



# **Advances in measuring healthcare productivity**

Research paper

**The Productivity Commission acknowledges the Traditional Owners of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to their Cultures, Country and Elders past and present.**

## **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](http://www.pc.gov.au)).

© Commonwealth of Australia 2024



With the exception of the Commonwealth Coat of Arms and content supplied by third parties, this copyright work is licensed under a Creative Commons Attribution 4.0 International licence. 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. The licence can be viewed at: <https://creativecommons.org/licenses/by/4.0>.

The terms under which the Coat of Arms can be used are detailed at: [www.pmc.gov.au/government/commonwealth-coat-arms](http://www.pmc.gov.au/government/commonwealth-coat-arms).

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.

ISBN 978-1-74037-786-7 (online)

ISBN 978-1-74037-785-0 (print)

An appropriate reference for this publication is:

Productivity Commission 2024, *Advances in measuring healthcare productivity*, Research paper, Canberra

Publication enquiries:

Phone 03 9653 2244 | Email [publications@pc.gov.au](mailto:publications@pc.gov.au)

# Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>Overview</b>	<b>1</b>
We've spent more on healthcare, but it's been worth it	2
Australia is getting good value for its healthcare dollar relative to international counterparts	3
Australia's relatively good performance is not grounds for complacency	4
<b>1. Why measuring healthcare productivity matters</b>	<b>5</b>
1.1 Why focus on improving measures of healthcare productivity?	6
1.2 What do existing measures of productivity tell us and how can they be improved?	8
1.3 What can't improved productivity estimates tell us?	10
<b>2. Estimating healthcare productivity growth</b>	<b>11</b>
2.1 Productivity growth estimates help us understand if additional health spending has been 'worth it'	12
2.2 Healthcare productivity growth has been robust	14
2.3 Results largely align with international and on-the-ground experience	16
<b>3. Benchmarking Australia's healthcare productivity</b>	<b>21</b>
3.1 Cross-country comparisons can highlight scope for improvement	22
3.2 Australia is getting good value for its healthcare dollar relative to its international counterparts	24
<b>Appendices</b>	<b>27</b>
<b>A. Consultation</b>	<b>29</b>
<b>B. Traditional approaches to estimating healthcare productivity growth</b>	<b>31</b>
<b>C. How we estimated healthcare productivity growth</b>	<b>35</b>
<b>D. How we estimated changes in quality of life</b>	<b>55</b>
<b>E. How we benchmarked Australia's healthcare productivity</b>	<b>65</b>
<b>References</b>	<b>75</b>

## Acknowledgments

The Commission acknowledges the extensive support it received from experts in healthcare productivity, data and measurement to undertake this work.

Our thanks for feedback on our approach go to:

- Professor David Cutler (Harvard University); Adjunct Associate Professor John Goss (University of Canberra); Dr Ralph Lattimore; Dr Son Nghiem (Australian National University); and Dr Calvin Ackley, Dr Abe Dunn and Dr Justine Mallatt (United States Bureau of Economic Analysis) for their insightful comments on drafts of parts of this paper
- Professor Tony Blakely (University of Melbourne); Professor Kevin Fox (University of New South Wales); Adjunct Associate Professor John Goss (University of Canberra); Mr Jason Annabel, Mr Derek Burnell and Ms Katherine Keenan (Australian Bureau of Statistics); and Mr Geoff Callaghan, Ms Melissa Goodwin, Ms Michelle Gourley and Mr Nick Mann (Australian Institute of Health and Welfare) for participating in a roundtable on methodologies for estimating healthcare productivity growth.

Our thanks for assistance with data go to:

- staff from the Australian Institute of Health and Welfare's Burden of Disease and Mortality Unit and Health Economics Unit for supplying early releases of data and unpublished data used in the paper
- staff from the Australian Bureau of Statistics' Data Integration Services Branch for facilitating access to data used in the paper
- staff from the Australian Data Archive for facilitating access to data used in the paper.

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the authors and should not be attributed to either DSS or the Melbourne Institute.

The Commission also acknowledges the staff who worked on this paper. The work was overseen by Commissioner Catherine de Fontenay and Assistant Commissioner Dominique Lowe. The team was led by Lawson Ashburner and comprised Ethan Li, Lachlan O'Neil, Anuraag Roy and James Smith.

# Overview

## Key points

- \* **Contrary to prevailing views, new Commission research suggests that parts of the healthcare sector have experienced robust productivity growth in recent years.**
  - Quality-adjusted productivity grew by about 3% per year across the subset of diseases we studied.
- \* **The biggest contributions to productivity growth haven't come from doing more with less, but rather from providing more effective healthcare services.**
  - Quality improvements, not cost reductions, were the big drivers of productivity growth, and the vast majority of these have come from advances in saving lives.
- \* **Not only has Australia experienced robust productivity growth, Commission research also finds that our healthcare sector is one of the world's most productive.**
  - Australia's healthcare productivity ranks third among 28 high-income countries once we account for behavioural and environmental risk factors and the age of our population.
- \* **Australia's relatively good performance is not grounds for complacency.**
  - Productivity gains have not been universal. While we have made big quality gains from advances in saving lives, we have made fewer, if any, gains in improving quality of life.
  - Although welcome, quality-driven productivity improvements have done little to ease healthcare's growing fiscal burden.
  - Reducing our sizeable risk factors, such as obesity and alcohol consumption, would enable our healthcare sector to do more with less.

Healthcare spending is on the rise – up from 8% of GDP in 2000-01 to around 10% today. Population ageing, societal expectations and the growing burden of chronic disease means this trend is unlikely to abate.

As healthcare spending continues to grow, governments and taxpayers alike want to know: Are we getting better health outcomes with the extra money we have spent, and is there scope to do better?

## We've spent more on healthcare, but it's been worth it

While we have spent more on healthcare, our research shows this has been more than outweighed by improvements in outcomes delivered by the healthcare system.

For the first time in Australia, the Commission has sought to measure healthcare productivity growth in a way that accounts for improvements in healthcare quality.

Our research suggests that parts of the healthcare sector have experienced robust growth. Quality-adjusted multifactor productivity grew by about 3% per year between 2011-12 and 2017-18 for the subset of the sector we studied, healthcare used to treat: cancers, cardiovascular diseases, blood and metabolic disorders, endocrine disorders and kidney and urinary diseases. Together, treatment of these diseases accounts for around one third of healthcare spending. To put this in perspective, multifactor productivity growth in the market sector was estimated to be around 0.8% per year over the same period.

Quality improvements, not cost reductions, were the big contributors to productivity growth. Health practitioners' ability to understand, diagnose and prescribe has been transformed by medical advances.

These quality improvements have made us vastly better off. Indeed, the biggest contribution to productivity growth has come from advances in saving lives.

But productivity gains have been far from universal. We have made fewer, if any, gains from reducing the burden of ill health for people living with studied diseases.

The incidence of the productivity gains gives us some insights into the drivers. Productivity grew particularly strongly in the treatment of cancers, suggesting that advances in treatments, rather than across-the-board healthcare reforms, have been the major drivers of growth. Studies from abroad likewise find that diffusion of new treatments is a big contributor to productivity.

## Australia is getting good value for its healthcare dollar relative to international counterparts

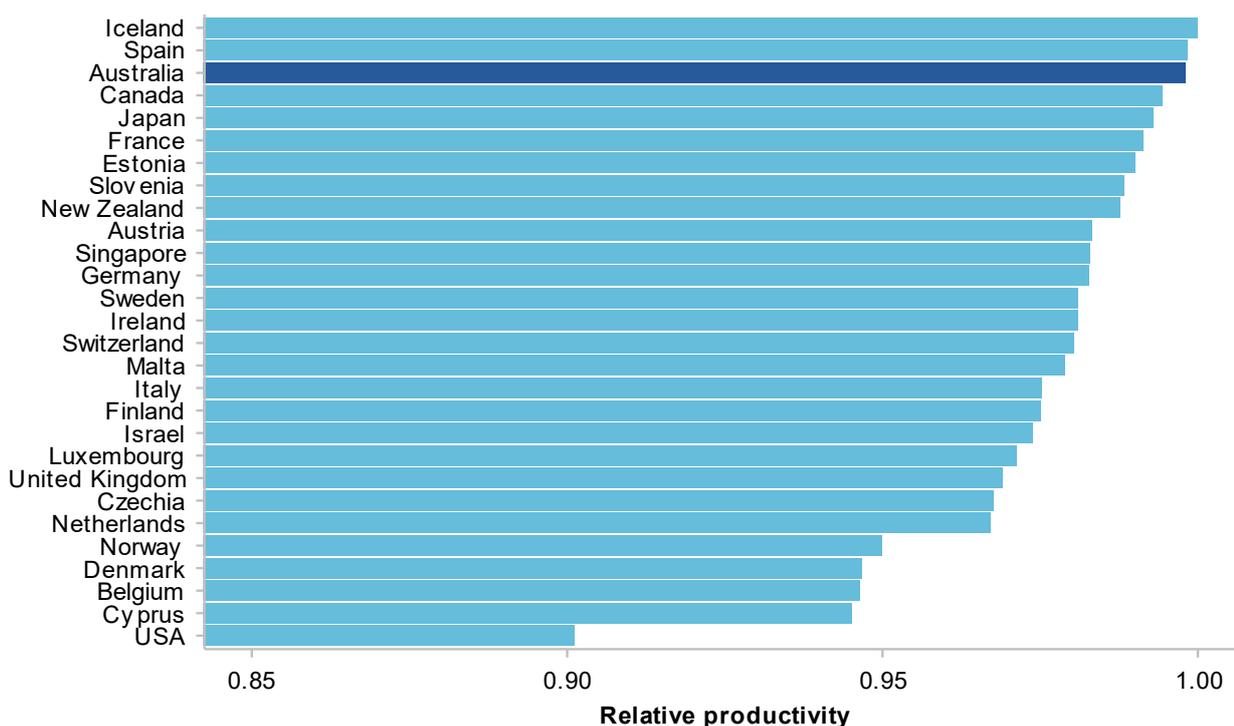
Not only has Australia experienced robust productivity growth, our healthcare sector is among the most productive in the world.

Our research shows that Australia’s healthcare productivity ranks third among 28 high-income countries, once we account for the age of our population and behavioural and environmental risk factors such as smoking, obesity, diet and alcohol consumption.

Adjusting for these risk factors makes a significant difference to our ranking. While Australia has relatively low smoking rates, we have the fourth highest obesity rate, the sixth highest level of alcohol consumption per capita and relatively low consumption of fruit and vegetables per capita. Of these risk factors, accounting for obesity is the most important – cross-country differences in obesity had larger effects on health outcomes than any other behavioural or environmental risk factor.

**Figure 1 – Australia’s healthcare sector is highly productive**

**Average estimated relative healthcare productivity level by country, 2010 to 2019**



Source: Productivity Commission estimates.

## **Australia's relatively good performance is not grounds for complacency**

The Commission's research suggests that Australia is getting relatively good, and increasing, value for its healthcare dollar.

But an ageing population, societal expectations and the growing burden of chronic disease all mean Australia can't sit on its hands when it comes to healthcare reform.

Our challenge lies not in catching up to the frontier but in shifting it, to make our healthcare dollars go further. Past work by the Commission on chronic care and mental health have laid out several productivity enhancing reforms.

This paper highlights three further candidates.

First, reducing our sizeable risk factors, such as obesity and alcohol consumption would enable our healthcare sector to do more with less. Australia has made significant gains in reducing rates of smoking and skin cancers, but we spend relatively little of our healthcare budget on preventive care and significantly less than our OECD counterparts.

Second, more timely approval processes for pharmaceuticals and other medical technologies would help ensure that the diffusion of new treatments remains a positive contributor to productivity growth.

Third, while advances in quality have been a welcome improvement, the growing fiscal burden from healthcare spending shows we must also identify avenues for cost saving. The Commission's forthcoming *Leveraging digital technology to lift healthcare productivity* paper considers the role digital technology can play in transforming the way we deliver healthcare, including to reduce costs.

# 1. Why measuring healthcare productivity matters

## Key points

- ✳ **Healthcare spending is on the rise. Australia now spends around one in every ten dollars on healthcare. This begs two questions: have the increases in spending been accompanied by commensurate gains in health outcomes; and is there scope to make our healthcare dollars go further?**
  - Measures of healthcare productivity can inform both these questions.
- ✳ **Measures of productivity growth can shed light on whether additional spending has been ‘worth it’. While existing estimates suggest there has been little improvement in healthcare bang-for-buck in Australia, they suffer from shortcomings.**
  - Existing estimates do not account for improvements in the quality of services (better health for patients) or the cost impacts when patients switch between different types of services or care settings.
  - Evidence from abroad suggests that these factors have large effects on healthcare productivity growth.
- ✳ **This paper provides Australia’s first estimates of healthcare productivity growth that account for changes in the quality of care that patients receive and how that care is provided.**
  - The estimates cover about a third of all healthcare and are intended as a complement to existing ABS estimates of healthcare productivity growth.
- ✳ **Benchmarking healthcare productivity across countries can tell us about our healthcare sector’s productivity relative to what is possible, but existing studies use data from nearly 20 years ago.**
- ✳ **This paper benchmarks Australia’s healthcare productivity using more contemporary data.**
- ✳ **But even improved measures of healthcare productivity can only ever tell us about certain aspects of our healthcare sector’s performance.**
  - Productivity measures do not directly consider access and equity issues or provide detailed information about the performance of particular services.
  - Considering productivity alongside more direct measures of system-wide responsiveness, safety and accessibility provides a more complete picture of performance.

## 1.1 Why focus on improving measures of healthcare productivity?

### Productivity measures help us understand the sector’s performance

Healthcare allows us to live longer, better and more productive lives.

It is also increasingly expensive. Spending on healthcare rose from 8% of GDP in 2000-01 to 10% in 2021-22 (AIHW 2012, 2023c; figure 1.1, panel a). As about 70% of healthcare is government-funded, this increase has had significant consequences for governments’ budgets. And the trend is unlikely to abate: the recent Intergenerational Report projected that Federal Government healthcare spending alone would rise from 4.2% of GDP in 2022-23 to over 6% by 2062-63 (Commonwealth of Australia 2023).

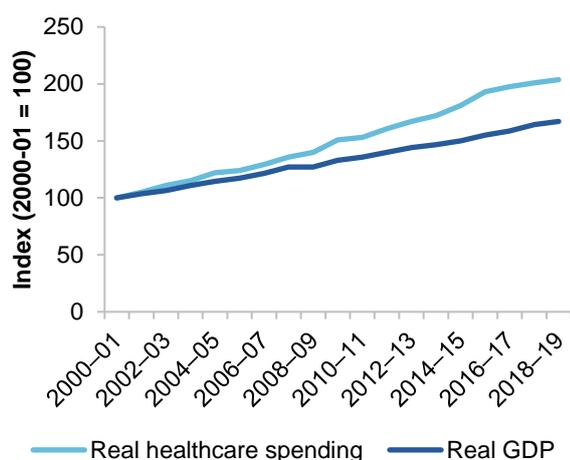
In part, this growth in spending reflects the pressures of an ageing population, as older people are more likely to be in poor health and in need of healthcare. Indeed, ageing accounted for about 20-30% of the growth in real per-capita healthcare spending between 2000-01 and 2018-19 (Goss 2022).

But by far the biggest driver of healthcare spending has been increases in the intensity of care provided. Increases in healthcare spending per case of disease, adjusted for changes in healthcare input costs accounted for about 60-70% of the growth in real per-capita healthcare spending between 2000-01 and 2018-19 (Goss 2022).

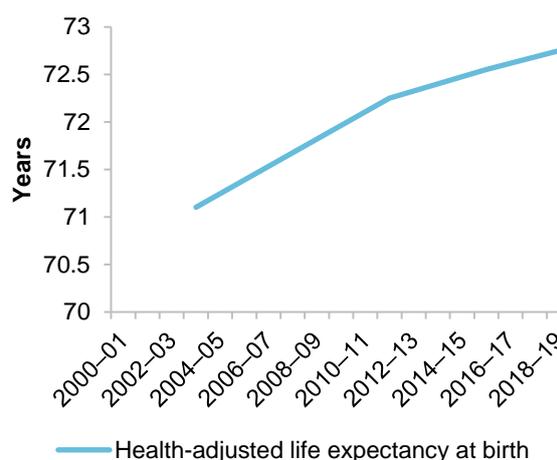
At the same time that spending has been increasing, health outcomes have been improving. Health-adjusted life expectancy at birth increased by over 1.5 years between 2003 and 2018 (figure 1.1, panel b). While better healthcare is not the only reason for this – changes such as lower smoking rates have helped too – it is part of the story.

**Figure 1.1 – We’ve spent more on healthcare and outcomes have improved**

**a. Healthcare spending has grown in absolute terms and as a share of GDP<sup>a</sup>**



**b. Health-adjusted life expectancy has improved<sup>b</sup>**



**a.** Both series have been deflated by the implicit GDP deflator. **b.** Average of male and female health-adjusted life expectancy at birth.

Source: AIHW (2022a); Productivity Commission estimates based on AIHW (2012, 2023c) and ABS (Australian National Accounts: National Income, Expenditure and Product, September 2023).

The combination of increasing health spending and outcomes raises two key questions:

1. Was the improvement in outcomes commensurate with the increase in spending?
2. How much scope was there to get better outcomes with the money we spent?

Measures of healthcare productivity can shed light on both these questions (box 1.1).

The first question can be informed by high-quality estimates of healthcare productivity *growth*. Healthcare productivity growth compares the growth in healthcare outputs (benefits for patients) with the growth in healthcare inputs (the costs of delivering those benefits) (chapter 2). Positive productivity growth would imply that the additional spending has been 'worth it'. But this in turn relies on capturing healthcare outputs in a way that reflects what patients value – changes in the quality of healthcare.

The second question can be informed by considering the gap between what is possible and what was achieved by benchmarking the *level* of productivity of Australian healthcare against that of other countries (chapter 3). The more productive Australia's healthcare sector was relative to those of other countries, the less the scope there was to get better outcomes with the money we spent.

### **Box 1.1 – A primer on productivity**

Productivity is a measure of bang for buck – it is defined as outputs divided by the inputs used to produce those outputs.

'Labour productivity' considers only the labour inputs that produced the output, while 'multifactor productivity' is a more complete measure that considers both labour, capital and intermediate inputs. This paper focusses solely on multifactor productivity, and refers to it simply as 'productivity' throughout.

For many other sectors, the notions of 'inputs' and 'outputs' are intuitive. For example, the output of a banana farm is bananas, the labour inputs are the hours of work undertaken by workers, the capital inputs are land and equipment, and the intermediate inputs are materials. When measuring productivity, these are each expressed as dollar values so that they can be compared using common units.

In the context of healthcare, the notion of inputs is similarly intuitive – labour inputs are hours worked by healthcare workers and administrators, capital inputs include hospital beds and magnetic resonance imaging machines, and intermediate inputs include drugs and bandages.

But healthcare outputs are less intuitive. The healthcare sector produces, among other things, hospital admissions and consultations with general practitioners. Some studies of healthcare productivity consider these to be outputs of the healthcare sector, but they are secondary to healthcare's overarching objective, which is to improve health. This paper considers better health itself to be the output of the healthcare sector.

## 1.2 What do existing measures of productivity tell us and how can they be improved?

### Current estimates suggest our healthcare sector has experienced little productivity growth ...

Existing estimates suggest that healthcare productivity growth has been positive, but small. Productivity estimates produced by the Australian Bureau of Statistics (ABS) (limited to hospitals) indicate that productivity grew by 0.1% per annum between 2008-09 and 2018-19 (ABS 2021). To put this in context, market sector productivity was estimated to grow at 0.7% per annum over the same period.<sup>1</sup>

But these estimates only capture some of the ways in which healthcare productivity can grow. They are based on what is commonly referred to as the 'traditional approach' to measuring healthcare productivity growth (appendix B). This approach captures productivity changes that occur due to changes in the inputs used to produce particular healthcare services, such as reductions in the average length of hospital admissions.

### ... but only provide a partial picture

There are, however, other ways that healthcare productivity can change which are not captured by the traditional approach.

*First, healthcare quality can change.* Innovation and advances in technology have improved the quality of healthcare by giving rise to more effective procedures and drugs. For example, innovations in lung transplantation procedures and donor procurement have steadily improved post-transplant outcomes for recipients in Australia (Paraskeva et al. 2018).

Studies from abroad have found that capturing changes in the quality of healthcare has a large effect on measured productivity.

- Two recent studies of healthcare productivity from the United States, while adopting different methodologies, both found that quality change added about 3 percentage points per annum to healthcare productivity growth between 2000 and 2017 (Dunn et al. 2022) and between 1999 and 2012 (Cutler et al. 2022).<sup>2</sup>
- Using a less comprehensive method, the United Kingdom Office for National Statistics found that accounting for quality change added 0.4 percentage points per annum to healthcare productivity growth in England between 2001 and 2019 (ONS 2021).

*Second, different types of healthcare services can substitute for one another.* As new practices or innovations emerge, the settings in which particular conditions are treated and the number of treatments needed can change. For example, kidney dialysis can now be administered at home, providing equivalent or better outcomes, more cost effectively than hospital-based dialysis (Walker et al. 2017).

Whether failing to account for these substitution effects increases or decreases measured productivity growth is unclear. Where treatment shifts from higher- to lower-cost settings, such as in the kidney dialysis example, failing to account for these shifts will lead to underestimates of productivity growth. But some

---

<sup>1</sup> The following industries comprise the market sector: agriculture, forestry and fishing; mining; manufacturing; electricity, gas, water and waste services; construction; wholesale trade; retail trade; accommodation and food services; transport, postal and warehousing; information, media and telecommunications; financial and insurance services; rental hiring and real estate services; professional, scientific and technical services; administrative and support services; arts and recreation services; and other services.

<sup>2</sup> This study is specific to healthcare provided to people aged 65 and over.

substitutions might be from lower- to higher-cost settings, or the number of treatments per case might grow. For example, the average number of sessions of focussed psychological treatment per mental health patient grew following changes to Medicare funding arrangements in 2020 (Pirkis et al. 2022).

Studies from abroad similarly highlight that capturing substitution effects has a large effect on measured productivity, but unlike quality adjustments they can both increase and decrease measured growth depending on the circumstances.

- A recent study from the United States indicates that failing to account for substitution would have led to healthcare productivity growth being *overstated* by about 1–1.5 percentage points per annum in the United States over 2000–2010 (Dunn et al. 2015).
- A lower quality study from the Canada indicates that failing to account for substitution would have led to healthcare productivity growth being *understated* by about 3 percentage points per annum in Canada over 2002–2010 (Gu and Morin 2014).

These two shortcomings suggest existing measures in Australia provide an incomplete answer to our first key question: *Have improvements in outcomes been commensurate with the increase in spending?*<sup>3</sup>

This paper represents the first attempt to capture both quality and substitution effects when measuring healthcare productivity growth in Australia. Chapter 2 provides those measures for about a third of the healthcare sector. These new measures are intended as a complement to, rather than a substitute for, the ABS's productivity estimates. In the United States and the United Kingdom, statistical agencies publish separate estimates that capture one or the other of quality and substitution effects which are helpful for understanding healthcare productivity.

## **Benchmarking results suggest Australia is getting good value for its healthcare dollar relative to its international counterparts ...**

While estimates of healthcare productivity growth can shed light on whether additional spending has been 'worth it', they tell us little about our healthcare sector's performance relative to what was possible. For example, positive productivity growth could arise because Australia's healthcare sector caught up to a relatively static technological frontier, as healthcare providers adopted the best available treatments or reforms improved the way the sector is organised. This would be cause to 'pat ourselves on the back'. Alternatively, positive productivity growth could arise because the global technological frontier improved rapidly and Australia's healthcare sector adopted improved practices, but not the best available practices. This would be cause to strive for better.

Benchmarking the level of productivity of Australian healthcare against that of other countries can provide an indicator of which of these scenarios is closer to the truth. If our healthcare sector had been less productive than those of other countries, this would suggest there was scope for substantial improvement by adopting more practices used abroad, while the converse would suggest relatively less scope for improvement in this way.

<sup>3</sup> In part, this is because the strictures of the National Accounts prevent the ABS from estimating healthcare productivity growth in a way that takes account of these shortcomings. Explicit adjustment for healthcare quality change violates the convention that there should be no return to capital on assets used in non-market production (ABS 2020). And accounting for substitution between different types of healthcare services requires an approach that pools all healthcare received by a patient (chapter 2), but the National Accounts further divide the healthcare and social assistance industry into sub-industries such as 'hospitals' and 'medical and other health care services'. The pooled approach does not align with these sub-industry boundaries.

Two studies have benchmarked Australia's healthcare sector against that of other countries, and both found it to be relatively productive.<sup>4</sup> Australia was ranked as:

- the 39<sup>th</sup> most productive of 191 member states (about the 80<sup>th</sup> percentile) by the World Health Organisation (2000) using data from 1993 to 1997
- the 7<sup>th</sup> most productive of 78 countries (Ogloblin 2011) using data from 2000 to 2007 (about the 90<sup>th</sup> percentile).

### ... but existing studies are outdated

While there is scope for some methodological improvements, the main issue is that these studies are now outdated. Australia's healthcare sector has undergone significant reforms since 2007. Evidence based on more recent data is needed to inform our second key question: *How much scope was there to get better outcomes with the money we spent?* Chapter 3 estimates the productivity of Australia's healthcare sector relative to that of other countries using data from 2010 to 2019.

## 1.3 What can't improved productivity estimates tell us?

### Measures of productivity are just one way of assessing our healthcare sector's performance

While this paper is focussed on how to improve measures of healthcare productivity, we recognise that this is just one way of gauging the performance of Australia's healthcare system. Other dimensions of concern to policy makers include the system's responsiveness, safety and accessibility, including for different groups within the community.

Productivity measures provide few if any insights into these other dimensions of performance. For example, questions about who benefits and whether certain groups miss out are not questions that productivity measures are directly intended to answer. Indeed, our measures only touch on accessibility insofar as there are diminishing health returns to additional healthcare inputs (Rodriguez and Sobrino 2015), such that healthcare systems that devote more resources to those in need of care are likely to be assessed as more productive. Moreover, our measures might not capture the impacts of changes in accessibility on healthcare productivity if they take many years to influence health outcomes (chapter 3 and appendix E).

Considering productivity alongside more direct measures of system-wide responsiveness, safety and accessibility provides a more complete picture of performance.

Healthcare productivity estimates are also largely intended to provide an overarching assessment of health sector performance. These overarching perspectives, particularly where estimated at a national level, can mask localised trends, such as productivity at a state or territory level, or the efficiency with which we deliver healthcare for certain groups such as Aboriginal and Torres Strait Islander people.

Specific performance indicators such as those reported under the Productivity Commission's Report on Government Services (SCRGSP 2024), or the Australian Institute of Health and Welfare's MyHospitals framework (AIHW 2023e), may be more appropriate for assessing the performance of particular services, including in addressing the needs of particular groups within the community.

---

<sup>4</sup> Other studies have focussed on countries in specific regions such as Europe (Asandalui et al. 2014) and Asia (Ahmed et al. 2019), and states within the United States (Darabi et al. 2021).

## 2. Estimating healthcare productivity growth

### Key points

- \* **For the first time in Australia, the Commission has sought to measure healthcare productivity growth in a way that accounts for improvements in healthcare quality.**
- \* **We found that healthcare productivity grew by about 3% per year between 2011-12 and 2017-18 in the parts of the sector that we studied – treatment of cancers, cardiovascular diseases, blood and metabolic disorders, endocrine disorders and kidney and urinary diseases.**
  - Together, the treatment of these diseases accounted for around one third of healthcare spending.
- \* **Nearly all of the productivity growth was due to quality improvements, not cost reductions.**
  - The vast majority of these quality improvements have come from advances in saving lives rather than advances in improving the quality of life of people living with studied diseases.
- \* **Productivity grew particularly strongly in the treatment of cancers, suggesting that advances in treatments, rather than across-the-board healthcare reforms, have been the major driver of growth.**
  - While patients benefit either way, not all of these improvements can be ascribed to the healthcare sector; some stem from improvements in domestic and offshore drug and medical equipment manufacturing industries.
  - More timely approval processes for pharmaceuticals and other medical technologies would help ensure that the diffusion of new treatments remains a positive contributor to productivity growth.
- \* **While advances in quality are a welcome improvement, the growing fiscal burden of healthcare means that we must also identify avenues for cost savings.**
  - The Commission's forthcoming *Leveraging digital technology to lift healthcare productivity* paper considers the role digital technology can play in transforming the way we deliver healthcare, including to reduce costs.

Spending on healthcare has risen and is expected to continue to rise (chapter 1), which begs the question: *Has the improvement in outcomes been commensurate with the increase in spending?*

Estimates of healthcare productivity growth can inform this question, as they compare the growth in healthcare outputs (benefits for patients) with the growth in healthcare inputs (the costs of delivering those benefits).<sup>5</sup>

But there are few estimates of healthcare productivity growth in Australia, and those that do exist do not capture some of the key ways in which healthcare productivity can increase or decrease, including through changes in the quality of care (chapter 1).

This chapter presents Australia's first estimates of healthcare productivity growth that take account of quality changes.

## **2.1 Productivity growth estimates help us understand if additional healthcare spending has been 'worth it'**

### **We estimate productivity growth on a disease-by-disease basis to capture improvements in healthcare quality ...**

Healthcare productivity growth depends on changes in healthcare inputs and healthcare outputs.

Healthcare inputs are the resources used by the healthcare sector; the hours of labour supplied by healthcare workers and administrative staff, the flow of services from equipment such as hospital beds and magnetic resonance imaging machines, and materials consumed such as drugs and bandages. These are measured in a similar way to inputs in other sectors – in dollar terms, adjusting for changes in the prices of healthcare inputs.

Healthcare outputs are less clear. The 'traditional approach' to healthcare productivity measurement (chapter 1; appendix B) frames healthcare outputs as services delivered such as hospital admissions and general practitioner consultations. Under this approach, healthcare productivity growth occurs when the healthcare sector delivers more of these services with the same inputs, the same services with fewer inputs, or a mix of both.

But healthcare services are secondary to healthcare's overarching objective, which is to improve health. And healthcare services have clearly gotten better at delivering on this overarching objective – given the choice, most people would likely prefer to be treated by today's healthcare system rather than the system of the past. Failing to account for these improvements can lead to underestimates of healthcare productivity growth.

For the first time in Australia, this paper estimates healthcare productivity in a way that accounts for improvements in healthcare quality.

We estimate healthcare productivity growth on a disease-by-disease basis, e.g. productivity growth in the treatment of cancer. This allows us to account for changes in the quality of care by tracking changes in disease survival rates and the quality of life of people living with disease (box 2.1).

The disease-by-disease approach also allows us to account for changes in the number and mix of services that patients receive, which traditional approaches cannot do (chapter 1). This is because the disease-by-disease

---

<sup>5</sup> Throughout this paper, 'productivity' refers to multifactor productivity.

approach is agnostic to the type of healthcare that patients receive; all that matters is the change in the real cost of treating a case of disease, which captures changes in the number and mix of services.

### **Box 2.1 – Measuring changes in healthcare quality**

To capture changes in healthcare quality, we estimate the change in healthcare-attributable health outcomes per case of disease per year between 2011-12 and 2017-18 in quality-adjusted life years (QALYs). A QALY is a generic unit of health, defined as a year of life lived in perfect health. By using QALYs, we can capture the effects of changes in healthcare on both quantity of life (by saving lives) and quality of life (by improving the standard of living of those who survive the disease).<sup>a</sup>

This task requires some simplifying assumptions. We assume that healthcare affects health outcomes among patients with a disease, but not the number of cases of that disease. And we assume that other factors influence the number of cases of a disease but not health outcomes among patients with that disease. Other studies have also made these assumptions, for example Highfill and Bernstein (2019), Wamble et al. (2019) and Weaver et al. (2022).

To see what these assumptions mean, consider the case of lung cancer. Smoking causes lung cancer and smoking rates have fallen in Australia (AIHW 2023a). Comparing health outcomes due to lung cancer *per capita* in 2011-12 and 2017-18 would be misleading because of the confounding influence of changes in smoking rates. But if we assume that changes in smoking rates (among other non-healthcare factors) influence the prevalence of lung cancer but not outcomes among prevalent cases, we can isolate the contribution of changes in healthcare by examining changes in health outcomes due to lung cancer *per case of lung cancer*.

Examining changes in health outcomes per case of disease requires some further assumptions. If disease survival rates change, the change in QALYs gained depends on how long those who survive the disease live for, and in what state of health. We assume that those who survive a disease in a given year revert to the average level of health for someone of their age (among those who do and do not have the disease) from the next year onwards. Because this is an optimistic assumption, we apply a high discount rate to these future QALYs – 7% per year in real terms.<sup>b</sup> We also account for the associated expected future healthcare costs because healthcare received in the future is a driver of these expected future health outcomes.

Appendix C discusses our approach to measuring changes in healthcare quality in more detail.

**a.** Because the value of an additional QALY to society has been widely studied, using QALYs allows us to express changes in health outcomes in dollar value terms. Our central estimate is that an additional QALY is worth \$110,000 in 2018 dollars. We also consider other plausible values of a QALY in appendix C. **b.** We also consider other discount rates – 3% and 10% – in appendix C.

## **... and compare the improvements in quality with the changes in costs**

To estimate productivity growth, we compare the changes in health outcomes (per case of disease) with changes in the average cost of treating each case of disease. The approach is similar to undertaking a cost-benefit analysis of the changes in the treatment of each disease. We place a dollar value on the change in outputs (health outcomes per case of disease) and contrast it with the dollar value change in healthcare inputs (real costs per case of disease). If the dollar value change in healthcare outputs exceeds the dollar value change in healthcare inputs, the change delivered a net benefit to society and the productivity of the

treatment of that disease grew. Conversely, if the dollar value change in healthcare inputs exceeds the dollar value change in healthcare outputs, the change represented a net cost to society and the productivity of the treatment of that disease fell. This way of accounting for quality change, known as the 'cost of living index' approach, is discussed in more detail in appendix C.

## **We study about one third of the healthcare sector, by spending**

Estimating productivity growth in this way requires substantial amounts of data – for each disease, we need data from 2011-12 and 2017-18 on healthcare costs, prevalence, mortality and quality of life. Data limitations (appendix C) mean we can only study about one third of all healthcare spending – healthcare used to treat three broad groups of diseases:

- Cancers – all cancers.
- Cardiovascular diseases – all cardiovascular diseases.
- Blood, endocrine and kidney diseases – all blood and metabolic disorders, endocrine disorders and kidney and urinary diseases.

## **2.2 Healthcare productivity growth has been robust**

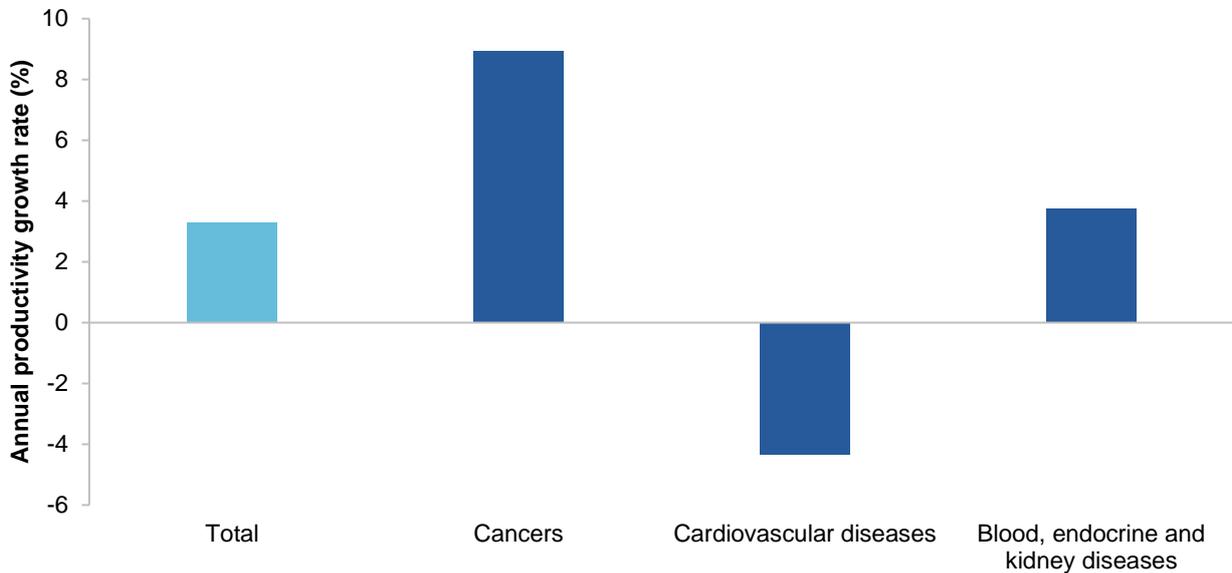
Australia's healthcare sector has enjoyed robust productivity growth. Quality-adjusted healthcare productivity grew by about 3% per year between 2011-12 and 2017-18 for the third of the healthcare sector we studied (figure 2.1). Our estimates based on alternative sources of sources of data similarly point to robust growth of about 3% to about 4% per year (appendix C).<sup>6</sup> This suggests that the additional spending in the parts of the healthcare sector we studied has been 'worth it' – it has delivered net benefits for society.

However, productivity growth was far from uniform – the productivity of the treatment of cancers and blood, endocrine and kidney diseases grew, while the productivity of the treatment of cardiovascular diseases declined. Estimates based on alternative data generally show similar patterns (appendix C).

---

<sup>6</sup> Appendix C also presents alternative productivity growth estimates based on alternative assumptions – higher and lower dollar values of a QALY and higher and lower discount rates. Overall productivity growth ranges from about 2% to about 6% per year with these alternative assumptions.

**Figure 2.1 – Overall healthcare productivity grew rapidly, but growth varied by disease**



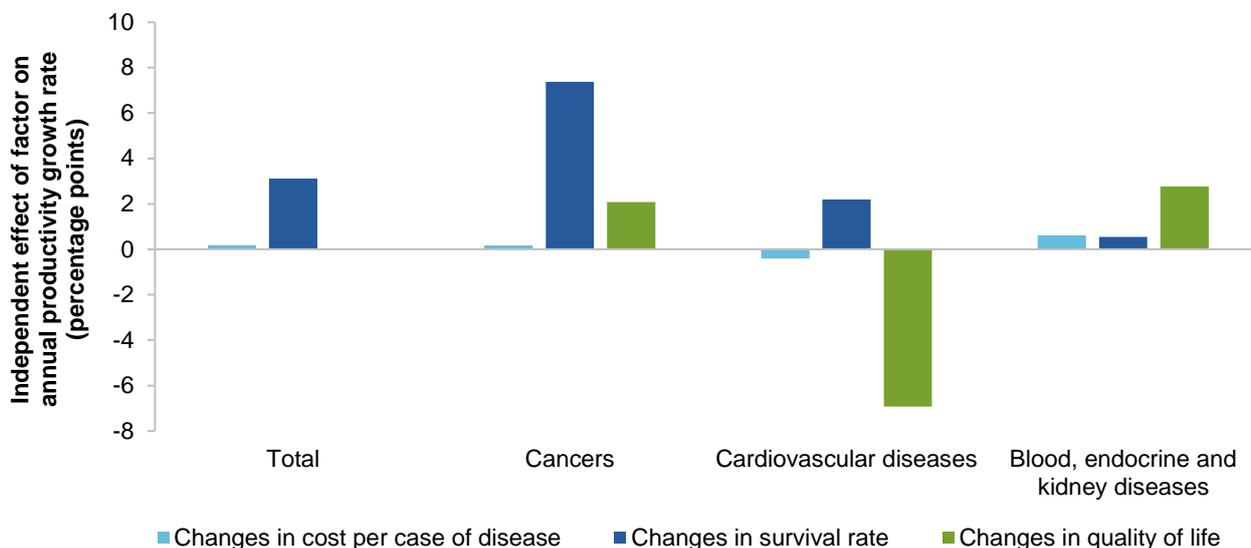
Source: Productivity Commission estimates.

### Improvements in healthcare quality drove nearly all of the gains

The biggest contributions to productivity growth have not come from doing more with less, but rather from more effective healthcare services. While there is some variability across diseases, including depending on the data source we use, in the main, quality improvements rather than cost reductions were the big drivers of productivity growth, and the vast majority of these have come from advances in saving lives. Indeed, the effects on overall productivity growth of changes in survival rates swamp the effects of changes in costs per case of disease and changes in quality of life (figure 2.2). This was because quality of life improved for people with cancers and blood, endocrine and kidney diseases but fell sharply for people with cardiovascular diseases.<sup>7</sup>

<sup>7</sup> Estimates based on alternative sources of data similarly show changes in survival rates to be the main contributor to overall productivity growth, but suggest more muted changes in quality of life for each disease – smaller improvements in quality of life for cancers and blood, endocrine and kidney diseases and smaller reductions or even improvements in quality of life cardiovascular diseases (appendix C).

**Figure 2.2 – Changes in survival rates underpinned most of the productivity growth<sup>a</sup>**



a. These independent effects do not exactly sum to the overall productivity growth rate because quantitatively minor interactions between the three factors also influence healthcare productivity growth.

Source: Productivity Commission estimates.

## 2.3 Results largely align with international and on-the-ground experience

### Improvements in quality are intuitive ...

Our results indicate that overall healthcare productivity for studied diseases grew at about 3% per year, almost entirely due to healthcare quality improvements.

Studies from abroad have found a similar impact on productivity growth from healthcare quality improvements. The results of one recent study from the United States suggest that healthcare quality improvements increased productivity growth by 3.3 percentage points per year between 2000 and 2017 (Dunn et al. 2022).<sup>8</sup> Another, while adopting a different methodology, similarly pointed to gains of around 3 percentage points between 1999 and 2012 (Cutler et al. 2022).<sup>9</sup>

<sup>8</sup> This is computed as follows: the study found that quality-adjusted healthcare prices declined by 1.33% per year (relative to aggregate GDP inflation), while disease-based (non-quality-adjusted) healthcare prices rose by 1.96% per year (also relative to aggregate GDP inflation). This suggests that adjusting for healthcare quality improvements reduced measured healthcare output price growth by about 3.29 percentage points per year (relative to a non-quality-adjusted disease-based measure) which is equivalent to increasing measured healthcare productivity growth by about 3.29 percentage points per year.

<sup>9</sup> This is because this study found that overall quality-adjusted productivity growth was 1.5 per cent per year, very similar to the finding of Dunn, Hall and Dauda (2022) that overall quality-adjusted productivity growth was 1.56 per cent per year. As the studied time periods largely overlap, this suggests a similar contribution of healthcare quality improvement.

Our results also broadly align with advances in treatments. Improvements in saving lives, particularly for cancers, was a key driver of growth. This accords with changes in cancer treatment between 2011-12 and 2017-18 – new immunotherapies and targeted therapies for cancer treatment were introduced during this period (Shahid et al. 2019). These breakthroughs have been heralded as among the most significant scientific advances of their time (Couzin-Frankel 2013).

### **... but our estimates may understate gains in the treatment of some diseases**

While an improvement on approaches that do not capture quality at all, it could be that our approach does not capture the full benefits of improvements in cardiovascular disease treatment. Overall, we found declining productivity in the treatment of cardiovascular diseases, whereas studies of productivity growth from earlier periods in the United States have found substantial improvements in the productivity of cardiovascular disease treatment. For example, Cutler et al.(2022) found that productivity growth in cardiovascular diseases treatment in the United States between 1999 and 2012 was higher than productivity growth in the treatment of any other group of diseases.

There are two potential sources of improvements in cardiovascular disease treatment that our approach may not have fully captured.

First, our estimates found that cardiovascular disease survival rates rose – a plus for productivity growth – but that these improvements were outweighed by declines in quality of life for those living with cardiovascular diseases. The extent of the decline in quality of life comes down to which source of data we rely on. Estimates from the Australian Burden of Disease Study show smaller declines in quality of life among people living with more severe cardiovascular diseases (stroke, coronary heart disease and cardiomyopathy) and our own alternative quality of life estimates show small improvements in quality of life for people with cardiovascular diseases (appendix C).

Second, our approach may not have fully captured the contribution of healthcare to reducing deaths from cardiovascular disease. This ultimately comes down to choice of methodology. Our approach assumes that healthcare affects health outcomes among patients with a disease but not the number of cases of that disease (box 2.1). If better healthcare led to declines in the prevalence of cardiovascular diseases (which in turn reduced deaths from cardiovascular disease), our estimates will understate the healthcare-attributable improvement in survival rates. And healthcare can prevent cardiovascular disease. For example, blood pressure medications, which are among the most commonly prescribed medications in Australia, can prevent hypertension (itself considered a cardiovascular disease for the purposes of this study and a significant risk factor for coronary heart disease and stroke) (ABS 2018).

### **Not all quality improvements can be attributed to the healthcare sector**

On its face, our estimates of robust growth in healthcare productivity appear to run counter to the view that productivity growth in non-market services is difficult to achieve. But these two perspectives are not as much at odds as they might appear.

Not all of the improvement in health outcomes can be attributed to the healthcare sector per se. The productivity gains we have found have been driven by quality improvements, not cost reductions, and it is likely that many of the quality improvements derive from better healthcare inputs (such as better drugs and

equipment) rather than improvements in the quality of the services provided by the healthcare sector.<sup>10</sup> The case of improved cancer therapies is one example of this. Where this is the case, the gains represent unmeasured productivity growth in the domestic and offshore drug and medical equipment manufacturing sectors rather than productivity growth in the healthcare sector. That said, some of the improvements in healthcare inputs might have required (and driven) complementary increases in the skills of healthcare workers to use those better inputs, and these complementary quality gains did originate in the healthcare sector and rightly contribute to healthcare productivity growth. Moreover, some innovations in treatment do not require new inputs, just better ways of using existing inputs. These quality gains should be attributed to the healthcare sector.

For the purposes of answering our question: '*Have outcomes improved sufficiently to justify the increase in spending?*', the origin of the quality gains matters less. Patients benefitted regardless of whether the quality improvements originated in the healthcare sector or in some other sector.

## **Better data would help improve future estimates**

While we can confidently conclude that productivity growth has been robust in the parts of the healthcare sector we studied, it begs the question, how representative are they of the wider healthcare sector?

The third of the healthcare sector (by spending) that we studied was chosen based on data availability alone, and so there is no apparent reason to expect that productivity growth in other parts of the healthcare sector has been faster or slower.

Absent better data, it is difficult to reach a conclusion one way or the other. In lieu of that data, we have looked at other studies to provide some insights. Two potentially relevant studies suggest that productivity growth may have been slower in other parts of the healthcare sector.

First, on the inputs side, Goss (2022) estimated that real healthcare spending per case of disease grew by 1.2% between 2011-12 and 2017-18. This suggests that changes in costs per case of disease had a negative impact on healthcare productivity growth across the entire healthcare sector, whereas our estimates suggest they had a very small, but positive, impact for the diseases studied. But this study faced some of the same data constraints we confronted – a lack of reliable sources on prevalence (including changes over time) for the full suite of diseases. And, data limitations aside, even if healthcare spending per case of disease had grown by 1.2%, it would not alter our overarching finding that healthcare productivity growth has been comfortably positive.

Second, on the outputs side, the diseases studied in this paper typically experienced larger reductions in the burden of disease due to premature death per capita (on an age-adjusted basis) between 2011 and 2018 than other diseases (figure 2.3). This might suggest that the quality of the treatment of other diseases did not improve as much as for the diseases studied. But again, there are reasons for pause. What matters for estimating productivity is not just the size of the reduction in the burden of disease, but how the reduction compares to the cost of treatment and how the prevalence of the disease in question changed.<sup>11</sup> Looking at one without the other provides an incomplete and potentially misleading picture. And reductions in premature death are just one, albeit important, way in which outcomes can improve – the data in figure 2.3 does not account for changes in quality of life of those living with disease.

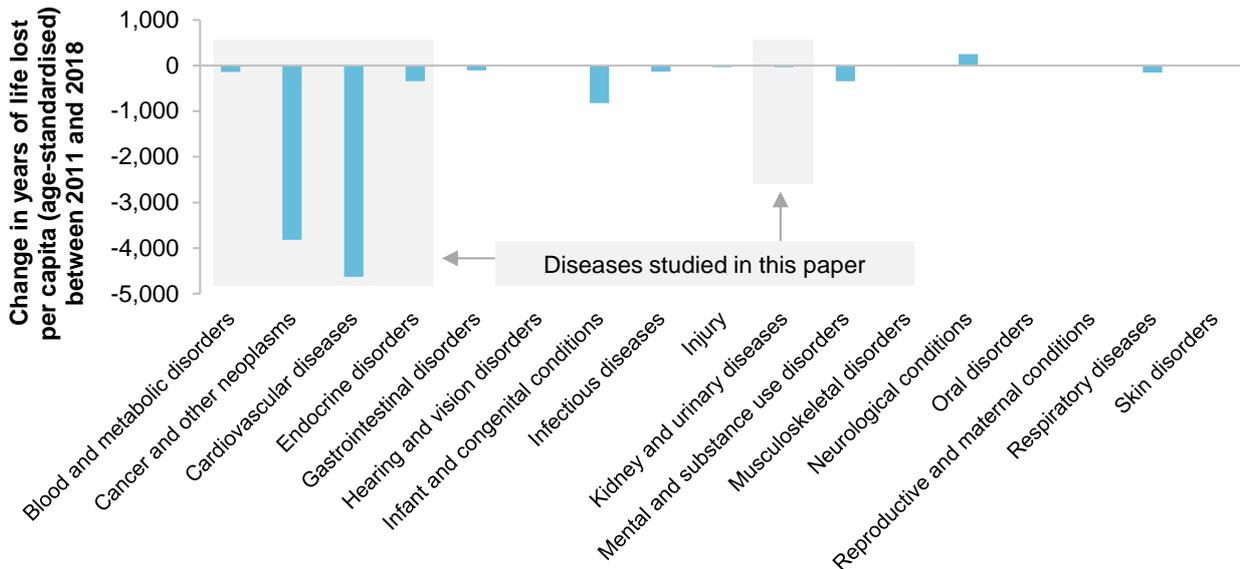
---

<sup>10</sup> At a more technical level, the reason for this is that we have not adjusted our measures of healthcare inputs for quality changes.

<sup>11</sup> It is also possible that changes in non-healthcare factors influenced the number of cases of disease per capita (on an age-adjusted basis) differently from the diseases we have studied.

Gaining insights into productivity growth in the remaining parts of the healthcare sector ultimately requires better data on the prevalence of other diseases and changes in the quality of life of those living with those diseases.

**Figure 2.3 – The burden of disease due to premature death typically declined by more among diseases studied in this paper<sup>a</sup>**



a. ‘Years of life lost’ is a QALY-denominated measure that assumes that people who die of the disease in question would have otherwise lived a very long life in perfect health. While this assumption is unrealistic, it is consistent across all diseases and so is suitable for providing an indication of the relative magnitude of the differences between diseases. The Australian Burden of Disease Study also reports a fuller measure (disability-adjusted life years) which also accounts for the loss in quality of life of those living with disease, but changes in the quality of life component of this measure are assumption-driven for some diseases and so differences between diseases are less comparable.

Source: Productivity Commission estimates based on AIHW (2021d).

## Implications for policy

Our results suggest that Australia is getting relatively good, and increasing, value for its healthcare dollar.

But these good results are not grounds for complacency.

Our results show the value of quality improvements in healthcare to Australia. It is likely much of the improvement is due to the adoption of better treatments as they have become available. More timely approvals processes for pharmaceuticals and other medical technologies would allow these new treatments to diffuse through the healthcare sector faster, boosting productivity growth.

The growing fiscal burden of healthcare means that we must also identify opportunities for cost savings. Spending ever more on healthcare will require ever higher taxes. The Commission’s forthcoming *Leveraging digital technology to lift healthcare productivity* paper considers the role digital technology can play in transforming the way we deliver healthcare, including to reduce costs. Preventive healthcare may also have a role to play here. While preventive healthcare can be highly cost-effective, Australia devotes less of its healthcare spending to public health than its OECD counterparts (chapter 3).



## 3. Benchmarking Australia's healthcare productivity

### Key points

- ✳ **Benchmarking Australia's level of healthcare productivity against that of other countries provides insights into whether we can get better outcomes for our healthcare dollar.**
- ✳ **Our research suggests that Australia's healthcare sector is among the most productive in the world.**
  - Australia's healthcare productivity ranks third among 28 high-income countries once we account for the age of our population and behavioural and environmental risk factors such as smoking, obesity, diet and alcohol consumption.
- ✳ **Adjusting for risk factors makes a significant difference to our ranking.**
  - Although Australia has relatively low smoking rates, we have comparatively poor diets, a high obesity rate, and high levels of alcohol consumption per capita.
  - Failing to account for these factors gives a misleading indication of our healthcare sector's productivity. On balance, they worsen our health outcomes but are mostly outside the control of our health system.
- ✳ **While welcome, these benchmarking results also highlight the importance of prioritising prevention.**
  - Australia spends relatively little of its healthcare budget on preventive care.
  - Greater and more timely adoption of cost-effective preventive measures would help reduce our sizeable risk factors and enable our healthcare sector to do more with less.

Chapter 2 found that Australia's healthcare sector experienced robust productivity growth between 2011-12 and 2017-18, suggesting that improvements in outcomes have been more than commensurate with the increases in healthcare spending.

While good news, this says little about the other key question this study has sought to answer: *How much scope is there to get better outcomes with the money we are spending?*

This question can be informed by comparing what was achieved to what was possible, by benchmarking the level of productivity of Australia's healthcare sector against that of other countries. If, for example, we found that our healthcare sector was less productive than that of many other countries, it would suggest there was potential to make our healthcare dollars go further.

Two studies have benchmarked Australia's healthcare sector against that of other countries, and both found it to be relatively productive.<sup>12</sup>

- The World Health Organisation (2000) found Australia's healthcare sector was the 39th most productive of 191 member states (about 80<sup>th</sup> percentile) based on data from 1993 to 1997.
- Ogloblin (2011) found Australia's healthcare sector was the 7th most productive of 78 countries (about 90<sup>th</sup> percentile) using data from 2000 to 2007.

However, these studies are now out of date. Australia's healthcare sector has undergone significant reforms since 2007.

This chapter uses more recent data to provide contemporary insights into Australia's relative healthcare productivity. In doing so, consistent with the approach adopted elsewhere in this paper, it treats the output of the healthcare sector as improved health, rather than healthcare services, as this provides a more meaningful measure of healthcare productivity.

### **3.1 Cross-country comparisons can highlight scope for improvement**

#### **Simple comparisons suggest Australia's healthcare sector is not overly productive ...**

One simple approach for ranking countries' healthcare sectors is to compare measures of population health outcomes, such as health-adjusted life expectancy (HALE), with standardised measures of healthcare inputs, such as per capita healthcare expenditure. HALE is a good measure of health outcomes because it captures the two key dimensions of health – quantity and quality of life. Appendix E also considers another measure of health outcomes, the Healthcare Access and Quality Index.

This simple comparison of health outcomes and inputs suggests that Australia's healthcare sector is not overly productive. Among the highest-income countries, countries that spent similar amounts on healthcare tended to have a similar HALE (figure 3.1).<sup>13</sup> And many countries appear to have more productive health sectors than Australia's – those that spent about the same as Australia but achieved a higher HALE (the dots

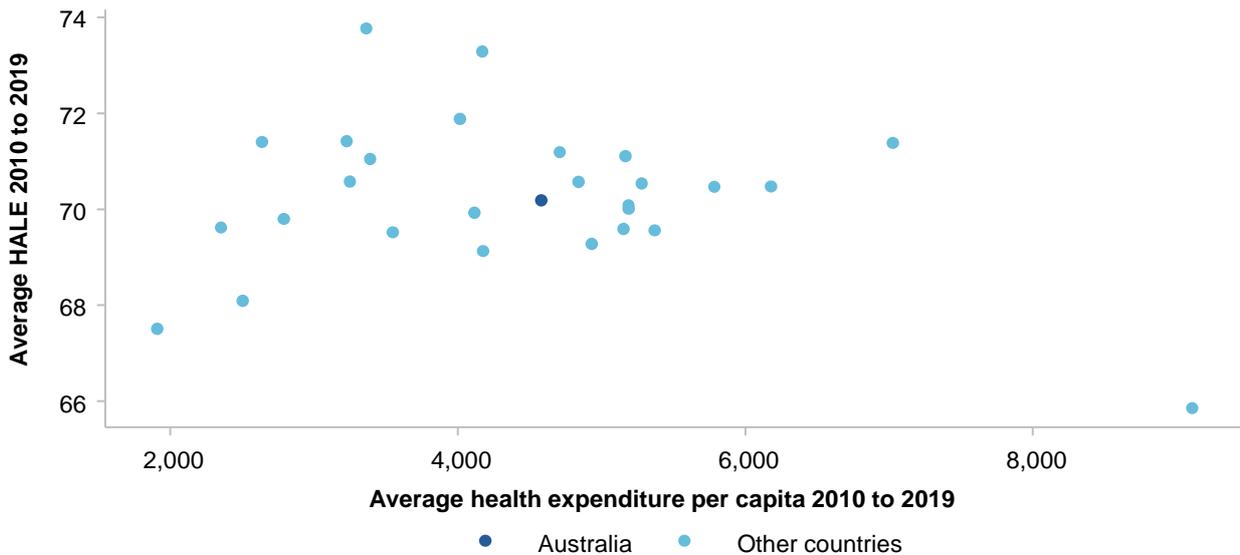
---

<sup>12</sup> Other studies have focussed on countries in specific regions such as Europe (Asandalui et al. 2014) and Asia (Ahmed et al. 2019), and states within the United States (Darabi et al. 2021).

<sup>13</sup> 28 countries; countries whose GDP per capita was above 30,000 USD in 2023 (on a purchasing power parity basis), excluding countries whose population was over 100,000 people in 2019 and countries with tobacco bans (Saudi Arabia and Oman). See Appendix E for a full list of countries.

directly above Australia's dot), and those that spent less than Australia but achieved a similar HALE (the dots directly to the left of Australia's dot).

**Figure 3.1 – Among high-income countries, Australia's healthcare spending and health-adjusted life expectancy is about average**  
**HALE versus per capita healthcare spending, by country, 2010 to 2019**



Source: Productivity Commission estimates based on IHME (2023b) and World Bank (2023a).

### ... but these comparisons ignore risk factors and demographics

However, these simple comparisons give a misleading indication of Australia's healthcare productivity relative to that of other countries, for two reasons.

First, the output of Australia's healthcare sector is not the sole determinant of its HALE. People's behaviour and the environment in which they live also affects their health. According to the Global Burden of Disease Study, the behavioural and environmental risk factors that most affect health in high-income countries are smoking, having a high body-mass index, dietary risks (overconsumption of unhealthy food and underconsumption of healthy food) and alcohol consumption (appendix E). Although Australia had a relatively low smoking rate (26<sup>th</sup> of the 28 countries between 2010 and 2019), it had the fourth highest obesity rate, the sixth highest level of alcohol consumption per capita and relatively low per-capita consumption of fruit and vegetables (19<sup>th</sup> of 28 countries).<sup>14</sup> Governments mostly aim to influence these factors through regulations and taxes, not healthcare services.<sup>15</sup>

Second, the age profile of countries' populations also affects whether countries that spend similar amounts on healthcare achieve a similar HALE. Countries with older populations need to spend more on healthcare to

<sup>14</sup> These statistics are Productivity Commission estimates based on IHME (2023b).

<sup>15</sup> Australia's experience with reducing smoking provides a good example. The share of Australian adults who smoke fell from more than 40% to about 11% since 1970 due to increases in cigarette taxes, advertising bans, plain packaging regulations, smoking bans in some environments, advertising campaigns and quit smoking services (Breadon et al. 2023). Only the last two of these factors count towards healthcare spending.

maintain a given HALE. Among the 28 high-income countries considered, Australia had a relatively small number of people aged 65 and over as a share of its total population (ranking 22<sup>nd</sup> of the 28 countries).<sup>16</sup>

## 3.2 Australia is getting good value for its healthcare dollar relative to its international counterparts

To get a more meaningful picture of Australia's relative performance, we need to isolate the effect of healthcare spending on outcomes by adjusting for cross-country differences in population age and key risk factors (box 3.1).

### Box 3.1 – Comparing like with like: controlling for cross-country differences in risk factors and population age

To better assess Australia's relative healthcare productivity, we need to isolate the effect of healthcare spending on HALE by adjusting for cross-country differences in population age and key behavioural and environmental risk factors that affect population health.

We do so by estimating a model that predicts each country's maximum achievable HALE given its healthcare spending and other key factors that affect HALE, using stochastic frontier analysis. We then rank countries by comparing each country's estimated maximum HALE to their actual HALE (on average over 2010 to 2019). The smaller the difference between a country's estimated maximum achievable HALE and its actual HALE, the more productive its healthcare sector.

Our approach takes into account that there are likely to be diminishing returns to healthcare inputs. Other things equal, a country that spends more on healthcare will likely obtain less HALE per dollar spent than a country that spends less on healthcare (Rodriguez and Sobrino 2015). This reflects the intrinsic realities of healthcare rather than the productivity of the healthcare sector, and so taking it into account allows for more reasonable comparison between countries that spend more and less on healthcare.

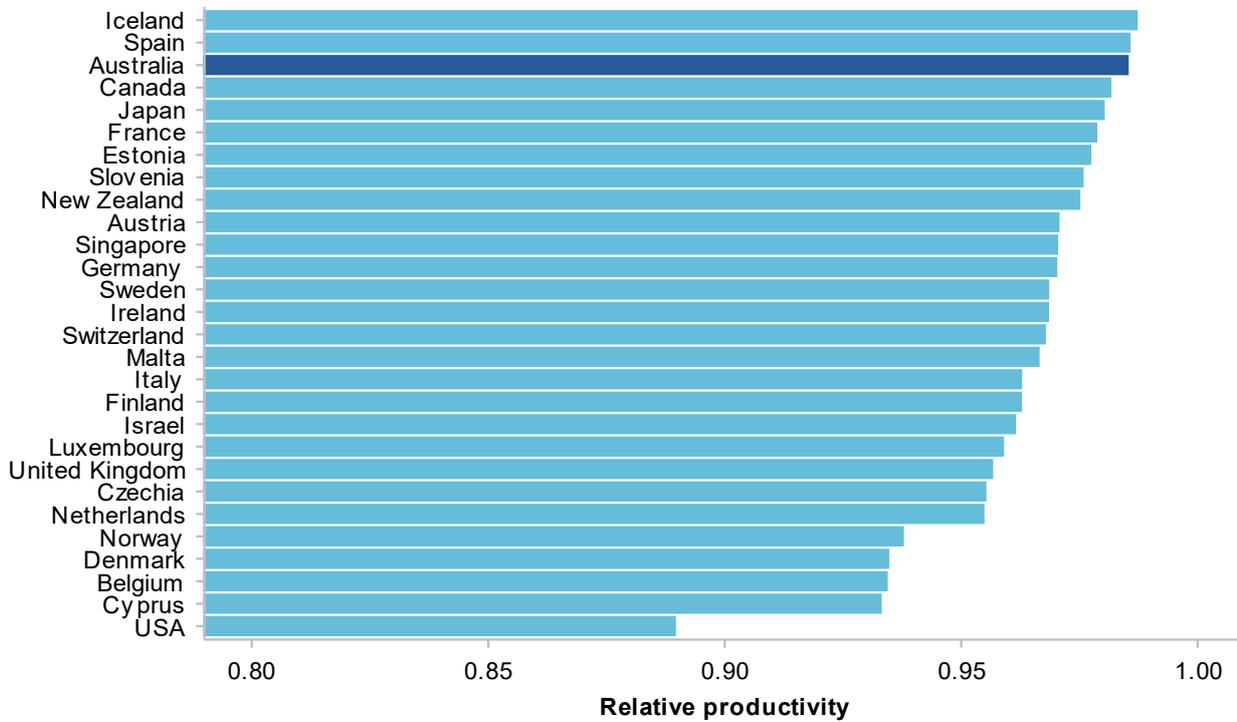
While a substantial improvement on simple comparisons of healthcare outputs and inputs, our model is not without its limitations. There are likely other important factors that influence HALE, such as culture, that the model does not account for. And it is unlikely to fully capture the complex relationship between risk factors and HALE – for example, reductions in healthcare spending that limit access to primary care would likely influence future HALE, rather than current HALE, but the model does not account for these intertemporal effects.

This modelling is outlined in detail in appendix E.

Once we adjust for risk factors and the age of our population, Australia's healthcare sector appears to be highly productive – ranking third amongst the 28 countries over the period 2010 to 2019 (about 90<sup>th</sup> percentile) (figure 3.2).

<sup>16</sup> Productivity Commission estimate based on World Bank (2023b). The indicator used here is a proxy measure of the age of the population constructed such that it is relatively independent of the healthcare available in that country. This makes it more suitable for the modelling approach outlined in section 3.2. Appendix E explains how the indicator is constructed and why it is more suitable for modelling than a direct measure of the age of the population.

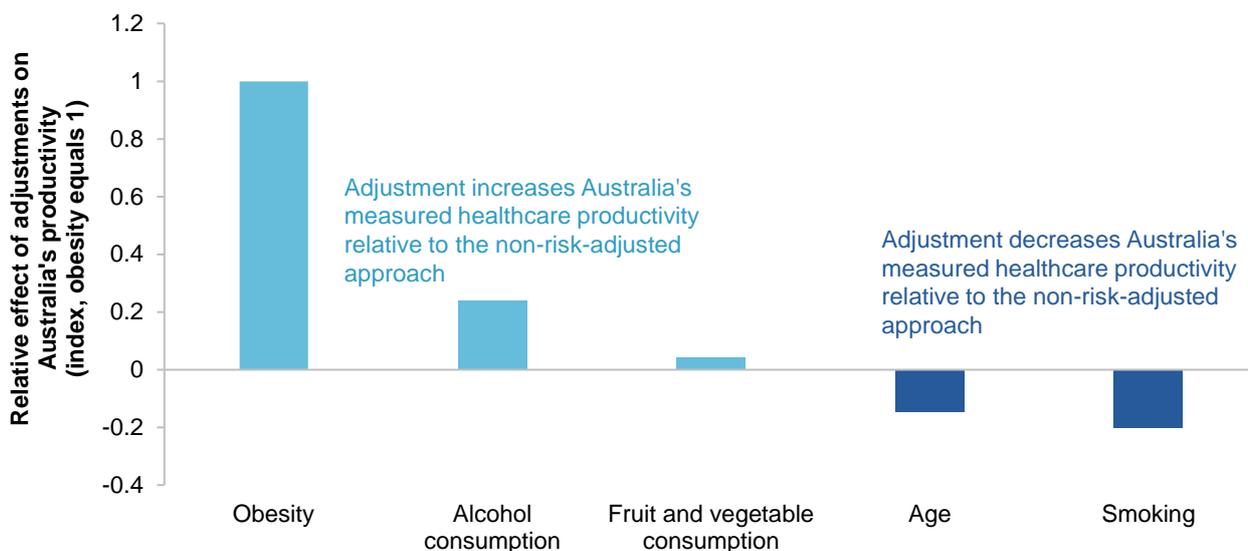
**Figure 3.2 – Australia's healthcare sector has been relatively productive**  
**Relative healthcare productivity by country, 2010 to 2019**



Source: Productivity Commission estimates.

Accounting for obesity makes the biggest difference to Australia's relative healthcare productivity. Australia has a high obesity rate and our modelling suggests that cross-country differences in obesity have larger effects on HALE than cross-country differences in any other behavioural or environmental risk factor considered (figure 3.3).

**Figure 3.3 – Adjusting for Australia's high obesity rate increases its measured healthcare productivity<sup>a</sup>**



a. For each (transformed) factor, this is computed as Australia's distance from the mean (on average, over time) multiplied by the estimated marginal effect of the transformed factor on transformed HALE and then converted into an index. The transformations are set out in appendix E.

Source: Productivity Commission estimates.

Alternative models likewise suggest Australia's healthcare sector is relatively productive.

- Using the Healthcare Access and Quality Index instead of HALE as a measure of population health suggests that Australia ranks fifth among the 28 countries for healthcare productivity (about 80<sup>th</sup> percentile).
- When the model is estimated using a larger set of countries – the 52 highest income countries – Australia ranks 10<sup>th</sup> of 52 countries for healthcare productivity (about 80<sup>th</sup> percentile). However, given their larger differences, comparisons across this larger set of countries are less meaningful.

## Implications for policy

The Commission's research suggests that Australia is getting comparatively good value for its healthcare dollar. But an ageing population, societal expectations and the growing burden of chronic disease all mean we can't be complacent when it comes to healthcare reform.

Reducing our sizeable risk factors, such as obesity and alcohol consumption would enable our healthcare sector to do more with less. Prevention can be highly cost-effective, with the costs of managing a disease sometimes twice that of primary prevention (Breadon et al. 2023).

While Australia has made significant gains in reducing rates of smoking and skin cancers, we spend relatively little of our healthcare budget on preventive care. At around 2 per cent of annual health expenditure, Australia spends significantly less on public health than its OECD counterparts (Australian Government Department of Health 2021). Greater and more timely adoption of cost-effective preventive measures would help our health system maintain its ranking as one of the most productive in the world.

# Appendices



## A. Consultation

In preparing this paper, the Commission consulted with a range of individuals and government agencies (table A.1) and convened a roundtable on 10 November 2023 on methodologies for estimating healthcare productivity growth (table A.2). Reviewers also provided comments on earlier drafts of sections of the paper (table A.3).

The Commission is extremely grateful to all those who participated.

### Table A.1 – Consultations

#### People and organisations we consulted with

Australian Bureau of Statistics

Australian Commission on Safety and Quality in Health Care

Australian Institute of Health and Welfare

Clarke, Prof. Philip (University of Oxford)

Cullen, Dr. David

Department of Health

Fox, Prof. Kevin (University of New South Wales)

New South Wales Treasury

The Treasury

United States Bureau of Economic Analysis

Victorian Department of Jobs, Skills, Industry and Regions

### Table A.2 – Roundtable participants

#### Participants at our roundtable on 10 November 2023

Australian Bureau of Statistics

Australian Institute of Health and Welfare

Blakely, Prof. Tony (University of Melbourne)

Fox, Prof. Kevin (University of New South Wales)

Goss, Adj. Assoc. Prof. John (University of Canberra)

**Table A.3 – Reviewers**

People who reviewed earlier drafts of sections of this paper

Cutler, Prof. David (Harvard University)

Goss, Adj. Assoc. Prof. John (University of Canberra)

Lattimore, Dr. Ralph

Nghiem, Dr. Son (Australian National University)

Staff at the United States Bureau of Economic Analysis

## B. Traditional approaches to estimating healthcare productivity growth

This appendix describes the ‘traditional’ approach to measuring healthcare productivity growth (section B.1) and outlines its strengths and weaknesses (section B.2).

### B.1 What is the traditional approach to measuring healthcare productivity growth?

The rate of growth of healthcare productivity is the rate of growth of healthcare outputs divided by the rate of growth of healthcare inputs.<sup>17</sup>

Measuring the growth of healthcare inputs – labour, capital and intermediate goods – is not hugely dissimilar from measuring the growth of inputs in other sectors. As with inputs in other sectors, healthcare inputs have prices, such as doctors’ salaries and the price of bandages. These prices can be used to infer how changes in the mix and quality of inputs result in changes in overall healthcare inputs.

But measuring the growth of healthcare outputs is challenging. Most healthcare outputs are not sold at market prices, they are provided for free or with large subsidies. Hence, changes in the mix and quality of healthcare outputs must be mapped to changes in the growth of overall output using something other than market prices.

The ‘traditional’ approach (as termed by Sheiner and Malinovskaya (2016)) used by some statistical agencies to measure healthcare output growth is as follows (or is analogous to what follows):

- Healthcare is divided into ‘units’ that are single occasions of service delineated by the setting in which the service is provided (for example, a hospital admission is considered to be different from a general practitioner (GP) consultation) and patient characteristics (for example, a hospital admission of an 85 year old man is considered to be different from a hospital admission of a 32 year old woman). The average cost of delivering each of these types of units in the base year is estimated.
- An index of healthcare output in each period is computed as the sum of the units of healthcare weighted by their base period average costs and the percentage change is taken. More formally, the percentage change in healthcare output is computed according to the following equation, where  $q_t^i$  is the quantity of healthcare unit type  $i$  delivered in period  $t$  and  $c_t^i$  is its average cost.

$$\% \Delta \text{Healthcare output}_{t,t+1} = 100 \times \left( \frac{\sum_i c_t^i q_{t+1}^i}{\sum_i c_t^i q_t^i} - 1 \right) \quad (1)$$

<sup>17</sup> Throughout this paper, ‘productivity’ refers to multifactor productivity.

Hence, the traditional approach assumes that the value of a 'unit' of healthcare in both periods  $t$  and  $t + 1$  is equal to the average cost of producing that unit in period  $t$ .

As an example of the traditional approach, consider the Australian Bureau of Statistics' experimental estimates of hospital productivity, which found that hospital productivity grew by 0.1% per annum between 2008-09 and 2018-19 (ABS 2021). The output growth estimates that fed into these productivity growth estimates were derived by using equation 1 with the units defined as follows:

- Admitted hospital patient 'separations' delineated by disease and patient age (five year bands).
- Non-admitted hospital patient 'events' delineated by type of service (e.g. 'chemotherapy treatment' or 'magnetic resonance imaging').
- Emergency department 'presentations' delineated by urgency-related group (e.g. 'urgent-injury' or 'emergency-neurological illness').

## B.2 What are its strengths and weaknesses?

The traditional approach has been popular in the past because it requires less data to implement than some other approaches and can be flexibly altered to make use of the data available. For example, in Australia, administrative records contain much richer information about the characteristics of patients admitted to hospital and the care they received than they do about the characteristics of patients attending a short consultation with a GP. With the traditional approach, it is possible to use this richer information to finely delineate units of admitted hospital care while taking a coarser approach to units of GP care. The traditional approach is also suited to national accounting, as it can classify healthcare services the same way they are classified in national accounts and does not violate conventions of national accounting (chapter 1).

But the tide is turning. Starting with a landmark 1998 study of the productivity of heart attack treatment in the United States (Cutler et al. 1998), health economists have sought to move away from the traditional approach. Initial efforts mostly focussed on the treatment of specific diseases (Sheiner and Malinovskaya 2016), but more recently there have been several studies of healthcare productivity from the United States that capture most, if not all, of the healthcare sector using non-traditional methods (Cutler et al. 2022; Dunn et al. 2022; Weaver et al. 2022).

There are two reasons that the traditional approach has fallen from favour. First, it does not capture improvements in unit quality. Second, it mismeasures productivity when consumers substitute between units. The remainder of this appendix explores these shortcomings and explains why it is necessary to target both of them simultaneously when seeking to move beyond the traditional approach.

### Shortcoming 1: Improvements in unit quality are not captured

Improvements in the quality of a given unit of healthcare without accompanying cost increases are genuine productivity improvements. For example, if emergency department care improves such that a larger share of patients presenting with an urgent injury survive and costs do not rise, these improvements should result in higher measured outputs and unchanging measured inputs, with the result being an increase in measured productivity.

However, these changes are not captured by the traditional approach. For example, there is no scope for an improvement in the average quality of any unit between period  $t$  and period  $t + 1$  to enter into equation 1 either explicitly or implicitly. The reason is that a quality improvement in some unit between period  $t$  and period  $t + 1$  violates the assumption that the average cost of producing that unit in period  $t$  is indicative of its value in both periods  $t$  and  $t + 1$ .

This is a key shortcoming. Studies from abroad have found that changes in the quality of units of healthcare have a large effect on healthcare output and productivity. For example:

- two recent studies of healthcare productivity from the United States, while adopting different methodologies, both found that quality change added about 3 percentage points per annum to healthcare productivity growth between 2000 and 2017 (Dunn et al. 2022) and between 1999 and 2012 (Cutler et al. 2022; specific to people aged 65 and over)
- using a less comprehensive method, the United Kingdom Office for National Statistics (ONS) found that quality change added 0.4 percentage points per annum to healthcare productivity growth in England between 2001 and 2019 (ONS 2021).

## Shortcoming 2: Substitution between units leads to mismeasurement

In cases where units of healthcare (as defined using traditional methods) are substitutes, mismeasurement can result.

Suppose, for example, that treatment in a lower cost setting (such as non-admitted hospital care) were to substitute for treatment in a higher cost setting (such as admitted hospital care) to achieve the same outcome for a patient. This would be a genuine productivity increase as the same outcome for the patient is achieved, but with fewer inputs.

However, under the traditional approach, measured output would decrease (see equation 1) and so the effect on measured productivity would be ambiguous.

Substitution can occur between settings, as in the example above, but also within settings. If a single occasion of service substitutes for multiple services (or vice versa) to achieve the same outcome for a patient, the traditional approach will similarly result in mismeasurement. For example, suppose that a single hospital admission was to substitute for multiple hospital admissions to achieve the same outcome for a patient. This would be a genuine productivity increase, as the same outcome for the patient is achieved but with fewer inputs.

However, by the traditional approach, measured output would again decrease (see equation 1) and so the effect on measured productivity would be ambiguous.

In both of these cases, the reason for the mismeasurement is that substitution toward either lower or higher cost units or more or fewer units to achieve the same outcome violates the assumption that the average cost of a unit in period  $t$  indicates its value in both periods  $t$  and  $t + 1$ .

Evidence from abroad indicates that this is also a key shortcoming, although likely not as significant as the previous shortcoming when the quality of the evidence is taken into consideration. Assuming that substitution between units of healthcare does not affect patient outcomes, studies find that:

- the traditional approach would have led to healthcare productivity growth being overstated perhaps by about 1–1.5 percentage points per annum in the United States over 2000–2010 (Dunn et al. 2015)
- the traditional approach would have led to hospital productivity growth being understated by about 3 percentage points per annum in Canada over 2002–2010, although this study was of lower quality (Gu and Morin 2014).

## Addressing one of these shortcomings but not the other is risky

While these two shortcomings are distinct, addressing them requires a coordinated solution.

- Addressing the first shortcoming but not the second can result in error. Changes in unit quality can be inferred from changes in patients' health outcomes after consuming healthcare, but better care might also involve more or different units of care, such as more follow-up visits with the GP or the hospital. It is difficult to disentangle the effects of changes in the quality of a given unit from the effects of consuming a different number or mix of units.
- Addressing the second shortcoming but not the first would require strong assumptions about how substitution between units affects outcomes. One approach would be to assume that substitution between units does not affect patient outcomes (as per the examples above). This may be reasonable in some cases, for example a pharmacist giving a vaccination instead of a general practitioner. But in other cases, the assumption implicit in the traditional method (that patient outcomes are proportional to unit costs) may be more appropriate. For example, if patients experiencing depression were to see a psychologist more often (substitution of fewer units for more units), their health outcomes might improve.

A recent study from the United States provides a cautionary tale of the risks of addressing one shortcoming but not the other (Dunn et al. 2022). It found that official healthcare productivity growth statistics (computed using the traditional method) *understated* true healthcare productivity growth by 1.76 percentage points per annum. Had the study accounted for substitution between units while assuming no changes in patient outcomes, it would have concluded that the official statistics *overstated* true productivity growth by 1.43 percentage points per annum (an error of 3.19 percentage points). And had it accounted for changes in patient outcomes while assuming no substitution between units, it would have concluded that the official statistics *understated* true productivity growth by 3.19 percentage points per annum (an error of 1.43 percentage points).<sup>18</sup>

The upshot is that improved productivity statistics should account for both changes in patient outcomes (due to both substitution between units and improvements in unit quality) and the cost implications of substitution between units.

---

<sup>18</sup> The alignment of the 1.43 percentage points and 3.19 percentage points between the comparisons is not coincidental.

## C. How we estimated healthcare productivity growth

This appendix provides further background on the healthcare productivity growth estimates presented in chapter 2 and presents more detailed results. Section C.1 discusses the methodology and data and section C.2 presents the detailed results.

### C.1 Methodology and data

#### Overarching methodology

The Commission has estimated multifactor healthcare productivity growth for about one third of the Australian healthcare sector (by spending) between 2011-12 and 2017-18, specifically healthcare devoted to treating cancers, cardiovascular diseases, blood and metabolic disorders, endocrine disorders and kidney and urinary diseases. Our methodology has two key features, both of which draw from recent academic studies of healthcare productivity growth in the United States – person-centred units of healthcare and adjustments for quality change.

#### Key feature 1: Person-centred units of healthcare

We divide the healthcare sector into person-centred ‘units’ of healthcare. We define a unit of healthcare to be ‘all healthcare received in a financial year to treat a case of disease in a person of a given age (in bands)’. An example of a unit of healthcare is ‘all healthcare received in 2018-19 to treat a case of cancer in a person aged 50-54’. This differs from the traditional approach, which is to divide the healthcare sector into service-centred units such as hospital admissions and general practitioner consultations (appendix B).

In our units of healthcare, ‘disease’ is defined at the level of three broad groups of diseases:

- *Cancers* – all cancers.
- *Cardiovascular diseases* – all cardiovascular diseases.
- *Blood, endocrine and kidney diseases* – all blood and metabolic disorders, endocrine disorders and kidney and urinary diseases.

This way of defining units of healthcare makes our productivity growth estimates robust to substitution between different types of treatment and between more and fewer treatments. As one unit of healthcare encompasses all treatment received for a case of a specific disease in a year, the number of units consumed is invariant to changes in the number of occasions of service received per case, and the setting in which those occasions of service are received. By contrast, the service-centred approach is not robust to these substitutions (appendix B). Our approach is more commonly known as a ‘disease-based’ approach to defining units of healthcare (Hall 2015; Sheiner and Malinovskaya 2016), although it is its patient-centred nature (rather than its disease-specific nature) that makes it robust to substitution between units.

Using patient-centred units of healthcare also allows us to adjust for healthcare quality change (key feature 2). Adjusting for healthcare quality change when using alternative service-centred units of healthcare risks introducing bias (appendix B).

## Key feature 2: Adjustments for healthcare quality change

We adjust for healthcare quality change by examining changes in health outcomes (morbidity and mortality) per unit of healthcare (i.e. per case of disease). Our approach for doing so is described in greater detail in a later section.

### Bringing the key features together

Equation 1 shows how we compute healthcare productivity growth on a units-by-units basis, where  $c_{i,t}$  is the average real cost of units of type  $i$  (i.e. cases of diseases, delineated by age) in year  $t$  and  $v(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18})$  is the real monetary value of the average change in healthcare outcomes associated with unit type  $i$  between 2011-12 and 2017-18 (with all values converted to real terms using a healthcare input price deflator).

$$\text{Healthcare productivity growth}_{i,2011-12,2017-18} = \frac{c_{i,2011-12} + v(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18})}{c_{i,2017-18}} \quad (1)$$

This equation incorporates the ‘cost of living index’ method to adjust for healthcare quality change, the approach taken by much of the academic literature (for example, Dunn, Hall and Dauda (2022), Cutler et al. (2022) and Weaver et al. (2022)) and shown to be more robust than alternatives to the distortions common to healthcare provision (Dauda et al. 2022). Specifically, we use the Paasche index variant of the cost of living index method (also known as the ‘Paasche reservation price index’ method (Dunn et al. 2022)). Box C.1 shows the derivation of equation 1.

To compute healthcare productivity growth for multiple types of units  $j$  (e.g. for a disease, averaging over ages, or for all diseases studied), we take averages of  $c_{i,2011-12}$ ,  $c_{i,2017-18}$  and  $v(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18})$  weighted by  $\text{Prevalence}_{i,2017-18}$ , the number of prevalent cases of the disease during 2017-18 in the age group corresponding with unit type  $i$ .<sup>19</sup> This is shown in equations 2-4.

$$\text{Healthcare productivity growth}_{j,2011-12,2017-18} = \frac{c_{j,2011-12} + v(\Delta\text{Healthcare outcomes}_{j,2011-12,2017-18})}{c_{j,2017-18}} \quad (2)$$

$$c_{j,t} = \frac{\sum_{i \in j} c_{i,t} \text{Prevalence}_{i,2017-18}}{\sum_{i \in j} \text{Prevalence}_{i,2017-18}} \quad (3)$$

$$v(\Delta\text{Healthcare outcomes}_{j,2011-12,2017-18}) = \frac{\sum_{i \in j} v(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18}) \text{Prevalence}_{i,2017-18}}{\sum_{i \in j} \text{Prevalence}_{i,2017-18}} \quad (4)$$

<sup>19</sup> This is equivalent to computing productivity growth on a unit-type-by-unit-type basis as in equation 1 and then taking an average weighted by the share of total healthcare costs devoted to each type of unit. But this is partly co-incident; for example, this equivalence does not hold if the Laspeyres (base-period-weighted) index formulation is used. The prevalence-weighted formulation is preferred for expositional purposes as it is more logical; all variables are expressed in ‘per case’ form and so the number of cases (the prevalence) is the logical variable to weight the components with.

### Box C.1 – Measuring healthcare productivity growth

Approaches to measuring quality-adjusted healthcare output price growth are well established, but quality-adjusted healthcare productivity growth is less commonly studied. This box outlines how we adapted approaches to estimating quality-adjusted healthcare output price growth to measure quality-adjusted healthcare productivity growth.

#### Measuring quality-adjusted healthcare output price growth

To measure healthcare output price growth we followed the ‘cost of living index’ approach, in line with recent healthcare price growth literature. Specifically, we used the ‘Paasche’ index variant of the cost of living index approach as it is robust to the large improvements in healthcare quality we observe for some diseases. The alternative ‘Laspeyres’ and ‘Fisher’ index variants of the cost of living index approach are not defined for some of the large improvements in healthcare quality that we observe.

The Paasche index variant of the cost of living index approach is specified by Dunn, Hall and Dauda (2022), who call it the ‘Paasche reservation price index’. It captures (quality-adjusted) healthcare output price growth<sup>a</sup> as follows, where  $c_{i,t}^t$  is the nominal average cost of healthcare unit type  $i$  in year  $t$  (in year  $t$  dollars) and  $v^{2011-12}(\Delta \text{Healthcare outcomes}_{i,2011-12,2017-18})$  is the nominal value of the change in average healthcare outcomes associated with unit  $i$  between 2011-12 and 2017-18 (in 2011-12 dollars):

$$\text{Healthcare output price growth}_{i,2011-12,2017-18} = \frac{c_{i,2017-18}^{2017-18}}{c_{i,2017-18}^{2011-12} + v^{2011-12}(\Delta \text{Healthcare outcomes}_{i,2011-12,2017-18})}$$

#### Measuring quality-adjusted healthcare productivity growth

Healthcare productivity growth is the ratio of healthcare output growth to healthcare input growth. For healthcare unit type  $i$ , healthcare productivity growth between 2011-12 and 2017-18 is defined as follows:

$$\text{Healthcare productivity growth}_{i,2011-12,2017-18} = \frac{\text{Healthcare output growth}_{i,2011-12,2017-18}}{\text{Healthcare input growth}_{i,2011-12,2017-18}}$$

Healthcare output growth (a real quantity) can be expressed as the ratio of nominal healthcare spending growth to (quality-adjusted) healthcare output price growth, defined by the Paasche index formula given above. These are artificial prices because much healthcare is provided free of charge and because the price changes incorporate quality change, but the link between these variables is clear by analogy with market-provided goods and services. With market-provided goods and services, for a given change in the nominal output of a sector (the spending by consumers on those goods and services), the faster prices rose the less was the growth in the real output of that sector. Likewise, healthcare input growth (also a real quantity) can be expressed as the ratio of nominal healthcare spending growth to healthcare input price growth.<sup>b</sup> The analogy to market-provided goods and services assists here too – disregarding taxes, spending by consumers is revenue that is paid to capital (profits), intermediate and labour inputs. Reinvested profits can be disregarded as our measures of nominal healthcare spending exclude spending on new capital (although they include capital depreciation). Again, for a given change in the nominal inputs consumed by a sector, the faster those input prices rose the less was the growth in the real inputs consumed by that sector.

$$\text{Healthcare output growth}_{i,2011-12,2017-18} = \frac{\text{Nominal healthcare spending growth}_{i,2011-12,2017-18}}{\text{Healthcare output price growth}_{i,2011-12,2017-18}}$$

### Box C.1 – Measuring healthcare productivity growth

$$\text{Healthcare input growth}_{i,2011-12,2017-18} = \frac{\text{Nominal healthcare spending growth}_{i,2011-12,2017-18}}{\text{Healthcare input price growth}_{i,2011-12,2017-18}}$$

This means that healthcare productivity growth can be expressed as a ratio of the growth in these prices.

$$\begin{aligned} \text{Healthcare productivity growth}_{i,2011-12,2017-18} &= \frac{\text{Healthcare input price growth}_{i,2011-12,2017-18}}{\text{Healthcare output price growth}_{i,2011-12,2017-18}} \\ &= \frac{c_{i,2017-18}^{2017-18}}{c_{i,2011-12}^{2017-18} + v^{2011-12}(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18})} \end{aligned}$$

Scaling by healthcare input price growth converts these nominal terms to real terms, so (dropping the superscripts, which are now redundant) healthcare productivity growth for unit  $i$  is:

$$\text{Healthcare productivity growth}_{i,2011-12,2017-18} = \frac{c_{i,2011-12} + v(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18})}{c_{i,2017-18}}$$

We then convert this to an annualised growth rate.

**a.** ‘Growth’ is expressed in decimal terms in this context. For example, growth of 1.03 means a 3% growth rate. **b.** In principle this input price deflator should account for changes in input quality too, but in practice no such deflator exists. The consequence is that our estimates attribute quality changes in inputs (such as better drugs and medical equipment) to the healthcare sector.

With equations 1–4 established, the task of estimating healthcare productivity growth reduces to one of estimating  $c_{i,2011-12}$ ,  $c_{i,2017-18}$  and  $v(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18})$ . The remaining subsections in this section discuss our approach to estimating these variables and the choice of diseases studied.

## Estimating the value of the change in healthcare outcomes

### Value of a quality-adjusted life year

We measure most components of  $v(\Delta\text{Healthcare outcomes}_{i,2011-12,2017-18})$  in quality-adjusted life year-denominated (QALY-denominated) units and convert them to dollar value terms by scaling by an estimate of the value of an additional QALY to society. We assume an additional QALY to be worth \$110,000 for our central estimate, \$65,000 for our lower-bound estimate and \$208,000 for our upper-bound estimate (in 2018 Australian dollars; box C.2).

### Box C.2 – Value of a quality-adjusted life year

QALYs are commonly converted into dollar values to allow comparison with the costs of achieving them. To estimate healthcare productivity growth, we used a central estimate of the value of a QALY and upper- and lower-bound estimates as robustness checks. In 2018 dollars, our central estimate values an additional QALY at \$110,000, our lower-bound estimate values an additional QALY at \$65,000 and our upper-bound estimate values an additional QALY at \$208,000.

### Box C.2 – Value of a quality-adjusted life year

- Our central estimate is based on a recent meta-analysis of the value of a QALY wherein the authors developed a model for valuing a QALY in 2018 based on characteristics of the population in question (income per capita, average years of schooling, the unemployment rate, the average age of the adult population and the share of the population who are female) (Kouakou and Poder 2022). We applied this model to the 2018 Australian population, which had: an average of 12.5 years of schooling among people aged 25 years and over (World Bank 2024), a gross national income per capita of \$51,000 (United States dollars; purchasing power parity basis) (OECD 2024), 50.4% female and an average age among the adult population of 47 years (Productivity Commission estimates based on ABS 2024e).
- Our lower-bound estimate is based on studies of the implicit additional-cost-per-QALY-gained threshold applied by the Pharmaceutical Benefits Advisory Committee (PBAC) when deciding whether to recommend a drug be listed on the Pharmaceutical Benefits Scheme. A study of 2010 to 2018 PBAC recommendations suggested that the threshold lay between \$45,000 and \$75,000 (in nominal terms) (Lybrand and Wonder 2020), while another study noted that statements made by the PBAC in 2014 suggested that the threshold lay between \$45,000 and \$75,000 (Carter et al. 2016). We took the mid-point of this range, \$60,000, assumed it applied to 2014, and adjusted it by the wage-price index to 2018 dollars.
- Our upper-bound estimate is the Australian Government Office of Impact Analysis's recommended value of a statistical life year in 2023 for cost-benefit analysis, adjusted by the wage-price index to 2018 (OIA 2023b). While a statistical life year is conceptually similar to a QALY, its value is estimated using different methods, which usually yield much larger estimates (Ananthapavan et al. 2021).

### Identifying assumptions and structure of the model

To estimate the value of the change in healthcare outcomes, we use an approach original to this study but inspired by approaches taken in recent studies of healthcare productivity growth from the United States, most notably the quality-adjusted life expectancy approach developed by Cutler et al. (2022) and the disability-adjusted life years framework used by Weaver et al. (2022).

We have data on the following indicators:

- The one-year prevalence of each disease in each age group each year:  $Prevalence_{i,t}$ .
- The average disease-attributable QALYs lost per case of disease in each year, known in burden of disease studies as 'years lived with disability' (YLD) per case of disease:  $\frac{YLD_{i,t}}{Prevalence_{i,t}}$ . It is more convenient to frame this as a measure of disease-attributable quality of life per case (the average number of disease-attributable QALYs gained from living with the disease for one year):  $Quality\ of\ life_{i,t} = 1 - \frac{YLD_{i,t}}{Prevalence_{i,t}}$ .
- The number of people who die from each disease each year (based on primary cause of death; deaths are not attributed to multiple diseases in this data):  $Deaths_{i,t}$ . Using this we derive an annual survival rate:  $Survival\ rate_{i,t} = 1 - \frac{Deaths_{i,t}}{Prevalence_{i,t}}$ .
- Life expectancy and health-adjusted life expectancy (HALE) at each age (single years of age), delineated by sex.

Given the complexity of the task and few available indicators of healthcare outcomes, we make several identifying assumptions. Loosely stated, these assumptions are that healthcare affects disease-attributable

outcomes but not the prevalence of the disease and that other (non-healthcare) factors affect disease prevalence but not disease-attributable outcomes (a more precise statement of these assumptions is given in box C.3). For example, smoking is assumed to affect the prevalence of lung cancer but not outcomes for those who have lung cancer, whereas treatment for lung cancer is assumed to not affect lung cancer prevalence. This allows changes in the lung-cancer-attributable outcomes (measured in QALYs) to be attributed to changes in lung cancer treatment.

### Box C.3 – Identifying assumptions

To estimate the value of the change in healthcare outcomes per case for each type of unit we make two identifying assumptions.

- Assumption 1: changes in healthcare (past and present) do not change the reported prevalence of each disease. Under this assumption, changes in the number of people in each age group who have each disease are assumed to be due to changes in behaviour and the environment. For example, changes in the prevalence of lung cancer are allowed to be due to changes in smoking rates, but are assumed to not be due to changes in the past treatment of lung cancer and any other disease (which might allow for faster recovery of earlier incident cases), changes in preventive care or changes in lung cancer screening.
- Assumption 2: changes in factors other than healthcare (past and present) and changes in past healthcare do not affect the severity of prevalent cases of each disease. Under this assumption, changes in health outcomes among prevalent cases in a given year are assumed to be due only to changes in the healthcare supplied in that year. For example, changes in health outcomes of prevalent lung cancer cases are assumed to be changes in lung cancer treatment in that year, and not due to changes in smoking rates or past treatment for lung cancer.

While clearly oversimplifications, these assumptions are necessary given the paucity of available data. The same identifying assumptions are also made in studies by Highfill and Bernstein (2019), Wamble et al. (2019) and Weaver et al. (2022).

A wrinkle is that many healthcare outcomes occur in the future and so are also influenced by healthcare that will be received in the future. For example, if healthcare improves such that fewer lung cancer patients die of lung cancer in the year in question, the additional survivors will live longer lives, in part because of the lung cancer treatment they received in that year and in part because they will receive more healthcare in the future. Counting the additional QALYs but not these additional healthcare costs would amount to counting the benefits but disregarding some of the costs that generate those benefits. For this reason our approach incorporates estimates of these future costs. Cutler et al. (2022) also incorporate future healthcare costs in their approach.

We discount both future QALYs and costs at 7% per annum in real terms, with 3% and 10% discount rates presented as sensitivities, as recommended by the Office of Impact Analysis (2023a). The appropriateness of this discount rate is discussed in the next section.

Equation 5 summarises our approach, where  $V_{QALY}$  is the dollar value of a QALY,  $QALY_{i,t}^h$  is the average number of QALYs per case for unit type  $i$  with healthcare received in year  $t$  ( $h = Healthcare$ ) and in the absence of healthcare received in year  $t$  ( $h = No\ healthcare$ ), and  $Future\ cost_{i,t}^h$  is the associated future healthcare costs per case:

$$v(\Delta \text{Healthcare outcomes}_{i,2011-12,2017-18}) = \left[ (V_{QALY} \times QALY_{i,2017-18}^{\text{Healthcare}} - \text{Future cost}_{i,2017-18}^{\text{Healthcare}}) - (V_{QALY} \times QALY_{i,2017-18}^{\text{No healthcare}} - \text{Future cost}_{i,2017-18}^{\text{No healthcare}}) \right] - \left[ (V_{QALY} \times QALY_{i,2011-12}^{\text{Healthcare}} - \text{Future cost}_{i,2011-12}^{\text{Healthcare}}) - (V_{QALY} \times QALY_{i,2011-12}^{\text{No healthcare}} - \text{Future cost}_{i,2011-12}^{\text{No healthcare}}) \right] \quad (5)$$

As noted previously, our assumptions imply that patients' health in the absence of healthcare would have been the same in both years, hence  $QALYs \text{ per case}_{i,2017-18}^{\text{Without healthcare}} = QALYs \text{ per case}_{i,2011-12}^{\text{Without healthcare}}$  so these terms cancel. Equation 5 thus reduces to:

$$v(\Delta \text{Healthcare outcomes}_{i,2011-12,2017-18}) = [V_{QALY} \times QALY_{i,2017-18}^{\text{Healthcare}} - \text{Future cost}_{i,2017-18}^{\text{Healthcare}} + \text{Future cost}_{i,2017-18}^{\text{No healthcare}}] - [V_{QALY} \times QALY_{i,2011-12}^{\text{Healthcare}} - \text{Future cost}_{i,2011-12}^{\text{Healthcare}} + \text{Future cost}_{i,2011-12}^{\text{No healthcare}}] \quad (6)$$

### Assumptions about the effects of healthcare on future health and future healthcare costs

Given the limited data we have, our approach necessitates assumptions about the effects of healthcare on future health and future healthcare costs. Our approach is to assume that if person survives the disease in year  $t$ , their health from period  $t + 1$  onwards will revert to the average level in the population among people of their age and sex. When considered alongside our choice of discount rate (7% per annum), our approach is at the conservative end of the approaches that existing studies have taken.

- The approach taken by Highfill and Bernstein (2019), Wamble et al. (2019) and Weaver et al. (2022) uses the 'years of life lost' measure from the Global Burden of Disease Study to estimate the future effects of healthcare supplied in year  $t$ . This amounts to assuming that survivors of a disease in a given year will go on to live a very long life in perfect health. These studies also do not discount future QALYs and do not account for corresponding future healthcare costs. Each of these factors, especially the lack of a discount rate, contribute to these approaches being less conservative than our approach. For example, whereas the death of a newborn baby results in the loss of 88 QALYs in these studies, in our approach the death of a newborn baby results in the loss equivalent to about 15 QALYs.<sup>20</sup>
- The approach taken by Cutler et al. (2022) is more conservative than our approach in some respects and less conservative in others. Cutler et al., who study only people aged 65 and over, assume that survivors of a disease in a given year will continue to have that disease for the remainder of their lives. This is a more conservative assumption than ours, as people living with disease tend to be less healthy than the generation population. But Cutler et al. do not discount future QALYs and costs, which is a less conservative choice.

This raises the question of whether our approach is too conservative. There are two points worth making here. First, if we set aside the issue of discount rates, our approach is likely to overestimate future QALYs and costs because people who survive chronic disease in a given year tend to be less healthy than the general population. While Cutler's approach could be considered too conservative in this regard, ours is likely to be insufficiently conservative. Second, our choice of discount rate balances against this. While 7% per annum is the rate recommended by Office of Impact Analysis, it is higher than that typically used in studies involving healthcare. For

<sup>20</sup> Using the years of life lost measure amounts to assuming that newborn babies of either sex who die would have otherwise lived for another 87.9 years in perfect health, resulting in a loss of 88 QALYs per death. Assuming an equal split of girls and boys, our approach would assume they would otherwise live for another 82.5 years and with a undiscounted loss of only 72.5 QALYs per death because of expected years lived in poor health. After discounting at 7% per annum, the present value of the loss in QALY terms is only about 15 QALYs per death. Discounted expected future healthcare costs are about \$50,000 per death which reduces the net benefit further by the equivalent of about half a QALY (with a QALY valued at \$110,000, our central estimate).

example, the PBAC uses a 5% discount rate and many other countries use a rate lower than this for health technology assessment (Centre for Health Economics Research and Evaluation 2022).

Following our assumptions,  $QALY_{i,t}^{Healthcare}$  among those who survive the disease in year  $t$  is the sum of QALYs per case in period  $t$  and discounted remaining HALE at one year more than the average age of those who receive unit type  $i$ .<sup>21</sup> Because HALE varies for reasons largely unrelated to treatment of the disease in question, we hold it constant at the average value taken between 2011-12 and 2017-18 and denote it  $HALE_i^d$ , with the  $d$  superscript indicating it is discounted. We have:

$$QALY_{i,t}^{Healthcare} = Survival\ rate_{i,t} (Quality\ of\ life_{i,t} + HALE_i^d) \quad (7)$$

Under these assumptions,  $Future\ cost_{i,t}^h$  can similarly be estimated by taking the average of 2011-12 and 2017-18 life expectancy at one year more than the average age of those in unit type  $i$  and combining this with data on the average cost of all healthcare per person at each age over 2011-12 and 2017-18 to estimate discounted expected future healthcare costs for those who survive the disease in year  $t$ , denoted  $Future\ cost_i^d$ . We have:

$$Future\ cost_{i,t}^h = Survival\ rate_{i,t}^h \times Future\ cost_i^d \quad (8)$$

Because our assumptions imply the survival rate absent healthcare is the same in both years,  $Future\ cost_{i,2017-18}^{No\ healthcare} = Future\ cost_{i,2011-12}^{No\ healthcare}$  and so these terms cancel in equation 6. We are left with:

$$v(\Delta Healthcare\ outcomes_{i,2011-12,2017-18}) = Survival\ rate_{i,2017-18} [V_{QALY} (Quality\ of\ life_{i,2017-18} + HALE_i^d) - Future\ cost_i^d] - Survival\ rate_{i,2011-12} [V_{QALY} (Quality\ of\ life_{i,2011-12} + HALE_i^d) - Future\ cost_i^d] \quad (9)$$

Note that  $V_{QALY} \times HALE_i^d - Future\ cost_i^d$  is positive for all  $i$  and values of a QALY that we are considering except for men and women aged 85 and over when the lower-bound value of a QALY is used; that is, survival is a net benefit to society in nearly all cases considered.

## Data used to measure the value of the change in healthcare outcomes

### Main dataset

Our main estimates, the estimates presented in chapter 2, are based on data from several sources:

- $Prevalence_{i,t}$  is self-reported disease prevalence from the 2011-12 and 2017-18 Australian Bureau of Statistics (ABS) National Health Surveys (NHS) (ABS 2024c, 2024a). As the NHS collects data on only long-term (chronic) diseases, this is taken to represent the one-year prevalence (the number of people who had the disease in question at any point during the year in question).
- $\frac{YLD_{i,t}}{Prevalence_{i,t}}$  is estimated using the 2011-12, 2014-15 and 2017-18 NHS. Our approach was to estimate 'health utility' for each survey respondent based on their age, sex, self-reported overall health, self-reported bodily pain and Kessler-10 psychological distress score and attribute health utility to diseases using a propensity score matching method adapted from Cutler et al. (2022). Appendix D outlines this modelling in detail.
- $Deaths_{i,t}$  is taken from the Australian Burden of Disease Study (ABDS) (AIHW 2021d). The ABDS estimates are based on administrative data on causes of death. In instances where the recorded cause of death is implausible (such as hypertension), is an intermediate cause that has an underlying precipitating cause (such as septicaemia), is a cause that occurs in the final stage of dying (such as cardiac arrest) or is ill-defined or ill-specified, these deaths are distributed to the more probable underlying cause (AIHW 2021c). Deaths are defined at the level of individual ABDS diseases (e.g. lung cancer), which is a

<sup>21</sup> Because HALE varies by sex, we take a weighted average of male and female HALE to match the share of men and women who receive unit type  $i$ . We use the same approach when using life expectancy to estimate future healthcare costs.

more refined categorisation than that used here (e.g. all cancers), but as deaths are attributed to at most one disease the data can be aggregated.

There are two limitations of this data.

First, coverage of the healthcare sector is limited to 34% of 2019-20 healthcare spending because the NHS only collects prevalence data for chronic diseases and because the self-reported prevalence statistics are expected to be unreliable for many chronic diseases. Our coverage is limited to diseases whose prevalence is expected to be reliably captured by the NHS data. This is discussed further in a later section.

Second, diseases are categorised differently in the ABDS data on deaths and the NHS. Our data on costs follows the ABDS classification. This necessitated aggregating diseases to three broad groups to facilitate comparison across the datasets. This is also discussed further in a later section.

### Secondary datasets

We also present two alternative estimates of productivity growth using alternative datasets as robustness checks.

The first alternative dataset is identical to our main dataset except that our estimates of  $\frac{YLD_{i,t}}{Prevalence_{i,t}}$  use a simpler multivariate regression approach to attribute health utility to diseases. This method is likely to be less robust as it requires stronger assumptions but it provides a robustness check for our preferred estimates derived using a propensity score matching approach. The alternative estimates of  $\frac{YLD_{i,t}}{Prevalence_{i,t}}$  are outlined in appendix D.

The second alternative dataset is more substantially different. It uses published ABDS estimates of  $YLD_{i,t}$  and unpublished estimates of  $Prevalence_{i,t}$  obtained from the ABDS. These data run from 2011-12 to 2018-19.

The advantage of the ABDS dataset over our main dataset is that it allows us to estimate productivity growth at more granular levels. It allows productivity growth to be estimated at the level of individual diseases (e.g. lung cancer) rather broader groups of diseases, and it allows for more refined age groups and delineation by sex. This means that it better controls for demographic changes and changes in the mix of diseases within disease groups.

However, there are two disadvantages of this dataset.

First, some ABDS estimates of changes in  $YLD_{i,t}$  and  $Prevalence_{i,t}$  between 2011-12 and 2018-19 are likely to be less accurate. The ABDS's method is sophisticated but data-intensive and for some diseases ideal data does not exist (box C.4). The ABDS estimates are generally of highest quality for cancers.

Second, and more importantly, due to related data quality issues, the coverage of this dataset is more limited – only 10% of healthcare spending (about one third of the coverage of the other datasets). All diseases studied with this dataset are contained within the three broad disease groups studied with our main dataset. The criteria for including and excluding diseases is discussed in a later section.

#### **Box C.4 – Method underpinning Australian Burden of Disease Study estimates of years lived with disability and disease prevalence**

##### **Years lived with disability**

The ABDS estimates years lived with disability using a sophisticated and data-intensive approach. For each disease, the ABDS devises a conceptual model of the health conditions that result from the disease (known as 'sequela'). These sequela are then mapped to one or more 'health states' and each health state is mapped to a disability weight that captures the per-case loss of health (in QALYs terms) that would occur if a person

### **Box C.4 – Method underpinning Australian Burden of Disease Study estimates of years lived with disability and disease prevalence**

were to live in that health state for one year. Some sequela are mapped to a single health state (and so the mapping effectively goes from the sequelae to the health state), while others are delineated by severity (and so each severity level is mapped to a different health state). The ABDS uses a mapping from health states to disability weights developed by the Global Burden of Disease Study.

With this structure in place, the ABDS estimates YLD by estimating the prevalence of each health state (i.e. the prevalence of each sequela for those with one severity level and otherwise the prevalence of each sub-sequela severity level) at a point in time and summing the associated disability weights. Hence, YLD changes because the estimated prevalence of the sequela and sub-sequela severity levels change. The mapping from health states to disability weights is held constant over time to permit comparison between periods.

Therefore, to accurately capture changes in YLD, the ABDS requires contemporary and accurate data on these prevalences. Unfortunately, in some circumstances this data do not exist and so proxies are used:

- For some diseases, the data are not contemporary in both 2011-12 and 2018-19. For example, for coronary heart disease, the prevalence of the sequela are based on contemporary administrative data but the severity distributions within the sequela are held fixed over time.
- For some diseases, the data are not nationally representative. For example, for type 2 diabetes, most of the prevalences are based on stage 2 of the longitudinal Fremantle Diabetes Study, which studies people with diabetes who live in or near to Fremantle, with adjustment to match the age-sex distribution of the wider Australian population.

Our assessment is that changes in YLD are most accurately measured for cancers. The cancer prevalences are generally based on contemporary administrative data from hospitals and cancer registries. A proviso is that these administrative records often capture the incidence of cancer sequela and assumptions about the duration of these sequela are necessary to derive prevalence estimates.

#### **Disease prevalence**

The unpublished ABDS estimates of disease prevalence used in this paper are based on the sequela prevalence estimates described above. The estimates are constructed by taking the sum of the prevalence of sequela associated with the disease in question, excluding sequela that are not mutually exclusive (as this would lead to double counting) and asymptomatic sequela.

As the ABDS disease prevalence estimates are constructed from the ABDS sequelae prevalence estimates, they are subject to the same strengths and weaknesses as the ABDS YLD estimates. The extent to which the exclusion of asymptomatic sequela from the estimates creates additional issues is unclear. For example, if changes in *current period* healthcare were to result in a larger share of cases of the disease in question being asymptomatic then excluding the asymptomatic cases from the prevalence measures would downwardly bias measured healthcare productivity. But if asymptomatic cases were to go untreated in all periods then excluding them would be appropriate.

Our assessment again is that changes in disease prevalence are most accurately measured for cancers, for the same reasons as with the YLD estimates.

Source: AIHW (2021a, 2021b).

The differences between each dataset are summarised in table C.1.

**Table C.1 – Characteristics of each dataset used to estimate the value of the change in healthcare outcomes**

	Main dataset	Dataset with alternative NHS estimates of $YLD_{i,t}$	Dataset with alternative ABDS estimates of $YLD_{i,t}$ and $Prevalence_{i,t}$
Data source for $Prevalence_{i,t}$	ABS NHS	Same as main dataset	Based on ABDS
Data source for $YLD_{i,t}$	Based on ABS NHS using propensity score matching method adapted from Cutler et al. (2022). See appendix D.	Based on ABS NHS using simpler multivariate regression approach. See appendix D.	ABDS
Data source for $Deaths_{i,t}$	ABDS	Same as main dataset	Same as main dataset
Data source for $HALE_i^d$	Based on ABDS	Same as main dataset	Same as main dataset
Data source for $Future\ cost_i^d$	Based on disease expenditure data (described in the next section) and ABS life expectancy estimates.	Same as main dataset	Same as main dataset
Age/sex stratification	Irregular age bands, not separated by sex. Age bands are: 15-39, 40-49, 50-59, 60-64, 65-69, 70-74, 75-79, 80+	Same as main dataset	5-year age bands (0 to 85+), separated by sex
Share of healthcare spending covered	34% in 2017-18	Same as main dataset	10% in 2018-19

## Measuring healthcare costs

To estimate average real healthcare costs per case  $c_{i,t}$  for each unit  $i$  and year  $t$  we take total real healthcare spending associated with each unit type and divide by the number of prevalent cases:

$$c_{i,t} = \frac{\text{Total real healthcare spending}_{i,t}}{\text{Prevalence}_{i,t}} \quad (10)$$

$\text{Total real healthcare spending}_{i,t}$  is adapted from the AIHW's 'disease expenditure database', which delineates healthcare spending by disease, age, and sex (AIHW 2023d). As this data only allocates about three quarters of healthcare spending, we allocate the remaining healthcare spending to units in proportion to their share of total healthcare spending. When doing so we exclude spending on:

- public health, because public health aims to prevent disease rather than treat existing diseases
- research, because spending on research does not treat existing diseases
- capital expenditure, because spending on capital is more appropriately captured via capital depreciation (which is included).

Box C.5 discusses the disease expenditure data's allocation of spending to diseases.

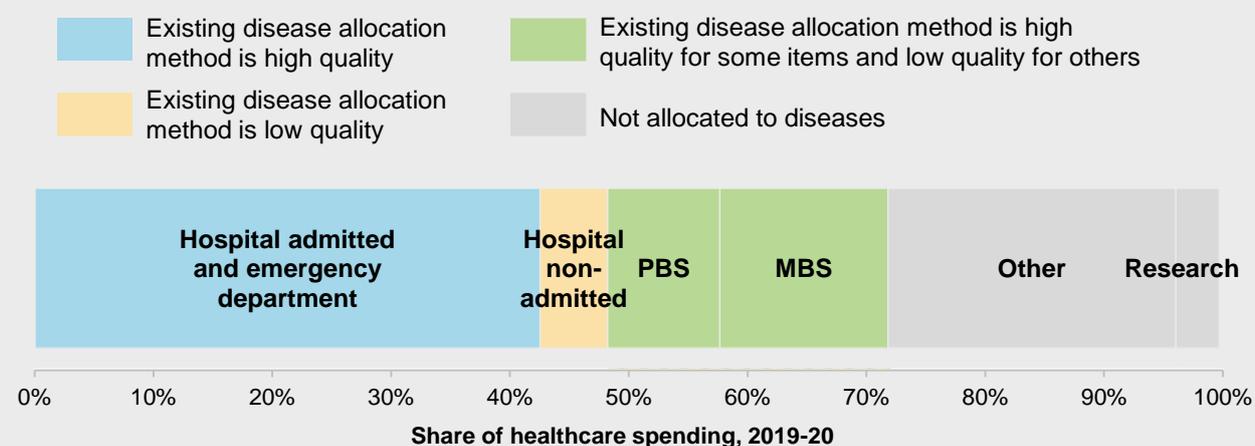
### Box C.5 – Disease expenditure data

The AIHW’s disease expenditure database allocates healthcare spending to disease-age-sex groups. The allocation to age and sex groups is high quality for all allocated spending because administrative records routinely collect the age and sex of the recipient. But allocating spending to the disease for which treatment was received is harder because:

- in some settings, for example, general practitioner consultations, administrative records do not routinely capture the disease being treated
- many patients have comorbidities and so may be treated for multiple diseases at once.

The figure below summarises the quality of the method used to map spending to diseases. Of the spending allocated to diseases, *at least* 68% was allocated using administrative records, which is a relatively high quality method for our purposes. The remainder was allocated using now-discontinued surveys of general practitioner prescription and referral patterns from 2006–2010 (the Bettering the Evaluation of Care and Health survey) applied under the assumption that these patterns remain unchanged aside from demographic shifts. While suited to estimating the *level* of spending on a particular disease (the primary purpose of the disease expenditure data), this method is less appropriate for estimating the year-on-year *change* in spending on a particular disease, as the change is assumption-driven. The change in spending is what matters for productivity growth.

#### Quality of disease expenditure data for the purposes of estimating productivity growth<sup>a,b,c</sup>



**a.** Excludes spending on oral disorders. **b.** ‘Other’ includes over-the-counter pharmaceuticals, public health, community health, and health practitioners who do not receive government subsidies. **c.** ‘MBS’ is the Medicare Benefits Schedule and ‘PBS’ is the Pharmaceutical Benefits Scheme.

Source: Productivity Commission estimates adapted from on AIHW (2022c, 2022b).

The disease expenditure data are converted to real terms using the AIHW’s healthcare input price deflators from its *Health expenditure in Australia* publication (AIHW 2023c). These deflators are specific to the type of expenditure; for example different deflators are applied to hospital services and MBS-rebated services.

There are two shortcomings of using these price deflators:

1. Some are computed using (non-quality-adjusted) output prices rather than input prices. For example, while price growth in community healthcare is measured using changes in health workers wages (an input price) and price growth in hospitals is measured using changes in the government consumption final expenditure deflator (which focusses on input prices), price growth in medical services is examined using changes in Medicare Benefits Schedule fees charged (an output price).
2. It does not adjust for input quality change; for example, improvements in the quality of drugs and medical equipment. This means that some of the productivity growth ascribed to the healthcare sector may be attributable to quality improvements in domestic and offshore drug and medical equipment manufacturing, although patients benefit from quality improvements regardless of their source. This is discussed in chapter 2.

## Diseases studied

### Diseases studies using quality of life and prevalence data from the NHS

When studying diseases using the NHS data on quality of life and prevalence (alongside ABDS data on deaths and disease expenditure study data) we aimed to study as many diseases as possible. However, several limitations of the NHS data restricted this.

- The NHS asks only about diseases present for 6 months or more – chronic diseases.
- The presence of some chronic diseases is likely to be unreliably reported to the NHS. Studies from abroad have typically found that musculoskeletal diseases are less reliably self-reported by patients than other chronic diseases (Galenkamp et al. 2014; Oksanen et al. 2010; Simpson et al. 2004), and consultation with stakeholders suggested that mental health conditions were less reliably captured by the NHS.

A further issue is that the ABDS and NHS classify diseases in different ways. This necessitated aggregating the data from each source to a level where they could be aligned. Among chronic diseases, we were able to align the classifications (with some reallocations and minor residual inconsistencies) at the ABDS disease group level. The exception is the blood and metabolic disorders, endocrine disorders and kidney and urinary diseases ABDS groups – we pooled these groups into a single group to align them with the NHS.

We ended up with three groups of diseases covering about 34% of healthcare spending. Table C.2 summarises these groups.

**Table C.2 – Classification of diseases<sup>a,b</sup>**

### Diseases studies using quality of life and prevalence data from the NHS

ABDS group	NHS group (ICD-10 chapter)	Share of 2017-18	
		healthcare spending (%)	Comments <sup>a</sup>
Blood and metabolic disorders, endocrine disorders and kidney and urinary diseases	Diseases of blood and blood forming organs and endocrine, nutritional and metabolic disorders and diseases of genito-urinary system	14.1	Necessary to pool three groups as many overlaps. Excludes 'high sugar levels in blood/urine' (R73, R81) from the NHS. Includes 'sarcoidosis' (D86.0, D86.2, D86.9) from the respiratory ABDS group, 'urinary tract infections' (N30, N34, N39.0) from the infectious diseases ABDS group, and gestational diabetes' (O24.4), 'endometriosis' (N80), 'polycystic ovarian syndrome' (E28.2), 'infertility' (N46, N97) and 'other reproductive conditions'

ABDS group	NHS group (ICD-10 chapter)	Share of 2017-18 healthcare spending (%)	Comments <sup>a</sup>
			(N43–N45, N47–N50, N60, N62–N64, N70–73, N74.8, N75–N77, N82–N83, N84–N90, N91–N96, N98–N99, O94) from the reproductive and maternal ABDS group.  The following diseases could not be mapped correctly: ‘sequelae of complication of pregnancy, childbirth and the puerperium’ (O94), ‘female pelvic inflammatory disorders in diseases classified elsewhere’ (N74.x), ‘tuberculous cystitis’ (N33.0), ‘late syphilis of kidney’ (N29.0), ‘haemolytic-uraemic syndrome’ (D59.3), ‘female genital prolapse’ (N81), ‘cystic kidney disease’ (Q61), and ‘hypertensive renal disease’ (I12).
Cancers and other neoplasms	Neoplasms	9.2	Includes ‘Uterine fibroids’ (D25) from the reproductive and maternal conditions ABDS group.
Cardiovascular diseases	Diseases of circulatory system	10.5	Excludes ‘symptoms and signs involving circulatory system’ (R00, R01) from the NHS. The following diseases could not be mapped correctly: ‘transient cerebral ischaemic attacks and related syndromes’ (G45), ‘oesophageal varices’ (I85), and ‘hypertensive renal disease’ (I12).

a. Codes are ICD-10 codes. b. Instances where diseases within these NHS groups map to the ‘infectious diseases’ ABDS group have been disregarded. These are generally variants of an infectious disease, for example ‘pneumococcal meningitis’. While these are technically inconsistencies, it is likely that NHS respondents who have the variant would report having the primary condition, e.g. ‘meningitis’ which would be classed as belonging to the ‘certain infectious and parasitic diseases’ NHS group.

Source: Productivity Commission estimates based on AIHW (2023d).

### Diseases studied using quality of life and disease prevalence data from the ABDS

When studying diseases using ABDS data on quality of life and prevalence (alongside ABDS data on deaths and disease expenditure study data) as a robustness check on our main estimates we aimed to study as many diseases as possible from within the three broad groups.

However, the scope to do so was limited by data quality issues. As noted in box C.4, the ABDS’s methods for estimating changes in quality of life and prevalence have significant data requirements and in many instances suitable contemporary data are not available. We reviewed the quality of the quality of life data for 16 diseases – the subset of the 50 diseases that induce the largest loss of health (as measured by disability-adjusted life years) that are contained in our three broad disease groups and are not catch-all categories (such as other blood and metabolic diseases). The data quality was high for 10 of these diseases (mostly types of cancer) and so we included them. We also included a further 4 diseases for which data quality was poorer on the grounds that most of their disease burden is from mortality rather than morbidity – coronary heart disease, chronic kidney disease, stroke and cardiomyopathy – leaving 14 diseases (table C.3).

Treatment of these diseases accounted for about 10% of healthcare spending; about 40% of spending on cancers, about 30% of spending on cardiovascular diseases and about 20% of spending on blood, endocrine and kidney diseases.

**Table C.3 – Diseases studied using quality of life and disease prevalence data from the ABDS**

Disease	Disease group	Share of total DALYs lost in 2018 (%)	Share of 2018-19 healthcare spending (%)	Share of DALYs lost to mortality (%)	Quality of life data relatively high quality (for our purposes)?
Coronary heart disease	Cardiovascular diseases	6.3	1.9	78.1	No
Type 2 diabetes	Endocrine disorders	2.3	1.4	39.3	Yes
Chronic kidney disease	Kidney and urinary diseases	1.0	1.4	70.6	No
Breast cancer	Cancer and other neoplasms	1.4	1.1	86.2	Yes
Bowel cancer	Cancer and other neoplasms	2.0	0.8	93.3	Yes
Stroke	Cardiovascular diseases	2.5	0.7	86.6	No
Lung cancer	Cancer and other neoplasms	3.2	0.5	97.9	Yes
Unknown primary	Cancer and other neoplasms	0.5	0.5	94.8	Yes
Non-Hodgkin lymphoma	Cancer and other neoplasms	0.5	0.4	92.9	Yes
Melanoma of the skin	Cancer and other neoplasms	0.6	0.3	87.7	Yes
Pancreatic cancer	Cancer and other neoplasms	1.1	0.1	98.4	Yes
Brain and central nervous system cancer	Cancer and other neoplasms	0.8	0.1	95.7	Yes
Cardiomyopathy	Cardiovascular diseases	0.5	0.1	86.6	No
Liver cancer	Cancer and other neoplasms	0.8	0.1	98.5	Yes
<b>Total</b>		<b>24</b>	<b>9.5</b>	<b>-</b>	<b>-</b>

Source: Productivity Commission estimates based on AIHW (2021d, 2023d).

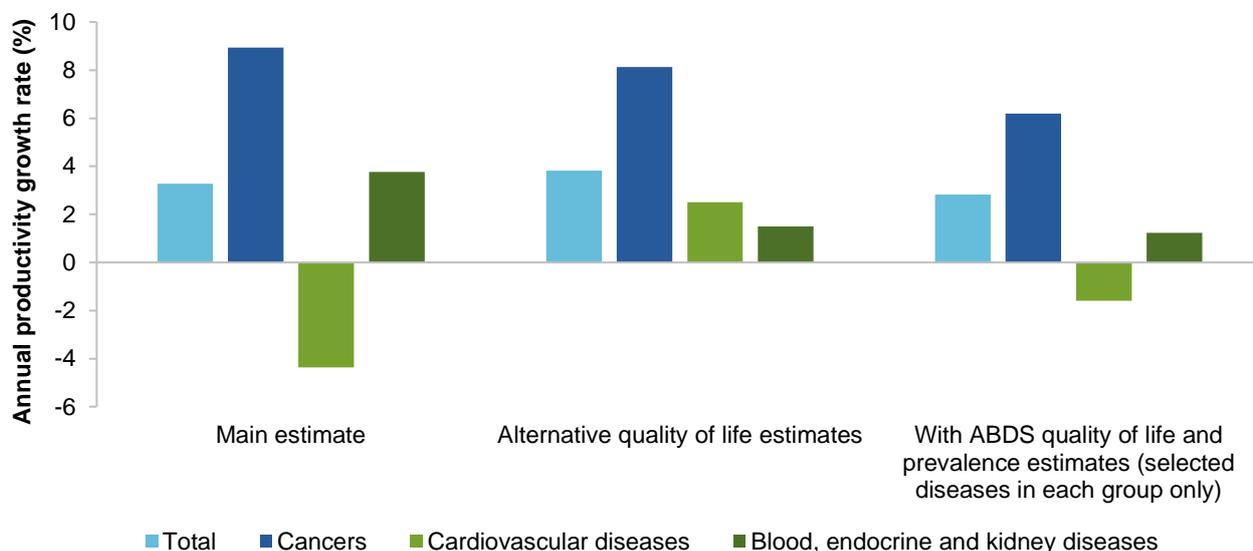
## C.2 Results

### Main results

Among studied diseases, overall healthcare productivity grew at about 3% per annum (figure C.1). Overall healthcare productivity growth was similar when alternative datasets were used; slightly higher with the alternative quality of life estimates from NHS data (appendix D) and slightly lower with the alternative quality of life and prevalence estimates from the ABDS.

At the disease level, the datasets showed similar results (in terms of direction and relative magnitude) for cancers and blood, endocrine and kidney diseases. But the productivity estimates based on alternative quality of life estimates showed rising productivity in the treatment of cardiovascular diseases whereas the other datasets showed declining productivity.

**Figure C.1 – Healthcare productivity growth**



Source: Productivity Commission estimates.

## Decomposition by changes in costs, survival rates and quality of life

We also decomposed the productivity growth estimates into the independent effects of:

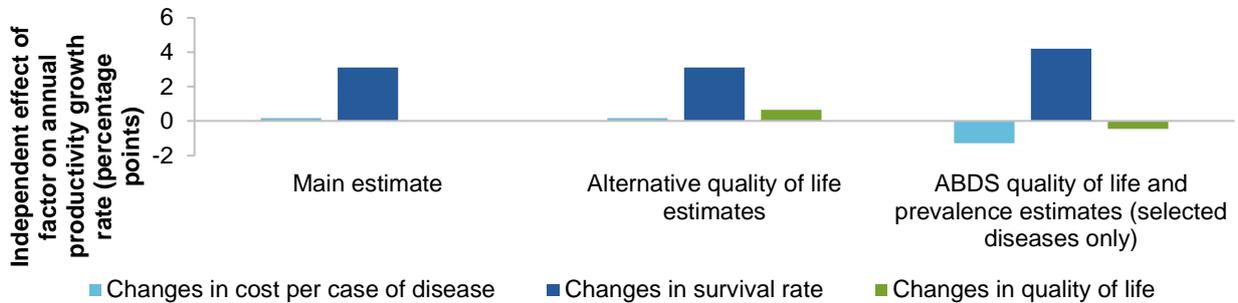
- changes in costs, by assuming no healthcare quality change (setting  $v(\Delta Healthcare\ outcomes_{i,2011-12,2017-18}) = 0$ )
- changes in survival rates, by holding costs and quality of life constant at the 2017-18 level (setting  $Survival\ rate_{i,2011-12} = Survival\ rate_{i,2017-18}$  and setting  $c_{i,2011-12} = c_{i,2017-18}$ )
- changes in quality of life, by holding costs and survival rates constant at the 2017-18 level (setting  $Quality\ of\ life_{i,2011-12} = Quality\ of\ life_{i,2017-18}$  and setting  $c_{i,2011-12} = c_{i,2017-18}$ ).

The sum of these independent effects approximates productivity growth but does not exactly equal it because interactions between the factors also influence productivity growth.

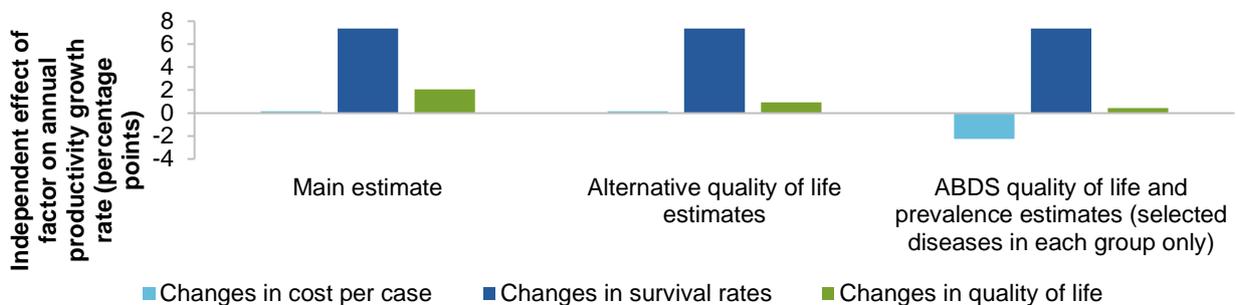
Overall, changes in survival rates were the main driver of productivity growth when estimated using all three datasets (figure C.2, panel a). At the disease level, this was also the case for cancers, but not for the other diseases studied. The main estimates show changes in quality of life to be the major driver for the other diseases, but the alternative datasets show much smaller changes in quality of life (figure C.2, panels b to d). And for cardiovascular diseases, the productivity estimates based on alternative quality of life estimates show increasing quality of life among people living with cardiovascular diseases, which differs from the main estimates and estimates based on ABDS data.

**Figure C.2 – Decomposition of healthcare productivity growth**

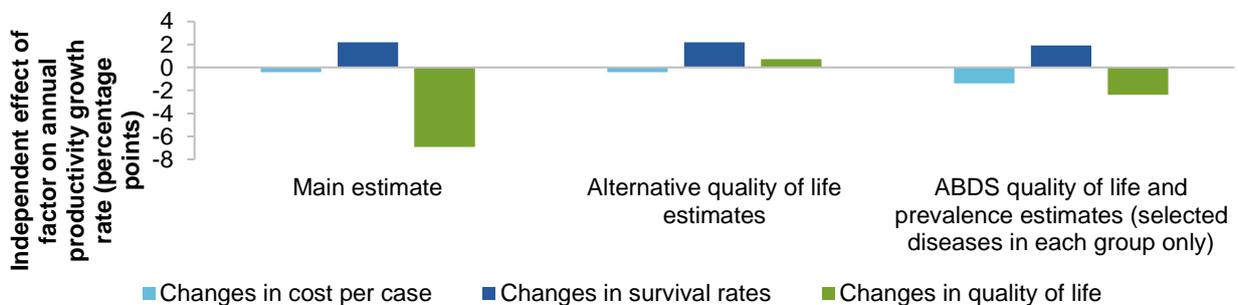
**a. Total**



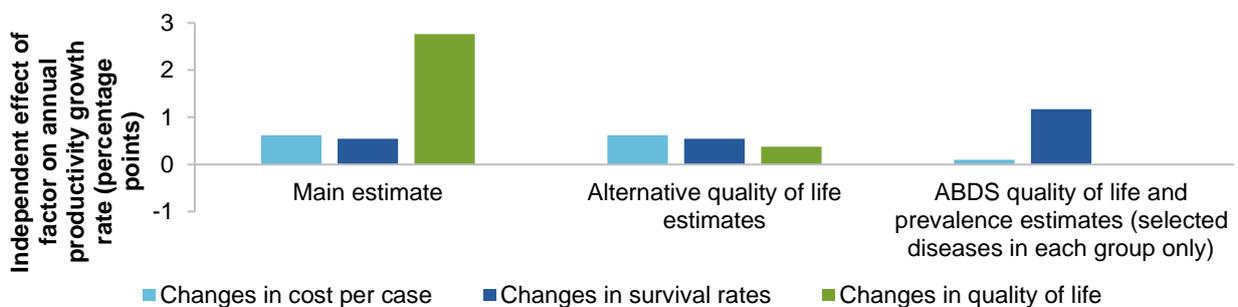
**b. Cancers**



**c. Cardiovascular diseases**



**d. Blood, endocrine and kidney diseases**



Source: Productivity Commission estimates.

These decompositions suggest that the magnitude of healthcare quality improvement has been similar in Australia (among the diseases studied) to the United States, albeit over a different time period. The effect of quality improvement (both via improvements in survival rates and quality of life) contributed about 3 percentage points to productivity growth for the main estimates and about 4 percentage points for both alternative estimates. One recent study from the United States suggests that healthcare quality improvement increased productivity growth by 3.3 percentage points per annum between 2000 and 2017 (Dunn et al. 2022),<sup>22</sup> while another does not provide this decomposition, its results suggests a similar magnitude of quality improvement between 1999 and 2012 (Cutler et al. 2022).<sup>23</sup>

An important sensitivity not well highlighted by comparing results from the alternative datasets is the rate of growth of the prevalence of each disease. Other things being equal, if the number of prevalent cases grew at a faster rate than indicated by our data, both measured changes in costs per case of disease and measured changes in survival rates would have a more positive impact on productivity, and vice versa.

Some alternative sources of data indicate that our estimates may overstate prevalence growth (and so productivity growth) for cancers and understate it for cardiovascular diseases.

- For cancers, the NHS data indicates that cancer prevalence grew at 4.7% per annum between 2011-12 and 2017-18 and the ABDS data (for a subset of cancers) indicates 3.1% per annum between 2011-12 and 2018-19. Alternative data from the Household Income and Labour Dynamics in Australia (HILDA) survey suggests cancer prevalence grew at 2.6% per annum between 2009 and 2017, and administrative data suggests that the incidence of new cancers grew at 2.4% per annum between 2011 and 2017 (AIHW 2023b).
- For cardiovascular diseases, the NHS data indicated that cardiovascular disease prevalence grew at 0.4% per annum between 2011-12 and 2017-18 and the ABDS data (for a subset of cardiovascular diseases) indicates a decline of 1.1% per annum between 2011-12 and 2018-19. Data from the HILDA survey suggests growth of 2.6% per annum between 2009 and 2017.

That said, the HILDA survey is a less authoritative source on disease prevalence than the NHS. The HILDA survey is conducted on a less random sample of the Australian population because it is a panel survey with non-random attrition of respondents from the sample and non-random inclusion of new respondents to the sample.

And while the administrative data on cancer incidence are very high quality, the incidence of new cases of cancer will grow at a different rate from the prevalence of cancer if survival rates change or if the average duration of illness changes.

## Sensitivity of results to the value of a QALY and discount rate

To test the sensitivity of our results, we also estimated productivity growth:

- using the high and low values of a QALY specified in box C.2 (our main results are based on the central value of a QALY)

---

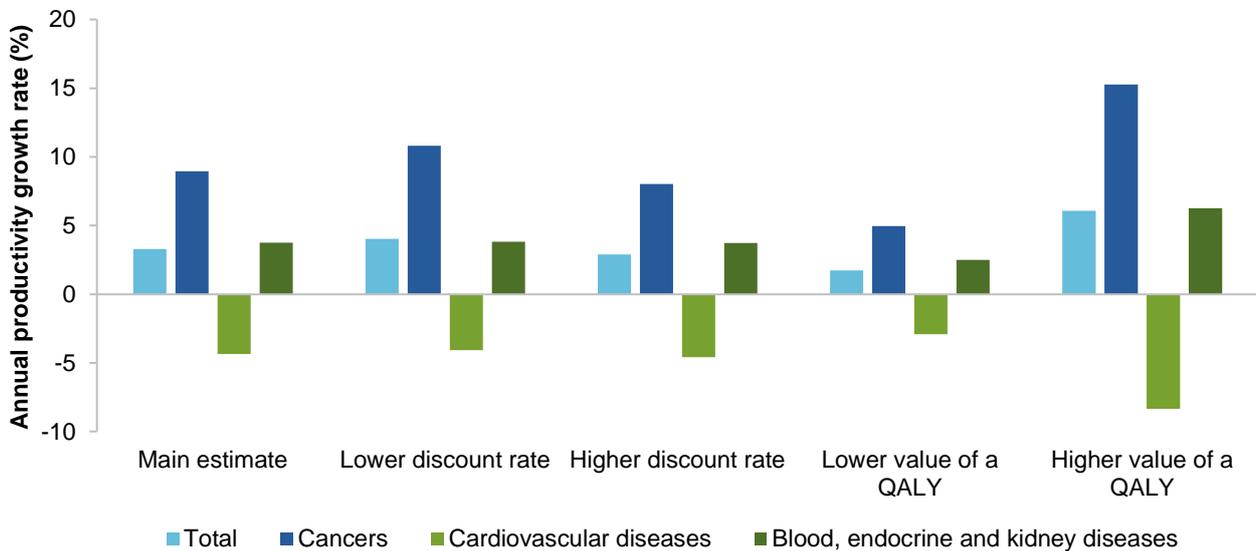
<sup>22</sup> This is computed as follows: the study found that quality-adjusted healthcare prices declined by 1.33% per annum (relative to aggregate GDP inflation), while disease-based (non-quality-adjusted) healthcare prices rose by 1.96% per annum (also relative to aggregate GDP inflation). This suggests that adjusting for healthcare quality improvements reduced measured healthcare output price growth by 3.29 percentage points per annum (relative to a non-quality-adjusted disease-based measure) which is equivalent to increasing measured healthcare productivity growth by 3.29 percentage points per annum.

<sup>23</sup> This is because this study found that overall quality-adjusted productivity growth was 1.5 per cent per annum, very similar to the finding of Dunn, Hall and Dauda (2022) that overall quality-adjusted productivity growth was 1.56 per cent per annum. As the studied time periods largely overlap, this suggests a similar contribution of healthcare quality improvement.

- using a lower (3% per annum) and higher (10%) discount rate. Our central results use a discount rate of 7% per annum (section C.1).

The results are sensitive to both factors, with overall productivity growth ranging from about 2% to about 6% (figure C.3). This is because the factors determine the magnitude of quality change, the main driver of productivity growth. Lower discount rates and higher values of a QALY both increase the importance placed on quality change when computing productivity growth. This leads to higher overall productivity growth because the quality change was positive (healthcare quality improved). With lower discount rates, changes in survival rates have more influence over productivity growth because changes in survival rates affect future QALYs. With higher values of a QALY, both survival rates and quality of life have more influence over productivity growth because dollar-denominated changes in QALYs are amplified when a QALY is worth more.

**Figure C.3 – Sensitivity of healthcare productivity growth to alternative assumptions**



Source: Productivity Commission estimates.



## D. How we estimated changes in quality of life

This appendix outlines how the Commission estimated changes in ‘years lived with disability’ (YLD) per case of disease using Australian Bureau of Statistics National Health Survey (NHS) data. Sections D.1, D.2 and D.3 set out the methodology and section D.4 presents the results.

### D.1 Conceptual approach

There are several steps to estimating YLD per case of disease. This section defines the relevant terms and sketches out the steps.

#### Definition of health utility

A health utility score is a value less than or equal to 1 such that, on average, a person would be indifferent between living for  $x > 1$  years with health utility  $y$  and living for 1 year with health utility  $\frac{y}{x}$ . For example, on average, a person would be indifferent between living for 1 year with health utility 1 and living for 2 years with health utility 0.5. Perfect health is represented by a health utility score of 1, conversely, death is represented by a health utility score of 0. Health utility scores of less than 0 indicate states that, on average, are considered worse than death.

#### Definition of a quality-adjusted life year and YLD per case of disease

A quality-adjusted life year (QALY) is a generic unit of health that draws on the notion of health utility. A QALY is defined as a year of life lived in perfect health. Using the definition of health utility, a person who lives for  $x$  years with health utility  $y$  obtains  $\frac{x}{y}$  QALYs. Because of the way health utilities are defined, living for  $x$  years with *average* health utility  $y$  also obtains  $\frac{x}{y}$  QALYs.

YLD per case of disease – the measure being estimated here – is the number of QALYs lost to the disease in question per case of disease in a one-year period.

#### From health utility to YLD per case of disease

We estimated YLD per case of disease in a given year by:

- using survey data to estimate each survey respondent’s health utility score at a point in time during that year (stage 1; section D.2). Provided that the point in time at which they were surveyed is independent of their health, an individual’s point-in-time health utility is an estimate of the *expected value* of their average health utility in that year and so can be used to compute QALYs over a one-year timeframe. We refer to

their point-in-time health utility as their 'expected health utility' for the year in question:  $E(\text{health utility})_{a,t}$ , where  $a$  indexes survey respondents and  $t$  indexes years

- apportioning some of the health utility of respondents living with disease to those diseases and averaging over all cases of disease (stage 2; sections D.3 and D.4).<sup>24</sup>

## D.2 Stage 1: Estimating expected health utility among survey respondents

### Data sources

Our primary sources of data for estimating expected health utilities are the 2011-12, 2014-15 and 2017-18 NHS, population-representative, cross-sectional surveys that ask about the chronic diseases respondents are suffering from (ABS 2024c, 2024b, 2024a).

Unfortunately, while the NHS contains some questions about quality of life, it does not contain a quality-of-life questionnaire that has been mapped to health utility scores. Fortunately, the Household Income and Labour Dynamics in Australia (HILDA) survey, a panel survey, does contain such a questionnaire (the SF-36 questionnaire) and contains many of the same questions as the NHS about quality of life. (However, the HILDA survey has less comprehensive information about the diseases that respondents are suffering from.)

Our approach was to use both data sources in concert.

### Estimating the expected health utility of HILDA respondents

To estimate the expected health utility of NHS respondents, we developed a model using the HILDA survey that estimates SF-36-derived expected health utility scores based only on responses to questions asked in both the HILDA survey and NHS. We then used this model to estimate expected health utility scores among NHS respondents. These questions are:

- 'In general, how would you rate your health?' (5 point scale)
- 'How much bodily pain have you had during the past 4 weeks?' (6 point scale)
- the Kessler Psychological Distress Scale (K10): a composite indicator based on a 10-item questionnaire
- the respondents' age and sex.

To derive expected health utility scores from responses to the SF-36 questionnaire, we used the Australia-specific correspondence supplied with the HILDA survey (Norman et al. 2014).

We estimated the following model using least squares weighted by survey weights on HILDA data from 2007, 2009, 2011, 2013, 2015, 2017, 2019 and 2021 (110,720 observations), where  $x_{a,t}$  is a vector of responses to the above questions ( $a$  indexes individuals and  $t$  indexes survey waves) and  $\varepsilon_{a,t}$  is a random error term. Self-reported health and bodily pain were included in  $x_{a,t}$  as categorical variables while K10 and the respondents' age were included as continuous variables in linear and quadratic terms.

$$E(\text{health utility})_{a,t} = x'_{a,t}\beta + \varepsilon_{a,t} \quad (1)$$

<sup>24</sup> Because YLD is framed in terms of QALYs lost, we actually apportion disutility (the difference between the utility of perfect health, which is 1, and actual expected health utility) to diseases.

Table D.1 shows the regression output. The  $R^2$  equals 0.73, meaning that the model explains 73% of the variation in health utility scores, and all coefficients are very strongly statistically significant and have the expected sign.

**Table D.1 – Equation 1 estimation results**

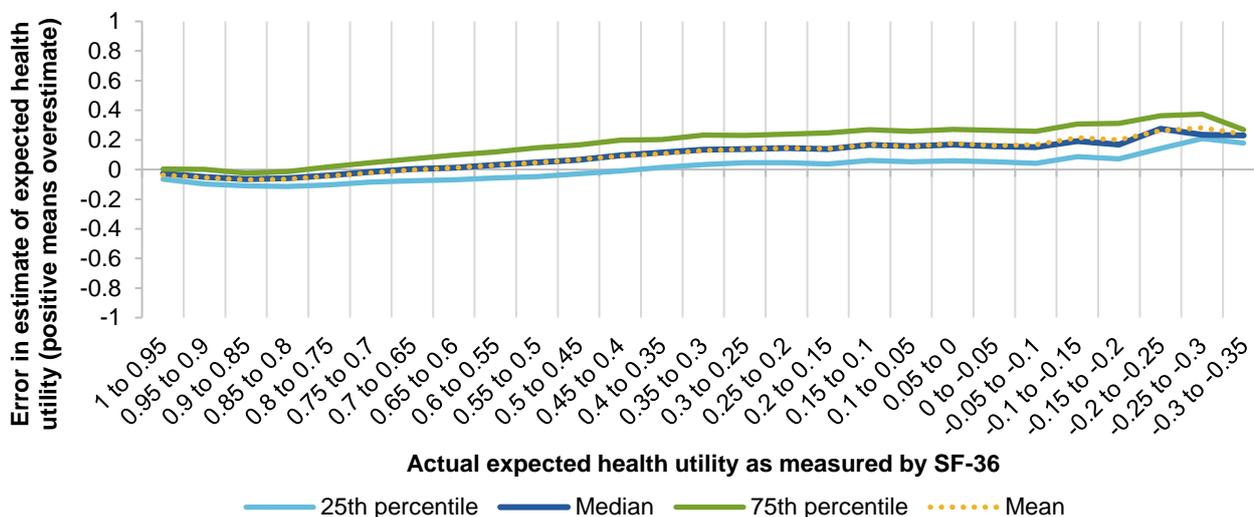
**Dependent variable is**  $E(\text{health utility})_{a,t}$

	Coefficient	P-value
<b>Age</b>	0.00205	0.000
<b>Age squared</b>	-0.00004	0.000
<b>Female (relative to male)</b>	-0.01101	0.000
<b>K10</b>	-0.00045	0.000
<b>K10 squared</b>	0.00001	0.000
<b>Bodily pain (relative to ‘no bodily pain’)</b>	-	-
<b>Very mild</b>	-0.12169	0.000
<b>Mild</b>	-0.18141	0.000
<b>Moderate</b>	-0.26192	0.000
<b>Severe</b>	-0.37444	0.000
<b>Very severe</b>	-0.41907	0.000
<b>Self-assessed health (relative to ‘excellent’)</b>	-	-
<b>Very good</b>	-0.00989	0.000
<b>Good</b>	-0.05109	0.000
<b>Fair</b>	-0.14578	0.000
<b>Poor</b>	-0.26076	0.000
<b>Constant</b>	1.14635	0.000

Source: Productivity Commission estimates.

The quality of the model can also be assessed by examining patterns in the error term  $\widehat{\varepsilon}_{a,t}$  at different levels of  $E(\text{health utility})_{a,t}$ . While the overall fit is good, the model compresses expected health utility scores – it tends to slightly underestimate the expected health utility of those in very good health and overestimate the expected health utility of those in very poor health (figure D.1). This appears to be because of omitted variables that capture other aspects of health rather than a misspecification of the functional form, as more complex functional forms (such as those that allow for interactions between terms) do not improve the model’s fit.

**Figure D.1 – Error in estimate of expected health utility**  
**Distribution of error term by SF-36-derived expected health utility, in intervals of 0.05**



Source: Productivity Commission estimates.

### Estimating the expected health utility of NHS respondents

We then used this model to estimate  $E(\text{health utility})_a$  for each NHS respondent in each edition of the NHS (an out-of-sample prediction). Table D.2 compares the distribution of predicted expected health utility scores by age among HILDA respondents to those of NHS respondents as a check on the quality of the out-of-sample prediction. The shape of the distributions closely match, although NHS respondents tend to report slightly better health than HILDA respondents, especially among people aged 75 and over. This may reflect differences in the sample or survey methodology – NHS respondents are surveyed in person whereas HILDA respondents complete a written survey (for these questions), which may elicit more negative responses.

**Table D.2 – Quality of out-of-sample prediction**  
**Distribution of predicted expected health utilities by age group, HILDA versus NHS**

Age group	Statistic	HILDA (within-sample prediction)	NHS (out-of-sample prediction)	Difference (NHS minus HILDA)
15-49	25 <sup>th</sup> percentile	0.83	0.85	0.02
	Median	0.72	0.73	0.01
	Mean	0.68	0.71	0.02
	75 <sup>th</sup> percentile	0.58	0.60	0.02
50-59	25 <sup>th</sup> percentile	0.77	0.80	0.03
	Median	0.66	0.68	0.02
	Mean	0.61	0.64	0.03
	75 <sup>th</sup> percentile	0.5	0.53	0.03
60-64	25 <sup>th</sup> percentile	0.74	0.77	0.03
	Median	0.63	0.65	0.02

Age group	Statistic	HILDA (within-sample prediction)	NHS (out-of-sample prediction)	Difference (NHS minus HILDA)
60-64	Mean	0.58	0.61	0.03
	75 <sup>th</sup> percentile	0.46	0.48	0.03
65-69	25 <sup>th</sup> percentile	0.73	0.76	0.03
	Median	0.62	0.63	0.01
	Mean	0.58	0.60	0.02
	75 <sup>th</sup> percentile	0.46	0.48	0.02
70-74	25 <sup>th</sup> percentile	0.71	0.74	0.03
	Median	0.58	0.60	0.02
	Mean	0.55	0.58	0.03
	75 <sup>th</sup> percentile	0.42	0.45	0.03
75-79	25 <sup>th</sup> percentile	0.68	0.72	0.04
	Median	0.55	0.58	0.03
	Mean	0.51	0.55	0.04
	75 <sup>th</sup> percentile	0.37	0.42	0.05
80+	25 <sup>th</sup> percentile	0.63	0.68	0.05
	Median	0.48	0.54	0.06
	Mean	0.46	0.51	0.05
	75 <sup>th</sup> percentile	0.32	0.37	0.05
Total	25 <sup>th</sup> percentile	0.79	0.82	0.03
	Median	0.68	0.7	0.02
	Mean	0.64	0.66	0.02
	75 <sup>th</sup> percentile	0.52	0.55	0.03

Source: Productivity Commission estimates.

### D.3 Stage 2: Estimating YLD per case of disease

To estimate YLD per case of disease we attributed a share of the expected health utilities (or, more precisely, disutilities) to the diseases that NHS respondents are suffering from.

The diseases considered comprise three broad disease groups (appendix C).

- *Cancers* – all cancers.
- *Cardiovascular diseases* – all cardiovascular diseases.
- *Blood, endocrine and kidney diseases* – all blood and metabolic disorders, endocrine disorders and kidney and urinary diseases.

To undertake this task, we used a propensity score matching method adapted from the method used by Cutler et al. (2022). This involved several steps.

## Stage 2a: Splitting the sample into age groups, by disease

For each of the three diseases and three surveys we split the sample into seven age groups: people aged 15-49, 50-59, 60-64, 65-69, 70-74, 75-79 and 80 years and over. Splitting the sample by age groups allows us to control for the confounding effects of population ageing in our productivity growth estimates. The age groups were chosen with sample size considerations in mind. Because disease is more prevalent among older people, we used wider age groups for younger survey respondents.

These disease-age groups are indexed by  $i$ . There are 21 disease-age groups: seven for each disease and three for each age group.

## Stage 2b: Estimating the probability of having each disease

For each disease-age group  $i$  and survey (indexed by year  $t$ ), we estimated the probability that survey respondents of the corresponding age had the disease in question. We did so by estimating the following logistic models (indexed by  $i$  and  $t$ ) where  $y_{a,i,t}$  is an indicator variable for whether the survey respondent (indexed by  $a$ ) had the disease corresponding with disease-age group  $i$  in year  $t$ ,  $L(\cdot)$  is the logistic transformation (bounded between zero and one; interpreted as a probability or propensity score),  $\mathbf{X}_{a,i,t}$  is a vector of indicator variables for each of the other diseases, demographic variables and a constant term, and  $\varepsilon_{a,i,t}$  is a random error term.<sup>25</sup>

$$y_{a,i,t} = L(\mathbf{X}_{a,i,t}'\boldsymbol{\beta}_{i,t} + \varepsilon_{a,i,t}) \quad (2)$$

For example, the model for 60-65 year olds with cancer in 2017-18 was estimated on all 60-65 year olds in 2017-18. We estimated 63 models in total (21 groups ( $i$ ) by three surveys ( $t$ )).

## Stage 2c: Comparing ‘similar’ survey respondents who did and did not have the disease to attribute YLD to cases of diseases

From stage 2b, we have propensity scores  $\widehat{y}_{a,i,t}$  and actual values  $y_{a,i,t}$ . For each  $i$  and  $t$ , respondents with similar  $\widehat{y}_{a,i,t}$  are similar in ways that are relevant to whether they have the disease in question. Propensity score matching exploits this by comparing differences in expected health utilities among respondents who actually have the disease in question ( $y_{a,i,t} = 1$ ) with those who do not ( $y_{a,i,t} = 0$ ) among respondents with similar  $\widehat{y}_{a,i,t}$ . This allowed us to attribute expected health utilities (and, thus, YLD) to the disease in question.

Following Cutler et al. (2022), we undertook ‘stratified’ propensity score matching. For each disease-age group and survey, we split the sample into strata (indexed by  $s$ ) based on quintiles of the propensity scores among those who had the disease in question. We then computed the expected health utility for each disease-age group and survey by taking the average of the within-strata differences between those who had the disease in question and those who did not (weighted by survey weights) and averaging over the strata (again weighted by survey weights). This is shown in equation 3 below, where  $i^{-1}$  indexes survey respondents of the same age as members of group  $i$  who did not have the disease corresponding with group  $i$ .

<sup>25</sup> The variables in  $\mathbf{X}_{a,i,t}$  are: a constant term, indicator variables for each disease including diseases not studied (cancers; cardiovascular diseases; blood, endocrine and kidney diseases; mental health conditions; neurological diseases; musculoskeletal diseases; hearing and vision diseases; respiratory system diseases; digestive system diseases; skin diseases and ‘other diseases’), age and age squared, sex, body mass index, household SEIFA decile, number of people in the household, presence of asthma, whether received a blood pressure check in the last 12 months, current smoker status, ever smoked status, marital status, education level, employment status, risky alcohol consumption (2009 guidelines), risky fruit and vegetable intake (2013 guidelines) and risky exercise level (2014 guidelines).

$$\frac{YLD_{i,t}}{\text{Number of people who had disease during period}_{i,t}} = 1 - \frac{\sum_s \left[ \sum_{a \in i, t, s} w_a \left( \frac{\sum_{a \in i, t, s} E(\text{health utility})_{a, w_a}}{\sum_{a \in i, t, s} w_a} - \frac{\sum_{a \in i-1, t, s} E(\text{health utility})_{a, w_a}}{\sum_{a \in i-1, t, s} w_a} \right) \right]}{\sum_s [\sum_{a \in i, t, s} w_a]} \quad (3)$$

We also examined the degree of ‘covariate balance’ for each disease-age group ( $i$ ) and survey ( $t$ ) to check for bias. Covariate balance tests assess, within each stratum, the degree to which those who have the disease in question differ from those who do not according to the covariates  $X_{a,i,t}$  in equation 2. The greater the degree of within-strata similarity between those who have the disease and those who do not (i.e. the greater the degree of covariate balance), the more that differences in health-related quality of life between those who do and do not have the disease can be attributed to the disease. For each disease-age group and survey, we checked the number of covariates for which the within-strata ‘standardised mean difference’ between those who have the disease and those who do not was less than 0.1, a commonly applied threshold (Greifer 2023).<sup>26</sup> On average, over the 63 regressions, 38.7 of 41 variables were adequately balanced according to this threshold.

## Stage 2d: Smoothing the results

To compute productivity growth, we need to know YLD per case of disease for each disease-age group in 2011-12 and 2017-18. One approach would be to obtain estimates for those years from the corresponding edition of the NHS, but this would disregard the estimates from the 2014-15 NHS. Instead, we incorporated the 2014-15 data by computing the results for 2011-12, 2014-15 and 2017-18, fitting a linear trend through these data, and taking the predicted values from that trend to be the YLD per case of disease for 2011-12 and 2017-18.

## D.4 An alternative stage 2

As a robustness check, we also estimated YLD per case of disease using a simpler alternative approach. This approach used a multivariate regression instead of propensity score matching.

With this approach, stages 2a and 2d were the same, but stages 2b and 2c differed.

After splitting the sample into age groups (stage 2a), we estimated the effect of having each of the diseases on expected health utility by estimating the following model for each survey ( $t$ ) and age group ( $i$ ) using ordinary least squares, where the disease variables are indicator variables for whether the respondent has the disease,  $X_{a,i,t}$  is a vector of the same demographic variables and a constant term that featured in equation 2 and  $\varepsilon_{a,i,t}$  is a random error term.<sup>27</sup>

$$E(\text{health utility})_{a,i,t} = \beta_{i,t,1} \text{Cancers}_{a,i,t} + \beta_{i,t,2} \text{Cardiovascular diseases}_{a,i,t} + \beta_{i,t,3} \text{Blood, endocrine and kidney diseases}_{a,i,t} + X_{a,i,t}' \beta_{i,t} + \varepsilon_{a,i,t} \quad (4)$$

With this approach,  $1 - \beta_{i,t,1}$ ,  $1 - \beta_{i,t,2}$  and  $1 - \beta_{i,t,3}$  are the estimates of YLD per case of disease for each disease, survey and age group. We smoothed these estimates using the same approach as in step 2d above.

We prefer the propensity score matching approach over this multivariate regression approach because the multivariate regression approach relies on stronger assumptions about the way in which disease impacts

<sup>26</sup> For a given covariate, standardised mean difference is the difference between the within-strata average value of that covariate among those who do and do not have the disease in question divided by the within-strata standard deviation of that covariate among those who have the disease in question.

<sup>27</sup> Here it suffices for the  $i$  subscripts to index age groups rather than disease-age groups because a single model can estimate disease-attributable expected health utility for each disease simultaneously.

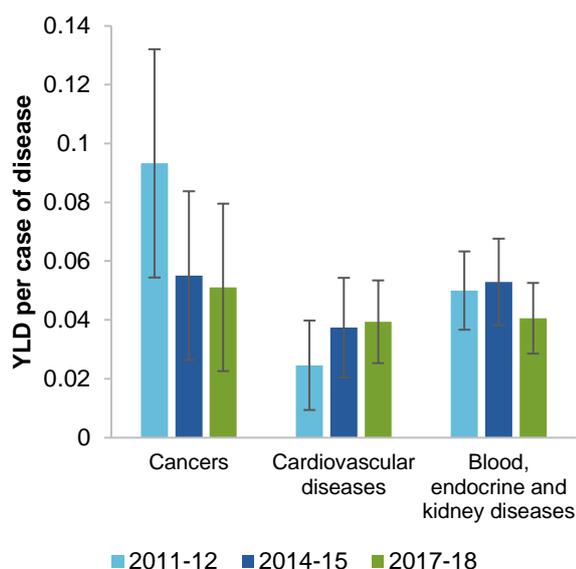
expected health utility. While these stronger assumptions allow for more precise estimation, they bias the results if they are violated.

## D.5 Results

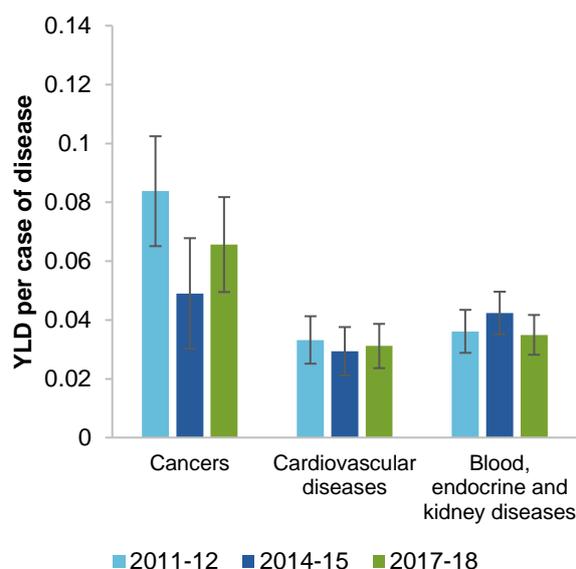
Figure D.2 shows the unsmoothed results of both sets of estimates aggregated to the disease level (weighted by 2017-18 prevalence shares). Both sets of estimates find that YLD per case of disease declined for cancers and blood, endocrine and kidney diseases. The main (propensity score matching) estimates find that YLD per case rose for cardiovascular diseases while the alternative (multivariate regression) estimates find that it slightly decreased. For all diseases, the confidence intervals are narrower for the multivariate regression estimates because this approach is premised on stronger assumptions about how having the disease in question influences YLD per case of disease.

**Figure D.2 – Unsmoothed disease-level results<sup>a</sup>**

**a. Main estimates (propensity score matching approach)**



**b. Alternative estimates (multivariate regression approach)**



a. Error bars show 95% confidence intervals. These do not account for uncertainty in the 2017-18 distribution of prevalent cases over the age groups.

Source: Productivity Commission estimates.

Figure D.3 compares the smoothed estimates with estimates for selected diseases within each disease group from the Australian Burden of Disease Study (ABDS; appendix C). This is not a like-for-like comparison as the ABDS estimates do not account for all the diseases within each group. By health expenditure share, the ABDS estimates shown here account for about 40% of all cancers, 30% of all cardiovascular diseases and 20% of all blood, endocrine and kidney diseases (appendix C).

Three key insights emerge.

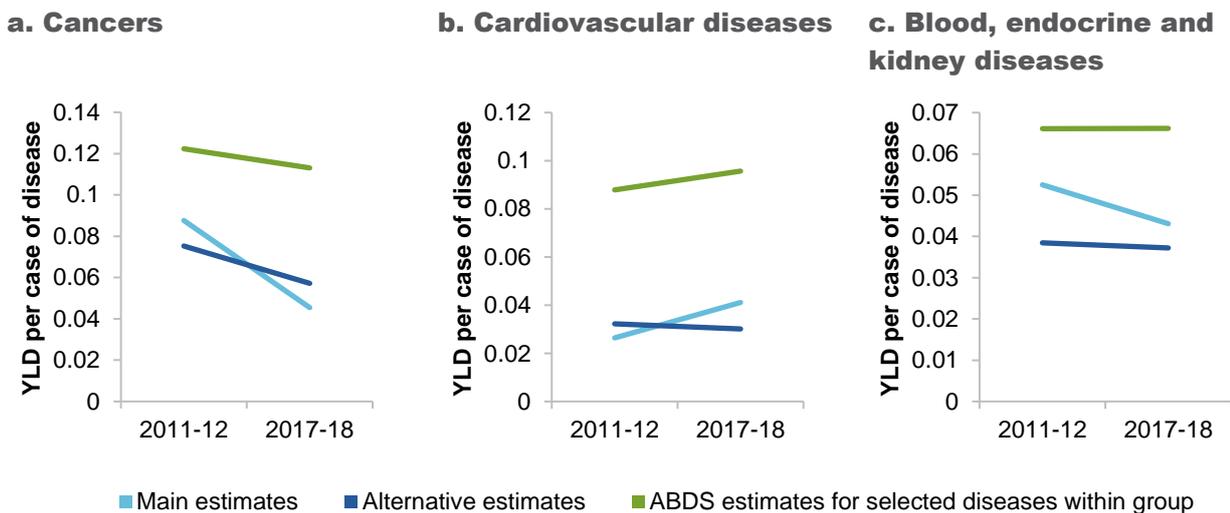
First, the ABDS estimates indicate larger YLD per case of disease. Setting aside differences in the diseases studied, this may be because the ABDS estimates exclude asymptomatic cases of diseases (which are assigned zero YLD per case of disease by the ABDS method). The main and alternative estimates are

based on self-reported prevalence, which may include some asymptomatic cases. However, because of the way that productivity growth is estimated, the *level* of YLD per case of disease only has a trivial influence on healthcare productivity growth – what matters far more is the *change* in YLD per case of disease between 2011-12 and 2017-18 (in level terms, not proportional terms) (box D.1).

Second, the direction of the change in YLD per case of disease is the same between the main estimates and ABDS estimates for all disease groups except for blood, endocrine and kidney diseases. For these diseases, the ABDS estimates show a negligible increase, and the main estimates show a decrease. The direction of the change for the alternative estimates is the same as the main estimates for cancers and blood, endocrine and kidney diseases but differs for cardiovascular diseases.

Third, the absolute magnitude (in level terms) of the change in YLD per case of disease is generally similar between the alternative estimates and the ABDS estimates, and larger for the main estimates. This may be due to the wider confidence intervals that apply to the main estimates.

**Figure D.3 – Smoothed disease-level results and ABDS estimates<sup>a</sup>**



a. ABDS estimates are for 2011 and 2018.

Source: Productivity Commission estimates.

**Box D.1 – The change in YLD per case of disease matters but the level much less so**

The level and change in YLD per case of disease affect productivity growth by affecting the change in the value of healthcare outcomes between 2011-12 and 2017-18.

This is captured by equation 9 in appendix C, which can be simplified to the following proportionality relationship, where  $v(\Delta Healthcare\ outcomes_{i,2011-12,2017-18})$  is the dollar value of the change in healthcare outcomes per case of disease,  $Survival\ rate_{i,t}$  is the share of cases who survived the disease in year  $t$  and  $\frac{YLD_{i,t}}{Prevalence_{i,t}}$  is YLD per case of disease in year  $t$ .

**Box D.1 – The change in YLD per case of disease matters but the level much less so**

$$v(\Delta \text{Healthcare outcomes}_{i,2011-12,2017-18}) \propto \left( \frac{YLD_{i,2011-12}}{Prevalence_{i,2011-12}} - \frac{YLD_{i,2017-18}}{Prevalence_{i,2017-18}} \right) - \left( \frac{Survival\ rate_{i,2017-18} - Survival\ rate_{i,2011-12}}{Survival\ rate_{i,2011-12}} \right) \frac{YLD_{i,2017-18}}{Prevalence_{i,2017-18}} + constant$$

Allowing for the constant term, this proportionality relationship shows that the level of YLD per case of disease influences the value of the change in healthcare outcomes by only a factor of the proportional change in the survival rate  $\left( \frac{Survival\ rate_{i,2017-18} - Survival\ rate_{i,2011-12}}{Survival\ rate_{i,2011-12}} \right)$  relative to the influence of the change in YLD per case of disease.

The average proportional change in the survival rate (weighted by 2017-18 prevalence shares) is only 0.04 for cancers, 0.001 for cardiovascular diseases and 0.0004 for blood, endocrine and kidney diseases. This means that the influence of the *level* of YLD per case of disease is trivial compared to the influence of the *change* in YLD per case of disease.

## E. How we benchmarked Australia's healthcare productivity

Chapter 3 benchmarks the productivity of Australia's healthcare sector against that of other countries. This appendix provides further detail to support chapter 3. It discusses:

- choices about variables and countries (section E.1)
- the modelling approach (section E.2)
- the results (section E.3).

### E.1 Variable and country selection

#### Healthcare output variables

When choosing a healthcare output variable, the primary goal is to capture the full effect of healthcare on health. For example, other things equal, a healthcare output variable that captures the effects of healthcare on both length (mortality) and quality of life (morbidity) is superior to one that captures only the effects of healthcare on mortality. A secondary goal is to choose a variable that is less influenced by factors other than healthcare, such as behavioural and environmental factors.

Variables commonly chosen in the previous literature include:

- measures of life expectancy, including health-adjusted life expectancy at birth (HALE) (Kumbhakar et al. 2020; Ogloblin 2011). Life expectancy at birth captures how long a newborn baby could expect to live if prevailing mortality rates were held constant in the future. HALE adjusts this to account for the expected share of that life that will be lived in poor health (IHME 2023c)<sup>28</sup>
- outcomes at birth, such as birth weight or other measures of infant health (Darabi et al. 2021)
- measures of preventive healthcare, such as immunisation rates (Kumbhakar et al. 2020).

Of these, HALE is our preferred output variable. It captures the effects of healthcare on both length and quality of life, whereas other measures capture only some of the effects of healthcare on health.

An alternative output variable, which has not been previously used in benchmarking studies, is the Healthcare Access and Quality Index (HAQI), developed by the Institute for Health Metrics and Evaluation (GBD 2016 Healthcare Access and Quality Collaborators 2018). The HAQI is based on mortality rates from 32 diseases that are considered amenable to healthcare (such as cancer and cardiovascular diseases), adjusted for cross-country differences in the age of the population and risk factors that affect those diseases.

<sup>28</sup> HALE uses the notion of a quality-adjusted life year to compare death with perfect health and various levels of poor health (chapter 2, appendix C).

For example, cancer mortality is adjusted for cross-country differences in smoking rates because smoking causes cancer. A strength of this measure is that it should, in principle, capture only variation that is within the control of the healthcare sector. However, as it is focused on mortality, it does not capture the effects of healthcare on quality of life.

Our approach was to use HALE as our primary healthcare output variable and HAQI as a robustness check. Both variables are sourced from the Global Burden of Disease Study (IHME 2019). For technical modelling reasons, we used an adjusted version of HALE in our modelling – this adjustment is explained in section E.2.

## Healthcare input variables

Healthcare sector benchmarking exercises commonly use total health expenditure per capita in United States dollars on a purchasing power parity (PPP) basis as the healthcare input variable (for example Ogloblin (2011)).

A problem with this approach is that this expenditure measure is adjusted for country-wide price differences, rather than price differences specific to the healthcare sector. The price of healthcare inputs relative to other goods and services will vary between countries, making a general PPP measure problematic. For example, lower income countries have been found to have lower relative wages for health workers on average than higher income countries (Serje et al. 2018).

There are two potential solutions to this issue.

1. Adopt a PPP measure based on the cost of healthcare inputs. However, no such measure currently exists. The World Bank and OECD have produced some estimates of price levels of healthcare *outputs* (OECD 2020; World Bank 2020), but these are only available for a few years (2011 and 2017).<sup>29</sup>
2. Use measures of actual capital and labour, rather than measures of expenditure. For example, the number of healthcare workers and hospital beds per capita as measures of actual labour and capital, respectively. But these measures are also imprecise. Measures such as ‘healthcare workers per capita’ do not distinguish between the mix of doctors, nurses and other health workers, and measures such as hospital beds per capita disregard capital other than hospital beds.

Because of these data limitations, our models used the same health expenditure per capita variable used in the previous literature. Data for this measure were obtained from the World Bank (2023). Were a healthcare input PPP measure to be developed, ideally it would be used in any future studies.

## Control variables

We included, as control variables, variables that are mostly outside the control of healthcare and either directly affect our measures of healthcare output or affect the relationship between our measures of healthcare inputs and outputs.

Failing to control for these factors will either:

- lead to biased estimates of healthcare sectors’ relative healthcare productivity (if they are correlated with healthcare inputs)
- lead to less precise estimates of healthcare sectors’ relative healthcare productivity (if they are not correlated with healthcare inputs).

---

<sup>29</sup> Using data from only 2011 to 2017 leads to a spurious negative relationship between expenditure and life expectancy, largely driven by a decline in HALE in the United States alongside increasing expenditure.

## Risk factors

Estimates from the Global Burden of Disease Study indicate that, for high income countries, tobacco use, high body-mass index, dietary risks (overconsumption of unhealthy food and underconsumption of healthy food) and alcohol consumption are the risk factors mostly outside the control of healthcare that have the largest effect on health (figure E.1). Based on this we include the following controls:

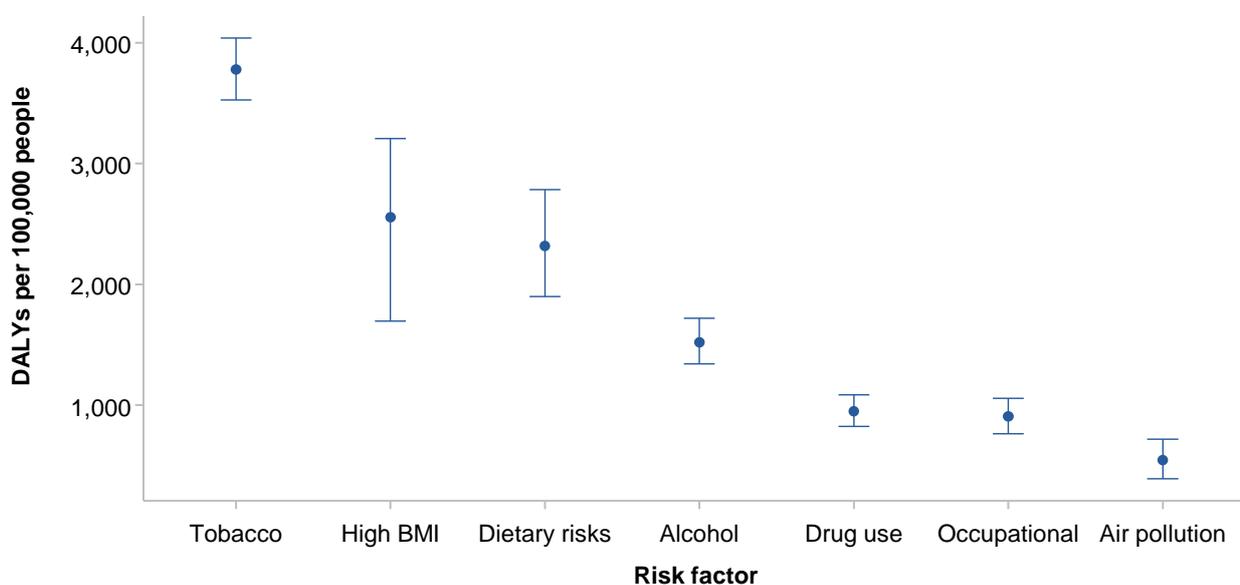
- Proportion of the population who smoke daily or occasionally.
- Litres of pure alcohol consumed per capita per year.
- Proportion of the population who are obese (which captures both high body-mass index and, indirectly, dietary risks due to overconsumption of unhealthy food).
- Grams of fruit and vegetables consumed per capita per day (which captures dietary risks due to underconsumption of healthy food).

Data on all of these variables was obtained from the Global Burden of Disease covariate dataset (IHME 2023b).

Next most important are drug use, occupational hazards and air pollution, but we were unable to locate suitably accurate indicators that capture these factors.

**Figure E.1 – Risk factors outside the control of the healthcare sector that induce the largest loss in health<sup>a,b,c</sup>**

### High income countries, 2019



**a.** DALYs are disability-adjusted life years. **b.** Error bars show 95% confidence intervals around the central estimate. **c.** Estimates are relative to a 'theoretical minimum risk exposure' level, which differs by risk factor.

Source: Institute for Health Metrics and Evaluation (2023a).

## Age of the population

The age of the population is likely to affect the relationship between healthcare spending per capita and HALE. All else equal, sustaining a given HALE requires more healthcare expenditure with an older population than with a younger population.

However, including a simple measure of the population age distribution, such as the proportion of the population aged 65 and over, is problematic as HALE affects population age distribution. For example, other

things being equal, a lower HALE will cause the population to be younger because people will not survive into older age. Including a simple measure of population age distribution as a control would violate the assumption that our controls are exogenous and bias our results.

To address this, we defined the *potential* share of the population aged 65 and over according to the following formula, where  $i$  indexes countries and  $t$  indexes time (in years):<sup>30</sup>

$$\text{Potential share of population aged 65 and over}_{i,t} = \frac{\text{Number of people aged 45 to 65}_{i,t-20}}{\text{Number of people aged under 65}_{i,t}} \quad (1)$$

As people aged 45–65 20 years prior would be aged 65–85 in the current period, this measure would indicate the demand on the healthcare sector from this group ageing into a more healthcare-intensive period of life in a way that is independent of the success of that healthcare at keeping them alive.

### Transforming variables so that they have a positive relationship with healthcare outputs

To meet the requirements of our modelling approach, each control variable needs to take on positive values and have a positive relationship with HALE (section E.3). But most of the controls described above have a negative relationship with HALE – for example, other things being equal, the larger the proportion of the population who smoke, the lower HALE will be.

For this reason, in our modelling we transformed the control variables so that they have a positive relationship with HALE. For three of the control variables there is a natural way to undertake this transformation.

- Proportion of the population who do *not* smoke daily or occasionally, denoted  $Smoking_{i,t}$  (where  $i$  indexes countries and  $t$  indexes time).
- Proportion of the population who are *not* obese, denoted  $Obesity_{i,t}$ .
- The potential share of the population *not* aged 65 and over, denoted  $Age_{i,t}$ .

Per capita fruit and vegetable consumption, denoted  $Fruit\ and\ vegetables_{i,t}$ , already has a positive relationship with HALE and so requires no transformation.

The remaining control variable is alcohol consumption per capita per year. There is no clear upper bound on this variable, although the maximum observed in the sample is 17 litres of pure alcohol per person per year. We assumed the upper bound to be 20 litres per person per year, and so defined the variable as 20 minus the number of litres of pure alcohol consumed per capita, denoted  $Alcohol_{i,t}$ .<sup>31</sup>

### Choice of countries

The choice of countries involves trade-offs. The more different a country is from Australia, the more different the relationship between its healthcare inputs, outputs and control variables is likely to be. As our approach is contingent on this relationship being the same for all countries (except for random variation and cross-country differences in healthcare productivity), including countries that are very different from Australia is likely to lead to biased results. However, the smaller the number of included countries, the fewer the number of observations and the poorer the statistical precision of our estimates.

<sup>30</sup> A shortcoming of this indicator is that it does not account for net migration flows among the population aged 45 to 65 20 years prior. However, migrants tend to be aged under 45. For example, in Australia in 2015, 97% of the net migration flow was of people aged under 45 (Productivity Commission estimate based on ABS 2024d).

<sup>31</sup> As a robustness check, we also tried an upper bound of 30 litres per person per year and found it did not substantially change the modelling results.

Our chosen group of countries included 28 countries.<sup>32</sup> Countries were selected where GDP per capita was above 30,000 USD in 2023 (on a PPP basis) and whose population was over 100,000 people in 2019 (because very small countries, such as Monaco, may have anomalous populations and healthcare sectors). Countries were further excluded based on partial residual plots from a linear regression of HALE against each covariate (this led to the exclusion of several Middle Eastern countries, for example, based on their low reported smoking rates).

To assess sensitivity to choice of country, we also estimated a model with a larger set of 52 countries.<sup>33</sup> This includes all countries whose GDP per capita was above gross world product per capita in 2023 (purchasing power parity) and whose population was over 100,000 people in 2019, with the same low-smoking countries excluded as the 28-country model.

Table E.1 summarises our choice of variables to use as outputs, inputs and covariates.

**Table E.1 – Cross-country benchmarking – variables**

Variable type	Variable description	Comments
Healthcare output (main model)	Health-adjusted life expectancy at birth, adjusted as described in section E.2: <i>Adjusted HALE</i> <sub><i>i,t</i></sub>	Covers both quality and quantity of life, but is affected by non-healthcare factors (e.g. smoking).
Healthcare output (robustness check)	Healthcare access and quality index: <i>HAQI</i> <sub><i>i,t</i></sub>	Index developed by the Institute for Health Metrics and Evaluation. Covers only quantity of life (mortality), but less affected by non-healthcare factors.
Healthcare input	Healthcare spending per capita (USD, purchasing power parity): <i>Healthcare spending</i> <sub><i>i,t</i></sub>	Standard measure, used in all studies.
Control	Proportion of population who <i>do not</i> smoke daily or occasionally: <i>Smoking</i> <sub><i>i,t</i></sub>	Reflects healthcare-invariant risk factor that induces the most DALYs in high income countries.
Control	Proportion of population who <i>are not</i> obese: <i>Obesity</i> <sub><i>i,t</i></sub>	Reflects 'high body-mass index' and 'dietary risks', the healthcare-invariant risk factors that induce the second and third most DALYs in high income countries.
Control	Fruit and vegetable consumption (grams per capita per day): <i>Fruit and vegetables</i> <sub><i>i,t</i></sub>	Reflects 'dietary risks', the healthcare-invariant risk factor that induces the third most DALYs in high income countries.
Control	<i>Lack of</i> alcohol consumption (20 minus the number of litres of pure alcohol consumed per capita per year): <i>Alcohol</i> <sub><i>i,t</i></sub>	Reflects healthcare-invariant risk factor that induces the fourth most DALYs in high income countries.

<sup>32</sup> Countries in the main model are Luxembourg, Ireland, Norway, Switzerland, Singapore, United States, Iceland, Denmark, Australia, Netherlands, Austria, Israel, Sweden, Finland, Belgium, Canada, New Zealand, United Kingdom, France, Malta, Italy, Japan, Cyprus, South Korea, Slovenia, Czechia, Spain, and Estonia.

<sup>33</sup> Additional countries in this model are the countries in the main model plus Saudi Arabia, Bahrain, Lithuania, Portugal, Latvia, Slovakia, Greece, Oman, Uruguay, Barbados, Guyana, Croatia, Poland, Trinidad and Tobago, Seychelles, Hungary, Romania, Antigua and Barbuda, Chile, Maldives, Panama, Bulgaria, Palau, Costa Rica, Russia, China, and Argentina.

Variable type	Variable description	Comments
Control	Potential share of population <i>not aged 65 and over</i> : $Age_{i,t}$	An indicator of the age of the population, constructed so as to be independent of health-adjusted life expectancy.

## E.2 Modelling approach

Our goal is to rank countries' healthcare sectors by the output they produced relative to the maximum amount feasible given their inputs and the control variables.

The two methods commonly used to perform tasks of this nature are stochastic frontier analysis (SFA) and data envelopment analysis (DEA). Of these, SFA is our preferred approach because it allows for the fact that healthcare outputs are measured with error and that factors not included in the model also affect healthcare outputs (provided these are not correlated with healthcare inputs and controls) (box E.1).

### Box E.1 – Stochastic frontier analysis or data envelopment analysis?

There are two approaches that are commonly used to measure the relative productivity of different units of production, such as firms or, in this case, countries' healthcare sectors – stochastic frontier analysis (SFA) and data envelopment analysis (DEA). Each has its advantages and disadvantages.

The primary advantage of SFA is that it allows for the fact that HALE is measured with error and that factors not included in the model also affect HALE (provided these are not correlated with healthcare spending, population ageing and the risk factors included in the model). Several sources advise against using DEA, which cannot account for these sources of uncertainty, for this reason (Cornwell and Schmidt 2008; Spinks and Hollingsworth 2009, p. 701).

The primary advantages of DEA are that it:

- can incorporate multiple output variables that capture different aspects of population health simultaneously instead of a single output variable, such as HALE. This can be useful when measuring, for example, the productivity of a factory that produces multiple different products. But it is less useful in this context. If this advantage is not being exploited, the case for using DEA is considerably reduced – Spinks and Hollingsworth (2009, p. 425) note that 'if there is only one output used in the model, it is difficult to argue beyond using econometric techniques [such as SFA]'
- requires fewer assumptions about the underlying 'true' statistical relationship between HALE, healthcare inputs and other factors that influence HALE.

This study uses SFA because its advantages are more suited to our needs than those of DEA.

That said, this is not a straightforward application of SFA. SFA requires that we specify the 'true' underlying functional form of the relationship between inputs, outputs and controls. The timing of this relationship poses difficulties. For many applications it is relatively safe to assume the timing of the relationship. For example, when modelling the productivity of factories one can assume that labour used during 2011 would influence output in 2011 but not output in 2012. In our context, however, the timing of the effects of inputs and controls on outputs is quite variable. For example, one might expect shocks to obesity or to healthcare spending due to the introduction of a new treatment to have immediate effects on HALE that then persist in future years (for an uncertain period). And one might expect shocks to smoking rates or to other types of healthcare

spending, such as due to a limiting of access to primary care, to affect HALE only in the future (but when in the future and for how long is unclear). It is plausible that the 'true' model would contain lags of varying length for each input and control variable, but these lags are highly uncertain.

Given this uncertainty, we appealed to the principle of parsimony and chose a simple model that assumes away all lags. One way to do this would be to collapse the time dimension entirely by taking averages for all variables over time for each country, but this would substantially reduce the sample size and would disregard that shocks to healthcare inputs likely do have immediate effects on HALE. Instead, we included annual observations for each country (a panel dataset) but avoided using panel-data-specific SFA approaches as these are sensitive to the timing of the underlying relationship.<sup>34</sup> We used a simpler pooled estimator with time fixed effects to allow for common global increases in the production frontier due to improvements in healthcare technology.

## Model set-up

The stochastic frontier approach is to model healthcare outputs as a function of healthcare inputs and controls (including time fixed effects and a random component) and a multiplicative relative healthcare productivity term:

$$\text{Healthcare outputs}_{i,t} = f(\text{Healthcare inputs and controls}_{i,t}) \times \text{Relative healthcare productivity}_{i,t} \quad (2)$$

Relative healthcare productivity, the variable of interest, ranges between zero and one with higher values indicating a more productive healthcare sector. As is standard in stochastic frontier analysis, it is modelled as a function of another term  $u_{i,t}$ , which is assumed to follow a positive half-normal distribution (i.e. a normal distribution centred on zero with the negative values removed) that is independent of healthcare inputs and controls:

$$\text{Relative healthcare productivity}_{i,t} = e^{-u_{i,t}} \quad (3)$$

$f(\text{Healthcare inputs and controls}_{i,t})$  is assumed to take on Cobb-Douglas functional form. This is a common choice in stochastic frontier analysis, although more complex functional forms such as translog and constant elasticity of substitution functions are sometimes used. We used a Cobb-Douglas functional form as it requires fewer degrees of freedom than these more complex forms (as there are fewer parameters to estimate) while still allowing for diminishing returns to healthcare spending. It is defined as follows, where  $\mathbf{x}'_t$  is a vector of time dummy variables and  $v_{i,t}$  is an independent and normally distributed error term:

$$f(\text{Healthcare inputs and controls}_{i,t}) = e^{\beta_0} \times \text{Healthcare spending}_{i,t}^{\beta_1} \times \text{Smoking}_{i,t}^{\beta_2} \times \text{Obesity}_{i,t}^{\beta_3} \times \text{Fruit and vegetables}_{i,t}^{\beta_4} \times \text{Alcohol}_{i,t}^{\beta_5} \times \text{Age}_{i,t}^{\beta_6} \times e^{\mathbf{x}'_t \boldsymbol{\beta}} \times e^{v_{i,t}} \quad (4)$$

While not imposed as a restriction in the model, the use of a Cobb-Douglas functional form and the logic underpinning our choice of inputs and controls requires that  $\beta_1, \dots, \beta_6$  are all positive (this is satisfied for all models; section E.3).

Combining these components and taking the log of both sides yields the following model, which is estimated using maximum likelihood on data for  $t = 2000, \dots, 2019$ :

$$\ln(\text{Healthcare outputs}_{i,t}) = \beta_0 + \beta_1 \ln \text{Healthcare spending}_{i,t} + \beta_2 \ln \text{Smoking}_{i,t} + \beta_3 \ln \text{Obesity}_{i,t} + \beta_4 \ln \text{Fruit and vegetables}_{i,t} + \beta_5 \ln \text{Alcohol}_{i,t} + \beta_6 \ln \text{Age}_{i,t} + \mathbf{x}'_t \boldsymbol{\beta} + v_{i,t} - u_{i,t} \quad (5)$$

<sup>34</sup> For example, a panel fixed effects SFA estimator exploits the timing of the underlying relationship between inputs, outputs and controls to reduce the effects of bias due to omitted controls that are correlated with healthcare inputs and included controls. If the timing is not correctly specified, this approach is invalid.

To estimate country-level relative healthcare productivity scores for the period 2010 to 2019, we took the within-country average of *Relative healthcare productivity* $_{i,t}$  over  $t = 2010, \dots, 2019$ .

## Adjustments to health-adjusted life expectancy

As stated earlier, our preferred measure of healthcare outputs is HALE. A complexity is that the output variable should take on an expected value of zero when all inputs and controls are set to zero,  $t = 2000$  and when relative healthcare productivity is equal to one. This is the case regardless of the functional form of  $f(\text{Healthcare inputs and controls}_{i,t})$ ; it is a consequence of multiplicative nature of the relative healthcare productivity term in equation 2. But HALE would still be positive if there were no healthcare spending, no obesity, no smoking and so on.

To resolve this issue, we estimated and use the following alternative measure of HALE, called ‘adjusted HALE’, that is designed to meet this criterion:

$$\text{Adjusted HALE}_{i,t} = \text{HALE}_{i,t} - \text{HALE}_0 \quad (6)$$

$\text{HALE}_0$ , which equals 44.2, is an estimate of the expected value of HALE when all inputs and controls are set to zero. We estimated  $\text{HALE}_0$  by estimating the following model for HALE using ordinary least squares and taking the predicted value of HALE when all variables were set to zero and  $t = 2000$ :

$$\text{HALE}_{i,t} = \beta_0 + \beta_1 \text{Healthcare spending}_{i,t} + \beta_2 \text{Smoking}_{i,t} + \beta_3 \text{Obesity}_{i,t} + \beta_4 \text{Fruit and vegetables}_{i,t} + \beta_5 \text{Alcohol}_{i,t} + \beta_6 \text{Age}_{i,t} + \mathbf{x}'_t \boldsymbol{\beta} + \varepsilon_{i,t} \quad (7)$$

We then estimated equation 5 with adjusted HALE as our measure of healthcare outputs. We also estimated models with HAQI in place of adjusted HALE and with a larger country set.

## E.3 Results

Our results suggest that Australia’s healthcare sector is relatively productive (table E.2). Using our primary model, Australia’s healthcare sector ranks as the third most productive between 2010 and 2019 and the fifth most productive between 2000 to 2019 (about 90<sup>th</sup> percentile). The other models suggest similar results – when HAQI is used as the output variable, Australia’s healthcare sector ranks as the fifth most productive between 2010 and 2019 (about 80<sup>th</sup> percentile), and when larger country set of countries is considered Australia’s healthcare sector ranks as the tenth most productive (about 80<sup>th</sup> percentile). This is consistent with the previous literature, which found that Australia’s healthcare sector ranked at about the 80<sup>th</sup> to 90<sup>th</sup> percentile for productivity, using data from prior to 2007.

In each model, all coefficients have a positive sign, as expected (noting that all control variables are transformed such that they have a positive relationship with healthcare outputs). All coefficients are statistically significant in the main model. The coefficients on the controls are smaller and less statistically significant in the HAQI model, which is to be expected as HAQI already controls for many of these factors.

**Table E.2 – Australia's healthcare sector is relatively productive<sup>a</sup>****Results for main model and alternative models**

	Output is ln Adjusted HALE <sub><i>i,t</i></sub> , 28 countries (primary model)	Output is ln HAQI <sub><i>i,t</i></sub> , 28 countries	Output is ln Adjusted HALE <sub><i>i,t</i></sub> , 52 countries
<b>Inputs</b>			
ln Healthcare spending <sub><i>i,t</i></sub>	.025*** (.005)	.075*** (.005)	.100*** (0.004)
<b>Controls</b>			
ln Smoking <sub><i>i,t</i></sub>	.132*** (.038)	.058* (.027)	.009 (.028)
ln Obesity <sub><i>i,t</i></sub>	.629*** (.040)	.185*** (.032)	.517*** (.035)
ln Fruit and vegetables <sub><i>i,t</i></sub>	.018* (.629)	.010 (.007)	.038*** (.008)
ln Alcohol <sub><i>i,t</i></sub>	.046*** (.008)	.011*** (.007)	.033*** (.008)
ln Age <sub><i>i,t</i></sub>	.136*** (.031)	.076** (.027)	.130*** (0.035)
<b>Error variance</b>	.0006	.0003	.0008
<b>Inefficiency variance</b>	.0027	.0022	.0153
<b>Australia rank</b>			
2010-2019	3	5	10
2000-2019	3	4	8

a. Standard errors presented in parentheses. \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$

Source: Productivity Commission estimates based on World Bank and Global Burden of Disease data.



# References

- ABS (Australian Bureau of Statistics) 2018, *Changing Patterns of Mortality in Australia, 1968 - 2017*, <https://www.abs.gov.au/statistics/health/causes-death/changing-patterns-mortality-australia/latest-release> (accessed 3 April 2024).
- 2020, *Non-market output measures in the Australian National Accounts: A conceptual framework for enhancements, 2020*, <https://www.abs.gov.au/statistics/research/non-market-output-measures-australian-national-accounts-conceptual-framework-enhancements-2020> (accessed 8 April 2024).
- 2021, *Experimental hospital multifactor productivity estimates*, <https://www.abs.gov.au/statistics/research/experimental-hospital-multifactor-productivity-estimates> (accessed 30 October 2023).
- 2024a, *Detailed microdata: National Health Survey 2011-12*, Accessed via ABS DataLab.
- 2024b, *Detailed microdata: National Health Survey 2014-15*, Accessed via ABS DataLab.
- 2024c, *Detailed microdata: National Health Survey 2017-18*, Accessed via ABS DataLab.
- 2024d, *Net overseas migration, Arrivals, departures and net, State/territory, Age and sex - Calendar years, 2004 onwards*, <https://explore.data.abs.gov.au/> (accessed 22 February 2024).
- 2024e, *Quarterly population estimates (ERP), by state/territory, sex and age*, <https://explore.data.abs.gov.au/> (accessed 4 January 2024).
- Ahmed, S, Hasan, MZ, MacLennan, M, Dorin, F, Ahmed, MW, Hasan, MM, Hasan, SM, Islam, MT and Khan, JAM 2019, 'Measuring the efficiency of health systems in Asia: a data envelopment analysis', *BMJ Open*, vol. 9, no. 3, pp. 1–12.
- AIHW (Australian Institute of Health and Welfare) 2012, *Health expenditure Australia 2010-11*.
- 2021a, *Australian Burden of Disease Study: Methods and supplementary material 2018, ABDS 2018 quality index*, <https://www.aihw.gov.au/reports/burden-of-disease/abds-methods-supplementary-material-2018/contents/abds-quality-framework/abds-2018-quality-index> (accessed 25 September 2023).
- 2021b, *Australian Burden of Disease Study: Methods and supplementary material 2018, Years lived with disability (YLD)*, <https://www.aihw.gov.au/reports/burden-of-disease/abds-methods-supplementary-material-2018/contents/estimating-burden-of-disease-measures/years-lived-with-disability-yld> (accessed 23 November 2023).
- 2021c, *Australian Burden of Disease Study: Methods and supplementary material 2018, Years of life lost (YLL)*, <https://www.aihw.gov.au/reports/burden-of-disease/abds-methods-supplementary-material-2018/contents/estimating-burden-of-disease-measures/years-of-life-lost-yll> (accessed 14 February 2024).
- 2021d, *Data tables: ABDS 2018 National estimates for Australia*.
- 2022a, *Australian Burden of Disease Study 2022, Health-adjusted life expectancy*, <https://www.aihw.gov.au/reports/burden-of-disease/australian-burden-of-disease-study-2022/contents/interactive-data-on-disease-burden/health-adjusted-life-expectancy> (accessed 9 February 2024).
- 2022b, *Disease expenditure in Australia 2019–20*, <https://www.aihw.gov.au/reports/health-welfare-expenditure/disease-expenditure-in-australia-2019-20/contents/summary> (accessed 18 October 2023).
- 2022c, *Disease Expenditure Study: Overview of Analysis and Methodology 2019-20*.
- 2023a, *Alcohol, tobacco & other drugs in Australia, Tobacco and e-cigarettes*, <https://www.aihw.gov.au/reports/alcohol/alcohol-tobacco-other-drugs-australia/contents/drug-types/tobacco> (accessed 10 April 2024).
- 2023b, *Cancer data in Australia*, <https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/overview-of-cancer-in-australia-2023> (accessed 8 April 2024).
- 2023c, *Health expenditure Australia 2021-22*, 25 October.
- 2023d, *Health system spending on disease and injury in Australia, 2020-21*, <https://www.aihw.gov.au/reports/health-welfare-expenditure/health-system-spending-on-disease-and-injury-in-au/contents/about> (accessed 10 April 2024).
- 2023e, *MyHospitals*, <https://www.aihw.gov.au/reports-data/myhospitals> (accessed 6 September 2023).
- Ananthapavan, J, Moodie, M, Milat, AJ and Carter, R 2021, 'Systematic Review to Update "Value of a Statistical Life" Estimates for Australia', *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, p. 6168.
- Asandalui, L, Roman, M and Fatulescu, P 2014, 'The efficiency of healthcare systems in Europe: A data envelopment approach', *Procedia Economics and Finance*, vol. 10, pp. 261–268.
- Australian Government Department of Health 2021, *National Preventive Health Strategy 2021-2023*.
- Breadon, P, Fox, L and Emslie, O 2023, *The Australian Centre for Disease Control (ACDC): Highway to health*, Grattan Institute.
- Carter, D, Vogan, A and Haji Ali Afzali, H 2016, 'Governments need better guidance to maximise value for money: The case of Australia's Pharmaceutical Benefits Advisory committee', *Applied Health Economics and Health Policy*, vol. 14, no. 4, pp. 401–407.
- Centre for Health Economics Research and Evaluation 2022, *Review of the Discount Rate in the PBAC Guidelines*, May, University of Technology Sydney.
- Commonwealth of Australia 2023, *Intergenerational Report 2023: Australia's future to 2063*, Canberra.

- Cornwell, C and Schmidt, P 2008, 'Stochastic Frontier Analysis and Efficiency Estimation', *The Econometrics of Panel Data*, Springer, pp. 697–726.
- Couzin-Frankel, J 2013, 'Cancer Immunotherapy', *Science*, vol. 342, no. 6165, pp. 1432–1433.
- Cutler, DM, Ghosh, K, Messer, KL, Raghunathan, T, Rosen, AB and Stewart, ST 2022, 'A satellite account for health in the United States', *The American Economic Review*, vol. 112, no. 2, pp. 494–533.
- , McClellan, M, Newhouse, JP and Remler, D 1998, 'Are medical prices declining? Evidence from heart attack treatments', *The Quarterly Journal of Economics*, vol. 113, no. 4, pp. 991–1024.
- Darabi, N, Ebrahimvandi, A, Hosseinichimeh, N and Triantis, K 2021, 'A DEA evaluation of U.S. States' healthcare systems in terms of their birth outcomes', *Expert Systems with Applications*, vol. 182, p. 115278.
- Dauda, S, Dunn, A and Hall, A 2022, 'A systematic examination of quality-adjusted price index alternatives for medical care using claims data', *Journal of Health Economics*, vol. 85, p. 102662.
- Dunn, A, Hall, A and Dauda, S 2022, 'Are medical care prices still declining? A re-examination based on cost-effectiveness studies', *Econometrica*, vol. 90, no. 2, pp. 859–886.
- , Rittmueller, L and Whitmire, B 2015, *Introducing the New BEA Health Care Satellite Account*.
- Galenkamp, H, Huisman, M, Braam, AW, Schellevis, FG and Deeg, DJ 2014, 'Disease prevalence based on older people's self-reports increased, but patient–general practitioner agreement remained stable, 1992–2009', *Journal of Clinical Epidemiology*, vol. 67, no. 7, pp. 773–780.
- GBD 2016 Healthcare Access and Quality Collaborators 2018, 'Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016', *The Lancet*, vol. 391, no. 10136, pp. 2236–2271.
- Goss, JR 2022, 'Health expenditure data, analysis and policy relevance in Australia, 1967 to 2020', *International Journal of Environmental Research and Public Health*, vol. 19, no. 4, p. 2143.
- Greifer, N 2023, *Assessing Balance*, <https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html#ref-ali2014> (accessed 21 December 2023).
- Gu, W and Morin, S 2014, 'Experimental Measures of Output and Productivity in the Canadian Hospital Sector, 2002 to 2010', *Measuring Economic Sustainability and Progress*, University of Chicago Press, pp. 575–594.
- Hall, A 2015, *Adjusting the Measurement of the Output of the Medical Sector for Quality: A Review of the Literature*, October, Bureau of Economic Analysis.
- Highfill, T and Bernstein, E 2019, 'Using disability adjusted life years to value the treatment of thirty chronic conditions in the U.S. from 1987 to 2010: a proof of concept', *International Journal of Health Economics and Management*, vol. 19, no. 3, pp. 449–466.
- IHME (Institute for Health Metrics and Evaluation) 2019, *Global Burden of Disease Study 2019 Data Resources*, <https://ghdx.healthdata.org/gbd-2019> (accessed 2 November 2023).
- 2023a, *Global Burden of Disease Results*, <https://vizhub.healthdata.org/gbd-results/> (accessed 4 October 2023).
- 2023b, *Global Burden of Disease Study 2019 Covariates 1980-2019*, 3 March.
- 2023c, *Global Burden of Disease Study 2019 Life Expectancy and Healthy Life Expectancy 1990-2019*, <https://ghdx.healthdata.org/record/ihmde-data/gbd-2019-life-expectancy-HALE-1990-2019> (accessed 21 February 2024).
- Kouakou, CRC and Poder, TG 2022, 'Willingness to pay for a quality-adjusted life year: a systematic review with meta-regression', *The European Journal of Health Economics*, vol. 23, no. 2, pp. 277–299.
- Kumbhakar, SC, Loayza, NV and Narambuena, V 2020, *International Benchmarking for Country Economic Diagnostics: A Stochastic Frontier Approach*, June, 9304, Policy Research Working Paper, World Bank.
- Lybrand, S and Wonder, M 2020, 'Analysis of PBAC submissions and outcomes for medicines (2010–2018)', *International Journal of Technology Assessment in Health Care*, vol. 36, no. 3, pp. 224–231.
- Norman, R, Viney, R, Brazier, J, Burgess, L, Cronin, P, King, M, Ratcliffe, J and Street, D 2014, 'Valuing SF-6D Health States Using a Discrete Choice Experiment', *Medical Decision Making*, vol. 34, no. 6, pp. 773–786.
- OECD (Organisation for Economic Cooperation and Development) 2020, *Healthcare prices and Purchasing Power Parities*, <https://www.oecd.org/health/health-purchasing-power-parities.htm> (accessed 2 November 2023).
- (Organisation for Economic Cooperation and Development) 2024, *National income - Gross national income*, <https://data.oecd.org/natincome/gross-national-income.htm> (accessed 4 January 2024).
- Ogloblin, C 2011, 'Health care efficiency across countries: A stochastic frontier analysis', *Applied Econometrics and International Development*, vol. 11, no. 1, pp. 5–14.
- OIA (Office of Impact Analysis) 2023a, *Cost Benefit analysis*, Guidance Note, July.
- 2023b, *Value of statistical life*, October.
- Oksanen, T, Kivimäki, M, Pentti, J, Virtanen, M, Klaukka, T and Vahtera, J 2010, 'Self-report as an indicator of incident disease', *Annals of Epidemiology*, vol. 20, no. 7, pp. 547–554.
- ONS (Office for National Statistics) 2021, *Public service productivity, healthcare, England*, <https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimateshealthcare/finalyearending2019> (accessed 13 September 2023).
- Paraskeva, MA, Levin, KC, Westall, GP and Snell, GI 2018, 'Lung transplantation in Australia, 1986–2018: more than 30 years in the making', *Medical Journal of Australia*, vol. 208, no. 10, pp. 445–450.
- Pirkis, J, Currier, D, Harris, M and Mihalopoulos, C 2022, *Evaluation of Better Access*, Main report, December, The University of Melbourne.
- Rodriguez, L and Sobrino, CR 2015, 'Public HEs, diminishing returns, and life expectancy in Latin American and Caribbean countries: a panel data analysis', *Brazilian Journal of Health Economics*, vol. Forthcoming.
- SCRGSP (Steering Committee for the Review of Government Service Provision) 2024, *Report on Government Services 2024 - Health (part E)*, Productivity Commission, Canberra.
- Serje, J, Bertram, MY, Brindley, C and Lauer, JA 2018, 'Global health worker salary estimates: an econometric analysis of

- global earnings data', *Cost Effectiveness and Resource Allocation*, vol. 16, art. 10, pp. 1–9.
- Shahid, K, Khalife, M, Dabney, R and Phan, AT 2019, 'Immunotherapy and targeted therapy - the new roadmap in cancer treatment', *Annals of Translational Medicine*, vol. 7, no. 20, p. 595.
- Sheiner, L and Malinovskaya, A 2016, *Measuring Productivity in Healthcare: An Analysis of the Literature*, The Brookings Institution.
- Simpson, CF, Boyd, CM, Carlson, MC, Griswold, ME, Guralnik, JM and Fried, LP 2004, 'Agreement Between Self-Report of Disease Diagnoses and Medical Record Validation in Disabled Older Women: Factors That Modify Agreement', *Journal of the American Geriatrics Society*, vol. 52, no. 1, pp. 123–127.
- Spinks, J and Hollingsworth, B 2009, 'Cross-country comparisons of technical efficiency of health production: a demonstration of pitfalls', *Applied Economics*, vol. 41, no. 4, pp. 417–427.
- Walker, RC, Howard, K and Morton, RL 2017, 'Home hemodialysis: a comprehensive review of patient-centered and economic considerations', *ClinicoEconomics and Outcomes Research*, vol. 9, pp. 149–161.
- Wamble, D, Ciarametaro, M, Houghton, K, Ajmera, M and Dubois, RW 2019, 'What's Been The Bang For The Buck? Cost-Effectiveness Of Health Care Spending Across Selected Conditions In The US', *Health Affairs*, vol. 38, no. 1, pp. 68–75.
- Weaver, MR, Joffe, J, Ciarametaro, M, Dubois, RW, Dunn, A, Singh, A, Sparks, GW, Stafford, L, Murray, CJL and Dieleman, JL 2022, 'Health care spending effectiveness: estimates suggest that spending improved US health from 1996 to 2016', *Health Affairs*, vol. 41, no. 7, pp. 994–1004.
- WHO (World Health Organisation) 2000, *The World Health Report*.
- World Bank 2020, *International Comparison Program 2017 Data*, <https://www.worldbank.org/en/programs/icp/data> (accessed 2 November 2023).
- 2023a, *Current health expenditure per capita, PPP (current international \$)*, <https://data.worldbank.org/indicator/SH.XPD.CHEX.PP.CD> (accessed 28 November 2023).
- 2023b, *Population estimates and projections*, <https://databank.worldbank.org/source/population-estimates-and-projections> (accessed 22 February 2024).
- 2024, *Education Statistics: Education Attainment*, [https://databank.worldbank.org/Average-years-of-schooling-of-adults-\(male-and-female\)/id/12d63977](https://databank.worldbank.org/Average-years-of-schooling-of-adults-(male-and-female)/id/12d63977) (accessed 4 January 2024).