved outcomes for consumers.

Improving outcomes for consumers

By encouraging individual choice ...

The publication of comparative performance indicators for government-funded services allows individual consumers to make better-informed choices between services and service providers if they choose (and are able) to exercise choice. In turn, this creates competitive pressures on suppliers to improve their services, to the potential advantage of all consumers (Berwick, James and Coye 2003).

In the health care area², there are, however, some subtleties in linking the availability of (high quality) performance indicators, choice and outcomes:

- The evidence that public divulgence of performance metrics *per se* makes a significant difference is relatively weak. In the United States, provision of high quality information to consumers has been identified as a strong determinant in the choice of high quality-rated health plans (Faber et al. 2009). However, a review of the few rigorous trials of the outcomes from public divulgence of performance data did not find substantial effects (Ketelaar et al. 2011).
- The existence of better performance indicators and a capacity to choose does not necessarily result in many people deciding to exercise that choice. In the case of health care, there is some evidence that proximity is a major basis for choice, and that published information has secondary impacts over advice from their GP or friends and family (Barratt 2011). Nevertheless, the international evidence suggests that patients strongly support the option of choice, and appear more likely to exercise it when it is likely to affect the outcomes of care (Dixon et al. 2010). For example, people tend to shift away from what they perceive as a poor local hospital for elective surgery. There is some evidence that people exercising choice have had better outcomes, at least in waiting times (Jones and Mays 2009; Ringard and Hagen 2011; Ringard, Saunes and Sagan 2016; Vrangbaek et al. 2007).
- There should be no presumption that choice only relates to the selection of a provider. Performance indicators that measure the efficacy and impacts of different treatment choices are also important, and some believe far more so than the choice of practitioner (Coulter 2010).

² Of the three industries in the non-market sector, considerably more research and experience on performance indicators has been accumulated in health care than in education or public administration. As such, much of this paper is reliant on insights from the health care industry.

-
- Patients need sufficient health literacy (as noted in chapter 2 of the main report), and it can take time for people's behaviour to change.
 - If supply constraints are high, then suppliers face lower risks from a failure to improve as patients may be unable to find alternative suppliers with spare capacity.
 - Health practitioners' behaviours and assumptions about patient capabilities can encourage or thwart choice (Harding et al. 2014).
 - Weak governance arrangements that bail out underperforming providers or that fail to discipline senior decision-makers reduce the potential impacts of information provision and choice on improving consumer outcomes.

... revealing comparative performance issues ...

Although comparative performance indicators can improve competitive incentives, there are limits to how much competition can be introduced into many areas of government-funded services. This is especially true in areas where there are artificial or natural barriers to consumer choices. For example, choice between government schools can be hindered by artificial regulations restricting students to the school district they reside in, as well as natural barriers (such as poor transport links) between districts. There are also limits on competition where there are government-run monopolies that do not have to compete.

In cases where competition between providers is impractical, it is still possible to use comparative information to create pressures similar to those that might exist in a competitive environment. More specifically, relative performance indicators can make providers more accountable to the communities they operate in and give them incentives to improve (AIHW 2017; Berwick, James and Coye 2003). In addition:

- Funders and regulators already use benchmarking to determine efficient prices for services in utilities and in hospitals (under activity-based funding).
- Poorly performing entities may be forced to remove their CEOs or boards.
- There may be circumstances where there is sequential competition *for* a regional market (such as through an infrequent tender for monopoly provision of services in a given area), which still provides scope for benchmarking against other regions to determine value for money in service tenders and, hence, whether an entity should be allowed to extend its temporary monopoly over government-funded services.

Indicators can also reveal to providers areas of poor performance, which may have been previously unknown. Many of the professions that dominate government-funded services — including nurses, doctors, teachers, police officers, magistrates, firefighters and policy advisers — are characterised by high levels of altruism and considerable regard for community service, so publication of comparative information may provide, in some cases, a sufficient prompt and pointers toward improvement. Of course, the incentives for and

extent of actual improvement may be affected by other constraints on change, whether these are resource, time, political or regulatory constraints.

... assisting the diffusion of best-practice services ...

In the market economy, the diffusion mechanism generally occurs organically, as firms on or near the frontier of productivity attract more resources and expand, while lagging firms are forced to improve, lest they shrink or exit.³ However, the diffusion of best-practice service delivery in areas where governments fund or provide services is often more problematic. This reflects the lack of price signals and that governments often cannot close or reduce the size of a poorly performing entity because they must guarantee supply. The lack of any market-based diffusion mechanism therefore has to be addressed through policies that encourage — and allow — service providers to continually improve consumer outcomes.

Transparent provision of performance indicators that identify leaders and laggards on a like-for-like basis can assist the diffusion of best-practice service delivery. Absent transparent indicators *and* incentives to act, large-scale inefficiencies can persist (AIHW 2017). Decision-makers can also use indicators to estimate the net gains from lifting the performance of laggards, which can prioritise the areas of reform.

Comparative indicators are only one component of creating performance disciplines and diffusion. Equally important is the diagnosis of *why* some entities perform much worse or better than others — because that information is the basis for providing specific advice on what must be diffused to lift performance. Handwashing in hospitals to manage avoidable infections is an exemplar (OECD 2017). Another well-known example is the use of surgical safety checklists to avoid adverse outcomes (such as ‘Wrong-Site Surgery’) (Bergs et al. 2014; Haugen et al. 2015; Lyons and Popejoy 2014; Panesar et al. 2009). Yet adoption across hospitals in Australia and other countries of checklists has been variable (Ragusa et al. 2016; Rajendram 2016; Swan 2015), suggesting that its effective diffusion should be a target for policy. Although it may not be ideal to force this through mere administrative fiat, in National Health Service (NHS) hospitals in the United Kingdom, uptake was improved by administrators refusing to allow surgery teams to take a patient to a recovery room after surgery unless the full World Health Organisation (WHO) Safe Surgery Checklist had been completed.

The ‘checklist’ example highlights why information about comparative performance is a necessary, but not sufficient, basis for achieving diffusion. This is apparent in widespread (and persistent) deviations from best practice in the health care industry (box 1).

³ Although there are indications that the market sector diffusion mechanism may have weakened in many OECD countries — see Andrews, Criscuolo and Gal (2015) and Conway (2016).

... and providing a basis for the selective use of financial incentives

Performance indicators are a prerequisite to using financial incentives to improve provider performance — and indeed the development of indicators is often motivated by this function. Financial incentives can either involve payments (or provider retention of financial savings) if a provider achieves a higher level of performance or involve the withdrawal of funding or the imposition of penalties if the provider deviates too far from some acceptable benchmark. Although there are concerns about the adequacy of some performance metrics and the environment in which they are used (see discussion in section 4 below), a necessary precursor to establishing any financial incentives is to create informative performance metrics.

Box 1 Evidence of diffusion problems in health care

A broad array of research provides evidence that the adoption of best-practice methods in the health care sector is less extensive than it could be:

- Runciman et al. (2012) found that 43 per cent of a sample of Australian adults had received inappropriate care in their recent health care encounters, according to evidence-based and consensus-based guidelines.
- The Australian Commission on Safety and Quality in Health Care's 2015 *Atlas of Healthcare Variation* showed dramatic unexplained variations in procedures and prescribing, and the use of procedures for which there is no favourable clinical evidence (ACSQHC 2015).
- A 2007 study by the Commonwealth Fund found that 15 per cent of Australians reported undergoing unnecessary repeat imaging (Russell and Doggett 2015).
- Western Australian researchers found that 23 of 47 medicines commonly dispensed to over 100 000 different pregnant women from 2002 to 2005 were associated with some form of birth defect (Colvin et al. 2010).
- The former National Institute of Clinical Studies (2003, 2005) identified gaps between evidence and practice in areas such as advising on smoking cessation, screening for lung cancer, and vaccinating against influenza.
- In 2013-14, about 30 per cent of people presenting to general practitioners in Australia for acute upper respiratory tract infection — the 'common cold' — were prescribed antibiotics, even though antibiotics are ineffective for treating viral infections (SCRGSP 2015).
- Paracetamol is commonly recommended and prescribed for back pain in Australia. However, a recent randomised trial of paracetamol for the treatment of acute lower back pain found no benefit versus a placebo (Carpenter et al. 2014).
- Approximately 6.5 per cent of separations in public hospitals in 2012-13 were associated with 'adverse events' — where patients are harmed during hospitalisation — in part due to poor practice methods, including injuries from falls, adverse drug effects and surgical errors (SCRGSP 2015).

Sources: PC (2013, 2015).

Social impact bonds are an example where the development of performance measures — at the outcome level — is essential. These bonds aim to attract finance to fund human services on the bases that the bond holder receives a bonus if the service achieves a

specified set of outcomes. The aim is to create incentives for the service provider to deliver real improvements over an agreed baseline, with the bondholders providing oversight of performance. Several social impact bonds have been undertaken in Australia in recent years, including a bond to tackle homelessness in South Australia, one focusing on drug and alcohol treatment programs in Victoria, and one to reduce recidivism in New South Wales.

The structure of social impact bonds requires agreement on the outcome measures — the baseline and the expected improvement required to trigger the returns. As measurement is central to the success of such bonds, efforts have gone into developing reliable and accepted measurement frameworks (see for example, Deloitte Access Economics 2016). However, such outcome measurement can also become excessive, with concerns recently that the reporting and measurement requirements of some social impact bonds are preventing take-up and hindering their benefits (Edmiston and Nicholls 2017).

Overcoming information gaps

In conventional (reasonably competitive) markets, prices often convey quality. However, such price signals are typically not present for many government-funded services as they are either provided free of charge or heavily subsidised. Equally, there is no equivalent in most government services of ‘warranties’, which, in the commercial sphere provide signals about the qualities of services.

This creates a greater imperative for information provision to consumers that provides them with some understanding of the variation in the quality of providers/services. Consumers can express their satisfaction or dissatisfaction with services formally (such as through complaint lines) and informally (such as through online forums), but these may not be representative of all users’ views or provide clear or very useful information. Provision of meaningful information that genuinely gives consumers some agency is non-trivial. This reflects that many governments services have effects that take years to surface (for example, there is considerable uncertainty about future earnings and employment prospects from different types of education) or involve complex technical matters (for instance, in making choices between various clinical procedures).⁴ As such, well-curated comparative information is needed, although, as discussed below, governments need not always themselves provide the curation service.

Examples of existing performance reporting

The Productivity Commission undertakes annual comparative performance reporting of many government services — on behalf of the Steering Committee for the Review of Government Service Provision (SCRGSP) — through the Report on Government Services (ROGS). These

⁴ Moreover, people often need sufficient technical capability (‘literacy’) in a given service area to make use of much of the information provided. For example, as discussed in chapter 2 of the main report, most Australians have poor health literacy, so may struggle with some measures of health care performance.

reports include performance indicators for services funded by the Commonwealth, States and Territories (such as emergency management, health, justice, community services, education and housing and homelessness) and measure their equity, effectiveness and efficiency across different states and territories (SCRGSP 2017).

Commonwealth, State and Territory governments have also progressed a range of different platforms for reporting and comparing the performance of different government-funded service providers (box 2).

Box 2 Existing performance reporting platforms

- **My Hospitals** — The *My Hospitals* website, currently managed by the Australian Institute of Health and Welfare (AIHW), was established in 2010 to ensure easy access to nationally consistent and comparable performance information for individual public and private hospitals. Examples of the types of performance indicators that are currently published for hospitals on the *My Hospitals* website include emergency department waiting times, the number of admissions by type, surgery length of stay and waiting time by different surgery types, the number of healthcare-associated *Staphylococcus aureus* infections, and expenditure per National Weighted Activity Unit (a common unit of hospital activity that accounts for differences in the complexity of conditions or procedures) (AIHW 2016c).
- **My School** — Launched in January 2010, the *My School* website contains performance data and other information on approximately 9500 public and private schools and is managed by the Australian Curriculum Assessment and Reporting Authority (ACARA). School-level information available on the site includes the type of school, student and staff numbers, student attendance rates, results from national literacy and numeracy tests, student demographic profiles, and school-level financial information (ACARA 2016; DET 2016a).
- **Quality Indicators for Learning and Teaching (QILT)** — The QILT website, maintained by the Social Research Centre and funded by the Commonwealth Government, provides prospective tertiary education students with information about Australian higher education institutions through the use of survey data from recent students and graduates. This includes surveys on student experiences (engagement, support, teaching quality, skills development), graduate employment (employment rates and median salaries) and graduate satisfaction, with surveys of employer satisfaction currently being trialled (DET 2015; SRC 2016).
- **My Healthy Communities** — Launched in 2013 and managed by the AIHW, the *My Healthy Communities* website compares a variety of health indicators between geographic areas around the country. Depending on the level of geographical detail chosen (from Primary Health Networks, to Medicare Local areas, Level 3 Statistical Areas (SA3) or individual postcodes), local data are available on life expectancy at birth, immunisation rates, GP attendance rates, the proportion of bulk-billed GP attendances, primary health expenditure per person and survey results of GP treatment quality (AIHW 2016b).
- **My Child** — The *My Child* website, run by the Commonwealth Government, can be used to find local child care service providers and shows information on their fees, available services, vacancies and quality — including educational practices, children's health and safety, the physical environment, staffing arrangements and more, as determined by the Australian Children's Education and Care Quality Authority and the National Quality Standard (Australian Government 2015; DET 2016b).
- **Know Your Council** — The Victorian Government established the *Know Your Council* website to improve the transparency and accountability of council performance to ratepayers through the regular reporting of 66 measures of service performance, including on animal management, waste collection, financial performance, library services and many others. Users can benchmark and compare each of Victoria's 79 local councils to councils that are similar in size and scope (Local Government Victoria 2015).

Although the considerable progress that has been made on many of these ‘My Service’ type sites is promising, most of them remain incomplete, with new data sources, performance indicators and providers still planned to be added by their managing organisations, as well as ongoing enhancements to the user interface.

The existing comparator sites are also not comprehensive in their reach across government-funded services, with significant gaps in public administration (particularly in the justice and emergency services sectors), primary health care and local government (outside Victoria).⁵

2 A basis for developing comparative performance indicators

The old computer adage ‘garbage in, garbage out’ equally applies to comparative performance data. Useful performance measures have seven characteristics: they should be relevant, valid, reliable, accurate, interpretable, accessible and cost-effectively collectable and storable. An exhaustive treatment of all of these aspects is beyond the scope of this paper, but there are several important general observations worth making, particularly in light of the ongoing development of comparative indicators in Australia.

Comparisons work best across the widest possible sample

Since a provider’s performance is relative, the most valuable performance indicators would cover the widest possible sample of comparable providers. This means comparative performance indicators should ideally seek to cover all comparable service providers across the widest catchment practicable (such as all public and private hospitals in Australia, as in the case of the *My Hospitals* website). This facilitates benchmarking and ensures that all consumers will have access to data that are relevant to their choice set.

There are several models that can support a national evidence base for government services, which will vary according to the context:

- collaborative models between jurisdictions or entities that use common definitions, consent provisions, methodologies and collection methods to assemble a national database
- a single body — most likely a Commonwealth agency — that collects the data (from the jurisdictions or directly) and builds it into an accessible database.

The desirable holder and disseminator of the data will also vary, depending on the form and nature of the use of the data. In many instances, comparative performance information

⁵ See SP 16 on local government administration.

is built up from micro information that should not be publicly available in an identified form. For example, data on comparative school performance needs to take account of the characteristics of students, since much of the variation in performance across schools reflects factors outside the school's control (PC 2016b). Such information — combined with other information about schools (their teaching methods, links to the community, teacher quality, and facilities) — can inform 'what works' to improve school performance, which is one of the key goals of gathering comparative information. Such micro data cannot be available publicly because of the need to protect privacy, but should be available to any research institution that has the capabilities to protect the data (see below).

There is a distinction between a database that contains all the comparative information and the platform that disseminates information to *consumers*. In the latter case, it would be usually desirable for Australian governments to agree to have only one point of access by consumers to comparative data (hence *My Hospitals*, not eight jurisdictionally-based portals for access to data on hospitals). This reduces confusion between the various portals and eliminates the costs of duplication.

The availability of a single government platform for informing consumers does not mean that national data should be reported only through that vehicle. Indeed, it would be desirable that other parties — for example, a consumer advocacy group — could add value to the data from such a platform, through analysis, data augmentation, interpretative tools and more accessible interfaces.

National comparisons (for benchmarking purposes) are less likely to be useful where there are substantive differences between jurisdictions that make comparisons difficult — one such example is local council performance reporting, as councils perform different tasks in different states. In these instances, where differences cannot be overcome, duplicated reporting frameworks between jurisdictions (with individual databases and access portals for each state or territory) are generally justified.

The level of performance granularity

The benefits of performance measurement for government-funded service providers vary depending on how a 'provider' is defined and, more particularly, how close the defined provider is to the unit or entity actually delivering the service. Comparisons can occur at the level of entire states and territories down to individual professionals providing government-funded services (box 3) — their value will depend on how closely reported results can be attributed to those deemed 'providers'.

More granularity is better, until it isn't

For many government services, more detailed data provide a better idea of how performance varies between different providers. Reporting data at higher levels can mask considerable performance differences between providers. For example, while secondary

schools in Queensland may perform at a similar level to other states *on average*, this may be hiding considerable differences between the comparatively wealthy Brisbane and Gold Coast communities and the remote areas of Far North Queensland. Even within the Brisbane area, there are likely significant differences between the performance of individual schools, while individuals teachers at a given school can also vary greatly in their comparative performance.

Box 3 **Different levels of performance reporting**

State and territory level reporting — One way to compare performance is at the level of entire states and territories — comparing, for example, the overall performance of Victoria’s correctional facilities with those of South Australia. This is the approach taken in the Productivity Commission’s Reports on Government Services. At this level, the provider of the services is taken to be the entire state or territory, which is most useful when there are monopoly or largely undifferentiated oligopoly providers and the actions of these providers can reasonably be viewed as those of the jurisdiction’s as a whole — such as some public transport and utility service providers.

Local level reporting — Some jurisdiction-wide services report indicators by local geographic area. As an example, this already occurs for much of the health care sector on the *My Healthy Communities* website, which compares performance data between local areas (from Primary Health Networks down to individual postcodes). This level of performance granularity is most appropriate where there are monopoly providers for specific areas, such as individual police precincts or Centrelink offices.

Facility level reporting — Reporting at the level of individual facilities, organisations or firms — depending on what is appropriate for the relevant sector — can provide a further detailed picture of how government-funded services are performing. Existing examples include the *My School* and *My Hospitals* websites. Facility-level reporting can extend to particular teams or units working within a particular facility, where that team’s work is sufficiently separate from others.

Individual level reporting — The most granular picture of performance possible defines a ‘provider’ as the individuals providing government-funded services (such as specific surgeons, GPs or teachers), rather than the facilities or institutions they work for. Public performance reporting at such a detailed level does not currently occur in Australia, but has existed for many years in some fields in the United Kingdom and United States, including to assess individual surgeon performance (PC 2016a).

Higher-level performance reporting can thus lead to distorted or dulled incentives, as good providers may know that the national- or state-level statistics do not apply to them, while bad providers know that their performance is camouflaged and not identifiable (Dunleavy 2016).

Although this would imply that individual level performance reporting is the ideal goal of comparative performance reporting, different reporting levels have different uses. As such, higher-level performance reporting is not mutually exclusive with lower-level reporting — there is often space for reporting at a variety of different levels through a range of different mechanisms. Higher-level monitoring (such as the annual ROGS reports compiled by the

Commission) can still hold governments and service providers to account, as well as improve transparency in performance.

In addition, higher-level indicators (such as comparisons between states and territories) have generally also been a first step towards more detailed future comparisons. For instance, the ROGS reports have been around for 22 years (although it took many years to achieve widespread recognition of their inherent value, given concerns and misgivings about how the data could be misused), whereas the *My School* and *My Hospitals* websites were not launched until 2010. Higher-level performance reporting can be particularly valuable through encouraging better and more timely data collection from different jurisdictions and service providers, prior to lower levels of granularity being developed.

Indeed, no Australian performance data are currently reported at the individual level (such as by individual surgeons, GPs or teachers). This is due in part to the significant challenges involved, including the cost of the additional data collection, the technical difficulties of ensuring comparisons are valid (discussed below) and concerns about the privacy of the individuals (also discussed below).

Individual-level performance indicators are also not practical for many professions, particularly where the outputs and outcomes of the individual are highly dependent on the performance of a team (fire and rescue workers being an obvious example). In these circumstances, performance measurement at the team level may therefore be more appropriate — as argued by RACS (2016) — although the existence of *any* team-related work should not be used as an excuse to avoid *all* individual reporting.

It is also sometimes claimed that public reporting at the individual professional level is unreliable due to typically small caseloads and the variations in the characteristics of customers. For instance, a recent New Zealand Government review concluded that the typical caseload of medical specialists is too small to have enough statistical power to identify poorer performers (HQSCNZ 2016). Some professional associations accordingly object to the publication of such data (RACS 2015). Clinician comparison websites, such as *Physician Compare* in the United States, has been met with considerable concern by physicians (Lowes 2015).

However, data at the clinician level may be more reliable for some areas of health — such as cardiologists or GPs — where the casemix is less variable. This is especially true where the data are combined over several years. Nor does the data necessarily need to relate to mortality or complications, but can also apply to the practices of physicians — such as their use of evidence-based diagnostics, like screening for osteoporosis for women aged 65 years and older (CMS 2016).

Warts and all? — reporting publicly

Unless there are strong counter-arguments, the default for disclosure of performance data should be full public transparency. When combined with consumer capabilities,

transparency allows informed choice. Regardless, sunlight has a cleaning effect because most parties do not want to be publicly exposed as poor performers.

If full public reporting is not possible, then other reporting should still take place

Where full public disclosure is not feasible for all data, this should not preclude its availability to some parties. Other parts of government, trusted third-party researchers and intermediaries often have a capacity to analyse the data, link it to other available information, experiment with it and discover new and valuable relationships between outcomes — all at little cost to government. Limits on access thus undermine any potential improvements to the diffusion of best-practice service delivery. In the health care sector for example, this can mean that patients receive ineffective (or even harmful) care, adverse effects of drugs go undetected, or significant money is spent on interventions that do not improve health outcomes (PC 2010a, 2015).

Where sensitive data are provided to third parties, there are a series of now well-known methods for ensuring that access is not abused, since trust is an important aspect of making data available. De-identification, perturbation of data, secured data storage, agreements about the scope of use, secure server access and ‘trusted user’ arrangements are all part of the repertoire of methods (PC 2017).

The power of private feedback

The incentives for improvement created by public disclosure could also be partially achieved if service providers or facilities were simply told of their own comparative performance, without public disclosure — for example, allowing an individual police station to know how its performance compared with that of similar stations over several time periods. Similarly, in health care, information provided to a clinician that indicated that they had the highest rate of post-operative complications among their peers allows the clinician to assess whether this reflected chance, the riskiness of their client base, or deficiencies in their practices. Many suppliers — as organisations or individuals — want to improve because of a strong belief in the public good goals of their activities or through a competitive spirit (both of these being cited as factors affecting the behaviour of clinicians in some of the Commission’s consultations).

Similarly, intermediaries (particularly funders) have strong interests in using individual performance metrics to improve consumer outcomes because these often also reduce costs.

Limited disclosure before full public disclosure can also be employed to ensure that new measures provide meaningful performance measures, as well as winning support from the entities or individuals concerned.

Measuring performance comprehensively

Similar to the quality-adjustment issues that exist for measures of productivity in the national accounts (see SP 2), there are many aspects of a provider's services that affect their performance, some of which are more difficult to measure and quantify. In particular, fuzzier notions of the *quality* of services provided can often not lend themselves to easy measurement. For example, in the education sector, although the nature of what makes a particular teacher 'better' can be difficult to identify, most people would recognise that it involves more than maximisation of test results, but also the development of students' broader analytical capabilities, non-cognitive skills and a positive ethos of learning (PC 2016b).

Given the difficulties of measuring these more intangible aspects, some comparative performance indicators may simply not include aspects of service quality, or may use imperfect proxies that are quantifiable instead. Where quality is not measured or proxy measures are poorly chosen, a focus on a narrow range of quantifiable indicators can create a risk that some providers will neglect unreported aspects of care or try to 'game' the metrics that are reported.

More broadly, some measures of performance can also mask considerable variation between providers — for example, the use of quality indicators based only on whether a provider is meeting required service standards does not allow for differentiating between providers who excel and those who only just achieve the standard.

To mitigate these issues, governments creating and administering platforms for indicators of comparative performance should also incorporate consumer views of service quality into the performance measures (including through surveys or feedback mechanisms that account for consumer experiences with individual service providers). There is hostility to consumer feedback in some settings — exemplified by the mixed views about the value of student evaluations of teacher performance.⁶ However, opposition in some quarters should not imply that consumer feedback mechanisms are wrong-headed. Context matters in many studies (is it a school or a university, how many students filled in the survey and over what period, what is the nature of student assessment questionnaire, what is the goal of the assessment, what is the subject?). More might need to be done to refine and interpret the tools in their various settings, but the notion that feedback from students has no value should not be accepted uncritically. The same applies to the feedback from the customers of other government services. Similarly, contextual information from service providers on significant extraordinary factors that have skewed performance is also important for interpretation.

⁶ The literature on student evaluations ranges from negative (Stroebe 2016; Uttl, White and Gonzalez 2017), to positive (Benton and Ryalls 2016; Hativa 2014; Marsh and Roche 1997), with many in between, depending on context and settings (Kelly 2012; Kornell and Hausman 2016; Ottoboni, Boring and Stark 2016; Spooren, Brockx and Mortelmans 2013; Stark and Freishtat 2014).

Valid comparisons — risk-adjusted and value-added measures

Comparisons between providers also need to be valid. For example, students in selective schools perform on average better than non-selective schools, which might suggest that high-ability students should move to selective schools. However, the empirical evidence suggests that high-ability students actually perform worse than they would have, had they not moved (Marsh and Hau 2003). The key problem is selection bias — schools perform differently, but much of it is due to the different traits of their students, which are outside the control of the school. A similar concern applies to the interpretation of the differences in outcomes for, say the crude exam scores of a wealthy, private metropolitan school with a remote school in a disadvantaged region — the two facilities have large disparities in their available resources and the inherent characteristics of their students and families.

One partial remedy is to adjust performance indicators to remove the key factors that affect performance, but that are outside the control of any given provider. These ‘risk-adjusted’ measures aim to control the part of the different outcomes explained by the underlying risk characteristics of the client group (for example, people receiving treatment for disease Y who also have disease X, compared with those with disease Y and no co-morbidity). In essence, risk-adjustment measures the effectiveness of any provider in achieving an outcome after controlling for the different characteristics of the population.

While mostly used in contexts outside health care — particularly in schooling — the concept of ‘value-added’ adjustment is very similar, but with adjustments for expected outcomes and growth over time. In particular, value-added analysis focuses on the value that a provider has given to consumers, over and above what would be expected given their backgrounds and prior circumstances. For schooling, value-added outcomes are most commonly measured through the equivalent years of education achieved each school year — thus students who achieve more than one year’s worth of learning in a given year are likely being taught well, even if their overall test scores are comparably worse (Kim and Lalancette 2013; PC 2016b).

Risk-adjusted or value-added measures are intended to simulate the outcome of randomised-control trials, assessing the impacts of any given service provider (school, hospital, physician and so on) as if customers were randomly assigned to them. Without proper risk-adjustment in published performance data, providers may have perverse incentives to ‘cherry-pick’ the easiest customers in order to influence their results. For example, surgeons may choose to avoid treating more complex cases (Fung et al. 2008; PC 2015; Totten et al. 2012). Examples of how risk-adjustment occurs on the *My Hospitals* and *My School* websites are in box 4.

However, such adjustment is imperfect for many reasons. First, some of the factors that affect performance are not easily observable. Second, a ‘like with like’ comparison only relates to average performance outcomes, whereas many customers want to know how a particular service will perform for them. For instance, the fact that school A produces better outcomes on average than school B after controlling for the influences of the characteristics of the students, does not answer the question relevant to a parent: ‘How well

will *my* child do at this school rather than that one?' Sophisticated performance measures would take into account the nature of the customer, producing bespoke comparisons. It may be that the application of data analytics will ultimately head in this direction, but such an approach is still a way off for those services where it is most likely to apply (such as education or career advice). In the meantime, imperfect risk adjustment, appropriately interpreted, is probably the best that can be done.

Box 4 Risk-adjustment in practice

For schools, some risk-adjustment already occurs on the *My School* website through the use of the index of community socio-educational advantage (ICSEA). The index, developed by Australian Curriculum, Assessment and Reporting Authority (ACARA), accounts for characteristics such as the geographic location of the school, and the occupations and education of students' parents. The average value of ICSEA scores is set at 1000, with lower scores denoting greater disadvantage. Test results on the *My School* website are then typically compared with the group of 60 schools that have the closest ICSEA values.

For hospitals, the New South Wales Bureau of Health Information (BHI) has developed a 30-day Risk-Standardised Mortality Ratio (RSMR) indicator to highlight outlier hospitals in the state. The measure calculates a ratio of expected deaths (based on condition specific indicators, including gender, age and co-morbidities) to the deaths that were actually observed in the 30-days following hospital admission for selected conditions (including acute myocardial infarction, ischaemic stroke, haemorrhagic stroke, pneumonia, and hip fracture surgery).

In health care more broadly, a hospital's performance can also be adjusted for differences in the mix of patients treated (including their demographics, procedure type, length of stay and other factors, known as the 'casemix') in order for results to be comparable across providers and across time. As a basic example of this adjustment, the Australian Institute of Health and Welfare (AIHW) publishes rates of cancer incidence and mortality that are age-standardised over time (rather than just crude rates) to reflect Australia's ageing population and that cancer incidence and mortality strongly depend on age.

Sources: ACARA (2015); AIHW (2016a); NSW BHI (2013).

Measure things that matter for people

Outcomes often lie on a continuum. Hip replacements provide an illustration. Reporting only the complete restoration of mobility following a hip replacement would conceal many other outcomes that people regard as worthwhile, such as improved mobility or the ability to undertake an especially valued activity (such as walking upstairs). Although the ability to deliver complete restoration of mobility may be correlated with other good outcomes, such a relationship is not perfect and likely overlooks many partial, yet beneficial, outcomes. In addition, other measures of the impact of service delivery (such as haemoglobin A1C in the case of a diabetes patient) might not correspond to a person's assessment of their own health status, but both are equally relevant for their ongoing treatment (Chen 2016, p. 17).

Accordingly, performance measures should take account of people's own appraisals of the impacts associated with service delivery, including their experiences of that delivery. The concept is most advanced in health care, where Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs) are now commonplace, including in England, the Netherlands and Sweden (Williams et al. 2016). PROMS ask patients for their assessment of how interventions have affected their quality of life, capacity to undertake activities, symptoms of pain, distress levels, and other aspect of their health (for which there already well-developed instruments, such as WHODAS 2.0). PREMS relate to people's perceptions of their health care — such as waiting times, involvement in decision-making, the quality of communication, and the support they receive to manage a long-term condition (Verma 2015).

Not only can PREMs and PROMs serve to provide better information on performance, but they can also involve the patient as a more active participant in their own health care. However, Australia has largely not adopted PREMs or PROMs. This is likely to change, as the NSW and Victorian Governments are currently running pilot programs to collect this data (see SP 5).

The balance between too many and too few indicators

There can often also be a balance between reporting too many indicators and reporting too few. Providing consumers and users with too many indicators can greatly increase compliance costs for providers and potentially make it difficult for consumers to determine (on balance) which provider is better or worse across an array of different performance measures.

On the other hand, however, limiting performance reporting to only one or two indicators for a given sector can be problematic if the measured areas do not provide an adequate sense of performance. This can also create incentives for service providers to focus unduly on those aspects of service delivery where performance is measured, resulting in perverse and unintended outcomes. For instance, hospitals may discharge patients too early to free up hospital beds as a way to improve performance against narrow waiting-time criteria, while neglecting the effect this has on patient outcomes (Dunleavy 2016; PC 2015).

Ideally, well-designed performance measures would be few enough in number to be comprehensible to consumers, but also broad enough to cover every important aspect of performance. Ongoing consultation with the providers and the trial and testing of new data variables can be useful in discovering the key factors relevant to judgements about performance.

Make prices as well as performance indicators visible

Providing performance metrics to consumers so they can exercise informed choice loses some of its potency if consumers are ex ante unaware of the magnitude of any payments

they may need to make when choosing between the various service providers. This reflects that, like all goods and services, people have trade-offs between prices and quality. To put it simply, a consumer choosing between a ‘5 star’ provider and a ‘4 star’ provider based on some performance metric would not necessarily choose the former if their cost was several times that of the latter. For instance, in the United Kingdom, the National Institute for Health and Care Experience provides advice on treatment options covered by the NHS, based on incremental cost effectiveness ratios (which are quality-adjusted life years obtained from a treatment — a performance metric — *per dollar*) and not just on QALYs alone.

For some government-funded services, pricing uncertainty for consumers is minimal or unlikely to present many problems:

- parents make co-contributions to private schools, but the prices are clearly posted
- where a person goes to a GP who does not bulk bill, patients may not know the fee charged beforehand. However, as GP services are usually repeat services, prices for standard consultations (the most common service) are revealed over time. In most Australian locations, competition in general practice is also relatively strong (as suggested by high bulk billing rates), which helps to limit premiums above the scheduled fee.

However, there are likely to be significantly greater problems associated with consumer uncertainty about co-payments for medical specialists (Sivey 2016). Most people do not see the same specialists frequently or for the same service, so there is little scope for learning about prices. Moreover, GPs are often the gatekeepers for specialist services and may not know the co-payments that patients will face with different specialists, and neither patients nor GPs may want to discuss this as part of clinical consultations. This could affect the genuine exercise of choice and, because uncertainty itself acts as a cost, may deter people from undergoing diagnostics or treatments. Further difficulties also arise because the specialist market is not as competitive as the GP market and the share of services with an out-of-pocket cost are much higher (Hillis et al. 2017).⁷ Further, the variations between out-of-pocket costs are very large and generally vary by specialty (see figure 1).

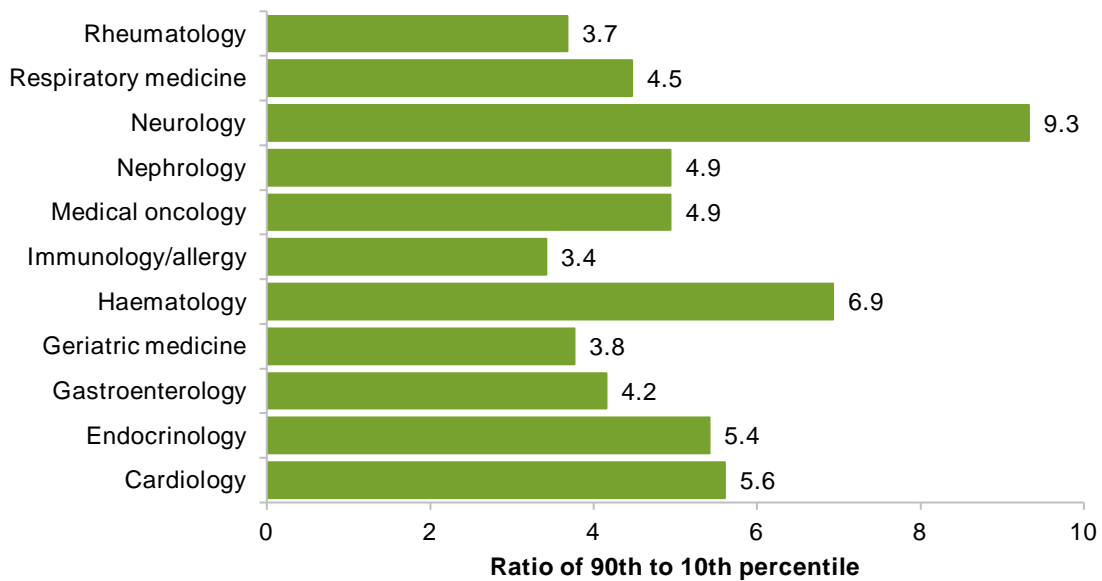
There are no websites that compare prices, as there are for many other consumer goods, a gap that some have recommended filling (McRae and Gool 2017; Sivey 2016; Taylor 2015). An editorial in the Medical Journal of Australia argued that:

Easier access to information may induce greater competition. Patients have little opportunity to verify claims of higher quality care by medical specialists, and it is difficult to shop around to find the best price. Unverified quality claims can lead to extensive price variation, despite there being little evidence that quality is correlated with price. (McRae and Gool 2017, p. 162)

⁷ The Australian Medical Association reports that in 2012-13, while only 18.9 per cent of GP attendances involved an out-of-pocket cost, this was 71.3 per cent for specialist attendances and 90.8 per cent for anaesthesia (AMA 2014).

While the analysis above applies to health care specialists, the general principle is that the public availability of performance indicators should be accompanied by transparency in pricing in any government-subsidised service where payments are expected from consumers and pricing variation between providers is significant.

Figure 1 **Variation in specialist out-of-pocket fees by specialty**
Australia, 2015



Source: Freed and Allen (2017).

3 Data availability and accessibility

The development of comparative performance indicators for individual government-funded service providers is a data-intensive task, requiring a range of different variables to be reported, collated, cleaned and published for each of thousands of service providers around Australia. As such, issues with the data sources underlying the performance indicators affect the quality of the indicators themselves.

Generally, issues with the underlying data fall in two camps: insufficient access to existing sources of data to determine specific indicators; and inadequate existing data sources (because they are not collected, are of low quality or are not comparable across suppliers or jurisdictions).

Making greater use of existing administrative data sources ...

Commonwealth, State and Territory Governments already collect vast quantities of information as by-products of its administrative functions. These data are collected for regulatory requirements (such as financial information for vocational education and training providers), program administration (for example, Centrelink and Medicare payments, school, university and vocational enrolments and completions, and hospital admissions) or as a byproduct of transactions (such as a purchase of health care services or fines and fees in the public administration system) (PC 2013, 2017).

Administrative data can be a rich vein of information because it is typically longitudinal and generally covers the full population of service-users instead of a sample, as well as largely avoiding non-response rates for individuals, participant attrition over time and many forms of under-reporting.⁸ Automated systems and routine collection also lower the cost of administrative data collection compared with standard labour-intensive survey methods (PC 2013).

There remain problems with such data for the development of performance indicators, mainly reflecting that the data are collected for administrative rather than evaluative purposes. Respondents and data collecting agencies make errors — so data requires cleaning to be usable. Data collections vary over time as policies and programs change. Variables are not always well-defined or documented (PC 2015). Administrative data also suffers from an inability for researchers to specify the scope of data that are collected in advance, as many crucial datasets for government-funded services are designed for purposes other than performance analysis. Their usefulness is therefore a welcome byproduct, but not always a planned outcome (PC 2015; Schreyer 2010).⁹

Further, different jurisdictions (particularly the states and territories) frequently have different definitions and collection standards, making the development of comparable performance indicators difficult — for example, there is considerable variation in the way hospitals code information about patient deaths (NHPA 2016).

Nevertheless, such data are a promising source of evidence on performance, and becomes more so if linked (by client or provider) across datasets. Arguably, the most significant barrier to the use of administrative data are accessibility — an issue that was a central concern of the Productivity Commission's inquiry into Data Availability and Use. While there is a need to meet privacy and confidentiality expectations of the community, Australian governments have tended to be overly risk-averse in providing access to administrative data, even on a private basis between government agencies or trusted users.

⁸ However, non-service-users are generally not included in administrative data (for example, the consumers of GP services are probably going to be less healthy than the general population), possibly limiting the usefulness for policy evaluations.

⁹ As discussed below, for example, MBS payments do not necessarily say much about the purpose of the visit to the GP.

Under one measure of accessibility, Australia’s provision of open access data lags that of comparable countries with similar governance structures — such as the United States, the United Kingdom and Canada. All outperform Australia in collecting and releasing health care data, including performance data on hospitals and administrative data on the use of health services (PC 2010a, 2013, 2015, 2016b, 2017). The United States government currently releases datasets containing over 100 measures of performance for over 4000 hospitals, ranging from operational measures to patient survey information, with nothing comparable existing in Australia (PC 2017).

Despite these challenges, there have been some improvements in utilising existing data sources. In particular, governments have gotten marginally better at ensuring that data collected by third-party, private providers of government-funded services are passed back to the regulating or contracting agency for further analysis (PC 2017). This largely occurs through standardised data sharing arrangements in contracts (such as under the guidelines published by TfNSW, noted in section 1).

However, further progress could be made, as agencies frequently collect large amounts of performance-related information from contractors, but only make use of a small portion when assessing performance and providing feedback. As this data collection can create considerable reporting burdens for contractors, there is room to improve through more accurate targeting of performance reporting requirements — collecting less data overall, but making better use of what is collected (PC 2010b). There are also concerns that excessive measurement and reporting requirements are hindering the benefits of social impact bonds (Edmiston and Nicholls 2017).

... while also covering gaps in data collections

There are also a range of areas where performance-related data either do not exist or the quality of what is collected is too low to be of any value. In health care, one such area is the nature of patient encounters in the primary health care sector (such as during GP consultations — box 5). For the education sector, there is also a lack of information about the characteristics of the education workforce and the student-level learning outcomes from early childhood education and the first year of primary school (PC 2016b).

Although additional data across the non-market sector would be useful, a key constraint of greater data collection is cost. To publish performance-related data for each individual service provider across the sector requires data to be collected from all of them. Often, doing this will not be practical or cost-effective — this is especially true for the collection of large-scale, time-consuming survey data to measure quality outcomes. Additional data collection requirements can also be seen as disruptive to the core activities of the service providers themselves, such as preventing doctors from attending to their patients or teachers from educating their students (PC 2017).

However, while governments often perceive additional data collection activities to be expendable relative to other functions, the savings can prove illusory when weighed

against the gains through better outcomes for consumers and greater cost-effectiveness for taxpayers. As such, increasing data collection may still be worthwhile in select, high-priority areas where the gains are likely to be large and outweigh the costs of collection, even if those costs may be considerable (PC 2010a). These cost constraints also mean that there is a need for careful consideration as to what sorts of information will be most useful for improving service delivery to customers, including what is most likely to inform policy design or prompt behavioural change.

Box 5 The black box of primary health care

In primary health care (such as visits to GPs and specialist clinics), there is very limited information available on the treatment and diagnostic details of individual patient encounters. This includes details of why an individual was visiting their primary health care professional, what their symptoms and diagnoses were, what treatment was prescribed, what the outcome was and whether follow-up treatments were needed. In 2008, the AIHW concluded that existing data sources in primary health care were severely limited and there was a pressing need for additional data collection to build a more comprehensive picture of activity to drive outcome improvements.

Although the Australian Government generally provides Medicare benefits for most patient encounters, there is no information collected on the contents of the encounter. Other data sources mostly rely on limited survey samples that are not broad enough to provide a comprehensive picture of primary health care activity and develop performance indicators for individual providers.

While there are genuine patient confidentiality concerns to be considered, many of the issues instead relate to individual health care providers maintaining their own siloed record-keeping arrangements and not sharing data. Historically, this has been exacerbated by the slow take-up of computerised patient records by the health care sector.

Further, since July 2016 one of the major datasets that shone a light on activities in the primary health care sector — the BEACH (Bettering the Evaluation and Care of Health) program — was discontinued after 18 years. The BEACH program randomly surveyed 1000 GPs a year on the details of 100 patient encounters, resulting in a dataset of 100 000 GP-patient encounters each year. Although not comprehensive enough to cover all GP encounters (about 127 million occur each year) and enable the development of robust performance indicators, the program nonetheless provided a sizable database of evidence in an otherwise largely unreported sector.

Despite these challenges, one area of recent progress for primary health care is the expansion of national eHealth records. Although adoption has been sporadic and there is not yet much existing research based on the data, the reinvigorated My Health Record system (see chapter 2 in the main report for further details) is likely to assist with providing additional details on primary health care treatments in coming years. In particular, as new patients sign up for the service and it becomes more widespread, it will become an increasingly useful source of data in the primary health care sector, as well as in hospitals and other health care areas more broadly.

Sources: AIHW (2008), Department of Health (2016a), FMRC (2016a, 2016b), PC (2017).

Use of technology can help to support the provision of credible information (as discussed in SP 13), particularly by minimising collection costs. For example, electronic online surveys are vastly cheaper to collect than paper ones (although response rates can be

lower), while automated email or text message notifications following interactions (such as through the MyGov portal once Medicare benefits are claimed for a GP appointment) can ensure low-cost delivery and notification (Couper 2011; Schuster and Perez Brito 2011). However, care is needed to ensure that such collection approaches do not result in a biased sample, that could lead to inferences that were not applicable to segments of the population for whom these collection methods do not apply.

Presentation and accessibility

Comparative performance indicators for use by consumers should be easily accessible. Recent New Zealand work recommended that the context of performance indicator publications must be explained, while results should be presented in a range of different formats to ensure that the information is not misinterpreted due to a failure to address different levels of technical literacy (HQSCNZ 2016). The *NHS Choices* website in the United Kingdom is an exemplar of a simple and accessible information source intended to provide informed choice (box 6).

As noted earlier, ready access to granular data should also be provided to third-party researchers (with the usual protections to ensure confidentiality and ethical use) so as to enable them to test and form their own conclusions.

Box 6 On Her Majesty's Surgical Service

NHS Choices provides information on health care services in England, including a complete list of all NHS providers across the country. This enables users to search for facilities that offer particular services nearby, such as accident and emergency departments, GPs, hospitals, dentists, pharmacies, specialists and care homes.

Besides general information on the listed providers — including their contact details, opening hours, services offered at the facility, available amenities (such as disability access and nearby parking) and the name of staff members (for GP clinics) — the *NHS Choices* site also provides details on the provider's performance based on survey responses and other metrics. For example, GP clinics are rated on whether patients would recommend the clinic to others, using results from the biannual National GP Patient Survey. The survey results are reported as a percentage rate, alongside a comparative score indicator, where the bottom 25 per cent of clinics are rated as 'among the worst', while the top 25 per cent are rated 'among the best'.

Another survey reports the proportion of hospital staff who would recommend their own facility, while others report on waiting times and convenience at GP clinics. Hospital ratings include measures on a procedure-by-procedure basis, such as waiting times and the results from Care Quality Commission inspections.

Source: NHS (2016).

4 Links to financial incentives

Adequate performance measures are a pre-requisite to payment-based incentives, such as financial rewards or penalties. The use of performance-based financial incentives has a long and chequered history, with a range of different types available, depending on the nature of the sector and the outcomes desired. Performance indicators linked to incentives do not necessarily have to be comparative indicators — creating financial incentives based on a provider’s performance against a given benchmark, rather than against competitors, can be equally useful (although under a basic design this can lead to uncontrolled costs, as the number of providers receiving payments is unknown *ex ante*¹⁰).

Experience in linking performance measures with financial incentives has been mixed, with some positive results, some negative, and a lot depending on context and design.

A key issue in the use of financial incentives linked to performance is the extent to which providers are motivated by financial rewards or sanctions. Much of this depends on the nature of the ‘business’ and market, governance structures, and what the financial rewards can be used for (such as personal income, spending to improve the work environment, or reinvestment in the business?).

Further, many of the motivations created by financial incentives occur in their absence anyway, as non-pecuniary initiatives (such as shaming through disclosure) and indirect financial measures (such as when consumers choose between providers based on their reported performance) act as their own incentives. As noted above, many government-funded services are also delivered by professionals whose primary motive may not be financial rewards from their work — these individuals are unlikely to be driven by performance-related pay.

The following section draws on evidence in the government administration, health and education sectors to examine how financial incentives linked to performance indicators have worked in practice.

Incentive regulations, capitation payments, activity-based funding and pay-for-performance

For regulated or government-owned natural monopolies, regulators often use ‘incentive regulation’ in which cost recovery from consumers cannot exceed some efficient benchmark level (such benchmarks are performance indicators by another name). In principle, providers have incentives to improve their efficiency because they retain any (or at least a share of) profits achieved from costs being below the benchmark level, and make losses if they exceed the benchmark. Penalties for not achieving a certain quality of service

¹⁰ For an example of how uncontrolled costs could occur, see SP 7 for discussion of performance-contingent funding for universities.

outcomes are also common in utility regulation (for example, for prolonged electricity network outages). The overall impacts of such incentive regulation on the efficiency of providers has been mixed, and has hinged on the exact regulatory design, and the capacity (and willingness) of governments to bail out poorly performing businesses. Governments can also issue directives that compromise efficiency (such as procurement rules, social obligations or quality standards). The same compromises often do not exist for private entities in markets without natural monopolies.

Capitation systems for remunerating health care providers share many characteristics with utility regulation, although the parallel is often not drawn. Payments under capitation are (intended to be) based on the efficient costs of providing services to a population with given risk characteristics. Australian Health Care Homes include this feature, and they are widespread in the United States health care system, particularly through health maintenance organisations. As discussed in chapter 2 of the main report, the design of capitation arrangements and the performance metrics that underpin them are critical to outcomes.

Activity-based funding (ABF) of hospitals is similar to capitation, but relates to the efficient costs of particular hospital services. In Australia, the Independent Hospital Pricing Authority makes an annual National Efficient Price determination for public hospitals for the coming year, which is the basis for the payments to hospitals. As with any financial incentive arrangements based on performance measures, a critical issue is that the measures be objective, precise, strongly correlated with good outcomes and resistant to cost shifting (such as pushing difficult clients onto other providers) or gaming (also known as ‘cherry picking’, where adequate risk adjustment has not occurred). Research overseas has found mixed results from ABF, such as reduced lengths of stay, shifts from high-cost inpatient care to outpatient care (which may be desirable), and a possible increase in readmissions (CIHI 2013, p. 5; Palmer et al. 2014). The most thorough (indeed best practice) meta study found that strong claims in favour or against ABF were not supported by the evidence, but that ABF had positive effects in some settings and not in others (Palmer et al. 2014). Context therefore clearly matters.

No system for managing health care (or any other non-market services) will provide perfect incentives for efficiency, which is hard to do even in market services where there is a clearer indicator of outcomes in terms of profits. The key question is whether the overall effect is positive or can be made to be so through finessing the model. Given the widespread adoption of ABF in Australia, monitoring and finessing is likely to be appropriate. This will involve decisions about payment levels, governance arrangements and the scope of performance indicators.

Pay-for-performance in health care

The potential value of disincentives for poor clinical outcomes has long been presumed. More recently, health purchasers have turned to financial incentives to encourage better clinical outcomes for patients. In particular, pay-for-performance (P4P) incentives have

been used to (ostensibly) encourage higher quality care in general practice — exemplified in Australia by the Australian Government’s Practice Incentives Program (PIP). This program has generally involved ‘performance’ measures of desirable processes¹¹ that the Government infers will enhance quality care, rather than measures of outcomes. The PIP has a range of deficiencies, being too complex, with high administrative costs and having inadequate data collection for the task. These problems have been recognised and it is currently subject to re-design (ANAO 2010; DoH 2016b). Nevertheless, there is some indication that it has influenced diabetes care (Oliver-Baxter et al. 2014). The international literature appears to suggest that P4P in health care ‘works’ or at least does not produce negative outcomes, but with the size of the effects dependent on context and the magnitude of the payment (Gee 2016; Ogundeji, Bland and Sheldon 2016; Partel 2014; Scott and Connelly 2011).

There also seems to be some promise for P4P in hospital settings, in which funders provide no payment for events that should never occur (sentinel events) and reduce payments for events that involve complications. Non-reimbursement for sentinel events in the United States appear to have been effective for some event types and no worse for others (Waters et al. 2015). The Council of Australian Governments (COAG) intends to introduce pricing incentives to reduce 16 Hospital Acquired Complications (HACs) in 2017, although stopping short of full non-reimbursement given that complications are often not fully avoidable (Gee 2016; Herkes 2016; IHPA 2016).

In a much more radical move, Medibank (Australia’s largest private insurer) has introduced non-reimbursement for 165 hospital-acquired complications. This initiative has proved controversial because of the large number of non-compensable complications, the cost of implementation, the way in which HACs have been determined, and the risk that, where complication risks are high, private hospitals may attempt to divert patients to public hospitals. In this case, the claim is that the performance indicators have been selected more to reduce pressures on premiums than to reduce adverse events — a claim the Commission has not tested — but which, in principle, illustrates another element of the complexities of linking performance measures and financial incentives.

Performance pay in teaching and the public service

While performance-related incentives are widely used in health care and utilities, internationally, their genuine adoption in the public service and in teaching is patchy. Prima facie, the contention that they should be used appears sound, and reputable parties have urged their adoption (Jensen and Reichl 2011). The use of performance pay in the Australian and State and Territory Government public services has waxed and waned (for example, being axed in Queensland in 2015). The Australian Government has recently

¹¹ Examples are payments to GPs for screening women between 20-69 years who have not had a cervical screen within the past four years.

announced plans to eventually provide funding to schools contingent on performance-based pay for teachers (Australian Government 2016).

Globally, the issue of performance pay for teachers and public servants is controversial, and accompanied by mixed and contested evidence about their benefits, sometimes infected with ideology. Most of the best (of a lot of bad) evidence relates to teachers. The results of performance pay depend on context, place and time. For example, in the United Kingdom, over a succession of studies, a researcher found that teacher performance pay in the United Kingdom appeared to produce positive results initially, which then subsided (Marsden 2009, 2015; Marsden and Belfield 2006). Teachers themselves have mixed views — with recent UK survey evidence suggesting that a (slight) majority supported some link between performance and pay (Ware et al. 2014).

The Organisation for Economic Cooperation and Development (OECD) found no average relationship between student performance on Programme for International Student Assessment (PISA) tests and the presence of performance pay. However, it did find a positive effect if teachers' base salaries were low and a negative effect if base salaries were high (OECD 2012). Australia falls into the latter category. Nevertheless, the result was based on very simple analytics for a single year of data, and ignored the possibility that countries that were concerned about their relative PISA standing might try to improve outcomes through performance pay. A recent comprehensive meta-study found a sufficient number of studies favourable to performance-based pay in teaching to warrant further trials and evaluations (Leigh 2013).

A lack of consensus

One of the key problems in appraising the impacts of pay for performance in teaching is that there is no consensus on:

- the form and size of financial incentive — such as a payment for high performance, variable bonuses for variable performance, progression in pay scales based on annual assessment, access to additional teaching resources, or providing non-personal financial rewards (such as additional school funding).
- the appropriate measure of performance — for example, test results, truant rates, student and parent appraisal results, achievement of some agreed standard of teaching, engagement in processes linked to performance such as professional development, or some mixture of the above.
 - It cannot be said that there is no mechanism for assessing the 'art of teaching'. After all, students training to be teachers are assessed on a proficiency rather than a competency-based standard (see chapter 3 of the main report for a discussion of the differences). There have been advances in the development of recognised teaching standards that more accurately reflect the nature of teachers' work and which could be a basis for new performance pay initiatives.

-
- the appropriate parties who should make the assessments (external to the school, peers, the principal)
 - the extent to which there needs to be agreement between the main actors in schools — governments, school administrators, principals, teachers, school boards, parents and children — about the legitimacy and accuracy of the measures of quality. At least one authoritative study has attributed failure of performance pay on the absence of buy-in by teachers and school administrators (Ingvarson, Kleinhenz and Wilkinson 2007)
 - the relative importance of performance-based pay compared with other initiatives that might improve student performance, bearing in mind that all school-based initiatives entail implementation and financial costs. Performance pay might produce a benefit, but it might be more or less than some other interventions.

Given the wide variety of outcomes from experiments in teacher pay performance around the world, any definitive declarations that they *do* or *do not* work are to be treated cautiously, as outcomes are highly contextual (that is, they depend on what incentives were used, for whom, and under what conditions).

References

- ACARA (Australian Curriculum, Assessment and Reporting Authority) 2015, *About ICSEA - fact sheet*.
- 2016, *My School*, www.myschool.edu.au (accessed 18 November 2016).
- ACSQHC (Australian Commission on Safety and Quality in Health Care) 2015, *Australian Atlas of Healthcare Variation*, November.
- AIHW (Australian Institute of Health and Welfare) 2008, *Review and Evaluation of Australian Information about Primary Health Care: A Focus on General Practice*, December.
- 2016a, *Australian Cancer Incidence and Mortality (ACIM) books*, January.
- 2016b, *MyHealthyCommunities*, www.myhealthycommunities.gov.au (accessed 1 November 2016).
- 2016c, *MyHospitals*, www.myhospitals.gov.au (accessed 16 November 2016).
- 2017, *Impacts of Local Level Public Reporting*, www.aihw.gov.au/health-performance/impacts-of-local-level-public-reporting/ (accessed 19 January 2017).
- AMA (Australian Medical Association) 2014, *AMA Submission on Out-of-pocket Costs in Australian healthcare*, Submission to the Senate Community Affairs References Committee's inquiry into out-of-pocket costs.
- ANAO (Australian National Audit Office) 2010, *Practice Incentives Program*, Performance Audit, 5 of 2010-11.
- Andrews, D., Criscuolo, C. and Gal, P. 2015, *Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries*, Productivity Working Paper No. 2, November, OECD Publishing, Paris.
- Atkinson, T. 2005, *Atkinson Review: Final Report - Measurement of Government Output and Productivity for the National Accounts*, United Kingdom.
- Australian Government 2015, *The National Quality Standard — My Child*, www.mychild.gov.au/agenda/quality-framework/quality-standard (accessed 11 January 2017).
- 2016, *Quality Schools, Quality Outcomes*, May.
- Barratt, H. 2011, 'Increasing patient choice', *The Lancet UK Policy Matters*, 7 May.
- Benton, S. and Ryalls, K. 2016, 'Challenging misconceptions about student ratings of instruction', *IDEA*, vol. 58, April.
- Bergs, J., Hellings, J., Cleemput, I., Zurel, Ö., De Troyer, V., Van Hiel, M., Demeere, J., Claeys, D. and Vandijck, D. 2014, 'Systematic review and meta-analysis of the effect

-
- of the World Health Organization surgical safety checklist on postoperative complications', *The British Journal of Surgery*, vol. 101, no. 3, pp. 150–158.
- Berwick, D., James, B. and Coye, M. 2003, 'Connections between quality measurement and improvement', *Medical Care*, vol. 41, no. 1, pp. 130–138.
- Carpenter, A., Elshaug, A., Turnbull, F., Wilson, A. and Jan, S. 2014, *Health policy, Administration and Expenditure*, Submission to the Senate Select Committee on Health.
- Chen, J. 2016, *Integrated Care: Patient-Reported Outcome Measures and Patient-Reported Experience Measures: A rapid scoping review*, NSW Agency for Clinical Innovation.
- CIHI (Canadian Institute for Health Information) 2013, *Evaluating the Impact of Activity-Based Funding on Health System Performance*.
- CMS (Centers for Medicare and Medicaid Services) 2016, *2015 Individual Clinician Measures*.
- Colvin, L., Slack-Smith, L., Stanley, F. and Bower, C. 2010, 'Linking a pharmaceutical claims database with a birth defects registry to investigate birth defect rates of suspected teratogens', *Pharmacoepidemiology and Drug Safety*, vol. 19, no. 11, pp. 1137–1150.
- Conway, P. 2016, *Achieving New Zealand's Productivity Potential*, New Zealand Productivity Commission.
- Coulter, A. 2010, 'Do patients want a choice and does it work?', *BMJ*, vol. 341.
- Couper, M. 2011, 'The future of modes of data collection', *Public Opinion Quarterly*, vol. 75, no. 5, p. 889.
- Deloitte Access Economics 2016, *A Practical Guide to Understanding Social Costs: Developing the Evidence Base for Informed Social Impact Investment*, February.
- DET (Department of Education and Training) 2015, *Upholding Quality — Quality Indicators for Learning and Teaching*, www.education.gov.au/upholding-quality-quality-indicators-learning-and-teaching (accessed 11 January 2017).
- 2016a, *My School*, www.education.gov.au/my-school (accessed 11 January 2017).
- 2016b, *MyChild*, www.education.gov.au/mychild (accessed 11 January 2017).
- Dixon, A., Robertson, R., Appleby, J., Burge, P., Devlin, N. and Magee, H. 2010, *Patient choice*, Picker Institute Europe; The Office of Health Economics; RAND Europe (funded by the Kings Fund).
- DoF (Department of Finance) 2016, *Resource Management Guide No. 130: Overview of the Enhanced Commonwealth Performance Framework*, July.
- DoH (Department of Health) 2016a, *Annual Medicare Statistics*, August.
- 2016b, *Consultation Paper: Redesigning the Practice Incentives Program*.

-
- Dunleavy, P. 2016, Public sector productivity: Measurement challenges, performance information, and prospects for improvement, presented at The 12th Annual Meeting of the OECD Senior Budget Officials Performance and Results Network, Paris, November.
- Edmiston, D. and Nicholls, A. 2017, 'Social impact bonds: The role of private capital in outcome-based commissioning', *Journal of Social Policy*, pp. 1–20.
- Faber, M., Bosch, M., Wollersheim, H., Leatherman, S. and Grol, R. 2009, 'Public reporting in health care: how do consumers use quality-of-care information? A systematic review', *Medical Care*, vol. 47, no. 1, pp. 1–8.
- FMRC (Family Medicine Research Centre) 2016a, Closure of the BEACH program after 18 years, Media statement, www.sydney.edu.au/medicine/fmrc/media/BEACH-closure-2016-04.php (accessed 1 December 2016).
- 2016b, *General Practice Activity in Australia 2015-16*, September, University of Sydney.
- Freed, G. and Allen, A. 2017, 'Variation in outpatient consultant physician fees in Australia by specialty and State and Territory', *Medical Journal of Australia*, vol. 206, no. 4.
- Fung, C., Lim, Y., Mattke, S., Damberg, C. and Shekelle, P. 2008, 'Systematic review: The evidence that publishing patient care performance data improves quality of care', *Annals of Internal Medicine*, vol. 148, no. 2, pp. 111–123.
- Gee, C. 2016, Pay-for-performance: Lessons for Australia, presented at the *The Royal Australasian College of Medical Administrators Conference, 12-14 October*, The Royal Australasian College of Medical Administrators.
- Harding, A., Sanders, F., Lara, A., van Teijlingen, E., Wood, C., Galpin, D., Baron, S., Crowe, S. and Sharma, S. 2014, 'Patient choice for older people in English NHS primary care: Theory and practice', *ISRN Family Medicine*, vol. 2014.
- Hativa, N. 2014, *Student Ratings of Instruction: Recognizing Effective Teaching*, 2nd edn, Oron Publications.
- Haugen, A., Sjøfteland, E., Almeland, S., Sevdalis, N., Vonen, B., Eide, G., Nortvedt, M. and Harthug, S. 2015, 'Effect of the World Health Organization checklist on patient outcomes: A Stepped Wedge Cluster Randomized Controlled Trial', *Annals of Surgery*, vol. 261, no. 5.
- Herkes, R. 2016, 'Developing a set of hospital acquired complications: - what's that got to do with Activity Based Funding?', presented at the *Developing a List of Hospital Acquired Complications*, Brisbane, 9 May.
- Hillis, D., Watters, D.A., Malisano, L., Bailey, N. and Rankin, D. 2017, 'Variation in the costs of surgery: seeking value', *Medical Journal of Australia*, vol. 206, no. 4.

-
- HQSCNZ (Health Quality and Safety Commission New Zealand) 2016, *Position paper on the Transparency of Information Related to Health Care Interventions*, Position paper, March, Wellington, New Zealand.
- IHPA (Independent Hospital Pricing Authority) 2016, *Consultation Paper on the Pricing Framework for Australian Public Hospital Services 2017-18*, September.
- Ingvarson, L., Kleinhenz, E. and Wilkinson, J. 2007, *Research on Performance Pay for Teachers*, 3–2007, Teaching and Leadership Research Program, Australian Council for Educational Research.
- Jensen, B. and Reichl, J. 2011, *Better Teacher Appraisal and Feedback: Improving Performance*, Grattan Institute.
- Jones, L. and Mays, N. 2009, *Systematic Review of the Impact of Patient Choice of Provider in the English NHS*, London School of Hygiene and Tropical Medicine, United Kingdom.
- Kelly, M. 2012, *Student Evaluations of Teaching Effectiveness: Considerations for Ontario Universities*, Wilfrid Laurier University, Waterloo, Ontario.
- Ketelaar, N., Faber, M., Flottorp, S., Rygh, L., Deane, K. and Eccles, M. 2011, 'Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations', *Cochrane Database of Systematic Reviews*, issue 11, no. CD004538.
- Kim, H. and Lalancette, D. 2013, *Literature Review on the Value-Added Measurement in Higher Education*, Organisation for Economic Co-operation and Development.
- Kornell, N. and Hausman, H. 2016, 'Do the best teachers get the best ratings?', *Frontiers in Psychology*, vol. 7.
- Leigh, A. 2013, 'The economics and politics of teacher merit pay', *CESifo Economic Studies*, vol. 59, no. 1, pp. 1–33.
- Local Government Victoria 2015, *About Know Your Council*, knowyourcouncil.vic.gov.au/about (accessed 11 January 2017).
- Lowes, R. 2015, 'Individual clinicians now get stars on Physician Compare Site', *Medscape*.
- Lyons, V. and Popejoy, L. 2014, 'Meta-analysis of surgical safety checklist effects on teamwork, communication, morbidity, mortality, and safety', *Western Journal of Nursing Research*, vol. 36, no. 2, pp. 245–261.
- Marsden, D. 2009, *The Paradox of Performance Related Pay Systems: 'Why Do We Keep Adopting Them in the Face of Evidence that they Fail to Motivate?'*, August, #946, CEP Discussion Paper, Centre for Economic Performance LSE, United Kingdom.
- 2015, *Teachers and Performance Pay in 2014: First Results of a Survey*, February, #1332, CEP Discussion Paper, Centre for Economic Performance LSE, United Kingdom.

-
- and Belfield, R. 2006, *Pay for Performance Where Output is Hard to Measure: the Case of Performance Pay for School Teachers*, August, #747, CEP Discussion Paper, Centre for Economic Performance LSE, United Kingdom.
- Marsh, H. and Hau, K. 2003, 'Big-fish-little-pond effect on academic self-concept: A cross-cultural (26 country) test of the negative effects of academically selective schools', *American Psychologist*, vol. 58, no. 5, pp. 364–376.
- and Roche, L. 1997, 'Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility', *American Psychologist*, vol. 52, no. 11, pp. 1187–1197.
- McGowan, M., Andrews, D. and Millot, V. 2017, *The Walking Dead? Zombie Firms and Productivity Performance in OECD Countries*, OECD Publishing.
- McRae, I. and Gool, K. van 2017, 'Variation in the fees of medical specialists: problems, causes, solutions', *Medical Journal of Australia*, vol. 206, no. 4.
- NHPA (National Health Performance Authority) 2016, *Towards public reporting of standardised hospital mortality in Australia*, Progress report, February.
- NHS (National Health Service) 2016, *NHS Choices*, www.nhs.uk (accessed 25 November 2016).
- NICS (National Institute of Clinical Studies) 2003, *Evidence-Practice Gaps Report*, 1.
- 2005, *Evidence-Practice Gaps Report*, 2.
- NSW BHI (NSW Bureau of Health Information) 2013, *Spotlight on Measurement: 30-day Mortality Following Hospitalisation, Five Clinical Conditions, NSW, July 2009 – June 2012*, December, NSW Government.
- OECD (Organisation for Economic Co-operation and Development) 2012, *Does Performance-based Pay Improve Teaching?*, #16, PISA in Focus, Paris.
- 2017, *Ineffective Spending and Waste in Health Care Systems: Framework and Findings*, Paris.
- Ogundeji, Y., Bland, J. and Sheldon, T. 2016, 'The effectiveness of payment for performance in health care: A meta-analysis and exploration of variation in outcomes', *Health Policy*, vol. 120, no. 10, pp. 1141–1150.
- Oliver-Baxter, J., Brown, L., Raven, M. and Bywood, P. 2014, *Quality Improvement Financial Incentives for General Practitioners*, PHCRIS Policy Issue Review, May, Primary Health Care Research & Information Service (PHCRIS).
- Ottoboni, K., Boring, A. and Stark, P. 2016, 'Student evaluations of teaching (mostly) do not measure teaching effectiveness', *ScienceOpen Research*.
- Palmer et al. 2014, 'Activity-based funding of hospitals and its impact on mortality, readmission, discharge destination, severity of illness, and volume of care: A systematic review and meta-analysis', *PLOS ONE*, vol. 9, no. 10.

-
- Panesar, S., Cleary, K., Sheikh, A. and Donaldson, L. 2009, 'The WHO checklist: a global tool to prevent errors in surgery', *Patient Safety in Surgery*, vol. 3, p. 9.
- Partel, K. 2014, *Can We Improve the Health System with Pay-for-performance?*, Deeble Issues Brief, May, 5, Australian Healthcare & Hospitals Association.
- PC (Productivity Commission) 2010a, *Annual Report 2009-10*, Annual Report.
- 2010b, *Contribution of the Not for Profit Sector*, Research Report.
- 2013, *Annual Report 2012-13*, Annual Report.
- 2015, *Efficiency in Health*, Research Paper.
- 2016a, *Introducing Competition and Informed User Choice into Human Services: Identifying Sectors for Reform*, Research Report.
- 2016b, *National Education Evidence Base*, Inquiry Report.
- 2017, *Data Availability and Use*, Inquiry Report.
- RACS (Royal Australasian College of Surgeons) 2015, *Public Reports on Surgical Outcomes and Performance*, Position paper, FES-PST-056.
- 2016, *Submission to Productivity Commission inquiry into Introducing Competition and Informed User Choice into Human Services: Identifying Sectors for Reform - Preliminary Findings Report*, November.
- Ragusa, P.S., Bitterman, A., Auerbach, B. and Iii, W.A.H. 2016, 'Effectiveness of surgical safety checklists in improving patient safety', *Orthopedics*, vol. 39, no. 2, pp. 307–310.
- Rajendram, R. 2016, *Refusing Access to Surgery Recovery Area at a UK Hospital Unless WHO Safe Surgery Checklist is Fully Complete*, May, European Society of Anaesthesiology.
- Ringard, Å. and Hagen, T. 2011, 'Are waiting times for hospital admissions affected by patients' choices and mobility?', *BMC health services research*, vol. 11, p. 170.
- , Saunes, I. and Sagan, A. 2016, 'The 2015 hospital treatment choice reform in Norway: Continuity or change?', *Health Policy*, vol. 120, no. 4, pp. 350–355.
- Runciman, W., Hunt, T., Hannaford, N., Hibbert, P., Westbrook, J., Coiera, E., Day, R., Hindmarsh, D., McGlynn, E. and Braithwaite, J. 2012, 'CareTrack: assessing the appropriateness of health care delivery in Australia', *Medical Journal of Australia*, vol. 197, no. 2, pp. 100–105.
- Russell, L. and Doggett, J. 2015, *Tackling Out-of-pocket Healthcare Costs*, Discussion Paper, January, Menzies Centre for Health Policy.
- Schreyer, P. 2010, *Towards Measuring the Volume Output of Education and Health Services: A Handbook*, OECD Statistics Working Papers, 2010/02, OECD Publishing, Paris.

-
- Schuster, C. and Perez Brito, C. 2011, *Lessons from a Cell Phone-Based Beneficiary Survey to Strengthen Guatemala's Conditional Cash Transfer Program*, En Breve, February, 166, World Bank.
- Scott, A. and Connelly, L. 2011, 'Financial incentives and the health workforce', *Australian Health Review*, vol. 35, no. 3, pp. 273–277.
- SCRGSP (Steering Committee for the Review of Government Service Provision) 2015, *Report on Government Services 2015*, Productivity Commission.
- 2017, *Report on Government Services 2017*, Productivity Commission.
- Sivey, P. 2016, 'How much?! Seeing private specialists often costs more than you bargained for', *The Conversation*, www.theconversation.com/how-much-seeing-private-specialists-often-costs-more-than-you-bargained-for-53445 (accessed 15 March 2017).
- Spooren, P., Brockx, B. and Mortelmans, D. 2013, 'On the validity of student evaluation of teaching: The state of the art', *Review of Educational Research*, vol. 83, no. 4, pp. 598–642.
- SRC (Social Research Centre) 2016, About this site — quality indicators for learning and teaching, www.qilt.edu.au/about-this-site (accessed 10 January 2017).
- Stark, P. and Freishtat, R. 2014, 'An evaluation of course evaluations', *ScienceOpen Research*.
- Stroebe, W. 2016, 'Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations', *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, vol. 11, no. 6, pp. 800–816.
- Swan, N. 2015, Many Australian surgeons not following international surgical checklist, 14 October, Health Report, ABC.
- Taylor, A. 2015, 'Time to empower patients in healthcare', *The Sydney Morning Herald*, 30 October.
- TfNSW (Transport for New South Wales) 2014, *Guideline for Construction Contractor Performance Reporting*.
- Totten, A., Wagner, J., Tiwari, A., O'Haire, C., Griffin, J. and Walker, M. 2012, *Public Reporting as a Quality Improvement Strategy, Closing the Quality Gap: Revisiting the State of the Science*, Evidence Reports/Technology Assessments, Evidence Report no. 208, Agency for Healthcare Research and Quality, United States.
- Uttl, B., White, C. and Gonzalez, D. 2017, 'Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related', *Studies in Educational Evaluation*, vol. 54, September, pp. 22–42.
- Verma, R.K. 2015, *Overview: What are PROMS and PREMs?*, NSW Agency for Clinical Innovation.

Vrangbaek, K., Østergren, K., Birk, H.O. and Winblad, U. 2007, 'Patient reactions to hospital choice in Norway, Denmark, and Sweden', *Health Economics, Policy, and Law*, vol. 2, no. 2, pp. 125–152.

WA DTF (Western Australian Department of Treasury and Finance) 2004, *Outcome Based Management: Guidelines for Use in the Western Australia Public Sector*, November.

Ware, G., Ryan, C., Stevenson, H. and Curran, T. 2014, 'Teachers remain divided on performance-related pay', *The Conversation*, www.theconversation.com/teachers-remain-divided-on-performance-related-pay-27664 (accessed 22 March 2017).

Waters, T., Daniels, M., Bazzoli, G., Perencevich, E., Dunton, N., Staggs, V., Potter, C., Fareed, N., Liu, M. and Shorr, R.I. 2015, 'Effect of Medicare's nonpayment for hospital-acquired conditions: Lessons for future policy', *JAMA Internal Medicine*, vol. 175, no. 3, pp. 347–354.

Williams, K., Sansoni, J., Morris, D., Grootemaat, P. and Thompson, C. 2016, *Patient-reported Outcome Measures: Literature Review*, ACSQHC.