MONASH University

Provider behaviour and quality of care in prostate cancer

Nicola Huxley

(MSci Mathematics with Statistics)

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in 2020
Centre for Health Economics, Monash Business School

Copyright notice

Abstract

Background

Prostate cancer is the second most commonly diagnosed cancer for men in Australia, with the cost of prostate cancer care expected to reach over $0.5billion by 2025. Concerns have been raised about the variability in prostate cancer care and its cost-effectiveness. Supported by models of quality measures, efforts to improve prostate cancer care include diagnosis and management guidelines, registries of clinician performance measures and patient outcomes, and peer-comparative feedback to clinicians. These efforts aim to provide clinical practice targets and identify and reduce unwarranted variation from the recommended best practice. Previous studies have found varying effectiveness for these interventions.

This thesis examines mechanisms to improve the quality of care through changes in provider behaviour. Specifically explored are whether GP guidelines improve the efficiency of prostate cancer detection; whether individualised clinician feedback improves treatment decisions and care outcomes (including when outcomes may be competing); and whether surgeons with increased experience of robotic surgery improve patient and surgical outcomes.

Methods

Patient data used in the analyses come from Australian cancer registries. Chapter 2 uses state-level prostate cancer case data from state cancer registries between 2000 and 2016. Chapters 3-5 use de-identified individual patient data from the Prostate Cancer Outcomes Registry-Victoria (PCOR-Vic), with the latest datacut from February 2019. In particular, the focus was on patient and hospital characteristics, treatment information and patient outcomes. The PCOR-Vic also provided the dates that comparative performance feedback was provided to clinicians. Population data was taken from the Australian Bureau of Statistics and the number of diagnostic tests from Medicare group reports.

Appropriate and robust statistical models are used to explore the specific questions of the thesis, including interrupted time series analysis, survival analysis, linear probability models, and inverse probability weighted regression analyses.

Results

In response to guidelines to reduce unnecessary testing in asymptomatic men, PSA testing rates and prostate cancer incidence reduced. There was limited evidence of improvements in the efficiency of testing (cases per test).

Feedback to urologists to encourage active monitoring in low-risk men and quicker active treatment in high-risk men was associated with longer time to treatment (reduction in overall treatment rates during the year following diagnosis) for men of low- and intermediate-risk but had no significant impact on high-risk men.

When feedback was on multiple competing behaviours, no change was found in outcomes at the aggregate level. However, there was evidence that individual clinicians appeared to prioritise their worse outcome for improvement.

Experience mattered for surgery performance; surgeons with higher levels of experience with robotic surgery had the most significant gains in patient and clinical outcomes from using the robotic surgery

Discussion

This study provides evidence that interventions designed to change provider behaviour in prostate cancer can improve adherence to best practice recommendations, but finds less evidence that this results in improved clinical or patient-related outcomes. Higher levels of surgery experience appear to be most beneficial to surgical outcomes.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: Nicola Huxley

Date: 30/10/2020

Acknowledgements

This thesis would not have been possible without the input and support of a large number of people.

I would first like to acknowledge the Prostate Cancer Outcomes Registry-Victoria who provided the thesis' data. I particularly wish to thank Prof. Sue Evans, Ms Melanie Evans and A/Prof Jeremy Millar, who helped me understand the data and the disease area.

I would also like to thank John Dowling (Western Australia Cancer Registry), Peta Gordon (Queensland Cancer Registry), Vicky Thursfield (Victorian Cancer Registry) and Niloofar Memari (New South Wales Cancer Registry) for providing state-level summary data from their respective cancer registries.

I also acknowledge the financial support of the 2016 Monash Business School Advanced PhD and Research Fellowship and the Monash University Graduate Research Completion Award that has supported me through the duration of this thesis.

I would also like to thank the Centre for Health Economics, particularly my fellow PhD students and Clare and Juliet, for whose pastoral and administrative support I have been most grateful.

I also thank my friends and family, who have been incredibly supportive throughout my thesis. Those of you who fed me, entertained me, and provided me with a space to vent, it has been much appreciated. Thanks to my housemates of the last four years, who helped me settle into Australia, and have this year tolerated me commandeering the dining room as my home office.

Most of all, thanks to my supervisors Prof Anthony Harris, A/Prof Dennis Petrie and A/Prof Arul Earnest who helped with all aspects of this thesis. This thesis would not be complete without their support in its design, development and finalisation.

Table of contents

Contents

List of figures

List of tables

Abbreviations

ACO	Accountable care organizations
AIHW	Australian Institute of Health and Welfare
CI	Confidence interval
DRE	Digital rectal examination
GP	General practitioner
IPW	Inverse-probability weighting
ITSA	Interrupted time series analysis
LRP	Laparoscopic radical prostatectomy
PCa	Prostate cancer
PSA	Prostate specific antigen
PSM	Positive surgical margin
ORP	Open radical prostatectomy
NCCN	National Comprehensive Cancer Network
RACGP	Royal Australian College of General Practitioners
RARP	Robot-assisted radical prostatectomy
RCT	Randomised control trial
RP	Radical prostatectomy

SE	Standard error
SEER	Surveillance, epidemiology and end results program
SF-12	Short-form 12

Glossary

PSA test	Prostate-specific antigen blood test, assessing the rate of the antigen in ng/ml
PSM	Positive surgical margin: cancerous tissue left behind after prostatectomy
NCCN risk	Cancer risk status established from staging and grading data: PSA, Gleason score, clinical T staging

1. Introduction

Prostate cancer is the second most common cancer diagnosis (after lung cancer), but only the sixth leading cause of cancer death for men (International Agency for Research on Cancer, 2018). Australia is the 15th highest-ranking country for prostate cancer rates, with an age-standardised rate of 85.6 per 100,000 men (International Agency for Research on Cancer, 2018) and prostate cancer care was estimated to cost $0.5billion in 2013 (Goldsbury et al., 2018). The total estimated cost of prostate cancer treatment to the Australian health system in 2016 was $383.6 million (0.21% of total health expenditure in 2016, https://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2016-17) and estimated to rise to $543.9 million in 2025, an increase of 42% (Gordon et al., 2018; Prostate Cancer Foundation of Australia & Griffith University, 2016).

With the risk of prostate cancer affecting many men, and its detection and treatment resulting in such a considerable health care cost, it is important to make cost-effective detection, diagnosis, and treatment decisions for all men.

Population-based screening for prostate cancer has never been recommended in Australia. Opportunistic testing of men aged 50-75 years-old was included in general practice guidelines between 2005 and 2012, before being restricted to on-demand from 2012 onwards (Royal Australian College of General Practitioners, 2005, 2009, 2012). General practitioners have also received more guidance on how to discuss testing for prostate cancer with men (National Health and Medical Research Council, 2014), with later recommendations highlighting problems with overdetection and additional risks of testing (e.g. physical harms of invasive testing and psychological harms such as an increased risk of suicide)(Royal Australian College of General Practitioners, 2009, 2012).

This change in practice has largely reflected international guidelines in the last ten years, which have also recommended against population screening for prostate cancer (Drazer, Huo, & Eggener, 2015; National Institute for Health and Care Excellence, 2019; Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015; US Preventative Services Task Force, 2008, 2012). One of the reasons for reducing testing and

overdetection of prostate cancer is to reduce unnecessary treatment. Asymptomatic men tested for prostate cancer are more likely to be diagnosed with low-risk disease that may never affect them within their lifetime. Further measures to avoid invasive treatments for low-risk men include care guidelines in Australia and internationally that recommend surveillance for men with low-risk disease (National Institute for Health and Care Excellence, 2019; Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015). There have also been advancements in surgery, with the uptake of robot-assisted surgery for prostate cancer. Robot-assisted surgery has been shown to add considerably to overall prostate cancer care costs (Basto et al., 2016; Bolenz et al., 2014; Cao, Yang, Qi, & Chen, 2019; Close et al., 2013; Forsmark et al., 2018; Hohwu, Borre, Ehlers, & Venborg Pedersen, 2011; Laviana et al., 2016; Leow et al., 2016; Nguyen et al., 2011; Ramsay et al., 2012; "Robotic Surgical System for Radical Prostatectomy: A Health Technology Assessment," 2017; Florian Rudolf Schroeck et al., 2017).

Despite expert agreement on testing and care guidelines, there remains considerable variation in testing and care received by patients across Australia and elsewhere (Calopedos et al., 2019; Cary, Odisho, & Cooperberg, 2016; Dasgupta et al., 2019; Filson et al., 2014; Hjertholm et al., 2015; Jayadevappa, Chhatre, Johnson, & Malkowicz, 2011; Lacey et al., 2016; Lawson et al., 2020; Löppenberg et al., 2017; Modi et al., 2018; Ong et al., 2019; Pollack, Weissman, Bekelman, Liao, & Armstrong, 2012; Pooli et al., 2019; Riedinger et al., 2014; Seidenwurm & Logsdon, 2014; Tran et al., 2015; Wang et al., 2014). While some variation in care across populations was inevitable due to differences in population characteristics and needs, unwarranted variation in prostate cancer care that cannot be explained by illness or patient preferences represents an opportunity to save resources, improve health, or both (Wennberg & Gittelsohn, 1973).

Some of the variation seen in the diagnosis and management of prostate cancer is considered acceptable. For example, low-risk patients may receive different management to high-risk patients, or some random variation in diagnostic testing is expected. There is also systematic variation that healthcare systems may actively try to reduce, such as inappropriate provider behaviour (e.g. overuse of diagnostic tests). Besides the uneven quality of care to patients, systematic differences in diagnostic

and management practices may drive the variation in prostate cancer care costs (Roehrborn & Black, 2011). In the United States of America, Wang et al. associated treatment intensity and modality with 21.2% and 31.2% of the variation in cost between the highest and lowest expenditure quintiles, respectively (as estimated from SEER data between 2005 and 2009)(Wang et al., 2014). There can also be variation in costly outcomes within treatment modalities such as length of stay in hospital after surgery (Gore et al., 2012).

1.1. Variation in diagnosis, treatment and outcomes of prostate cancer

There are several areas in the prostate cancer care pathway where previous studies have identified systematic variation in prostate cancer diagnosis, management and outcomes.

1.1.1. Variation in prostate cancer diagnosis

In Australia, prostate cancer is mostly diagnosed in asymptomatic men. A general pathway to diagnosis usually begins with a prostate-specific antigen (PSA) blood test ordered by a man's general practitioner (GP). PSA testing is conducted for suspected prostate abnormalities and is not used to diagnose prostate cancer. However, if the results are indicative of prostate cancer, men are referred to a specialist clinician for further tests, usually ending with a confirmatory biopsy. The biopsy is also used for staging and grading of the cancer. Similar methods of testing are used globally, with PSA testing as the initial test for prostate cancer.

There is some evidence that there is potential for both overuse and underuse of PSA testing. A study of Danish general practitioner (GP) registry data found that patients of GPs with higher PSA testing rates did not have reduced prostate cancer mortality, but had more downstream diagnostic and surgical procedures (Hjertholm et al., 2015). On the other hand, Calopedos et al. suggested that consistently lower testing rates in Australia's rural areas between 2002 and 2017 have resulted in more prostate cancer missed for men in rural areas (Calopedos et al., 2019). Concerns over time to prostate cancer diagnosis, particularly in rural Western Australia, have led to GP-level

interventions (resource cards of risk assessment charts and referral pathways) (Emery et al., 2014).

There have been several attempts to understand GP attitudes to PSA testing in Australia. Surveys of Australian GPs found evidence that PSA testing rates are related to a GP's perception of whether too few (underdiagnosis) or too many (overdiagnosis) prostate cancer cases are being detected (Pickles, Carter, & Rychetnik, 2015). GPs in Australia do not routinely receive information about their PSA testing (e.g., prostate cancer cases detected per PSA test). They may also be uninformed of the additional tests and treatments a patient receives once referred to a specialist. GPs may also be concerned about the outcomes of undertesting (e.g. litigation).

Delays to subsequent prostate cancer care can be a source of stress to prostate cancer patients (Tran et al., 2015). Referral to a prostate cancer specialist took at least three months for a third of patients diagnosed with prostate cancer in one of 5 Victorian Comprehensive Cancer Centre hospitals between Oct 2012 and April 2013 (Lacey et al., 2016). These delays suggest there is some variation not only in who is tested, but how long it takes to complete testing.

1.1.2. Variation in prostate cancer management

One well-documented area of differences in prostate cancer care is the variation in prostate cancer management. There are several approaches to the management of localised prostate cancer: observation (e.g., surveillance), curative treatment (e.g., prostatectomy, radiotherapy) and non-curative treatment to manage symptoms (e.g., chemotherapy). Risk related variation in patient management is efficient. Recent international guidelines recommend surveillance instead of curative treatment for men with low-risk prostate cancer, or asymptomatic men for whom prostate cancer is unlikely to affect them in their lifetimes (National Institute for Health and Care Excellence, 2019; Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015).

However, there is also some evidence of unwarranted variation. US longitudinal studies of registry data (SEER, National Cancer databases) have found differences in management across:

- institution characteristics (lower treatment rates in academic centres (Löppenberg et al., 2017; Pooli et al., 2019), reduction of prostatectomy in patients of facilities with groups of clinicians vs solo physicians (Satkunasivam et al., 2018)) and location (regional variation in treatment decisions) (Burt, Shrieve, & Tward, 2018; Cary et al., 2016; Dasgupta et al., 2019; Pooli et al., 2019);

- by patient insurance status (higher likelihood of treatment in patients with health insurance) (Burt et al., 2018; Pooli et al., 2019), race (Caucasian patients more likely to receive treatment(Pooli et al., 2019);

- physician attitudes (physician social networks associated with variation in prostatectomy (Pollack et al., 2012), physician behaviour patterns associated with overuse of imaging for staging/surveillance in low-risk disease (Lipitz-Snyderman et al., 2016)). Physician attitudes to treatment may also vary information they gather before recommending treatment (when considering a hypothetical patient, different oncologists considered between 5 and 69 questions about the patient essential) (Feldman-Stewart et al., 1998).

Differences in treatment by location, insurance status, and race are potentially inefficient and raise equality issues, e.g., where not all patients with equal need are treated equally. Studies in Australia found similar differences in treatment variation. Men residing in more rural areas or more socioeconomic disadvantaged areas of New South Wales are less likely to receive radical prostatectomy (Hayen, Smith, Patel, & O'Connell, 2008). Furthermore, hospital location was associated with variation in the rate of observation in Victoria (Ong et al., 2019) and remote residence in South Australia was associated with a reduction in access to certain types of radiotherapy (patients less likely to get brachytherapy) (Morias et al., 2020).

The thesis focuses on the effect of interventions that reduce variation in prostate cancer care, but these may not address all unwarranted variation.

1.1.3. Variation in outcomes

While there is evidence of variation in care, it is important to know whether this variation is likely to result in poorer clinical and patient outcomes. There have been

attempts to estimate the causal relationship between differences in prostate cancer outcomes and variation in care. However, it can be challenging to isolate the effect of one particular aspect of care to prostate cancer-related outcomes, such as quality of life and mortality where the disease trajectory and care pathway are so prolonged.

Instead, prostate cancer organisations have generally relied on quality indicators measured as the quality of the process or short-term patient and clinical outcomes. These can include the reporting of staging and prognostic factors before treatment; perioperative outcomes for patients receiving prostatectomy (e.g. blood use, length of hospital stay); surgical outcomes (e.g. whether some cancer tissue remains after prostatectomy, a positive surgical margin); short term patient-reported outcomes (sexual, urinary and bowel function following treatment); and cost of treatment.

There is some evidence that these outcomes may also be subject to unwarranted variation: individual hospitals in Washington state had different lengths of stay for prostatectomy (where hospital choice was associated with 26.7% of the variation in the length of stay) (Gore et al., 2012); US academic institutions reported lower positive surgical margin (PSM) rates following prostatectomy than other institution types (Pooli et al., 2020); and quality indicators related to reporting varied by treatment modality (EBRT patients more likely to have staging data and pre-treatment urinary and sexual function recorded than surgical patients) (Miller et al., 2007). Within treatments there can also be variations of outcome: there is some international evidence that robot-assisted radical prostatectomy is associated with higher costs, but the shorter length of hospital stay compared to open radical prostatectomy (Basto et al., 2016; Close et al., 2013; Forsmark et al., 2018; Leow et al., 2016; Liberman, Trinh, Jeldres, & Zorn, 2012; Medical Advisory, 2010; Mouraviev et al., 2007; "Robotic Surgical System for Radical Prostatectomy: A Health Technology Assessment," 2017; Florian Rudolf Schroeck et al., 2017; Tang et al., 2017). These short-term differences however may not translate to different outcomes. A recent Australian RCT (comparing 2 surgeons with 163 patients each) demonstrated similar oncological and patient outcomes for both robot-assisted and open surgery (Coughlin et al., 2018).

To understand differences in quality and variation of care for a given patient, influences of prostate cancer care behaviours must be understood.

1.2. Assessing the quality of care

Differences in the management of similar patients have consequences for the quality of care they receive. Measurement of the quality of care is often categorised using (Donabedian, 2005): 1) process (e.g., appropriate medical care); 2) outcomes (e.g., patient-reported outcomes after treatment); and 3) structure (e.g., availability of equipment, qualifications/skill of the physician).

Measures of process quality in prostate cancer often develop from guidelines and outcome quality from prostate care providers (Kötter, Blozik, & Scherer, 2012). Schroeck et al. used the Donabedian structure to establish a framework to assess the quality of care in prostate cancer and found that compliance with many measures was less than 80%. Measures included recording clinical staging, family history, baseline urinary/sexual/bowel function; avoiding overuse of bone scan for low-risk prostate cancer; and scheduling two or more follow up visits in the year after initial treatment. Therefore, if achieving these quality indicators represents good quality care, there is the potential for improving patient care by improving compliance with these indicators (F. R. Schroeck, Jacobs, & Hollenbeck, 2013).

Similarly, the thesis used this Donabedian structure to identify and assess potential influences and interventions on prostate cancer care quality. The measures of prostate cancer care are considered as such: 1) process: reducing overuse of PSA tests in asymptomatic men, avoiding overtreatment in low-risk-men; 2) outcomes: surgical quality measured through PSM, patient-reported outcomes; and 3) structure: surgeon experience with robotic surgery. There is evidence, discussed in the sections below, that interventions to improve quality of health tend to address one of the three categories and that interventions may be better at addressing process measures than outcome (which may rely on process or structure changes to address) or structural quality indicators (which may require more financial or leadership support than individual process changes). In prostate cancer care, process quality indicators also tend to be the easiest to collect data for, e.g., numbers of PSA tests, whereas structural and outcome measures often require reporting by patients or clinicians.

1.2.1. What influences provider behaviour?

1.2.1.1. Influences to process and outcomes

There are several common approaches to influence provider behaviour in terms of process quality (e.g. treatment decisions) and outcome (e.g. patient-reported quality of life or mortality). One common technique is using consensus or evidence-based guidelines, published by healthcare organisations or governing bodies, such as Cancer Council or National Health and Medicine Research Council in Australia or National Institute of Health and Care Excellence in the UK. There are also specific recommendations for clinicians, such as the Royal College of General Practitioners *Guidelines for preventative activities in general practice* series. Guideline production may involve an implementation strategy. For example., NHMRC provided a clinician guide for discussion (National Health and Medical Research Council, 2014) to go with PSA testing recommendations (Royal Australian College of General Practitioners, 2012). Historically, however, guidelines have been expected to change clinical behaviour primarily through knowledge provision. This knowledge may be adequate to overcome barriers such as changes in consensus and evidence over time. However, patient preferences and inability to implement recommendations (e.g., lack of support from peers and supervisors, lack of access to technologies) may prevent clinicians from responding to guidelines. There is also some evidence that guidelines may not provide enough knowledge to clinicians to judge their performance, as clinicians may overestimate their (and their peers') clinical performance (Gude et al., 2018).

One way to provide clinicians with information on their performance is through auditing and feedback. Auditing and feedback refer to reporting quality indicators and provider behaviours to clinicians, their departments/hospitals, or publicly. The target audience of the feedback may also alter the mechanism by which it affects change. Mechanisms can include fear of reputation loss, direct financial rewards through pay-for-performance strategies. For peer comparison feedback, clinicians may respond to the pressure of being at least as good as their peers, but may be dismissive if the choice of indicators and evaluation of their work is conducted by researchers external to their field (i.e. non-peers). This is the expectation of the peer comparison feedback

presented to urologists on the Prostate Cancer Outcome Registry-Victoria, which compares how successful urologists were at meeting prespecified quality indicators. Feedback may be limited by its quality, such as choice of indicators or presentation of feedback. Two common presentations of feedback are league tables (where providers are given a summary score for quality and compared to their peers, often using ordered bar plots) and funnel plots (where success rate of individual behaviours for providers are compared to their peers including both an average and a confidence interval to compare the behaviour to). Funnel plots are discussed further in Chapter 4 to describe their use in registry feedback. Because they tend to focus on individual behaviours rather than an aggregate score, and are constructed to show an individual providers variation from the average of all providers, funnel plots are considered more statistically sound than league tables (Spiegelhalter, 2005). Feedback may also be limited by competing or unclear elements, and the clinician perception of the importance of a performance measure. Clinicians were historically not involved in quality improvement (Audet, Doty, Shamasdin, & Schoenbaum, 2005). Methods for engagement are relatively new, resulting in quality indicators of differing relevance to patients and the public. A focus on process indicators (e.g. reporting of staging data) might have fewer benefits than outcome indicators (PSM, biochemical recurrence, patient-reported outcomes), which may be more relevant to patients and clinicians (Pross, Geissler, & Busse, 2017).

1.2.1.2. Influences to structure

There are also approaches to affect structural elements (e.g. facilities, clinician qualifications) of quality assessment, and these are often linked to changes at the organisation level. One of these includes clinician collaborations who seek to improve quality of care through shared knowledge, networking, peer comparison and collective power to direct and dedicate resources. These collaborations can take the form of multi-disciplinary teams, clinician networks and clinical quality registries. Clinical quality registries collect data on treatment and outcomes and use them to provide feedback to clinicians who take part in the registry. For example, Prostate Cancer Outcomes Registry- Victoria collects prostate cancer patient data in Victoria to provide peer comparison feedback to clinicians. Some barriers to change through these

collaborations may include financial incentives to overuse resources or not change financially-beneficial behaviours; or lack of support by management and peers.

Another structural element that can influence provider behaviour is the training, skills, and qualifications clinicians may have. Additional training or skills help increase clinician knowledge. They may come with professional pressure to improve or financial incentive, where clinicians may negotiate higher pay based on specialised services they can offer. The uptake of robot-assisted radical prostatectomy (RARP) is an area of prostate cancer care where additional training is anticipated.

RARP is also an example of another structural component that can change provider behaviour: changes to available technology and clinician skills. It is likely that new technologies have new costs and benefits, may improve clinician morale, or encourage high-quality clinicians' recruitment. There may also be a financial incentive when patients are more willing to pay for new technologies for perceived benefits. However, new technologies may not always be financially viable, and not all clinicians may have access to new technologies. Furthermore, should the new technologies fail to provide benefits, this may reflect poorly on the quality of care from the chosen provide. New technologies may also introduce new variations in care, where clinicians do not have similar experience or skill with using the latest technology. In the case of prostate cancer, surgeons may try to have skills in multiple surgery types to offer to their patients, but this could mean they are less able to specialise. RARP is chosen as the focus of the exploring the effect of experience with new technology as surgery method and outcomes are well-documented within the available dataset. Other new technologies in prostate cancer are available, such MRI and PSMA-PET, but these are not well recorded in the data, are hard to allocate to as one clinician’s responsibility, and the feedback that clinicians receive does not currently provide direct consequences of these scans. Therefore, it would be difficult to isolate the effect of these technologies within the context of data currently available.

Often behaviour modifiers are used in combination. For example, feedback can promote adherence to guidelines, and registries can collect data for the feedback reports.

In the following chapters, the focus is on factors that have been used in Australia to influence provider behaviour in prostate cancer:

- Guidelines to reduce PSA testing in asymptomatic men (the expectation that PSA testing was overused)
- Feedback to promote adherence to localised prostate cancer management recommendation (the expectation of overtreatment in low-risk patients)
- Feedback on localised prostate cancer outcomes (the expectation that PSM rates and patient-reported outcomes can be improved)
- Experience of technological innovation and adoption decisions using the example of robot-assisted radical prostatectomy as an alternative to open radical prostatectomy and its effect on prostate cancer outcomes (frequent use of RARP, but not recommended for all patients by international guidelines)

This thesis evaluates the effectiveness of these means to change provider behaviour, reduce variation in care, and improve prostate cancer patients' outcomes.

1.2.1.3. Evidence of the effect of guidelines

National evidence-based guidelines exist for the treatment of localised prostate cancer in Australia (Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015) and general practice screening handbooks are revised every few years to provide GPs with the latest consensus and evidence on PSA testing (Royal Australian College of General Practitioners, 2018)

Several reviews have examined interventions to increase adherence to guidelines across several populations. They have included interventions such as external inspection of adherence to recommendations (Flodgren, Gonçalves-Bradley, & Pomey, 2016), payment for services (Flodgren et al., 2011), specialised personnel involvement (Bighelli et al., 2016; Flodgren et al., 2013), rewriting guidelines (Bighelli et al., 2016). The studies included randomised control trials (RCTs) and observational studies. There is some evidence that guideline dissemination alone is not sufficient to change healthcare professionals' behaviour (Tzortziou Brown, Underwood, Mohamed,

Westwood, & Morrissey, 2016). Financial incentives are also not shown to be a strong driver of compliance with guidelines (Flodgren et al., 2011). Generally, guidelines are developed to affect process changes, i.e. modifications to provider behaviour.

Deciding to test men for prostate cancer can be a complex decision for men and their GPs, and a consensus guideline does not mean that there will be consensus in testing. Therefore, the variation in response to the guidelines in different states representing different professional environments and attitudes is tested.

1.2.1.4. Evidence of the effect of registries and collaborations

Formal clinical collaborations have existed in some form since at least the 18th Century (Shaw, 1968), often as societies that could share knowledge and establish clinician networks. A more recent development has been collaborations to build clinical registries where patient data can be collected to represent current practice, identify areas where the care may be improved, and provide a structure to implement changes. These are becoming increasingly common, with 31 disease- and state-specific registries currently listed by the Australian Commission on Safety and Quality in Health Care (Australian Commission on Safety and Quality in Health Care, 2019)

One of the most extensive international prostate cancer collaborations has been the PRIAS Project (Prostate Cancer Research International Active Surveillance, https://www.prias-project.org/), coordinated by the European Randomized Study of Screening of Prostate Cancer (https://www.erspc.org/). The PRIAS Project examines the effectiveness of active surveillance for low-risk patients as an alternative to curative treatment with data from medical centres located in Europe, Japan, Australia and New Zealand, among others. In the US, accountable care organizations (ACOs) are collaborations formed to encourage clinician engagement in quality improvement. Newly diagnosed prostate cancer patients (between 2012 and 2014) with a high risk of mortality within ten years had a significantly lower likelihood of potential overtreatment if their urologist was more engaged with the ACO (top quartile compared to bottom quartile). No significant difference was seen in other treatments or costs in general (Modi et al., 2018). ACOs are also expected to encourage coordination of care but could discourage sending patients to specialised care centres (Hohn, 2012).

One specific collaboration, the Michigan Urological Surgery Improvement Collaborative (MUSIC) is a state-wide physician-led collaboration designed to improve prostate cancer care (Riedinger et al., 2014). It provides physician-led learning opportunities such as video review of RARP to understand variation among surgeons (Wu et al., 2020). From 2014 it began reporting comparative performance feedback using the Qualified Clinical Data Registry Quality reporting system and in 2017 introduced a merit-based incentive payment system 2017 (Michigan Urological Surgery Improvement Collaborative, 2020). These measures have potentially improved surgery outcomes because surgery complications dropped by 2.6% (Share et al., 2011).

Cancer registries have existed in Australia since at least 1972 (Cancer Institute NSW, 2020). These registries record cancer incidence by cancer type and patient demographics. Their original purpose was to provide descriptive statistics as a resource for planning care services and provide opportunities to collaborate in research such as case-control studies (McCredie, Coates, Churches, & Taylor, 1991). Cancer registries have been used to track prostate cancer incidence over time and reflect upon practice changes. For example, prostate cancer incidence rose sharply after the introduction of PSA testing in NSW (Bird et al., 2005; Smith, Supramaniam, Marshall, & Armstrong, 2008).

As well as general cancer registries, specific localised prostate cancer registries have been introduced. South Australian Prostate Cancer Clinical Outcomes Collaborative has recorded data on localised prostate cancer diagnosed and treated in South Australia in greater depth than the general cancer registries since 1998. In late 2008, Victoria established the first Australian clinical quality registry for prostate cancer in Australia, providing peer comparison feedback to hospitals and urologists: The Prostate Cancer Outcomes Registry- Victoria (PCOR-Vic) (The Australian Commission on Safety and Quality in Health Care, 2016). PCOR-Vic provides a potentially large evidence base to demonstrate an impact of the registry upon clinical practice and patient outcomes, and examine provider behaviours towards emerging technologies and changing guidelines. From 2008 to 2019, the number of contributing hospitals increased from 3 to 82, with over 26,000 patients included between 2008 and 2019 [data on file].

Despite the expectation that clinical quality registries aim to improve clinical and patient outcomes by incentivising provider behaviour changes, few studies have rigorously tested the use of registries as an intervention on provider behaviours and health outcomes. A recent systematic review of clinical quality registries identified 17 studies, including 6 for diabetes care, 2 for cardiac disease, 2 for lung disease, and the rest on organ transplantation, rheumatoid arthritis, ulcers, surgical complications and kidney disease (Hoque et al., 2017). The majority of studies were non-experimental (11 studies) and used statistical techniques to adjust longitudinal data for biases to estimate the intervention's effect. The effects examined differed across the registries but included changes to the quality of care, care processes, treatment outcomes, adherence to guidelines and survival. Sixteen studies reported improvements in their interested outcomes after implementing the registry, suggesting a positive impact on clinical outcomes (Hoque et al., 2017). The different aims of the registries and methods for adjusting the observational data make it challenging to extrapolate the findings from one registry to another. However, in general, registries seem to improve the outcomes they set out to improve (Hoque et al., 2017).

Clinical registries can also be used as a research tool, to conduct studies that collect additional primary data or analyse the data already collected and assess how process guidelines can translate to patient outcomes. In Germany, providers must engage in quality assurance procedures, and clinical cancer registries collect the data (Inwald, Klinkhammer-Schalke, Koller, & Ortmann, 2014). Breast cancer quality indicators in this setting showed that adherence to treatment guidelines improved patient outcomes (higher overall survival) (Inwald et al., 2014). Similarly, in the US, CEASAR (Comparative Effectiveness Analysis of Surgery and Radiation) used registry data on 2,601 men localised prostate cancer between 2011-2012 to assess how quality measures translated to patient outcomes. No quality measure (avoidance of bone scan in low-risk tumours, ADT for high-risk patients, documentation of clinical T (cT) stage and Gleason score at diagnosis, documentation of digital rectal examination (DRE), Gleason score before initial treatment, documentation of discussion of treatment options, documentation of pathological T (pT) stage, pN, positive surgical margin status for men undergoing RP) was associated with changes in patient-centred outcomes (quality of life, satisfaction, or complications) (Sohn et al., 2016).

Clinical registries can also be used to assess structural quality indicators, e.g., clinician experience of technologies within the registry. Registry data have previously been used to compare surgical outcomes (e.g., PSM) for RARP and open radical prostatectomy (ORP) (S. M. Evans et al., 2014; Lowrance et al., 2010). These studies have not compared surgical technique at the system level (e.g. the average effect of RARP versus ORP on surgical and patient outcomes), with adjustment for different levels of surgeon experience of RARP (e.g., number of RARPs conducted). This thesis uses the PCOR-Vic to assess how RARP experience affects patient and surgical outcomes collected by the registry.

1.2.1.5. Evidence of the effect of feedback

Registries can also be used to collect data that is disseminated as feedback to clinicians and hospitals. For example, the PCOR-Vic seeks to alter behaviours and outcomes by providing benchmarking reports of 11 indicators (in areas such as diagnosis; management, and outcomes) to urologists and hospital stakeholders every six months, as well in the annual public reports (Sampurno F and Evans SM (eds) for the Victorian Prostate Cancer Clinical Registry Steering Committee, 2015). The registry is concerned with reducing unnecessary treatment (in patients for whom monitoring would be more beneficial) and surgical outcomes (reducing the likelihood of cancerous tissue missed during prostatectomy: positive surgical margins).

The PCOR-Vic began producing individual peer-comparison reports for urologists and hospitals in late 2012, after working closely with clinicians to choose the quality indicators that would be most relevant to them and help them change behaviours. The development of these reports remains a continuing process. On request, participating urologists receive 6-monthly reports online and in hardcopy as these were more likely to be read (personal communication with registry, SE). However, data on urologist engagement with the reports is not clear and should be a consideration as the reports are developed.

Currently the reports contain peer comparison funnel plots of specific quality indicators and from 2019 introduced cumulative sum plots for individual urologists and hospitals to track their own outcomes. Reports to individual urologists also detail specific patients who have received management unexpected for their risk status (e.g., high

risk but did not receive treatment). There have been investigations into how to best present feedback to clinicians and stakeholders (Koh, 2017); possible considerations of expanding individualised reports (currently only urologists receive individualised feedback); and whether other areas of patient care could be improved (M. A. Evans et al., 2018). This thesis builds on this knowledge to identify areas where provider behaviour varies and assess interventions that may change clinician practice and promote better care outcomes.

Previous assessment of PCOR-Vic between 2009 and 2013 using ARIMA methods suggested that Victorian prostate cancer care improved across three quality indicators. These indicators aim to reduce treatment in low-risk men, increase treatment with a year of diagnosis in higher-risk men, reduce positive surgical margins in men with pT2 staged disease (Sampurno et al., 2016). However, this restriction to specific patient characteristics does not reflect a complete analysis of prostate cancer care quality in Victoria or interventions to change provider behaviour.

A Cochrane review of RCTs (Ivers et al., 2012) (an update of (Jamtvedt, Young, Kristoffersen, O'Brien, & Oxman, 2006)) assessed the impact of feedback and audit as interventions to changes in the behaviour of healthcare professionals in primary and secondary care. Ivers et al. found that audit and feedback generally improved compliance to desired practice and patient outcomes (Ivers et al., 2012). In particular, they found that feedback was most effective when clinicians had low baseline performance; received feedback from a colleague or supervisor (rather than external source); received feedback more than once; received feedback in both written and verbal formats; and received feedback which included targets and an action plan (Ivers et al., 2012). There was more evidence that feedback could improve clinical practice outcomes than patient outcomes. Four studies were identified that assessed the effectiveness of peer comparison feedback. One study showed improvement in diabetes care and included peer comparison and benchmarks for clinical outcomes (Kiefe et al., 2001). Three studies provided peer comparison feedback to clinicians treating asthma and showed mixed improvement following feedback. Methods of feedback were mixed and included group feedback in-person and written feedback in aggregate tables and guideline information. None of these studies included comparative feedback presented in funnel plots or were in prostate cancer. Indeed,

while there have been some studies committing an audit of prostate cancer treatment in recent years (Alghamdi et al., 2016; Martin, Persaud, Corr, Casey, & Pillai, 2018), these have been used to establish compliance with guidelines, but not how practice changes after the audit.

Comparative feedback is expected to promote behaviour change in clinicians primarily through a peer effect. Bevan (2009) believes that the most significant change in response to comparative feedback comes from inferior quality healthcare providers responding to reputational damage (Bevan & Hamblin, 2009). Whether ambulance services met targets (most notably to respond to 75% of life-threatening emergency calls with 8 minutes) in England was reported to the public via a star rating. Scotland and Wales had the same target but no public reporting of whether the target was met. England had better rates of meeting the performance targets, believed to be a result of reputational damage to receiving a low star rating. Bevan also discussed some problems with implementing star ratings, including the appropriate selection of indicators, nature of quality measures, and aggregation of ranking. There were also concerns with practitioner response, such as the potential for gaming (providers improving targets set, to the detriment of other health care measures/healthcare overall), or damaging staff morale to the detriment of patient outcomes (due to absence of staff through sick leave, redundancy, quitting or less efficient staff because of their poorer psychological health) (Bevan & Hamblin, 2009).

Clinician attitude to feedback can act as a modifier to the effectiveness of feedback. Gude et al. found that audit and feedback to 72 intensive care professionals (21 Dutch institutions) helped care professionals correct estimations of clinical performance (peer and self-performance were previously overestimated) and increased intentions to improve (although many intended to improve before feedback). Where professionals did not intend to improve (8.3%), clinicians did not consider indicators important, did not trust the data, or deemed benchmarks unrealistic (Gude et al., 2018).

Bird recommends repeated collection and dissemination of feedback as a facet of quality improvement (Bird et al., 2005). Feedback would hopefully result in an increased response over time as repeated feedback on comparative underperformance is received both as the average performance improves and

individual experience of feedback increases, in line with the theory of healthcare improvements described by Prochaska et al., 1984 (Prochaska & DiClemente, 1984). However, repeated feedback exposure may also induce desensitisation and self-justification for underperformance. Furthermore, for individuals who respond quickly to feedback, there may be less significant behaviour changes with repeated feedback (e.g., (Lenderink, Spreeuwers, van der Klink, & van Dijk, 2010)]

The thesis investigated whether feedback stratified by risk leads to changes in the expected direction (reduce overtreatment in low-risk prostate cancer, increase treatment in high-risk prostate cancer). The thesis also considered how clinicians respond to surgical and patient outcomes that may result in competing clinician practices, for which no evidence in prostate cancer was identified.

1.3. Aim of the thesis

This thesis examines the effect of changes in prostate cancer clinical guidelines, the implementation of a system of audit and feedback to specialists, and the adoption of new technology on provider behaviour and patients with localised prostate cancer.

We focus on three specific questions and highlight which of the 3 Donabedian categories each chapter uses (*process, outcome, structure*).

1. Do guidelines result in changes in provider behaviour and improvements in prostate cancer care?

   Specifically: Is the introduction of general practitioner (GP) guidelines restricting PSA use associated with a reduction in PSA testing rates (*process*) and improvement in the efficiency of prostate cancer detection (*outcome*)? (Chapter 2)

2. Does individualised feedback to clinicians result in improvements in the quality of prostate cancer care

   Specifically: Does feedback on treatment decisions and surgical quality reduce overtreatment in low-risk patients (*process*), reduce time to curative treatment

for high-risk patients (*process*), and improve patient outcomes indicators (*outcome*)? (Chapters 3 and 4)

3. Does clinician experience of new technology improve patient outcomes in prostate cancer?

    - Specifically: Are patient outcomes (*outcome*) improved following surgery with robot-assisted radical prostatectomy compared to open radical prostatectomy (*structure*)? Do surgeons with more experience of RARP (*structure*) see more improvement (*outcomes*) than ORP? (Chapter 5)

2. State-level changes in prostate cancer detection and response to guidelines: an interrupted time series analysis

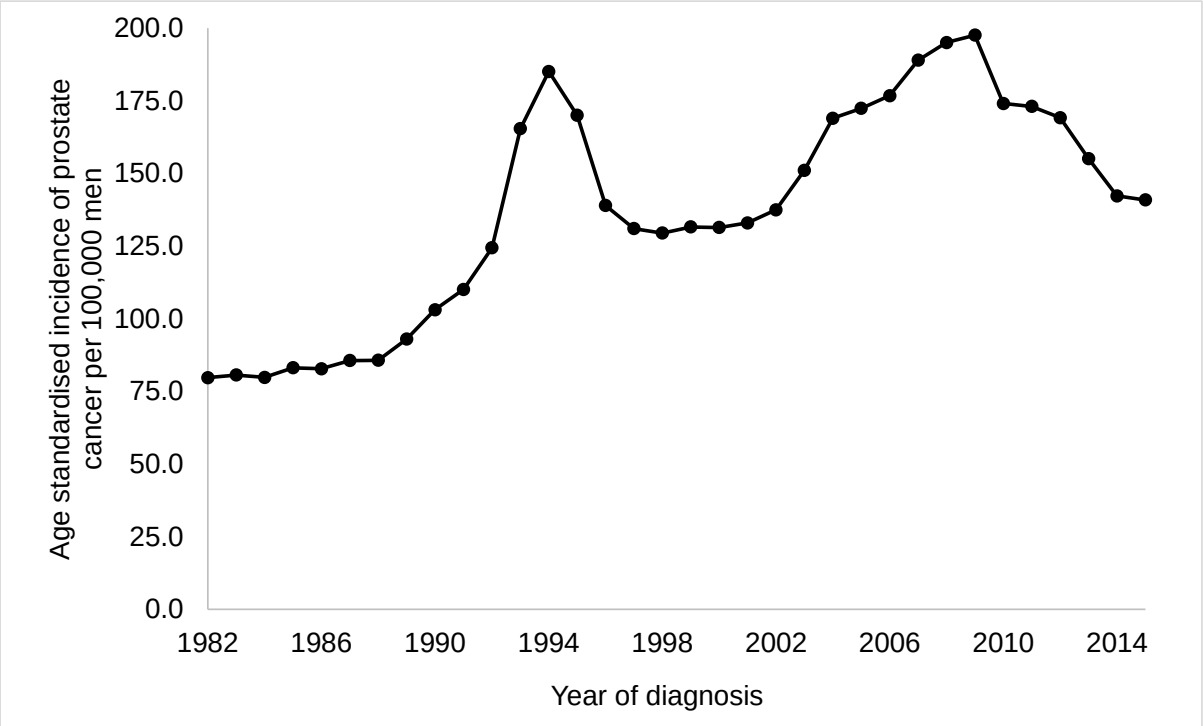
2.1. Introduction

Prostate cancer is the most commonly diagnosed cancer in men in Australia, with 19,305 new cases diagnosed in 2016 and 16,741 new cases predicted for 2020 (Australian Institute of Health and Welfare, 2020). The majority of prostate cancers are diagnosed in asymptomatic men, through PSA blood tests followed by a biopsy. Diagnosis in asymptomatic men results in cancer usually being confined to the prostate (localised) when diagnosed. In Victoria, localised cancer accounts for 70-80% of all prostate cancer diagnoses each year (73% in 2014; data provided by PCOR-Vic July 2017). Localised cancers can take many years to spread, become symptomatic or cause death. Their management with prostatectomy or radiotherapy may cause unnecessary harm (Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel., 2016). Therefore, early diagnosis and treatment of localised prostate cancer may have little economic value.

Unnecessary prostate cancer treatment may arise through overdetection, false-positive test results, or low risk of cancer progression. Exact figures are unknown, but Loeb et al. report over detection rates between 1.7% and 67% (Loeb et al., 2015). Records of age-standardised rates of prostate cancer detected (usually by cancer registries) have shown fluctuations in the rates of prostate cancer over time (Australian rates given in Figure 1), which is expected to reflect changes in the testing pathways for prostate cancer. For example, PSA blood testing first appeared in reimbursement schedules in Australia in 1989 (Smith & Armstrong, 1998) and during the 1990s PSA testing in asymptomatic men under the age of 65 years saw a sharp increase (Smith & Armstrong, 1998). PSA test introduction is likely to be why the rapid increase in prostate cancer detection in men peaked in 1994 (Feletto et al., 2015; Smith et al., 2008) in Australia (Figure 1). This peak is far higher than the projected estimates of prostate cancer incidence (Smith et al., 2008). The cause of the initial rapid increase and drop off in incidence between 1990-1998 has been partially attributed to a larger

proportion of prostate cancer detection in earlier stage disease due to PSA testing (Smith et al., 2008). This trend of earlier detection of prostate cancer has continued (Smith et al., 2008). If men are identified in early-stage disease, there may be a larger proportion of men receiving unnecessary treatments.

Figure 1 Prostate Cancer incidence in Australia, 1982-2015



Source: Reproduced from ©2020 Cancer Australia - https://ncci.canceraustralia.gov.au using data sourced from AIHW 2018. Australian Cancer Incidence and Mortality (ACIM) books. Accessed October 2020

Currently, the general diagnostic pathway for prostate cancer is given in Figure 2. To diagnose prostate cancer, patients first receive at least one PSA blood test and a physical examination. A significantly high PSA result leads to a confirmation test, usually biopsy.

Globally, there have been efforts in the last ten years to reduce what are viewed as inappropriately high rates of localised prostate cancer diagnoses (overdiagnosis). Overdiagnosis may result from false positive tests and from diagnosing very low-risk cancers that are unlikely to affect a man within his lifetime. These efforts to reduce overdiagnosis have been in part through the dissemination of new guidelines for opportunistic testing (National Health and Medical Research Council, 2014; National Institute for Health and Care Excellence, 2019; Prostate Cancer Foundation of

Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015; US Preventative Services Task Force, 2008, 2012). Guidelines are considered effective in reducing testing of men over 50 years old in the US, but overuse of PSA testing in older age groups is still a concern (Drazer et al., 2015).

In Australia, national diagnostic recommendations have aimed to reduce the number of PSA blood tests in asymptomatic men over the past ten years. The 6th and 7th editions of the *Guidelines for preventive activities in general practice* (released 2005 and 2009 respectively) recommended opportunistic testing for men between 50 and 75, but emphasise the disadvantages of testing: unproven survival benefit, and urinary and bowel incontinence and erectile dysfunction resulting from treatment (Royal Australian College of General Practitioners, 2005). The 7th edition also recommended digital rectal examination (DRE) alongside PSA testing. It was also the first edition that explicitly described the possibility of overdetection associated with PSA screening (Royal Australian College of General Practitioners, 2009). The 8th edition of the Guidelines for preventive activities in general practice (2012) required patients to request testing. It included additional disadvantages compared to previous editions, including increased risk of suicide and cardiovascular disease following prostate cancer diagnosis and the chance of sepsis following biopsy (Royal Australian College of General Practitioners, 2012). A timeline of potential modifiers to prostate cancer testing is given in Appendix 7.1, p132.

States were anticipated to have different approaches and attitudes to reducing low-value prostate cancer care and encouraging guideline uptake. Some states introduced clinical registries for prostate cancer (Victoria [PCOR-Vic] in 2008 and South Australia in 1998), which provide additional information and encouragement for clinicians to implement guidelines. However, neither registry directly interacts with GPs, and though PCOR-Vic provides reports on prostate cancer care to urologists, it does not report to GPs. But developing a clinical registry, particularly one which provides feedback, may also be indicative of a broader attitude to quality improvement. For this reason, Victoria may have behaved differently to states with no registry in both the periods before and after the introduction of the prostate cancer registry.

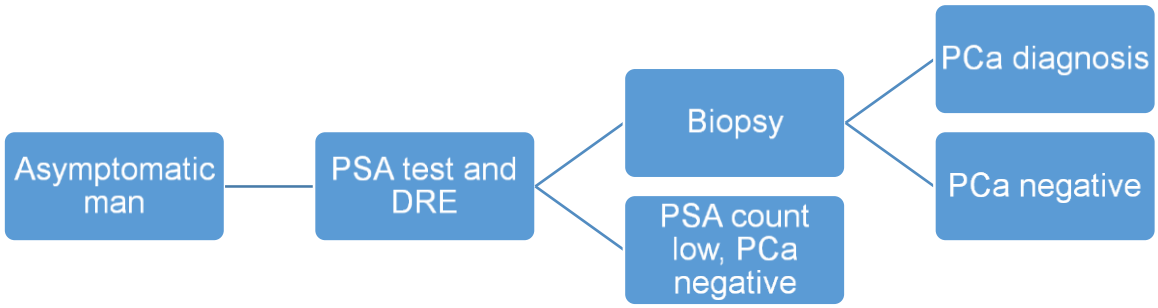
Incidence of prostate cancer was expected to be primarily driven by testing rates, but may also differ across states because of population characteristics including age, race,

family history, and socioeconomic status. Interpretation of patient risk (e.g., result of PSA) may also differ across states, with some more focused on reducing diagnosis in low risk men (e.g., repeat PSA testing may occur for men with borderline PSA levels). This analysis did not adjust for many of these characteristics due to a lack of data, but are important to interpreting the results.

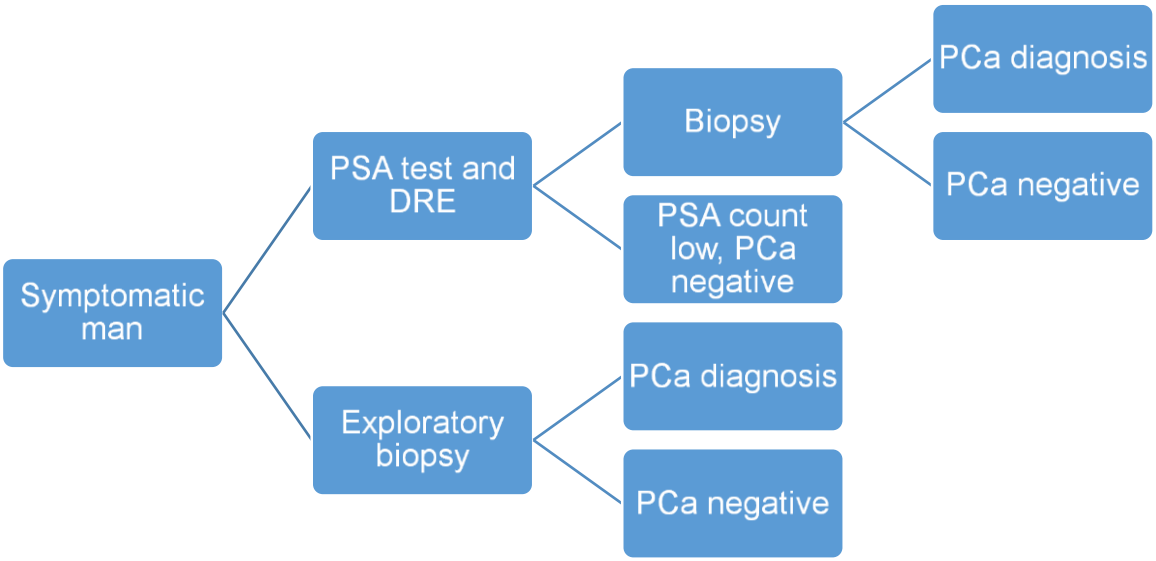
Also expected were differential effects of the guidelines by age because the risk of prostate cancer increases with age. However, the benefit of being prescribed curative treatment reduces with age, suggesting a prostate cancer diagnosis may be harmful in asymptomatic older patients and the cost of missing a prostate cancer diagnosis through asymptomatic testing is reduced. In older patients (75+ years old), the rate of cancer per test should not to change as much as other groups, as most cases are detected symptomatically.

This study aimed to assess how clinicians in different states responded to the national GP guidelines for restricting PSA testing, by describing the changes in behaviours (diagnostic testing) and system effectiveness (cancer detected per test) and the speed at which the changes may occur. Changes to guidelines were expected to reduce testing rates and cases detected across all states, with the cases detected per test increasing for men under 75 years old. For men aged 75 and over, a decrease in testing may not yield an increase in cases detected per test. A secondary aim was to compare the effect of guidelines in states with anticipated different approaches to guideline adherence (e.g., as evidenced by the existence of a registry for prostate cancer that provides feedback to specific clinicians). Victoria may differ from the other states either through the indirect effect of the registry, or the wider attitude that resulted in the registry's development, decreasing testing and incidence from 2009 either faster, or in a more pronounced manner than the other states. Similarly, Victoria may have higher testing rates prior to 2009, in line with the guidance that recommended opportunistic testing of asymptomatic men. South Australia was excluded from this analysis as their prostate cancer registry was established in 1998 but did not produce feedback.

Figure 2 Testing pathways for diagnosing prostate cancer



a)



b)

Notes: DRE= Digital rectal examination; PCa= prostate cancer; PSA= prostate specific antigen

2.2. Methods

2.2.1. Data

State-level cancer registries collecting incidence and mortality data have been established since at least 1972. They provide some information on prostate cancer incidence and detection changes over time. They also collect some demographic data, such as patient age to establish age-standardised trends. State-level prostate cancer case data, stratified by age group was received from Victoria (VIC), New South Wales (NSW), Queensland (QLD), and Western Australia (WA) cancer registries. Latest case data was from 2015 and each state and total number of cases ranged from 9,306 in 2000 to 19,238 in 2009. Prostate cancer incidence per 100,000 men was calculated using state-level population data from The Australian Bureau of Statistics, 2017 (Australian Bureau of Statistics, 2017).

Resource use on PSA testing grouped by age was sourced from Medicare claims (accessed through the (Australian Government Services Australia, 2017) http://medicarestatistics.humanservices.gov.au/statistics/mbs_item.jsp). The PSA test under code 66655 was chosen as the PSA test of comparison as it is restricted to men with no previous diagnosis of prostate cancer and one per man per year. Cases per test can, therefore, be interpreted as cases per man tested. As this test was introduced in 2001 (and the description of follow up tests updated in 2002), analyses restricted to 2002 onwards to adjust to the new code. As with cancer cases, PSA testing numbers were converted to PSA testing rates per 100,000 men for each year using state-level population data from The Australia Bureau of Statistics as published in 2016. Prostate cancer incidence and PSA testing rates were then used to calculate prostate cancer cases per 10,000 PSA tests, to measure the efficiency of testing in each state. Cases per test may be overestimated where cases are detected without prior PSA.

Patients were grouped into three age brackets (<45 years, 45-74, and 75+ years), to reflect that PSA testing guidelines and the likelihood of prostate cancer differ with age. The numbers for <45 years old age group are small and therefore were not included in the main analysis.

Staging of cases would help identify whether men were being diagnosed later, but this was excluded from the analysis and staging data was only available for NSW.

2.2.2. Interrupted time series analysis

The response to the guidelines in different states were compared through cancer detection rates, PSA testing rates, and cases detected per PSA across time between Victoria and other states using comparative interrupted time series analyses (ITSA). An ITSA compares a treatment state to comparison states over time and with respect to a time cut-off. ITSA has previously been used to assess the impact of interventions to change provider prescribing and reporting behaviour (Ansari et al., 2003; Chhapola, Tiwari, Brar, & Kanwal, 2016), including the clinician response to guidelines (Curtis, Walker, & Goldacre, 2018; Dickson et al., 2017). In the analyses, 2009 was chosen as the cut-off, as it represents the first guideline change after 2002 that introduced an additional testing step (DRE), which may discourage unnecessary testing. 2009 was also the year that international policies suggested evidence was not strong enough to recommend opportunistic PSA testing. The GP guidelines did not explicitly mention the international policies, but did cite similar sources of evidence used to develop the guidelines.

In this analysis, Victoria was compared to the average of all states with no prostate cancer registry during the period of interest (the comparison group). No single state was identified as being similar enough to Victoria in testing and detection in pairwise comparisons prior to 2009.

Some age groups may be more affected by policy changes than others: men aged between 45 and 74 years will be most affected by the changes, as most diagnoses are made at these ages, and men over 75 are likely to be detected in different ways.

For each outcome, Victoria was compared to the other states using a Stata program developed by Linden (2015) from the previously defined interrupted time series formula (Linden, 2015; Simonton, 1977, 1979):

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 X_t T_t + \beta_4 Z + \beta_5 Z T_t + \beta_6 Z X_t + \beta_7 Z X_t T_t + \epsilon_t \quad (1.1)$$

Where T=year, X= cut-off (0 if prior to cut-off (2009), 1 post cut-off (2009 onwards)), Z= state (0=average of comparison group [states other than Victoria], 1=VIC). Y is the outcome variable, which in the first instance will be incidence

Table 1 gives outcomes of interest from the ITS analyses that explore the states' trends in each time-period and the comparison across states or time-periods. Trends are compared prior to and from 2009 and between Victoria and the other states. Because Victoria could also differ from the other states prior to 2009, the overall trend changes pre-2009 to post-2009 for Victoria to the other states were not compared. One of the reasons a difference-in-difference analysis is not employed is that trends pre-2009 were not parallel. The results of equation (1.1) were presented visually in the results section, and described in full in Table 2.

Also estimated were the average absolute rates of cases and testing in the population (and cases per test) to determine how the trends affect the population over time.

Table 1 Formulae for outcomes of interest from interrupted time series analysis

Outcome	Formula
Pre-2009 trend: NSW, QLD, WA	β_1
Pre-2009 trend: VIC	$\beta_1 + \beta_5$
Difference in trends pre-2009: VIC vs NSW, QLD, WA	β_5
Post-2009 trend: NSW, QLD, WA	$\beta_1 + \beta_3$
Post-2009 trend: VIC	$\beta_1 + \beta_3 + \beta_5 + \beta_7$
Difference in trends post-2009 VIC vs NSW, QLD, WA	$\beta_5 + \beta_7$
Difference in trends pre and post-2009: NSW, QLD, WA	β_3
Difference in trends pre and post-2009: VIC	$\beta_3 + \beta_7$
Difference between VIC and NSW, QLD, WA trends pre- and post-2009	β_7

Based on (Linden, 2017)

2.2.3. Robustness analyses

Several GP guideline changes occur between 2002 and 2016: 2005 first introduced the requirement to inform men of the disadvantages of PSA testing, 2009 introduced

the DRE and 2012 no longer recommended opportunistic testing in asymptomatic men. These additional time cut-offs explored the assumption that 2009 is the primary time period of change. Because two of the time periods (2002-2005 and 2009-2012) only contain three time points, where one outlier can completely alter the direction of the trend, general trends are presented compared to the base case analysis, rather than quantified trendlines.

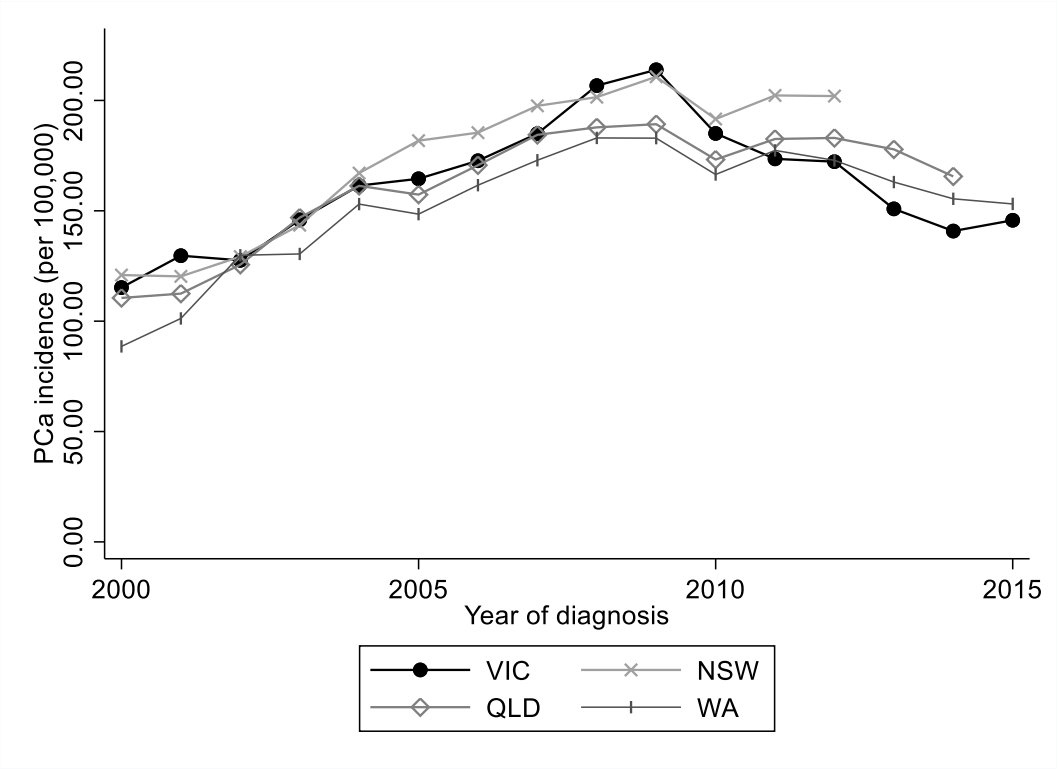
In younger patients (<45 years old), no change in trends of diagnoses per test or biopsies per PSA tests is expected as these are a high-risk group where only symptomatic testing occurs. The guidelines have also remained relatively unchanged for this group, expanding family history requirements in 2009. Absolute rates of cases and testing are also expected to be much lower in this group (Royal Australian College of General Practitioners, 2009).

2.3. Results

2.3.1. Prostate cancer incidence

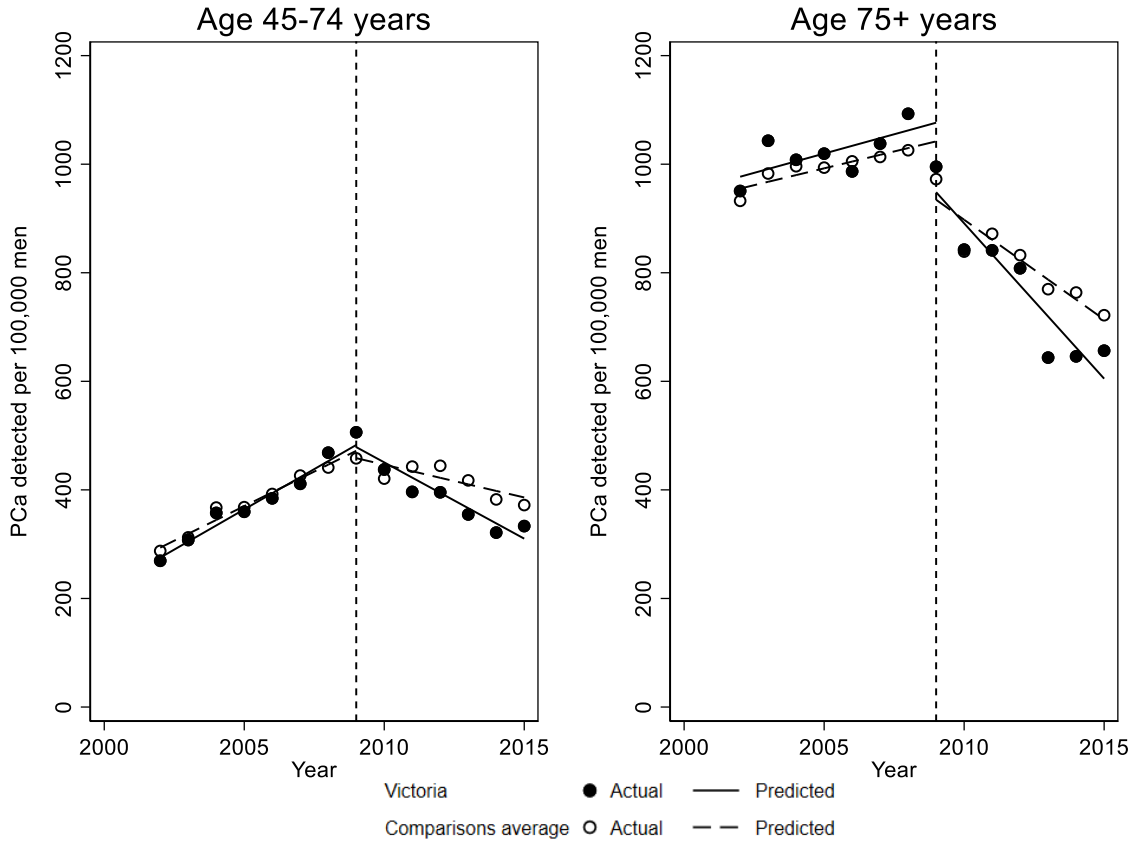
Figure 3 shows the crude incidence rate of prostate cancer over time in Victoria (VIC), New South Wales (NSW), Queensland (QLD) and Western Australia (WA) unadjusted for age. In all four states, an upward trend in prostate cancer incidence from 2000 resulted in an additional 50-100 prostate cancer detected per 100,000 men by 2009. From 2009 prostate cancer incidence declined for all states, with the sharpest decline seen in Victoria, where the reduction appears similar in magnitude to increase in prostate cancer incidence prior to 2009. Indeed, the other states appeared to be trending towards constant prostate cancer incidences compared to the large reduction seen in Victoria. The increase in prostate cancer appeared to slow earlier for QLD and WA, with incidence rates similar for both 2008 and 2009. It was unclear why this was the case, but may have been an early response to the developing evidence around PSA testing guidelines. Total case numbers in QLD and WA were highest in 2009 (3,958 QLD, 2,004 WA), so incidence may also reflect a fluctuation in population numbers around this time.

Figure 3 Prostate cancer incidence by state



Notes: PCa= prostate cancer incidence; NSW= New South Wales; QLD= Queensland; VIC= Victoria; WA= Western Australia

Figure 4 Estimated prostate cancer cases detected per 100,000 men, over time by age group, VIC vs comparison states (NSW, QLD, WA)



Notes: PCa= prostate cancer incidence; NSW= New South Wales; QLD= Queensland; VIC= Victoria; WA= Western Australia

Figure 4 onwards were produced using the ITSA formula in equation (1.1) and the full results that correspond to Table 1 are presented in Table 2, p51. Figure 4 shows the graphical result of the ITS analysis for prostate cancer cases detected per 100,000 men (predicted incidence). The trends for NSW, QLD and WA were calculated from their average predicted incidence at each time point. Prior to 2009, all states had similar increases in cases detected, with an additional average 25.5 cases per 100,000 men aged 45-74 diagnosed each year for NSW, QLD and WA; and 29.6 for Victoria. A smaller increase was seen in men aged 75 and over, 12.4 additional diagnoses per 100,000 men each year on average in NSW, QLD and WA; 14.2 in Victoria. Absolute prostate cancer detection rates were much higher for men 75 years old and above in this period (prostate cancer was detected in around 1 in 100 men 75 and over, and

varied between 1 in 400 and 1 in 200 for those aged 45-74 years). For both age groups trends in prostate cancer detection rates were not statistically significantly different between Victoria and the other states from 2002 to 2009 (p>0.10).

From 2009 onwards, all states and age groups saw a reduction in prostate cancer detection rates. For men aged 45-74 years a change in prostate cancer detection of -12.1 (95%CI -17.7, -6.5) cases per 100,000 men each year was estimated for the comparison states (NSW, QLD, WA) and -28.1 (95%CI -39.3, -16.9) in Victoria. These reductions were significantly different from the increases in prostate cancer cases incidence recorded from 2002 to 2009. Victoria's reduction was significantly faster than the other states, with cases changing by an additional -15.7 (95%CI -27.9, -3.55) cases per 100,000 men per year than the comparison states. Absolute estimates of prostate cancer remained around 1 case per 250 men in the comparison states and dropped from 1 case per 200 men to less than 1 case per 300 men in Victoria over the same time period. For men aged 75 and over, there was a significant reduction in prostate cancer incidence between 2009 and 2015 with absolute incidence estimates dropping as low as 3 cases per 400 men in the comparison states and 3 cases per 500 men in Victoria. These trends translated to -36.8 (95%CI -50.3, -23.2) cases per 100,000 men per year for the comparison states and -57.2 (95%CI -76.9, -37.6) for Victoria. These reductions equated to an additional -20.5 cases per 100,000 men per year for Victoria compared to the other states, significant at a 10% threshold.

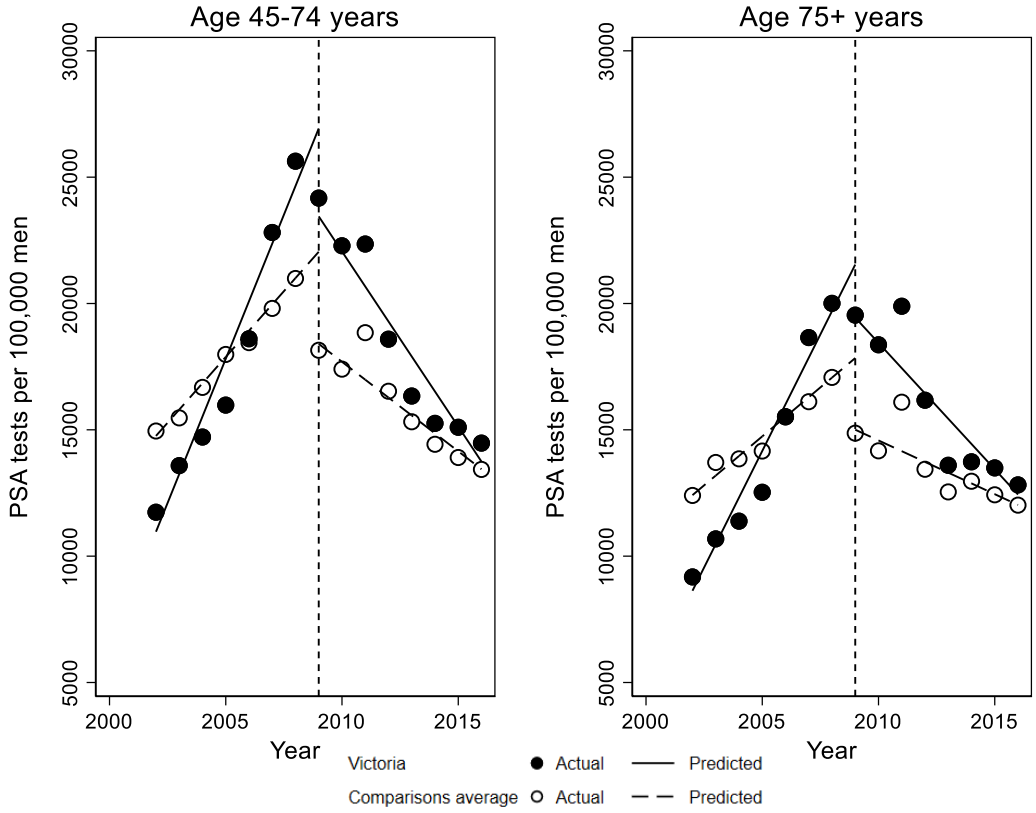
2.3.2. Prostate specific antigen testing

Figure 5 shows the ITS analysis result for the number of PSA tests (Medicare item 66655) per 100,000 men over time. As item 66655 allows only one PSA test per man per year, the PSA testing incidence should represent the number of men being tested each year (i.e. no repeat tests should be included). Full results of the trendlines are given in Table 2, p51. Prior to 2009, all states had increases in the rate of PSA tests conducted, with an additional 1,041 (95%CI 569, 1514) tests conducted per 100,000 men aged 45-74 each year for NSW, QLD and WA; and 2,280 (95%CI 1818, 2742) for Victoria. A slower but similar increase was seen in men aged 75 and over (775 additional tests per 100,000 men each year on average in NSW, QLD and WA; 1,844 in Victoria). Absolute rates of PSA testing were also similar across the age groups in

this period. Roughly 1 test per 10 men at the beginning of the period up to 1 test in every 4 men by 2009. For both age groups, trends in PSA testing rates were statistically significantly different between Victoria and the other states from 2002 to 2009, with Victoria increasing PSA testing by >1,000 PSA tests per 100,000 men every year compared to the comparison states ($p>0.10$).

From 2009 onwards, all states and age groups saw a reduction in PSA testing rates. For men aged 45-74 years the annual change in PSA tests was -1,386 (95%CI -1700, -1,073) compared to -708 (95%CI -1,119, -298) per 100,000 men in other states (NSW, QLD, WA). Compared to the increase in prostate cancer cases incidence recorded from 2002 to 2009 in all states, Victoria's reduction was significantly faster than the other states by -678 (95%CI -1,195, -162) tests per 100,000 men per year. The rate of PSA testing was similar in Victoria and the other states by 2015, with around 1 PSA test per 8 men for both men aged 45-74 and men aged 75 and over. For men aged 75 years and over, this translated to a change of -426 (95%CI -805, -46) PSA tests per 100,000 men per year for the comparison states, and -1,000 (95%CI -1,244, -756) for Victoria. This faster reduction in Victoria meant the trend saw 575 fewer tests per 100,000 men per year compared to the other states, significant at a 1% threshold.

Figure 5 Estimated rates of PSA testing (per 100,000 men) by age group, Victoria vs comparison states (NSW, QLD, WA)



2.3.3. Cases per test

Figure 6 shows the graphical result of the ITS analysis for the number of prostate cancers detected per 10,000 PSA tests (Medicare item 66655) over time. Full results of the trendlines are given in Table 2, p51. Prior to 2009 Victoria saw a statistically significant decline in prostate cancer cases detected per 10,000 tests per year for both men between 45 and 74 (-9.36 cases per 10,000 tests per year, 95%CI -14.3, -4.43) and men aged 75 and over (-87.6 cases detected per 10,000 tests per year, 95%CI -109, -66.6). By comparison, the other states saw a statistically non-significant increase in cases per 10,000 tests per year for men aged 45-74 (+2.03 cases per 10,000 tests per year, 95% CI -2.81, 6.87) and a slower, statistically significant decline in cases detected per 10,000 tests per year in men aged 75 and over (-26.8, 95%CI -43.1, -10.5). Absolute estimates of cancer detected per 10,000 tests were similar prior to 2009 for men aged 45-74 years (reducing from 1 in 40 to 1 in 60 in Victoria between

2002 and 2009 and staying around 1 in 50 for the other states). For men aged 75 years and over cases detected per 10,000 PSA tests were higher in Victoria (approx. 1 case per 10 tests) than the other states (approx. 1 case per 13 tests) in 2002 and reduced to 1 case detected per 25 tests in Victoria and 1 case in 17 tests in the other states by 2009.

From 2009 onwards, all states and age groups saw no statistically significant trends in the rate of prostate cancer detected per 10,000 PSA tests. For men aged between 45 and 74 years old both Victoria and the other states saw non-significant increases in prostate cancer cases detected per 10,000 PSA tests per year: an increase of 2.34 cases per 10,000 tests per year (95% CI -2.00, 6.68) in Victoria, 1.09 (95%CI 9.54, 11.7) in the comparison states. For men aged 75 and over the rate of reduction in cases per 10,000 PSA tests slowed in both Victoria and the other states: -0.4 (95%CI -13.5, 12.7) cases per 10,000 PSA tests per year in Victoria and -18.1 (95%CI -40.7, 4.4) cases per 10,000 tests per year in the other states. The trend in Victoria from 2009 was not significantly different from the other states in either age group. Overall, Victoria saw statistically significant changes in prostate cancer cases detected per 10,000 tests per year when pre- and post-2009 time periods were compared. The other states did not see a significant change in trend before and after 2009. Absolute prostate cancer rates detected per 10,000 tests were higher in the other states than Victoria (1 case in 30 tests compared to 1 case in every 50 tests) for men aged 45-74 years from 2009. For men aged 75 years and over, absolute rates of prostate cancer detection per 10,000 PSA tests differed between the states in 2009 with 1 case per 20 tests for Victoria compared to around 1 case per 15 tests for the other states. By 2016 Victoria remained at about 1 case per 20 tests and the other states reduced to 1 case per 18 tests.

Figure 6 Estimated rates of prostate cancer detection per 10,000 PSA tests by age group, VIC vs comparison states (NSW, QLD, WA)

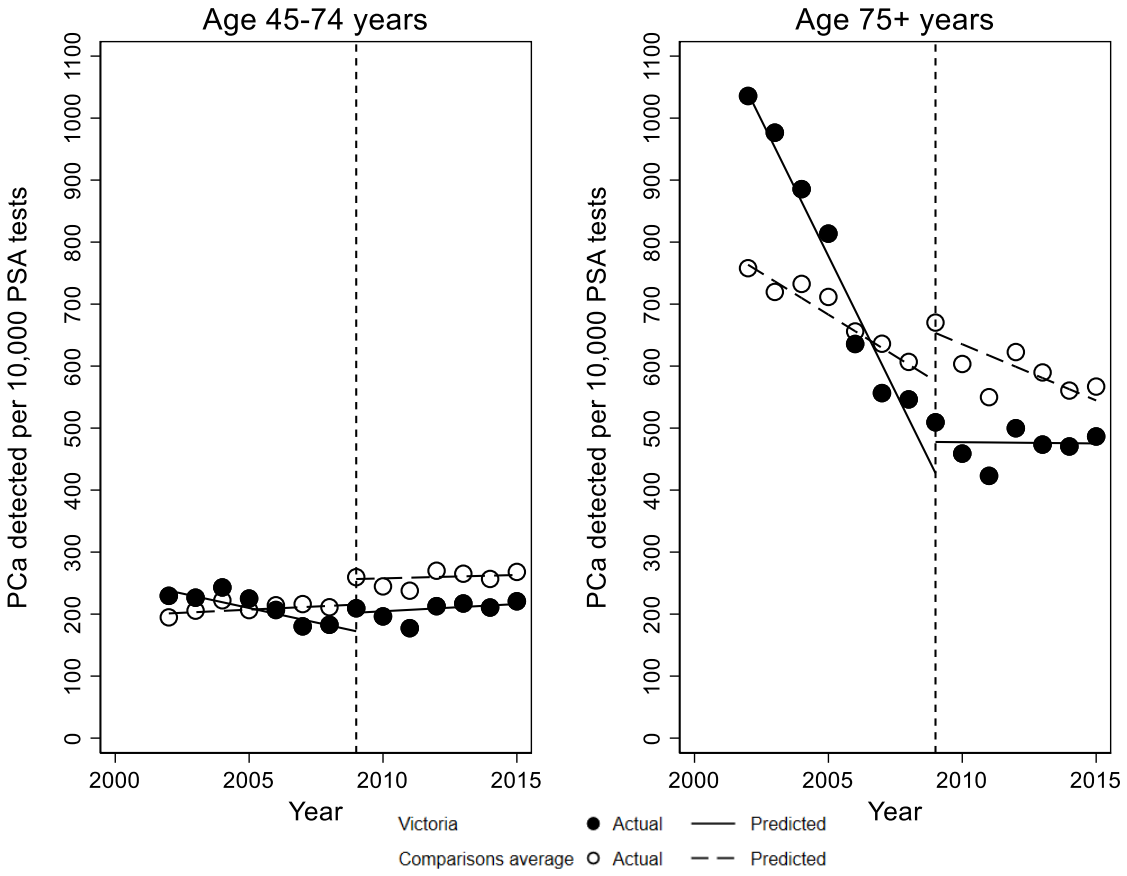


Table 2 Comparative interrupted time-series analyses VIC versus comparison states (NSW, QLD, WA), stratified by age. Intervention year 2009

Analysis: VIC versus states with no registry (NSW, QLD, WA)	Measure of interest	Age group (years)	
		Point estimate [95% CI]	
		45-74	75+
Prostate cancer cases detected per 100,000 men	Pre-2009 trend: NSW, QLD, WA	25.5*** [21.2, 29.8]	12.4** [2.8, 22.0]
	Pre-2009 trend: VIC	29.6*** [23.7, 35.6]	14.2** [0.78, 27.7]
	Difference in trends pre-2009: VIC vs NSW, QLD, WA	4.2 [-3.1, 11.6]	1.79 [-14.7, 18.3]
	Post-2009 trend: NSW, QLD, WA	-12.1*** [-17.7, -6.5]	-36.8*** [-50.3, -23.2]
	Post-2009 trend: VIC	-28.1*** [-39.3, -16.9]	-57.2*** [-76.9, -37.6]
	Difference in trends post-2009 VIC vs NSW, QLD, WA	-15.7** [-27.9, -3.55]	-20.5* [-44.4, 3.4]
	Difference in trends pre and post-2009: NSW, QLD, WA	-37.5*** [-44.9, -30.2]	-49.2*** [-65.8, -32.6]
	Difference in trends pre and post-2009: VIC	-57.8*** [-71.0, -44.5]	-71.5*** [-95.4, -47.6]
	Difference between VIC and NSW, QLD, WA trends pre- and post- 2009	-20.2** [-35.4, -5.1]	-22.3 [-51.4, 6.8]

Analysis: VIC versus states with no registry (NSW, QLD, WA)	Measure of interest	Age group (years) Point estimate [95% CI] 45-74	75+
PSA tests per 100,000 men	Pre-2009 trend: NSW, QLD, WA	1041.0*** [568.5,1513.5]	775.2*** [449.5, 1100.8]
	Pre-2009 trend: VIC	2280.1*** [1818.2, 2742.0]	1844.1*** [1505.2, 2183.0]
	Difference in trends pre-2009: VIC vs NSW, QLD, WA	1239.2*** [578.4, 1899.9]	1068.9*** [598.9, 1538.9]
	Post-2009 trend: NSW, QLD, WA	-708.2*** [-1118.5, -297.8]	-425.5** [-805.0, -45.5]
	Post-2009 trend: VIC	-1386.2*** [-1699.8, -1072.7]	-1000.1*** [-1244.3, -756.1]
	Difference in trends post-2009 VIC vs NSW, QLD, WA	-678.0** [-1194.5, -161.6]	-574.9** [-1026.4, -123.5]
	Difference in trends pre and post-2009: NSW, QLD, WA	-1749.2*** [-2385.8, -1112.5]	-1200.4*** [-1720.9, -679.9]
	Difference in trends pre and post-2009: VIC	-3666.4*** [-4329.7, -3003.0]	-2844.2*** [-3329.7, -2358.8]
	Difference between VIC and NSW, QLD, WA trends pre- and post- 2009	-1917.2*** [-2836.7, -997.8]	-1643.8*** [-2355.6, -932.0]
Prostate cancer per 10,000 PSA tests	Pre-2009 trend: NSW, QLD, WA	2.03 [-2.81, 6.87]	-26.8*** [-43.1, -10.5]
	Pre-2009 trend: VIC	-9.36*** [-14.3, -4.43]	-87.6*** [-109, -66.6]
	Difference in trends pre-2009: VIC vs NSW, QLD, WA	-11.4*** [-18.3, -4.48]	-60.8*** [-87.3, -34.2]
	Post-2009 trend: NSW, QLD, WA	1.09 [-9.54, 11.7]	-18.1 [-40.7, 4.41]
	Post-2009 trend: VIC	2.34 [-2.00, 6.68]	-0.397 [-13.5, 12.7]
	Difference in trends post-2009 VIC vs NSW, QLD, WA	1.25 [-10.2, 12.7]	17.7 [-8.32, 43.8]
	Difference in trends pre and post-2009: NSW, QLD, WA	-0.940 [-12.3, 10.4]	8.67 [-19.0, 36.3]
	Difference in trends pre and post-2009: VIC	11.7*** [4.21, 19.2]	87.2*** [55.6, 119]
	Difference between VIC and NSW, QLD, WA trends pre- and post- 2009	12.6* [-0.992, 26.3]	78.5*** [36.5, 120]

Notes: *p<0.10, **p<0.05, ***p<0.01
Key: PSA=prostate specific antigen; NSW=New South Wales; QLD= Queensland; WA= Western Australia; VIC= Victoria

2.3.4. Robustness Analyses

2.3.4.1. Additional time cut-offs

The ITS analyses of prostate cancer detection rates, PSA testing rates and cases per PSA tests are rerun with additional time cut-offs in 2005 and 2012 and presented in Appendix 7.2, p134. The shortening of the time periods makes the ITS analyses less likely to be accurate. Both in Victoria and the other states, prostate cancer incidence increased in both the period between 2002 and 2005 and between 2005 and 2012. The increasing trend is similar across time periods. Prostate cancer incidence also decreases from 2009 (both 2009 to 2012 and 2012 to 2016) for both Victoria and the comparison states. PSA testing rates provided a similar pattern (increasing prior to 2009 for both time periods and both Victoria and the comparison states) and reducing after 2009, with a negative trend occurring in Victoria and the comparison states from 2012. The trend in PSA testing rates from 2009-2012 appears to be primarily affected by a high rate of testing in 2011 compared to 2009-2010. The trends in cases per 10,000 PSA tests appear to differ with each time period, but as with prostate cancer incidence and PSA testing rates, 2009 seems to be cut-off where the most significant change occurs.

2.3.4.2. Men aged <45 years old

The ITS analyses of prostate cancer detection rates, PSA testing rates and cases per PSA tests for men aged below 45 years old are presented in Appendix 7.3, p137. In summary, absolute numbers of incidence and testing rates are much lower than for older patients, with prostate cancer detected less than 1 case per 50,000 men, PSA testing rates between 1 and 3 tests per every 200 men, and prostate cancer detected per 10,000 PSA tests ranged from 1 to 5 cases in every 2,000 tests between 2002 and 2016.

Over time, an increase in prostate cancer detection rates was seen for both Victoria and the other states, slowing the increasing trend after 2009 (more so for the other states than Victoria). PSA testing in men below 45 saw similar testing patterns prior to

and after 2009 to the other age groups. PSA testing rates increased prior to 2009 and decreased from 2009 in all states with similar trends and absolute testing rates in both Victoria and the comparison states. This similarity in trends of the time periods suggests that expansion of the family history criteria in the guidelines did not result in more tests for this group.

Prostate cancer incidence per 10,000 PSA tests increased after 2009 for all states, in absolute terms for the comparison states (where the trend remained relatively constant) and increasing in both absolute and trend terms for Victoria.

2.4. Discussion

We have compared patterns of behaviour in Victoria to other states recording prostate cancer data. There is some evidence that cancer detection rates for Victoria reduced faster than the other states after 2009, and some evidence that Victoria had faster growth in PSA testing prior to 2009 when GP test guidelines became more restrictive, and faster reductions in testing rates after the guideline change compared to the other states. In both Victoria and the other states increased PSA testing was associated with increased prostate cancer detected. The fall in PSA testing from 2009 was associated with a decrease in prostate cancer detected. However, the rate of prostate cancer detected per PSA test differed across the states. In the comparison states, the rates of prostate cancer detected per 10,000 PSA tests per year remained relatively constant for men aged 45-74 and decreased for men aged 75 and over across both time periods, suggesting that the reduction in PSA tests did not improve the efficiency of the testing (and in the over 75s testing efficiency continued to worsen). In Victoria, trends in prostate cancer detected per 10,000 PSA tests were negative in both age groups prior to 2009 (suggesting test efficiency worsened over time). However, there was some evidence of a flattening from 2009 suggesting some efficiency improvements. Absolute testing rates and rates of cases per test appeared to be converging between Victoria and the other states from 2009 to 2016, which suggests a reduction in the variation in testing and efficiency across states that were seen prior to 2009.

Comparing additional time cut-offs at the introduction of other guidelines confirmed that 2009 was when the most significant change in testing and incidence occurred.

The robustness analysis for men aged under 45 showed similar trends in PSA testing for both Victoria and the other states (with an increase in tests prior to 2009 and a decrease in tests from 2009) and Victoria saw increased prostate cancer rates between 2002 and 2016 (the other states and increase from 2002 to 2009 and a constant incidence from 2009). For men aged under 45 there appeared to be an improvement in absolute efficiency (increased cases per test) from 2009 for all states, suggesting a there has been a reduction in unnecessary testing since 2009, particularly for Victoria.

The difference in PSA testing rates for Victoria both prior to and post-2009 for all age groups suggests Victoria has different testing practices to the other states. There are a few reasons that Victoria may differ from the states before and after 2009. One difference may be a difference in the downstream/additional testing prior to diagnosis. The more considerable reduction in prostate cancer incidence from 2009 may have resulted from a more significant reduction in false-positive tests in the confirmation tests. Additional testing can include further PSA tests, compliance with DRE required in the guidance, and the number of biopsies conducted to confirm prostate cancer diagnosis. These tests are often hard to isolate to prostate cancer diagnosis in the MBS data, and therefore the differences in downstream tests across states were not explored.

A further reason for Victoria's lower prostate cancer incidence may be a lower average population risk of prostate cancer or a lower rate of symptomatic prostate cancer. Family history and staging of prostate cancer data are not available for all diagnosed men, or at the population level. These could indicate prostate cancer risk and the likelihood of symptoms, as later-stage prostate cancer is likely to result in more symptoms.

This analysis uses an interrupted time series approach, which allows for comparing longitudinal data from different states over the same time periods and exploring how they respond to the same interventions.

The ITS is unadjusted for population characteristics, and there are differences between the states prior to the chosen time cut-off that might confound any causal inference. Moreover, the analysis produces an average across the other states, assuming each state's weight is equal and that these states can be combined. Other major clinical changes or international guidelines would also undermine the validity of the ITSA, although it is not clear how states may have responded to these differently. For these reasons, there are no strong claims about the impact of registry in Victoria on prostate cancer diagnosis and treatment.

In conclusion, this approach helps describe how Victoria differs from other states regarding prostate cancer diagnosis. There is some evidence that guidelines have reduced PSA testing for all states, and testing efficiency appears to have improved or stopped worsening in Victoria. There is some evidence that Victoria had different prostate cancer testing practices from the other states prior to 2009. This means the difference between Victoria and the other states after the guideline change cannot be isolated to the guideline change in 2009.

3. How does feedback from a clinical quality registry impact the overtreatment of patients?

3.1. Introduction

Clinical quality registries can be used to identify variation in treatments and outcomes across institutions and clinicians and provide feedback on performance that potentially motivates improvements in care quality. In several clinical areas reporting and feedback have been shown to influence prescribing behaviours, primarily in general practitioners (Avery et al., 2010; Naughton, Feely, & Bennett, 2009; Soleymani et al., 2012; Steele, Bess, Franse, & Graber, 1989) and registries have been associated with improvements in clinical or patient outcomes, (Dinh et al., 2015; Ruseckaite et al., 2016; Stey et al., 2015). The current study aims to evaluate the impact of clinician feedback on the rates of intensive treatments for localised prostate cancer. Treatment rates are relevant for exploring the role of feedback as guidelines have changed over the last decade to strongly advise against the surgical treatment of low-risk localised prostate cancer.

PCOR-Vic was the first clinical quality registry for localised prostate cancer established in Australia that conformed to the Australian Commission on Safety and Quality in Health Care standards (The Australian Commission on Safety and Quality in Health Care, 2016). The systematic collection of localised prostate cancer patient data on PCOR-Vic, allows for variation in care between hospitals and clinicians to be identified. As a result, the registry can provide feedback to both hospitals and clinicians to improve treatment decisions and clinical and patient outcomes. Currently, individual clinician feedback is only offered to urologists, who are the specialists who most frequently manage patients with prostate cancer. Implementing this feedback may have changed practice over time, specifically if it may have changed treatment decisions.

Improving outcomes for prostate cancer patients is of great value, as prostate cancer is the most diagnosed cancer in men in Australia. An estimated 1 in 7 men is diagnosed with prostate cancer by their 85th birthday (Australian Institute of Health and Welfare, 2016). In Victoria, localised prostate cancer accounts for 70-80% of all

prostate cancer diagnoses each year (73% in 2014; data provided by PCOR-Vic July 2017).

Men with localised prostate cancer can be managed in several ways, with guidelines recommending curative treatments such as prostatectomy or radiotherapy for higher-risk patients. Non-curative management is recommended for patients where curative treatment may not be appropriate. For example, active surveillance is recommended for men with low-risk prostate cancer, and watchful waiting for men with life expectancies shorter than ten years (Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel., 2016). Curative treatment for prostate cancer is associated with erectile dysfunction (20–70%) and urinary incontinence (15–30%) (Royal Australian College of General Practitioners, 2018). These potential complications have contributed to non-curative management recommendations in appropriate populations, mainly low-risk patients (Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015). The population of patients who receive treatment when recommendations advise non-curative management are considered to be overtreated. Non-curative management may also come with side effects, particularly with regard patient psychological well-being, and may be a reason for low-risk men to undergo treatment even against recommendations (Chamie et al., 2015; Eredics, Dorfinger, Kramer, Ponholzer, & Madersbacher, 2017; Hefermehl, Disteldorf, & Lehmann, 2016; Lang et al., 2017; Loeb et al., 2015).

Previous economic analyses of localised prostate cancer have shown that surveillance in low-risk men can be a cost-effective strategy for treatment in Australia and internationally (Gordon et al., 2018; Hayes et al., 2013; Keegan, Dall'Era, Durbin-Johnson, & Evans, 2012; Koerber, Waidelich, Stollenwerk, & Rogowski, 2014; Prostate Cancer Foundation of Australia & Griffith University, 2016). One previous economic analysis of PCOR-Vic in 2016 estimated a 21% increase in the uptake of active surveillance in low-risk men associated with the registry's introduction (The Australian Commission on Safety and Quality in Health Care, 2016). A previous time series analysis using the PCOR-Vic showed statistically significant reductions in active treatment for low-risk patients (reduction of 9%) and a non-statistically significant increase in high-risk patients receiving treatment in the 1st year (an increase of 5%)

between January 2009 and December 2013 (Sampurno et al., 2016). It is unknown to what extent the change in treatment decisions in these previous analyses was due to individual feedback to urologists.

As the PCOR-Vic is the first clinical quality registry for localised prostate cancer in Australia providing individual feedback to urologists, no studies have yet assessed the effectiveness of the individualised feedback. Therefore, this study aimed to use the registry data to how feedback changed the treatment decisions for patients.

Urologists received feedback in 2 key areas related to management decisions:

1. low-risk patients who receive curative treatment;
2. intermediate and high-risk patients who receive curative treatment in the 12 months following diagnosis;

Though the feedback does not explicitly refer to the non-curative management low-risk patients get, low-risk patients are defined as patients meeting the PRIAS criteria, which recommends active surveillance (regular surveillance with intention to perform curative treatment in the future).

This analysis focuses on how these clinical quality indicators affect probability of curative treatment (radical prostatectomy, radiotherapy or brachytherapy) following diagnosis. The methods used also allow for exploration of time to treatment in the first year following diagnosis.

In line with previous evidence on the impact of feedback in other clinical areas (Avery et al., 2010; Naughton et al., 2009; Soleymani et al., 2012; Steele et al., 1989), if the registry is effective in its feedback there should be improved adherence to treatment guidelines, and therefore:

- an increase in non-active management in low-risk patients (e.g., active surveillance, AS), increasing time to curative treatment (treatment is deferred to a later date); and

- an increase active-management in intermediate and high-risk patients (e.g., surgery or radiotherapy), decreasing time to curative treatment (higher rates of treatment occur within the year following diagnosis); and
- a smaller difference in the proportion of treated intermediate-risk patients as guidance on whether they should have active management is less clear.

The next section outlines the registry data, and the nature of the feedback received, along with the analytical methods used to estimate the impact of the feedback on treatment uptake.

3.2. Methods

3.2.1. Data

PCOR-Vic individual-level data were restricted to patients diagnosed between August 2008 and January 2018. Information collected included age, National Comprehensive Cancer Network (NCCN) risk category at diagnosis (Table 3), diagnosis date, de-identified information on diagnosing clinician and hospital, and first treatment received, including information regarding the treating clinician and hospital. Patient postcode information was linked to SEIFA data from census data collected in 2016 to approximate socioeconomic status (Australian Bureau of Statistics, 2018).

Table 3 NCCN risk categories

NCCN risk	PSA level (ng/ml)	Gleason Score	T staging
Low	<10	2-6	T1, T2 (subcategories T1a, T1b, T1c, T2a)
Intermediate	10-20	7	T2 (subcategories T2b, T2c)
High	>20	8-10	T3, T4* (subcategories T3a, T3b*)

Notes: * If T staging is T3b and T4, patients are very high risk/locally advanced but are combined here with high-risk patients. Low-risk patients must fulfil all criteria, whereas intermediate and high risk may fulfil only one. NCCN risk definitions have remained consistent in registry and feedback for the period of the analysis.
Key: NCCN=National Comprehensive Cancer Network; PSA=prostate specific antigen
Source: (Mohler et al., 2012)

Since December 2012, participating urologists could receive individualised feedback, with regular six-monthly reports from February 2015. Across the whole registry, most urologists (41) began receiving reports in 2012-2013, but others began receiving feedback from 2014 onwards (8 in 2015, 8 in 2016, 10 in 2017, 2 in 2018) . To receive feedback, they must have diagnosed or treated at least ten patients between report. Clinician speciality is not listed within the registry. Clinicians were excluded who never performed surgeries or who had fewer than five patients, as these were not likely to be clinicians the feedback targeted. Both the feedback and analysis assumed that the clinician who managed the patient (either with curative or non-curative management) was ultimately responsible for the management decision. Registry experts determined that it was a generally held view that the ultimate responsibility falls to the managing clinician when patients are low-risk, and the diagnosing clinician when patients are intermediate or high-risk (personal communication). If a separate managing clinician was not recorded, diagnosing clinician was assumed to be the managing clinician. For most included patients (60%), their diagnosing clinician is their managing clinician. Diagnosing clinicians are usually defined as the specialist who confirms prostate cancer diagnosis in patients the GP refers. Frequently this specialist is a urologist. Patients may change the clinician due to their clinician or hospital's expertise or availability, cost of treatment, or patient preference. Another factor to consider is that feedback to clinicians and hospitals may have an unintended consequence of increasing or decreasing patients' movement between clinicians; for example, an increase in active surveillance recommendations may have reduced the need for referrals.

Individual clinician level feedback allows the opportunity to identify the added value of providing feedback to clinicians. The value of feedback will produce a conservative estimate of the overall registry effect as it only considers one component of the value of the PCOR-Vic. Feedback presented to clinicians aims to encourage two directions of behaviour: increased active treatment for high-risk patients and decreased active treatment (overtreatment) for low-risk patients. Low-risk patients are a primary focus of feedback: these patients are likely to benefit most from treatment avoidance, and national and international guidelines reinforce the feedback for this group.

There are two broad treatment groups in the first year after diagnosis: patients who receive surveillance (AS or watchful waiting, WW) and non-curative treatment (androgen deprivation therapy, ADT; chemotherapy); and patients who receive curative treatment (surgery, radiotherapy or brachytherapy). For higher-risk patients, clinicians received information on which patients did not receive active treatment and the proportion of patients who received active treatment within a year following diagnosis. Clinicians also received information on the proportion of radical prostatectomy patients who meet PRIAS criteria for active surveillance. Therefore, the feedback for the treatment of low and higher risk patients is framed slightly differently within the reports, such that low risk they receive information on who was treated and high risk they receive information on who was not treated.

Within the context of survival analysis, the analysis modelled time to curative treatment for patients. According to guidelines and the clinical quality indicators reported by the registry, patients who are of low risk should receive active surveillance until their cancer progresses; and patients who are of higher risk should receive curative treatment within the first year (Australian Cancer Network Management of Metastatic Prostate Cancer Working Party, 2010; Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015). Some patients may receive adjuvant ADT in advance of curative treatment (expert opinion, ME). Robustness analysis considered the effect of including ADT in the curative treatment group.

Inclusion flow diagrams for the analysis are given in Figure 7 and Figure 8. Patients were excluded if they were diagnosed before the establishment of the registry, after death or treatment, or had less than a year of data available. They were also excluded if data were missing because the patient did not consent, treatment status was unknown, diagnosing clinician and hospital were unknown, or patients were diagnosed or treated outside of Victoria. The patients were then divided into subgroups based on their risk category, and clinicians with <5 patients were removed. Clinicians with low numbers were expected to be outliers, less likely to be urologists, and therefore not the target group for feedback. A total of 11,125 patients were included across the three risk groups.

3.2.2. Analytical approach

This analysis assessed the effect of feedback on whether a patient receives treatment and the timeliness of that treatment. A survival analysis approach was used to explore time to curative treatment according to whether the patient's managing clinician received feedback in the year following diagnosis. Each NCCN risk level was modelled separately to account for the feedback and the other covariates' different expectations. Patients were censored if they died. The analysis was restricted to one year following diagnosis as this is the most critical treatment period for high-risk patients, and has the most accurately recorded data. Feedback was time varying, such that a urologist could move from no feedback to feedback according to the analysis time. All urologists had periods with and without feedback and feedback status was determined by the date of the earliest report received by the urologist.

Initially, Kaplan-Meier graphs are used to compare the time to curative treatment for each patient by the status of their urologist's feedback. Cox proportional hazard models are then used to explore the effect of feedback after controlling for time-varying and time-invariant patient, hospital and managing clinician (urologist) characteristics, similar to (Jayadevappa et al., 2011). Proportional hazard assumption testing is conducted using the Schoenfeld residuals to identify any variables where the proportional hazards assumption does not hold. Variables that violate the proportional hazard assumption are controlled through stratification instead. The following model is estimated:

$$\lambda(t|z) = \lambda 0(t) e^{F_j\beta_1 + Age_i\beta_2 + SEIFA_i\beta_3 + PSA_i\beta_4 + cT2_i\beta_5 + Gleason_i\beta_6 + Metro_k\beta_7 + Private_k\beta_8 + Y_i\beta_9 + u_j} \quad (2.1)$$

where the time-invariant variables are: patient (i)'s age at diagnosis (Age_i); SEIFA decile ($SEIFA_i$); patient PSA (PSA_i), cT2 ($cT2_i$) and Gleason score ($Gleason_i$) at diagnosis; managing hospital (k) characteristics (metropolitan versus regional [$Metro_k$], private versus public [$Private_k$]); date of diagnosis (Y_i); managing clinician fixed effect (u_j). Time-varying variables: feedback for clinician j at time t (F_{jt}). Where variables did not follow the proportional hazards assumption, the analysis was re-run stratified over these variables.

3.2.2.1. Control variables

The model controls for age (Age=[<55, 55-74, 75+]) and SEIFA decile estimated from postcode (SEIFA=[Lowest 20%, Lowest 21-40%, Lowest 41-60%, Highest 61-80%,

Highest 81-100%, Unknown]) as patient characteristics that are likely to affect patient management (Ruseckaite et al., 2016). There is some expectation that changes to prostate cancer diagnosis guidelines may have affected the severity of cancer within each risk group. Therefore PSA, Gleason score and T stage levels are included for each patient. An annual time trend in the model controlled for treatment time trends common across clinicians with or without feedback, such as those that may result in response to guideline changes. This annual time trend is based on the diagnosis date adjusted for the earliest feedback (1st December 2012).

Because urologists and hospitals are expected to have different treatment practices as a result of unobserved characteristics, the model controlled for clinician (u_j) fixed effects. Hospital characteristics are included as metro versus regional and public versus private. Metropolitan and regional hospitals are expected behave differently based on availability of technologies and expertise of clinicians. Public and private hospitals are expected to differ due to the financial incentives of treatment over surveillance, but also the perception of quality from the patient. Private patients may perceive surveillance as poorer quality of care (the perception of not actively treating the disease) and may have more influence on the decision to treat. As there is high potential for interaction between hospital and clinician effects, hospital specific effects are not included.

3.2.3. Robustness analyses

Several further analyses explored how the results differed by subgroups and tested the robustness of the conclusions to alternative model specifications. Urologists may respond to feedback differently for some patients. Clinical quality reporting of treatment decisions does not account for patient characteristics such as age or comorbidities that may affect treatment decisions. Feedback was interacted with patient age group to assess how the impact of feedback may differ by patient characteristics.

There is the possibility of a broader registry effect of feedback on patients that may be captured in the underlying time trend used in the base model. Robustness analysis is performed where the changes over time are split into two separate time trends before and after the earliest date feedback was available for urologists (1st December 2012).

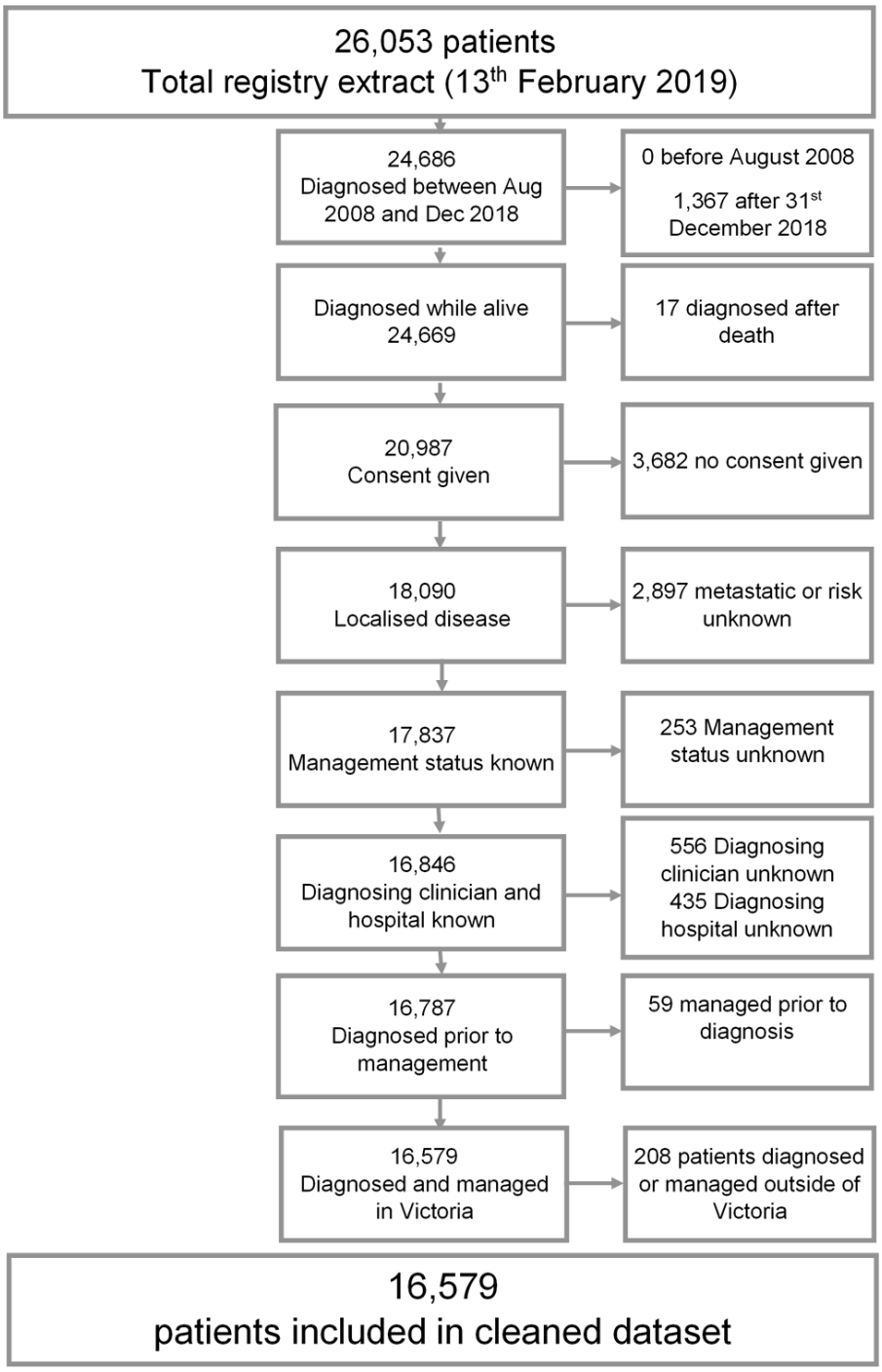
ADT was considered as a non-curative treatment in the base case. However, adjuvant ADT may occur before curative treatment to assist in treatment success. Robustness analysis explored the impact of ADT as a signifier of curative treatment in the model.

Patients may have different diagnosing and managing clinicians. Clinicians may change through necessity (clinician/hospital expertise and availability), but may also result from clinician expectations (e.g. they expect the patient requires a treatment they cannot provide) or patient preference (e.g. wanting a second opinion). Therefore, the diagnosing clinician's treatment decisions, e.g., referring a patient on for treatment, may influence the managing clinician's treatment decision. Feedback to the diagnosing clinician, rather than managing clinician is explored in the robustness analysis.

3.3. Results

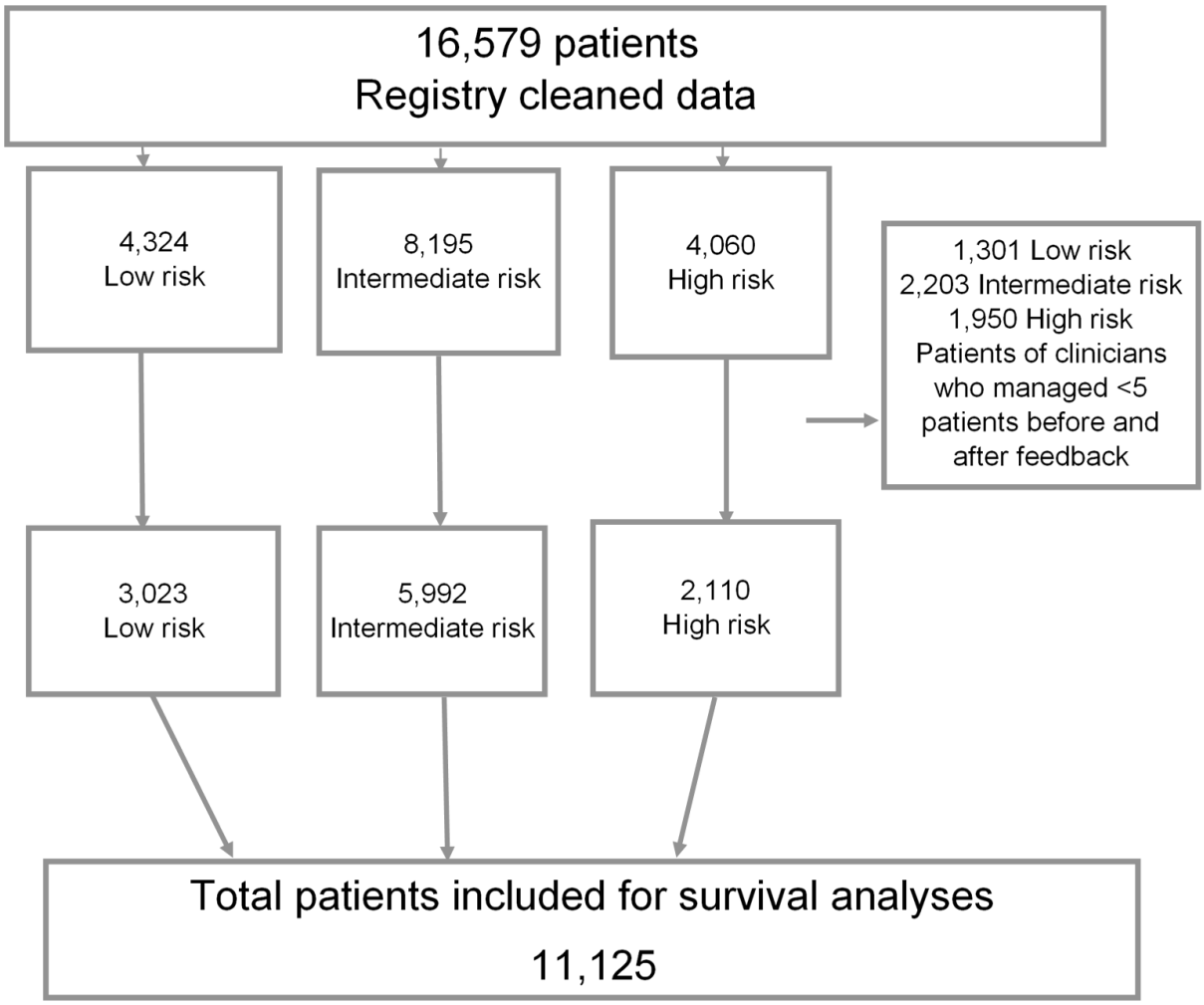
A summary of patient inclusion criteria is presented below and characteristics are summarised in Appendix 7.4, p139. Without adjustment, feedback is associated with lower rates of treatment for patients in low and intermediate-risk groups ($p<0.001$), and no difference in rate of curative treatment for the high-risk group.

Figure 7 Patient inclusion criteria for cleaned data extract



Notes: Cleaned dataset used in all PCOR-Vic analyses (Chapters 2-4)

Figure 8 Patient inclusion by risk stratification



3.3.1. Time to curative treatment, raw data

Initial results from the Kaplan-Meier plots suggested that among low- and intermediate-risk patients, those whose managing clinician received feedback were less likely to receive curative treatment within the first year following diagnosis. Lower rates of curative treatment are most pronounced for low-risk patients. At 8 months, 50% of low-risk patients had not received curative treatment when their clinician did not receive feedback, and 75% of patients had not received curative treatment if their clinician received feedback. Intermediate and high-risk patients were more likely to receive curative treatment than low-risk patients, regardless of their clinician's feedback status, with over 75% of patients receiving treatment by 6 months. For high-

risk patients there appeared to be no effect of feedback in the unadjusted Kaplan-Meier plots

Figure 9 Time to curative treatment for low risk patients, by clinician feedback status (Kaplan-Meier)

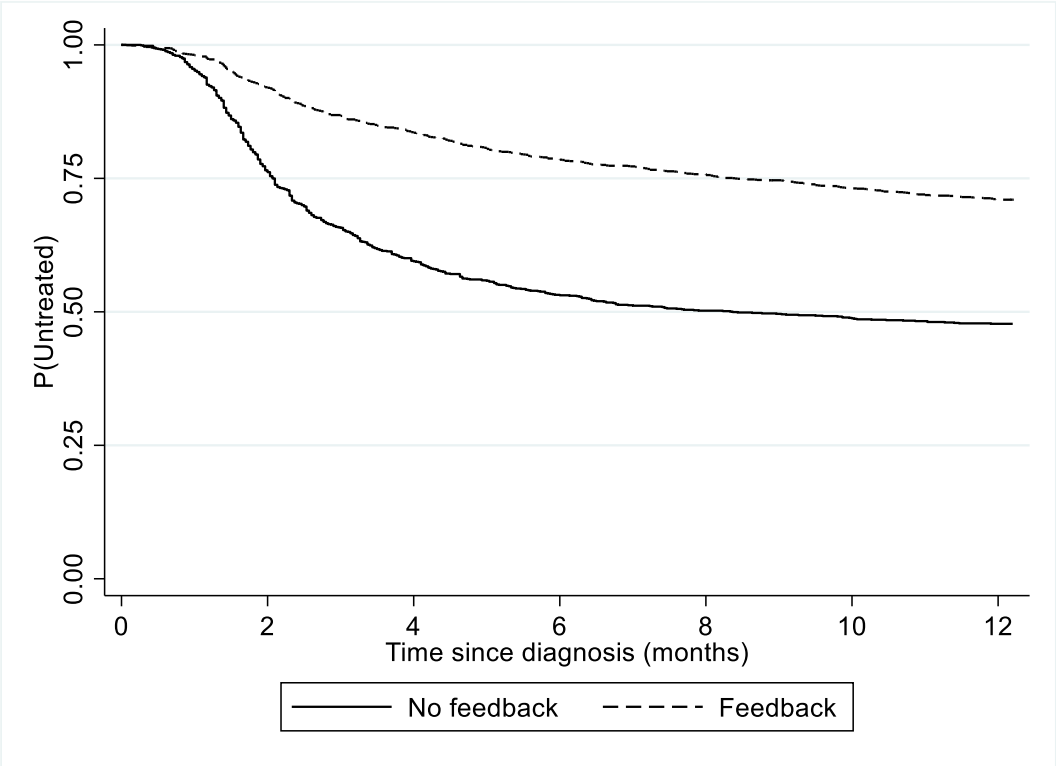


Figure 10 Time to curative treatment for intermediate risk patients, by clinician feedback status (Kaplan-Meier)

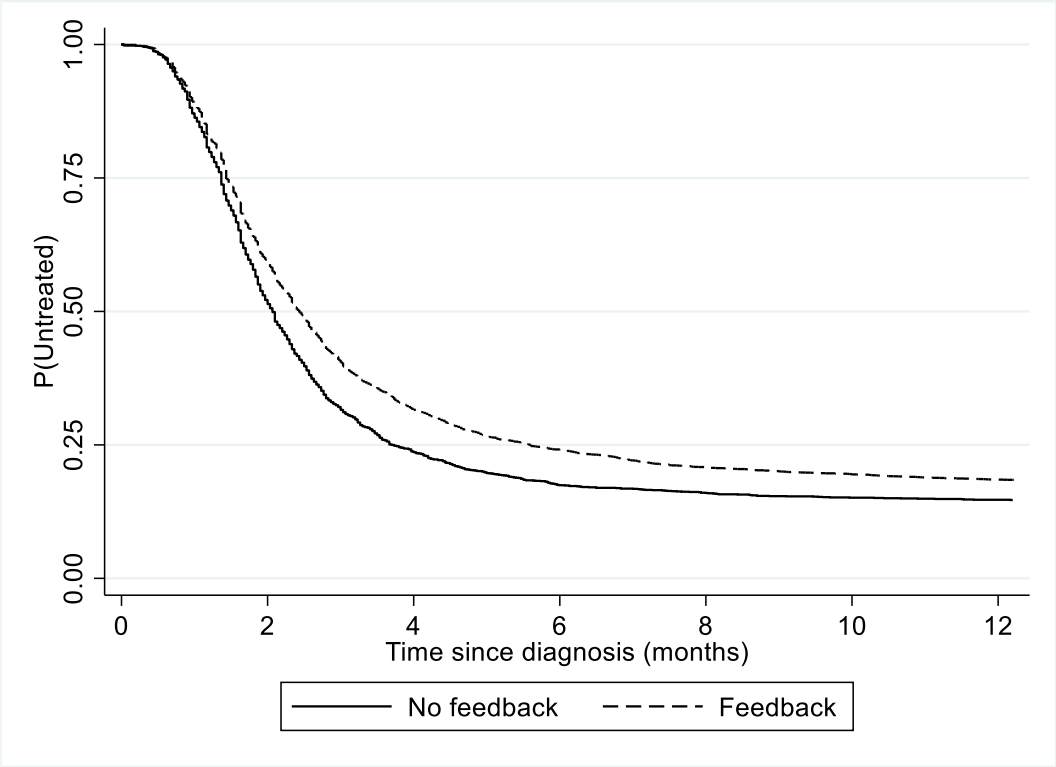
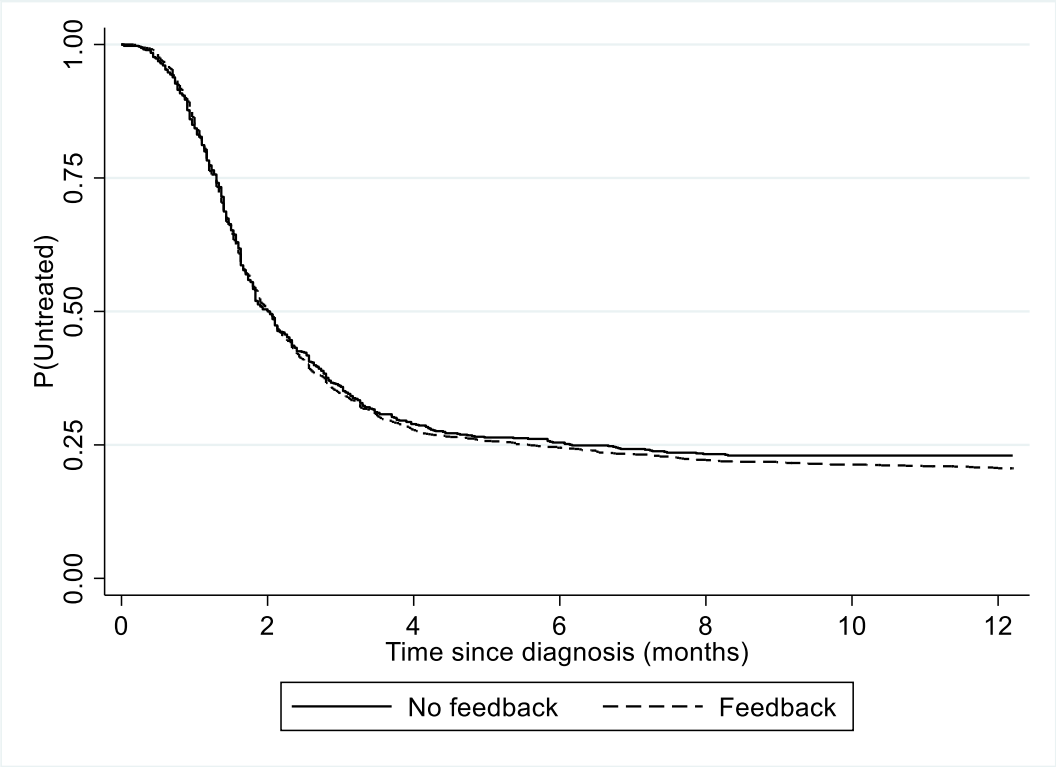


Figure 11 Time to curative treatment for high risk patients, by clinician feedback status (Kaplan-Meier)



3.3.2. Time to curative treatment, cox-regression adjusted results

A full model is run containing all covariates and is reported in Appendix 7.5, p146**Error! Reference source not found.**. Proportional hazard assumption testing for the intermediate group using the Schoenfeld residuals identified some variables for which the proportional hazard assumption did not hold: highest age group, private hospitals and several individual clinician effects (not reported here). The model was therefore stratified over these variables rather than including as them covariates. The feedback variable broke the proportional hazard assumption in the low-risk model (Schoenfeld results: rho=0.158, p<0.001). Inspection of the log-log plots suggests the proportional hazard assumption was only broken in the first month following diagnosis (lines are mostly parallel after this point). The observed versus predicted plot suggests that the Cox proportional hazard model underestimated the effect of feedback between month 2 and 10.

Table 4 Low risk, cox regression hazard ratios

Variable	Hazard ratio	SE	p-value	Lower 95%CI	Upper 95%CI
N=2,192					
Feedback	0.37***	0.0463	<0.001	0.29	0.48
PSA Level at diagnosis (ng/ml)	1.04**	0.019	0.017	1.01	1.08
cT stage (T2 vs T1)	1.65***	0.154	<0.001	1.37	1.98
SEIFA Decile					
Lowest 21-40%	0.81	0.122	0.157	0.6	1.09
Lowest 41-60%	1.07	0.161	0.638	0.8	1.44
Highest 61-80%	0.9	0.121	0.414	0.69	1.17
Highest 81-100%	0.72**	0.095	0.012	0.55	0.93
Unknown	0.52	0.382	0.373	0.12	2.2
Regional Hospital (vs Metro)	0.07***	0.041	<0.001	0.02	0.22
Annual time trend (1 Dec 2012 base)	0.998	0.029	0.935	0.94	1.06

Notes * p<0.10, ** p<0.05, *** p<0.01. HR<1 longer time to curative treatment (reduced likelihood of treatment in the first year following diagnosis), HR>1 shorter time to curative treatment (increased likelihood of treatment in the first year following diagnosis). Gleason score is not included in the low-risk analysis as all patients must have a score of less than 7 to qualify as low risk. Model stratified by age group, private vs public hospitals and treating clinician, as these variables violated the proportional hazards assumptions.

For low-risk patients, feedback to their clinician was associated with a reduction in the likelihood of treatment in the first year following diagnosis (HR 0.37; 95%CI 0.29,0.48; p<0.001). This was equivalent to a 63% reduction in treatment in the year following diagnosis compared to patients whose clinicians had not received feedback. Low-risk patients with higher risk indicators, e.g. higher PSA level and higher T staging at diagnosis were associated with an increased likelihood of treatment. Regional hospitals are also associated with a 93% lower likelihood of treatment in the first year following diagnosis for low risk patients compared to metropolitan hospitals (HR 0.07; 95%CI 0.02,0.22; p<0.001).

Table 5 Intermediate-risk, cox regression hazard ratios

Variable	Hazard ratio	SE	p-value	Lower 95%CI	Upper 95%CI
N=4,476					
Feedback	0.50***	0.03	<0.001	0.44	0.56
PSA Level at diagnosis (ng/ml)	1.005	0.01	0.345	0.99	1.02
cT stage (T2 vs T1)	1.28***	0.05	<0.001	1.19	1.38
Gleason score (7 vs <7)	3.22***	0.23	<0.001	2.80	3.72
SEIFA Decile					
Lowest 21-40%	1.01	0.08	0.929	0.87	1.17
Lowest 41-60%	1.14*	0.08	0.069	0.99	1.32
Highest 61-80%	1.03	0.07	0.608	0.91	1.18
Highest 81-100%	0.98	0.06	0.728	0.86	1.11
Unknown	1.13	0.31	0.650	0.66	1.93
Regional Hospital (vs Metro)	0.64**	0.13	0.030	0.43	0.96
Annual time trend (1 Dec 2012 base)	1.12***	0.01	<0.001	1.09	1.15

Notes: * p<0.10, ** p<0.05, *** p<0.01. HR<1 longer time to curative treatment (reduced likelihood of treatment in the first year following diagnosis), HR>1 shorter time to curative treatment (increased likelihood of treatment in the first year following diagnosis). Model stratified by age group, private vs public hospitals and treating clinician, as these variables violated the proportional hazards assumptions

For intermediate risk patients, feedback to their clinician was associated with reducing the likelihood of curative treatment by 50% in the first year following diagnosis compared to no feedback (HR 0.50; 95%CI 0.44,0.56; p<0.001). Patients with higher risk indicators, e.g. higher PSA level and higher T staging and higher Gleason score at diagnosis were associated with an increased likelihood of treatment. Regional hospitals were also associated with 36% less treatment in the first year following diagnosis for intermediate-risk patients compared to metropolitan hospitals (HR 0.64; 95%CI 0.43,0.96; p<0.001). There was also some evidence of a time trend for the

intermediate-risk group, with an annual increase in the likelihood of treatment of 12% (HR 1.12; 95%CI 1.09,1.15; p<0.001).

For high-risk patients, feedback to clinicians was associated with a not statistically significant reduction in curative treatment of 11% in the first year following diagnosis (HR 0.89; 95%CI 0.72,1.11; p=0.306). Higher risk factors, such as Gleason score were associated with 67% increase in treatment (p<0.001). An increase in PSA level was associated with a 1% reduction in treatment for every unit increase in ng/ml (p<0.001). There was some evidence that patients with median wealth were 34% more likely to receive treatment than the poorest groups (p=0.019).

Table 6. High risk, cox regression hazard ratios

Variable	Hazard ratio	SE	p-value	Lower 95%CI	Upper 95%CI
N=1,610					
Feedback	0.89	0.098	0.306	0.72	1.11
PSA level at diagnosis (ng/ml)	0.99***	0.002	<0.001	0.99	0.99
cT stage (T3 or 4 vs T2 or T1)	1.07	0.090	0.436	0.90	1.26
Gleason score (<8 vs 8+)	1.67	0.151	<0.001	1.40	1.99
SEIFA Decile					
Lowest 21-40%	0.99	0.127	0.944	0.77	1.27
Lowest 41-60%	1.34**	0.168	0.019	1.05	1.71
Highest 61-80%	1.27**	0.149	0.043	1.01	1.60
Highest 81-100%	1.22*	0.129	0.057	0.99	1.50

Variable	Hazard ratio	SE	p-value	Lower 95%CI	Upper 95%CI
Unknown	1.03	0.360	0.938	0.52	2.04
Regional Hospital (vs Metro)	0.90	0.345	0.773	0.42	1.90
Annual time trend (1 Nov 2012 base)	1.04	0.025	0.129	0.99	1.09

Notes * p<0.10, ** p<0.05, *** p<0.01. HR<1 longer time to curative treatment (reduced likelihood of treatment in the first year following diagnosis), HR>1 shorter time to curative treatment (increased likelihood of treatment in the first year following diagnosis). Model stratified by age group, private vs public hospitals and treating clinician, as these variables violated the proportional hazards assumptions

3.3.3. Robustness analyses

Table 7 Estimated hazard ratios for the effect of feedback on time to curative treatment, robustness analyses

Model	Feedback Hazard Ratio (95%CI)		
	Low risk	Intermediate risk	High risk
Base	0.37*** (0.29,0.48)	0.50*** (0.44, 0.56)	0.89 (0.72,1.11)
Feedback and age group interacted (vs age group with no feedback)	*<55:* 0.29*** (0.20,0.42)	*<55*: 0.57*** (0.45, 0.73)	*<55*: 0.74 (0.36, 1.54)
	55+: 0.40*** (0.31, 0.52)	*55-74*: 0.49*** (0.43, 0.55)	*55-74*: 0.86 (0.69, 1.08)
		75+: 0.85 (0.45, 1.61)	*75+*: 1.61 (0.83, 3.12)
Time trend before and after feedback introduced to the registry	0.33*** (0.26, 0.42)	0.50*** (0.44, 0.57)	0.86 (0.69, 1.08)
ADT as treatment	N/A	0.50*** (0.45.0.57)	0.87 (0.68, 1.12)
Diagnosing clinician assumed in charge	0.32***	0.46***	0.92

Model	Feedback Hazard Ratio (95%CI)		
	Low risk	Intermediate risk	High risk
	(0.24, 0.42)	(0.40, 0.52)	(0.69,1.22)

Notes: * p<0.10, ** p<0.05, *** p<0.01. HR<1 longer time to curative treatment (reduced likelihood of treatment in the first year following diagnosis), HR>1 shorter time to curative treatment (increased likelihood of treatment in the first year following diagnosis). ADT intermediate analysis time variable violates the proportional hazards assumption at the 95% confidence interval (rho=0.0346, p=0.0416). Diagnosing clinician for intermediate risk, feedback (rho =0.041, p=0.022) and time (rho=0.059, p=0.0013) violated PH assumption by Schoenfeld stats log-log plots and observed vs predicted for feedback effect were similar.

The feedback interacted with age results suggested that as the patient's age increased feedback has less effect on the probability of curative treatment in the year following diagnosis (hazard ratios trend towards 1) compared to patients whose clinicians have not received feedback. For high-risk patients, there was an increase in treatment associated with feedback to patients over 75+ compared to patients over 75 whose clinicians have not received feedback. Due to limited data, the low-risk analysis looks only at only two age categories (under 55 years old and 55 and above).

All other robustness analyses presented in Table 7 produced similar hazard ratios associated with clinician feedback. No robustness analysis found a statistically significant effect of feedback for high risk patients.

3.4. Discussion

Feedback to clinicians appears to be associated on average with a reduction in active treatment in low and intermediate risk patients by the end of the first-year post-diagnosis, once adjusted for observed patient, clinician and hospital characteristics and unobserved factors that trend over time. Feedback is also associated with a reduction in curative treatment for high-risk patients. The reductions suggest that with feedback low, intermediate and high-risk patients were 63%, 50% and 11% less likely to receive curative treatment than if there was no feedback. However, the reduction estimated for high risk patients was statistically non-significant.

As well as a total reduction in treatment in the first year, the results also suggest a longer time to treatment in response to feedback. Feedback may increase time to treatment by influencing the time to referral for treatment or influencing the managing clinician where they want to be certain curative treatment is appropriate.

Our results were robust to several variations in the model assumptions. The most considerable changes were seen in the interaction of feedback with age for the over 75s where a statistically non-significant reduction in treatment for the intermediate-risk group (15% less likely to receive treatment than 75+ without feedback) and a statistically nonsignificant increase in treatment for high-risk patients (61% more likely to receive treatment than 75+ without feedback).

There are some important limitations to the analysis. One potential source of bias may come from the identification of the clinician. The managing clinician was considered to be the decision-maker in the model. However, in practice, initial management decisions, particularly for referrals, may be driven by patient choice (e.g., locality, health insurance coverage) and the diagnosing clinician. Furthermore, assumptions were made to define who the managing clinicians were. Where the managing clinician was not recorded, the diagnosing clinician was assumed to be the managing clinician. A change in clinician is likely to dilute the effect of feedback as patients may move between clinicians who have and have not received feedback. In robustness analysis of feedback to diagnosing clinicians, the feedback was associated with a greater reduction in curative treatment versus no feedback for low and intermediate-risk groups than the base case. Therefore, the managing clinician decision appeared to be affected by the decision made at the diagnosing stage. Individual clinician effects could not be determined as the models were stratified over these.

To receive individual feedback on the registry, individual urologists must enrol on the registry. Participating clinicians joined as private practising urologists, though their public patients were then included in their reports. Therefore, results may not represent urologists who work mostly or entirely in the public sector.

The time constraint to one year after diagnosis may miss some of the effects of feedback. The clinical quality indicator for high-risk patients specifies treatment in the first year. However, the indicator for low-risk patients does not specify a time period.

A key benefit of surveillance in low-risk patients is to act as a long-term management strategy to avoid unnecessary treatments. Further research is needed to consider the role of feedback on the long-term avoidance of unnecessary treatments. Feedback may also increase treatment in the second year for higher-risk patients identified as untreated in the first year following diagnosis, where they are identified in the feedback reports.

Feedback is incorporated as a binary input; therefore, the analysis did not demonstrate how regular feedback or time spent receiving feedback might affect patients' management.

The effect of guidelines may also mitigate feedback to clinicians for the intermediate-risk group. Clinicians were told of intermediate-risk patients who do not receive treatment within a year post-diagnosis. However, guidelines are less clear on whether treatment is appropriate for intermediate-risk patients. Further analysis could explore the difference in feedback effect for intermediate-risk patients of lower risk (i.e. favourable) and intermediate-risk patients of higher risk (i.e. unfavourable) as this stratification is recommended by NCCN guideline update 2019 (Mohler & Antonarakis, 2019).

Within this analysis, the effect of feedback is only considered upon the effect of a process change (i.e. change in decision to treat). The results suggest a reduction in overtreatment (for low-risk patients), but also an increase in undertreatment for higher-risk groups, particularly for intermediate risk group who are the largest group of patients and for whom choice to treat is more complex.

3.4.1. Conclusion

There is strong evidence that individual feedback to clinicians was associated with reduced curative treatment for patients, particularly those diagnosed with low or intermediate-risk disease. The analysis does not provide a benchmark for the treatment rates in the first year following diagnosis and cannot comment on absolute treatment rates prior to feedback. Therefore, the analysis cannot claim whether clinicians responded correctly to feedback, particularly for intermediate-risk patients where treatment guidance is less clear, but there is potential for undertreating some

patients. Further research should consider how potential undertreating could be addressed in guidance or feedback. It could also consider the effect of repeated feedback to clinicians or the longer-term effect of feedback.

4. How do clinicians respond to feedback on competing outcomes? Evidence from a prostate cancer clinical quality registry

4.1. Introduction

4.1.1. Changing behaviour in clinicians

Routine collection and systematic reporting of healthcare delivery quality information is a widely used strategy to improve patient outcomes through improved clinical practice. Auditing of practice with or without feedback to the clinician effectively influences prescribing behaviours, primarily in general practitioners (Avery et al., 2010; Naughton et al., 2009; Soleymani et al., 2012; Steele et al., 1989). In a 2012 review of 140 randomized trials by the Cochrane collaboration (Ivers et al., 2012), feedback was shown to have a small to moderate effect on desired care practices, but patient outcomes were less clear. Clinical registries which collect a range of data from clinicians and their patients have also been associated with improvements in clinical or patient outcomes. Registries provide reports on clinician performance at an aggregate or institutional level, and some provide direct feedback of relative performance to individual clinicians. The use of registries is expected to improve overall performance in terms of clinically relevant outcomes. The precise mechanisms of improvement have rarely been explicitly explored (Dinh et al., 2015; Ruseckaite et al., 2016; Stey et al., 2015).

Though auditing and feedback demonstrated effectiveness in specific settings, there is limited evidence on which methods and presentations of feedback are most effective. Ivers et al.'s review of audit and feedback found that feedback was more effective when clinicians had a low baseline performance; the feedback came from a colleague or supervisor (rather than an external provider); feedback was provided more than once; feedback was provided in both verbal and written formats; and feedback included both specific targets and an action plan (Ivers et al., 2012). Peer comparison studies have found a mixed response to feedback, even within the same disease area (Schneider et al., 2008; Sondergaard et al., 2002). One specific area

where there is little evidence is how clinicians respond when they receive multiple pieces of feedback that may result in competing clinical practice, especially where clinicians are not provided with a specific target or action plan. Previous studies have shown that when multiple indicators are reported to healthcare providers, there can be improved responses to some feedback elements, but not others (Vratsistas-Curto, McCluskey, & Schurr, 2017).

4.1.2. Measures of surgical quality in prostate cancer

In prostate cancer, the positive surgical margin (PSM) is an important measure of clinician performance. PSM refers to any cancerous tissue remaining following surgery. A PSM is associated with a poorer prognosis with a higher likelihood of biochemical recurrence and additional treatment (Oh, Hong, Byun, Choe, & Lee, 2013; Ploussard et al., 2011; Swindle et al., 2005; Zhang et al., 2018). There is evidence that PSMs have been reducing globally over time. For example, overall PSM rates decreased in the US, from 18.7% of radical prostatectomy patients in 2004 to 9.7% in 2015 (Preisser et al., 2019). Similarly, previously published results for Victoria have shown an average reduction in PSM rates of 9% between 2009 and 2013 for clinical T stage 2 (cT2) patients (Sampurno et al., 2016). The cause of this is likely to be due to a combination of changes in patient case-mix (e.g., patient demographics, prognostic clinical factors), technology and reporting over time. PSM rates are also used to assess surgical quality, through peer comparisons presented to clinicians. This is described further below.

Other measures of surgical quality (length of hospital stay, blood usage, disease recurrence) are harder to isolate to the surgeon behaviour, and data for these was not available from the PCOR-Vic.

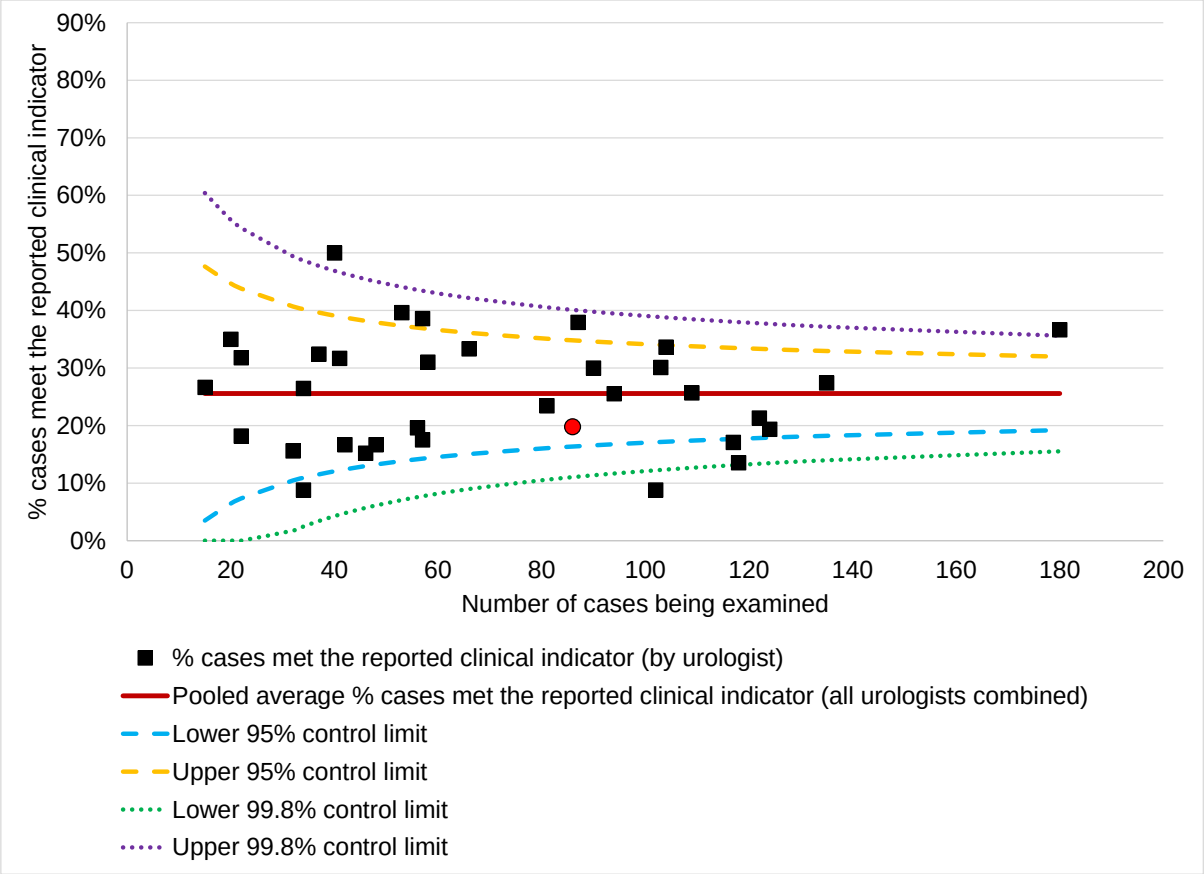
4.1.3. Feedback to prostate cancer surgeons in Victoria

PCOR-Vic aims to improve surgical outcomes for men with localised prostate cancer, through individual, peer-comparative feedback to hospitals and urologists. This feedback includes the reporting of peer comparison funnel plots of multiple quality indicators. Funnel plots are widely used in meta-analysis (Sterne et al., 2011), and

have been adopted to measure and report systematic heterogeneity amongst clinicians or hospitals (Koh, 2017; Spiegelhalter, 2005). In funnel plots, clinicians or hospitals can be compared to their peers for specific quality indicators, by identifying individuals whose behaviour differs considerably from the average, i.e., those who lie outside the established control limits (Spiegelhalter, 2005). These plots can prove useful to emphasising peer comparisons and therefore encouraging changes in behaviours for clinicians with the poorest performance. However, for better-performing clinicians, particularly those with the best outcomes, there is less information about their gap to best practice beyond their nearest peers, without a pre-determined benchmark.

Since December 2012, participating urologists have received individualised feedback, with regular six-monthly reports available from February 2015. Urologists must have consented to participate and have diagnosed or treated at least ten patients between reports to receive feedback. Hospital department heads also receive hospital-level feedback, which may also be disseminated to individual clinicians. Urologists and hospitals receive feedback on multiple quality indicators, including risk-stratified funnel plots of PSMs for patients undergoing radical prostatectomy when they have treated at least 20 patients who meet inclusion criteria. An example of these funnel plots is reported in Figure 12.

Figure 12 Sample funnel plot provided to urologists



Notes: red dot identifies the specific urologist who received the report

Also included in the PCOR-Vic feedback reports are funnel plots for quality of life indicators: percentage of patients reporting sexual, urinary, and bowel bother at their follow-up meeting, collected 12 months after their last treatment. For most patients undergoing surgery (>97%), this follow-up occurs less than two years from their surgery date. Sexual bother is described as the patient reporting a big problem with sexual function (achieving and maintaining an erection) in the four weeks prior to their follow-up. Urinary and bowel bother are collected similarly; however, bowel bother is only included in the feedback for patients who also underwent radiotherapy. Urinary bother is an important outcome for men, but within the context of the PCOR-Vic, only 88 men report a big problem with urinary bother (of 3,295 reporting urinary bother status). As urinary bother rates are low, this analysis focuses on sexual bother, which has larger incidence rates (>37% of men reporting sexual bother status report a big problem). Therefore, it is an outcome where a more considerable improvement can be gained.

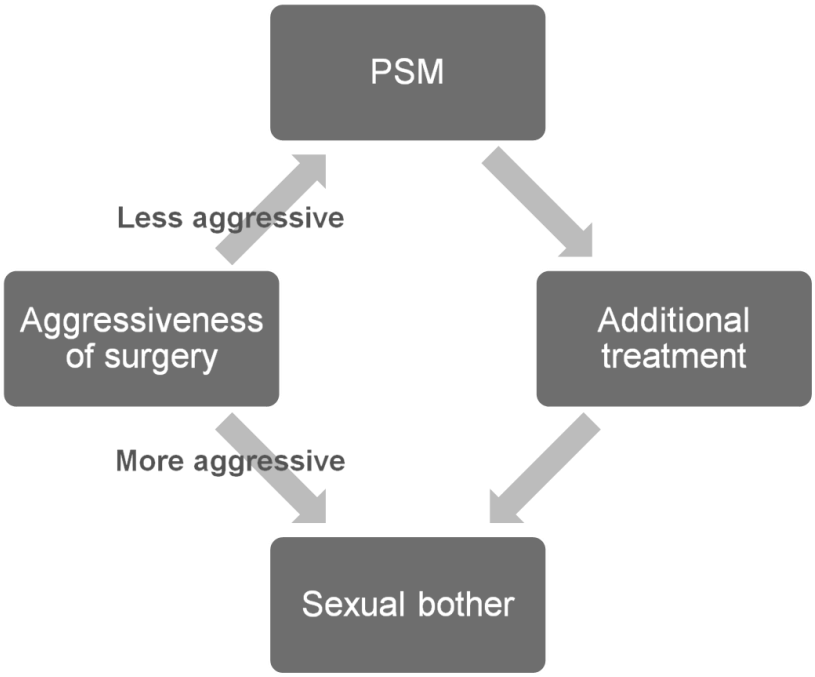
4.1.3.1. The effect of feedback on surgeon behaviour

The PCOR-Vic presents PSM and sexual bother quality indicators to encourage urologists to take steps to reduce the likelihood of patients having a PSM and sexual bother following surgery. Hospitals and stakeholders are more likely to encourage urologists who perform worse in these quality indicators to improve. However, there is an expectation that PSM and sexual bother rates may be related. Should urologists have a greater than average PSM rate compared to their peers, they may increase their surgery's aggressiveness. More aggressive surgery likely leads to an increase in sexual bother for their patients as a result of nerve damage. On the other hand, higher than average PSMs might be associated with increased additional active treatment (radiotherapy, brachytherapy) after surgery to combat remaining cancerous tissue (Expert opinion, SE)(S. M. Evans et al., 2014). Additional treatment may also increase the likelihood of sexual bother through nerve damage (Alsinnawi et al., 2019). Alternatively, urologists seeing reports of a greater than average sexual bother rate among their patients may reduce the surgery's aggressiveness, which in turn may lead to an increase in PSMs among their patients (Damani, Van Hemelrijck, Wulaningsih, Crawley, & Cahill, 2017).

A flow diagram of how sexual bother likelihood may be increased through treatment is given in Figure 13. Other mechanisms may be correlated with both surgery and treatment approaches and the likelihood of PSM and sexual bother, including patient characteristics and surgeon experience. In particular, sexual bother can result from ageing.

Previous work within the PCOR-Vic indicated that urologists believed PSM reporting to be the most important quality indicator, suggesting this may be the quality indicator urologists respond to if they have higher than average rates of both PSM and sexual bother. However, PSM rates may be more challenging to respond to due to uncontrollable factors, such as case-mix (Koh, 2017). Koh also indicates that PSM rates are an outcome of which urologists are already conscious. Therefore, more knowledge is gained from the feedback on sexual bother rates. Additional treatment rates are not reported to urologists and are therefore not expected to change due to the introduction of feedback directly.

Figure 13 Possible mechanisms to increased PSM & sexual bother



Key: PSM, positive surgical margin

In general, this study attempts to understand the mechanism of feedback on multiple quality indicators where one indicator may improve at the expense of the other within the group of patients who receive surgery. For example, whether feedback resulted in reduced PSMs at the expense of the more patient-focused outcome, sexual bother; or reduced sexual bother at the expense of higher PSM rates. Also explored is how urologists may alter their clinical practice (i.e., rates of additional treatment) in response to feedback on surgical quality.

4.2. Methods and data

4.2.1. PCOR-Vic

PCOR-Vic patient medical records are accessed every six months for up to 2 years after treatment, to provide information on surgical outcomes such as PSMs and patient-reported outcomes such as sexual bother as recorded at time of follow-up (usually one year after treatment). Baseline sexual bother status is not recorded. Patient responses to follow up queries are voluntary, and reporting of sexual bother

status can be low, even when follow up is completed. Additional treatment dates are collected from medical records within 2 years after surgery. In this analysis additional treatment referred to additional curative treatments: brachytherapy and radiotherapy. If no date was recorded for these treatments, patients were assumed to have received no additional treatment.

15,275 patients were previously identified on the PCOR-Vic for whom complete diagnosis and initial management data were available. Of these, 8,625 received surgery. After excluding patients with no reported surgeons or hospitals; and surgeons and hospitals with small numbers of patients (as the PCOR-Vic provides reports only for urologists with a minimum of 10 patients,), the final analysis included 7,401 patients who received surgery and whose urologists saw at least 10 patients both prior to and following feedback. Of the 7,401 patients included, 7,232 (97.7%) reported PSM status, and 3,261 (44.1%) sexual bother status. This smaller figure partially occurs from incomplete post-feedback follow-ups (10.7% of patients had no recorded follow-up in the feedback period compared to 2.1% prior to urologists receiving feedback). There also appears to be a reduction in reporting of patient sexual bother status for every urologist after feedback began (median reporting of sexual bother status 77% [IQR, 72-80%] across urologists prior to feedback, 30% after feedback [IQR, 26-36%]). As this reduction happens to all urologists, it is unlikely this was a response to their individualised feedback (e.g. poorer performing urologists failing to report patient sexual bother status to avoid scrutiny). It was more likely to do with the change in reporting of patient-reported outcomes around 2014. Very few patients reported sexual bother status had surgeries performed after 2014 (67 out of 3,261).

There are some differences between men who did or did not have their sexual bother status recorded. Patients with recorded sexual bother status were less likely to have received less invasive surgery: 59.0% versus 73.1% who do not report sexual bother status. This was likely driven by the surgery date (later surgeries are more likely to be less invasive). Men who report sexual bother status are also slightly younger (14.8% vs 10.7% under 55 years; 2.1% vs 4.2% 75 years and over), and slightly lower risk (18.4% NCCN low risk compared to 12.8%; 84.5% with PSA level ≤ 10ng/ml compared to 81.1%) than those who do not report sexual bother status, but rates of highest risk were similar across groups (20.7% for those reporting sexual bother status

compared to 21.9% without sexual bother status reported were classified as NCCN high risk prior to surgery), as was PSM status (25.7% for those reporting sexual bother compared to 25.2% for those who do not). These patient differences are likely to drive sexual bother rates in opposite directions. Younger men are more likely to be affected by sexual bother (~38% men under 75 years report sexual bother to be a big problem compared to ~19% for men 75 years and over). However, men of lower risk are less likely to be affected by sexual bother (as they receive less aggressive treatment). On average, sexual bother rates will likely be overestimated for the total group of patients after urologists begin feedback: the subgroup who reported sexual bother status is younger than the before feedback group.

However, there is little evidence that urologists are systematically selecting the men who receive surgery to affect their patient reported sexual bother rates due to feedback directly. If urologists were systematically choosing their patients (or the patients they report on) to improve their quality indicators after feedback, fewer complex cases should be chosen. Instead, men are older and higher risk on average after feedback begins, suggesting cases may be more complex. Alternatively, recording of sexual bother should occur for the subgroup of less affected men, e.g. more older men should have sexual bother status recorded. However, sexual bother status is recorded for younger men. Therefore, though the drop in sexual bother status reporting after feedback begins is not ideal, it is unlikely a direct effect of introducing feedback. By controlling for characteristics that may affect both sexual bother status and sexual bother reporting, the remaining variation in whether sexual bother status is recorded should be random variation.

4.2.2. Assessing the effect of feedback on the performance of competing outcomes

This study estimates how feedback might improve performance of two quality indicators of surgery, where the two indicators may act as substitutions, i.e. one improves at the expense of another, and whether feedback may alter this substitution rate. First, an estimate of how each indicator changed for each urologist with the introduction of feedback was made, controlling for urologist performance prior to

feedback. The urologist performance before and feedback estimated how the trade-off between PSM and sexual bother changed with feedback.

The relationship between feedback and each outcome is estimated separately based on linear probability models (equation (4.1)) for individual patients with clustering on urologists via a fixed effect for each urologist pre- and post- feedback ($v_{\mathrm{jf,q}}$) where q is the quality indicator (PSM, sexual bother), for urologist j with feedback status f (f=1|feedback, f=0|no feedback).

$$\begin{aligned}\mathrm{P}(\mathrm{q}_i = 1) = {} & \beta_{0,q} + \beta_{1,q}(\mathrm{NCCN\ int})_{\mathrm{i}} + \beta_{2,q}(\mathrm{NCCN\ high})_i + \beta_{3,q}\mathrm{S}_{\mathrm{i}} + \beta_{4,q}(\mathrm{Age} < 55)_{\mathrm{i}} \\ & + \beta_{5,q}(\mathrm{Age\ 75+})_{\mathrm{i}} + \beta_{6,q}\mathrm{PSA}_{\mathrm{i}} + \beta_{7,\mathrm{q}}T_i + \beta_{8,q}\mathrm{Metro}_k + \beta_{9,q}\mathrm{Private}_k \\ & + v_{jf,Q} + \epsilon_{ij,Q} \qquad (4.1)\end{aligned}$$

Notes: quality indicator, q:[PSM, sexual bother]; patient i; urologist j; hospital k; feedback f is (f=1|feedback, 0|no feedback), NCCN risk prior to surgery (NCCN int=1| intermediate risk, NCCN high=1| high risk, NCCN int and NCCN high=0| low risk), surgery type S=1| open surgery, S=0|less invasive surgery; age of patient at surgery (Age<55=1| age less than 55 years, Age 75+=1| age 75 years and over, Age<55 and Age 75+=0| patient aged between 55 and 74 years); PSA is PSA level prior to surgery (ng/mL), T is date of surgery, adjusted to an annual time trend (T=0|1st November 2012); Metro=1|metropolitan hospital, Metro=0|regional hospital; Private=1|private hospital k, Private=0| public hospital k; v is the urologist fixed effect; ϵ is the error term.

Surgical outcomes may change as urologists respond to feedback. However, these outcomes are not independent of each other, and they are not necessarily one-directional. For example, the likelihood of sexual bother may be increased if a patient had a PSM and therefore went on to further treatment or increase if a patient receives more aggressive surgery and does not have a PSM (see Figure 13). A potential way to adjust for this interdependence between outcomes would be to use a bivariate probit model, which would allow for a correlation in the errors of the PSM and sexual bother models. However, separate linear models also allow for adjustment of unobserved time-invariant urologist characteristics pre- and post-feedback. These can demonstrate whether conditional outcomes change with feedback. The relationship between PSM and sexual bother may also change with the introduction of feedback, as urologists change their behaviour. For example, the correlation may change from positive (high PSM rate results in more additional treatment results in more sexual bother) to negative (PSM results remain unchanged, but urologist avoids more additional treatment, resulting lower sexual bother). A bivariate probit model would not

capture this change. Linear probability models are also chosen over probit models to avoid the incidental parameter problem (Kunz, Staub, & Winkelmann, 2017; Lancaster, 2000).

Predicted means of outcomes were compared for each of the urologists pre- and post-feedback. The means of the fitted values $\widehat{X\mathrm{B}}$ were estimated from equation (4.1) and adjusted for the urologist and feedback fixed effect, as in equation (4.2). A separate linear relationship was created between sexual bother and PSM (equation (4.3)).

$$\mu_{jf}^{q} = \widehat{X\mathrm{B}} + \mathrm{v_{jf,q}} \qquad (4.2)$$

$$\mu_{\boldsymbol{jf}}^{SB} = \gamma_0 + \gamma_1 \mu_{\boldsymbol{jf}}^{PSM} \qquad (4.3)$$

The change in each quality indicator is estimated using equation (4.2) for each urologist after they begin to receive feedback (equation (4.4))

$$\mu_{j,change}^{q} = \mu_{jf=1}^{q} - \mu_{jf=0}^{q} \qquad (4.4)$$

The effect of feedback on change in sexual bother rates is then estimated as a linear combination of the estimated change in PSM for urologist j after feedback plus a constant (equation (4.5)). This linear model was used to explore changes in the correlation between PSM and sexual bother rates with the introduction of feedback

$$\mu_{j,change}^{SB} = \gamma_{0,change} + \gamma_{1,change} \mu_{j,change}^{PSM} \qquad (4.5)$$

Graphically plotting the relationship between sexual bother and PSM rates recognises movement patterns across the urologists. A linear fit is assumed, but the relationship may be more complex. In particular, as both the estimates for PSM and sexual bother rates are uncertain, and the urologist error terms likely correlated, the line of best fit is likely biased towards zero. This analysis may, therefore underestimate the size of the relationship between changes in PSM and sexual bother.

4.3. Results

4.3.1. Patient characteristics

Summary patient characteristics are presented in Table 8. When comparing patients whose urologists received feedback ('feedback' patients) to those whose urologists had not ('no feedback' patients), patients appeared to be older, of higher risk, received less invasive surgery, had a similar likelihood of PSM, and a reduced likelihood of sexual bother. Sexual bother status was not reported for 73.4% of patients of urologists receiving feedback, compared to 30.6% of the 'no feedback' group. The majority of 'no feedback' patients underwent surgery prior to 2013, whereas most 'feedback' patients received their surgery from 2013.

Table 8 Patient characteristics, by urologist feedback status

Characteristic	No urologist feedback	Urologist feedback	p-value (Pearson's chi-squared)
N	3028	5282	
Age group			<0.001
<55	461 (15.2%)	588 (11.1%)	
55-74	2520 (83.2%)	4471 (84.6%)	
75+	47 (1.6%)	223 (4.2%)	
NCCN risk at surgery			<0.001
Low Risk	636 (21.0%)	605 (11.5%)	
Intermediate Risk	1766 (58.3%)	3516 (66.6%)	
High Risk	626 (20.7%)	1161 (22.0%)	
PSA level at surgery (ng/mL)			<0.001
<=10.0	2514 (83.0%)	4206 (79.6%)	
10.01-20.0	383 (12.6%)	789 (14.9%)	
>20.0	116 (3.8%)	200 (3.8%)	
NR	15 (0.5%)	87 (1.6%)	
Surgery type			<0.001
Less invasive surgery	1593 (52.6%)	3836 (72.6%)	
Open surgery	1415 (46.7%)	1393 (26.4%)	
NR	20 (0.7%)	53 (1.0%)	

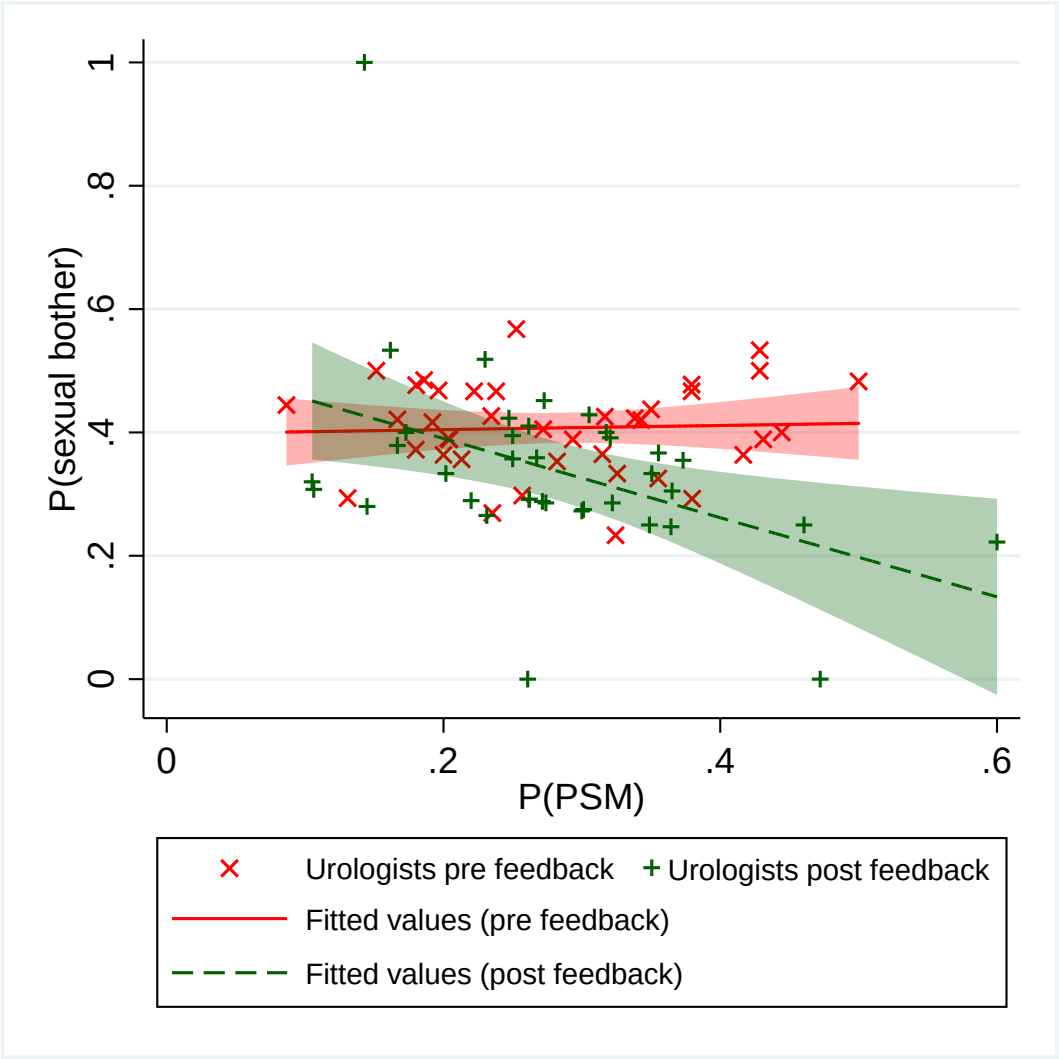
Characteristic	No urologist feedback	Urologist feedback	p-value (Pearson's chi-squared)
Surgery year			<0.001
2009	345 (11.4%)	0 (0.0%)	
2010	420 (13.9%)	0 (0.0%)	
2011	977 (32.3%)	0 (0.0%)	
2012	978 (32.3%)	60 (1.1%)	
2013	54 (1.8%)	986 (18.7%)	
2014	76 (2.5%)	937 (17.7%)	
2015	95 (3.1%)	993 (18.8%)	
2016	75 (2.5%)	992 (18.8%)	
2017	8 (0.3%)	1097 (20.8%)	
2018	0 (0.0%)	217 (4.1%)	
Surgical margin			0.16
No PSM	2165 (71.5%)	3865 (73.2%)	
PSM	781 (25.8%)	1295 (24.5%)	
NR	82 (2.7%)	122 (2.3%)	
Sexual Bother			<0.001
None or little sexual bother	1264 (41.7%)	933 (17.7%)	
Sexual Bother	838 (27.7%)	473 (9.0%)	
NR	926 (30.6%)	3876 (73.4%)	

In total, 3,508 patients reported sexual bother status and their characteristics are summarised in Appendix 7.6, Table 21, p151. They tended to be slightly younger (~4% more patients in <55 years old group), lower risk (~6% more patients in the low-risk group, PSA level lower overall), were more likely to have received an open surgery (~14% more than those not reporting sexual bother status) and showed no statistical difference in PSM rates. Sexual bother status was reported for very few patients beyond 2014. When restricting surgery year to before 2015, age group, NCCN risk, PSA level and PSM status become non-significantly different between the two groups.

4.3.2. Estimating the average effect of feedback on PSM and sexual bother for each urologist

Figure 14 presents the unadjusted relationship between PSM and sexual bother, explored at the urologist level and stratified by feedback status. The probability of a patient reporting PSM or sexual bother is clustered around 20-60% for each urologist. Mean rates of PSM at the urologist level reduce from 26.2% to 25.0% after urologists begin feedback. Mean rates of sexual bother at the urologist level decrease from 39.6% to 34.5% after feedback. There is some evidence of a negative correlation between PSM and sexual bother after feedback is introduced, whereas prior to feedback, this appeared to be a near-constant relationship. This relationship means that after feedback was introduced, as the PSM rate decreased, the rate of sexual bother increased for each urologist. Alternatively, as sexual bother decreased, the likelihood of PSM increased. From the observed data, it is not clear if the change in the relationship between the probability of PSM and sexual bother after feedback resulted from clinical practice or a result of the unadjusted patient characteristics.

Figure 14 Observed rates of PSM versus sexual bother by urologist pre- and post-feedback



Notes: Observed mean rates per urologist, unadjusted for patient characteristics

4.3.3. Predicted probability of PSM and sexual bother

Table 9 and Table 10 give the linear regressions for PSM and sexual bother, as described in equation (4.1). As urologists by pre- and post-feedback were included as fixed effects, they are not presented in these tables. In general, patient and hospital characteristics had the same direction of effect for PSM and sexual bother rates. For example, higher NCCN risk was associated with a higher probability of PSM (intermediate-risk associated with an increase in P(PSM) of 7.1% versus low-risk,

95%CI 4.2, 10.0%; high-risk associated with an increase of 21.7% versus low-risk, 95%CI 18.3, 25.1%). Similarly, NCCN risk was associated with a higher probability of sexual bother (intermediate-risk associated with an increase in sexual bother of 4.8% versus low-risk, 95%CI 0.3, 9.4%; high-risk 6.2% versus low-risk, 95%CI 0.4, 12.1%). Later dates were also associated with lower probability of PSM (-1.2% annual reduction in PSM, 95%CI -1.9, -0.4%) and sexual bother (-3.6% annual reduction in sexual bother, 95%CI -5.7, -1.5%).

Table 9 Linear probability model for PSM, urologist feedback status fixed effect

Variable **N=7,150 (34 urologists)**	**Change in Prob (%)**	**SE (%)**	**p-value**	**Lower 95% CI (%)**	**Upper 95% CI (%)**
Constant probability of PSM	39.6	5.6	<0.001	28.5	50.7
NCCN risk (vs low risk)					
Intermediate risk	7.1***	1.5	<0.001	4.2	10.0
High risk	21.7***	1.7	<0.001	18.3	25.1
Annual time trend (surgery date)	-1.2***	0.4	0.002	-1.9	-0.4
PSA level (increase in ng/ml)	0.003	0.014	0.849	-0.025	0.030
Age group (base 55-74)					
<55	-2.1	1.5	0.168	-5.1	0.9
75+	-4.8	2.9	0.101	-10.5	0.9
Open surgery (vs less invasive surgery)	1.5	2.0	0.462	-2.4	5.3
Hospital Characteristics					
Metropolitan (vs regional)	-12.9**	6.0	0.032	-24.7	-1.1
Private (vs public)	-12.3***	1.6%	<0.001	-15.4	-9.1

Notes: * p<0.10, **p<0.05, ***p<0.01. Urologist fixed effects interacted with feedback status not reported in table.

Table 10 Linear probability model for sexual bother, urologist feedback status fixed effect

Variable N=3,223 (32 urologists)	Change in Prob (%)	SE (%)	p-value	Lower 95% CI (%)	Upper 95% CI (%)
Constant probability of sexual bother	48.6	10.6	<0.001	27.9	69.4
NCCN risk (vs low risk)					
Intermediate risk	4.8**	2.3	0.038	0.3	9.4
High risk	6.2**	3.0	0.035	0.4	12.1
Annual time trend (surgery date)	-3.6***	1.1	0.001	-5.7	-1.5
PSA level (increase in ng/ml)	-0.03	0.13	0.826	-0.29	0.23
Age group (base 55-74)					
<55	1.7	2.4	0.496	-3.1	6.4
75+	-16.0***	6.1	0.009	-28.0	-4.0
Open surgery (vs less invasive surgery)	-0.2	3.5	0.947	-7.1	6.6
Hospital Characteristics					
Metropolitan (vs regional)	-11.1	11.2	0.319	-33.0	10.8
Private (vs public)	-7.4**	3.0	0.014	-13.4	-1.5

Notes: * p<0.10, **p<0.05, ***p<0.01. Urologist fixed effects interacted with feedback status not reported in table.

4.3.4. Estimating the relationship between PSM and sexual bother pre and post feedback for each urologist

Including urologist and feedback status fixed effects in equation (4.2) estimated the mean rates of PSM and sexual bother before and after feedback. On average, PSM and sexual bother increased by a small amount after feedback (predicted mean PSM rates rose from 24.3% to 26.2% after feedback; predicted sexual bother rates rose from 24.5% to 27.1% after feedback).

The adjusted PSM and sexual bother rates for each urologist pre and post feedback are estimated using equation (4.2). These adjusted estimates, along with the linear relationship between the probability of sexual bother and PSM (equation (4.3)) are shown in Figure 15. Figure 15 demonstrates a similar relationship between PSM and sexual bother pre-feedback compared to post-feedback, with a small overall increase in the absolute probability of sexual bother. The change between PSM and sexual bother (Figure 16) suggests a slight negative trend between the two outcomes due to feedback introduction. For a 1% increase in PSM probability, there is a 0.58% decrease in the probability of sexual bother.

Figure 15 Relationship between PSM and sexual bother, by urologist (adjusted to full sample)

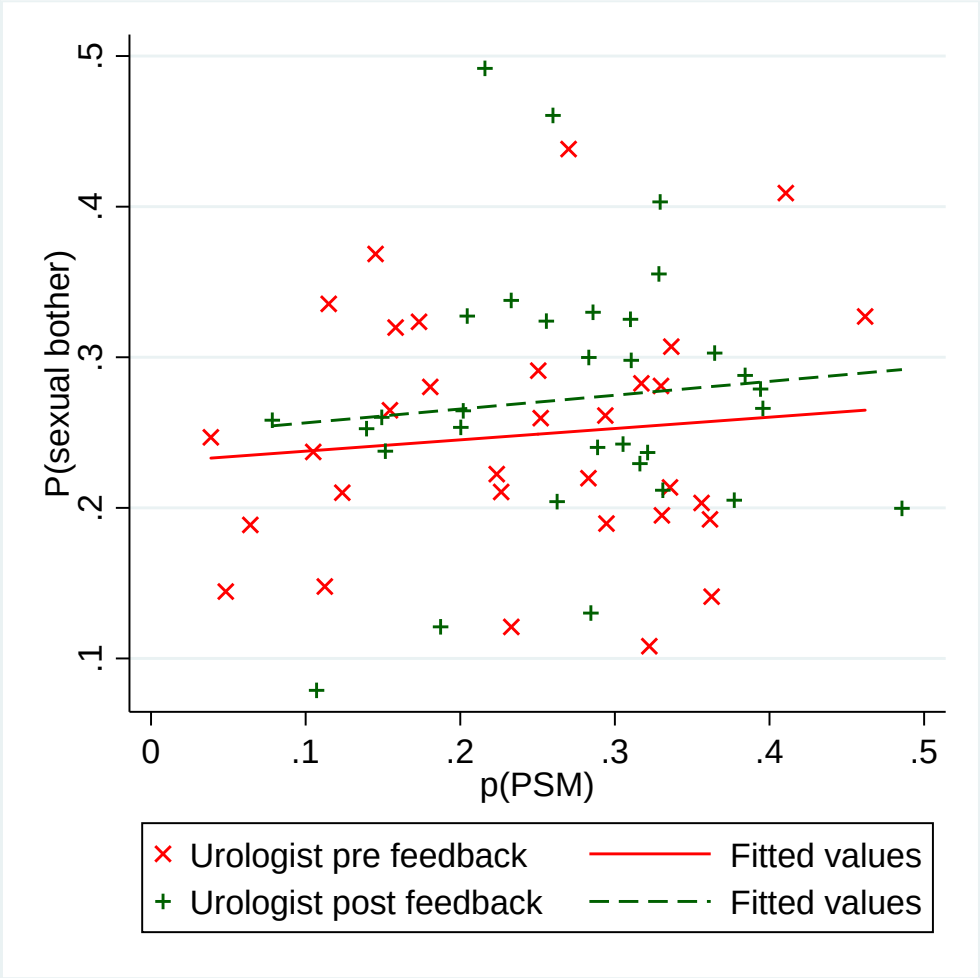
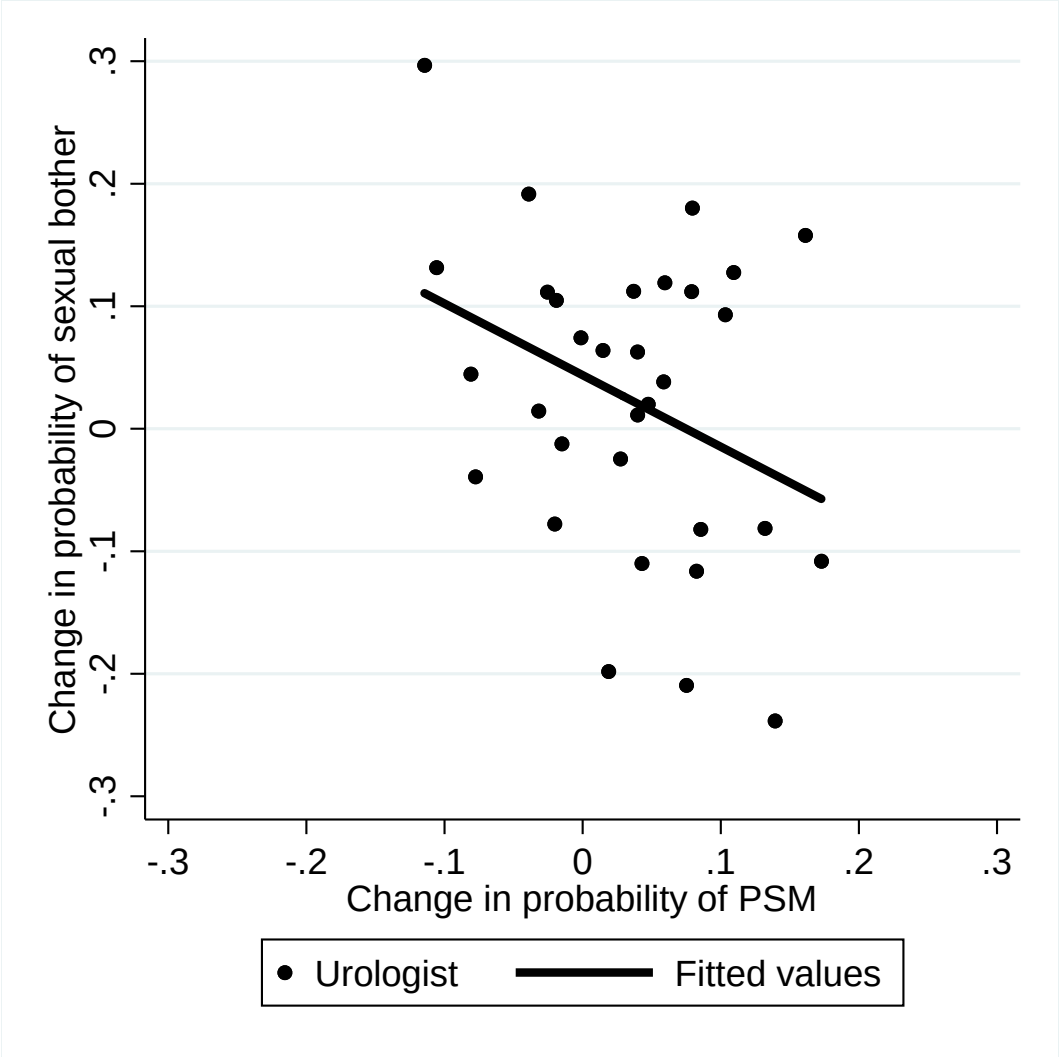


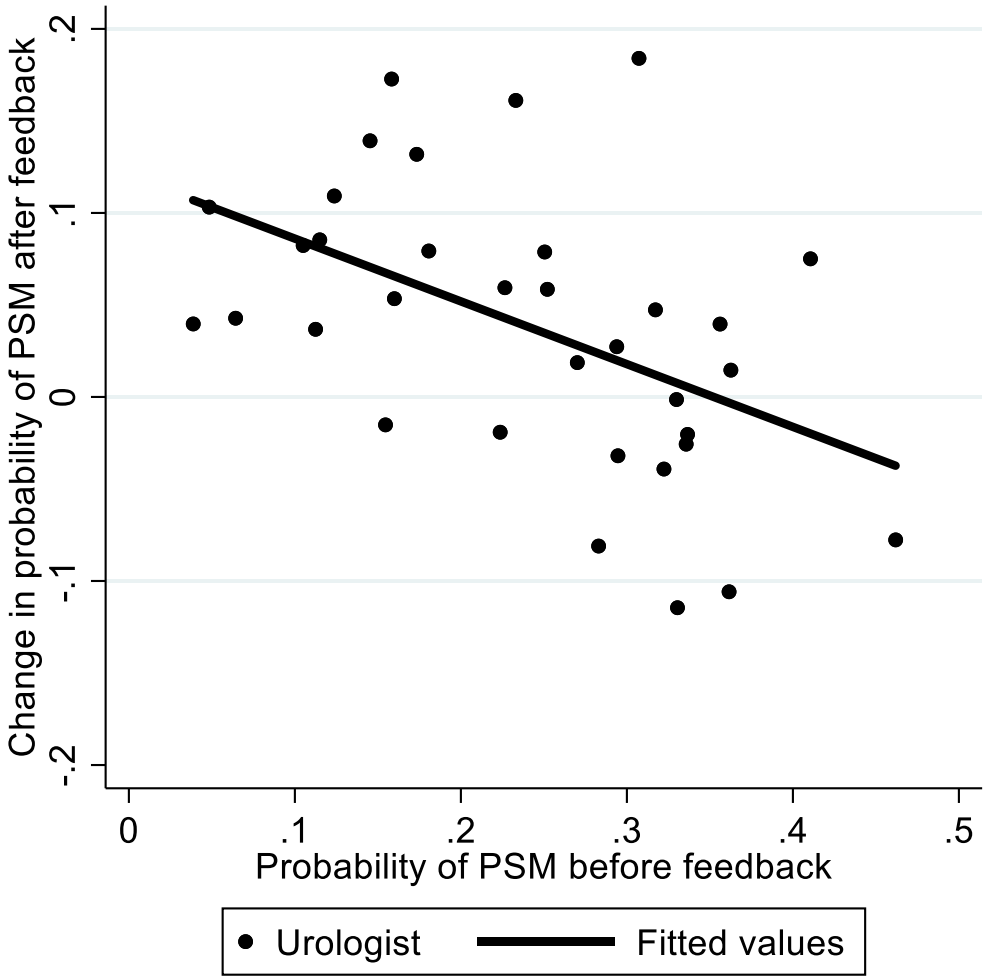
Figure 16 Relationship between change in PSM and sexual bother, by urologist



Note: For a 1% increase in PSM probability, there is a 0.58% decrease in the probability of sexual bother

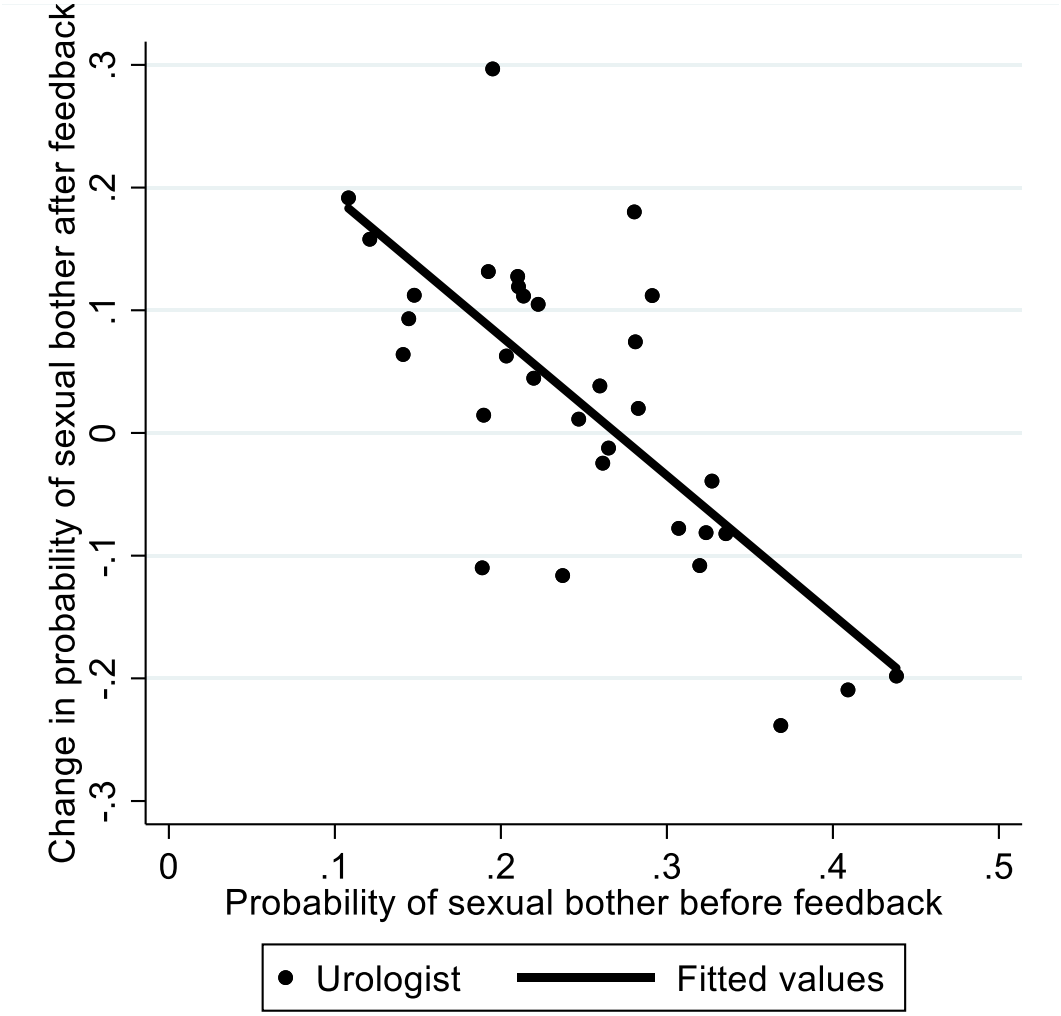
There appeared to be a negative trend in the change in PSM or sexual bother against each urologist's pre-feedback rates (Figure 17 and Figure 18, respectively). The higher the PSM rate or sexual bother prior to feedback, the greater the reduction following feedback. For urologists with the lowest rates of PSM or sexual bother, this appears to increase the likelihood of these outcomes following feedback. This negative trend is more substantial for sexual bother, where for every 1% higher a urologist's sexual bother rates were prior to feedback, their sexual bother rates reduced by 1.14% following feedback.

Figure 17 Change in probability of PSM following feedback compared to PSM rates prior to feedback, by urologist



Notes: for every 1% higher probability of PSM before feedback, reduce the likelihood by 0.34% following feedback. Also indicated that urologists reporting rates of PSM <26% may increase PSM rate following feedback

Figure 18 Change in probability of sexual bother following feedback compared to sexual bother rates prior to feedback, by urologist



Notes: for every 1% higher probability of sexual bother before feedback, reduce the likelihood by 1.14% following feedback. Also indicated that urologists reporting rates of sexual bother <26% may increase sexual bother rate following feedback

4.3.5. Role of additional treatment

The relationship of PSM and sexual bother to additional treatment (patients for whom surgery is not their only treatment) is briefly explored in Appendix 7.7, p153. This analysis examined the relationships between PSM, additional treatment and sexual bother at the urologist level, stratified by the urologist's feedback status.

The results demonstrate a positive relationship between PSM and additional treatment, particularly where an increase in PSM probability also results in a

probability of additional treatment for each urologist (Figure 29, p156). There was a slight increase in the probability of sexual bother as the probability of additional treatment increase for each urologist (Figure 30 p157). However, the overall rates of sexual bother changed from positively correlated with the overall rates of additional treatment prior to feedback to non-positively correlated following the introduction of feedback. Urologists with high rates of additional treatment prior to feedback appear to have more considerable reductions in additional treatment following feedback (Figure 31, p158).

4.4. Discussion

Before adjusting for patient and hospital characteristics, there was some evidence in the observed data of a small reduction in the mean probability of PSM and sexual bother across urologists. There was also some evidence from the raw data that PSM rates were negatively correlated with the likelihood of sexual bother if a urologist received feedback. However, once adjusted for patient characteristics (age, risk status) and hospital characteristics (whether the hospital is metropolitan or regional, public or private) the relationship between PSM and sexual bother was similar both pre- and post-feedback, with a small positive trend where sexual bother increase as PSM increase. However, a negative relationship was found between changes in PSM and changes in sexual bother probabilities. When the relationship between the changes in the likelihood of PSM and sexual bother was considered, results showed a negative relationship, suggesting that urologists responded to one quality indicator at the other's expense. Worst performing urologists (those with the highest rates of PSM or sexual bother) also see the most considerable improvements in outcomes. Urologists with the best performance prior to feedback appear to not improve following feedback, with some of them increasing their probabilities of patient sexual bother or PSM.

This analysis's magnitudes of effect sizes are biased towards zero, as the urologist error terms from the PSM and sexual bother analyses likely correlate. This analysis may, therefore underestimate the size of the relationship between changes in PSM and sexual bother.

One mechanism that may explain the relationship between PSM and sexual bother's probability is their relationship to additional treatment. The relationship between the likelihoods of additional treatment and PSM by each urologist appears unaffected by feedback and is positive in both periods. However, the relationship between the probability of additional treatment and sexual bother did appear to change after the introduction of feedback, with a negative relationship between the probability of additional treatment and sexual bother after feedback was introduced. Individual changes in additional treatment find that as urologists increased additional treatments, they also increased sexual bother and PSM rates. Therefore, additional treatment is unlikely to be the primary driver of changes to the probability of sexual bother. Instead, sexual bother is more likely influenced by the primary surgery's aggressiveness.

One of this analysis's strengths is that it uses recent individual patient data collected by a disease-specific registry, which captures up to 90% of all newly diagnosed prostate cancer cases in Victoria. However, it is limited by what is reported for each patient. It cannot account for the true patient case-mix for each urologist, such as complete data on comorbidities or family history.

This study also does not explicitly identify urologist characteristics that might explain the limited changes in patient outcomes seen in response to the introduction of feedback, as the data do not capture urologist characteristics. Also, data from urologists and their surgeries are restricted from the time they join the registry. Urologists and hospital stakeholders have identified PSM as an important quality indicator, but the results do not find any change in response to feedback. One barrier may be an individual urologist’s access to technology or perceived inability to improve technique. One change in practice that may facilitate the improvement of sexual bother rates rather than PSMs is the refinement of counselling to patients’ prior treatment, which may set lower expectations of sexual function following treatment (expert opinion, JM). Patients with lower expectations may be less likely to report sexual bother following treatment. The study also has limited extrapolation to older men, who are less likely to report sexual bother status, either through discomfort of discussing the topic, or acceptance that it is part of ageing.

This study also only accounts for feedback to the registry. All patient care and urologist behaviour are likely affected by the existence of the registry regardless of feedback.

Therefore, some of the influence of the registry may also be reflected in the time trends. To thoroughly explore whether this is the registry or just general time trends, Victoria would need to be compared to a state without a registry. These data are not available. The outcomes of interest, probability of PSM and sexual bother, are only modelled as binary outcomes. The analysis does not account for the severity of PSMs or sexual bother, which may also be important to patient prognosis and quality of life.

Using a linear probability model has many advantages, such as producing consistent estimates of the average marginal effects that are easy to interpret. However, linear probability models are not bounded and can produce probability estimates above 1 or below 0 for extreme values. However, this does not significantly affect this analysis where estimated average probabilities are not near 0 or 1. This study also does not consider the likely delay between receiving feedback on PSMs (data recorded at time of surgery) and receiving feedback on sexual bother (collected at follow up, a year after treatment). However, it is unlikely that this will have a large effect, as registry data collection is not automatic: specific staff members access the medical records to collect the registry data.

4.4.1. Conclusion

This study finds no change in the average likelihood of PSMs and limited evidence of a reduction in the likelihood of sexual bother as a response to urologist feedback. There was some evidence of a change in the relationship between PSMs and sexual bother rates following the introduction of feedback, suggesting some urologists respond to one quality indicator at the expense of another. There was also some evidence that urologists with the worst performance prior to feedback respond most strongly after feedback.

4.4.2. Further work

Further work should explore how individual urologists respond to feedback, to establish patterns of behaviour across individuals and identify the subgroups of urologists who may benefit most from feedback.

Additionally, work should explore urologist perception on multiple quality indicators, how they may consciously (or unconsciously) choose between them, and what is required to enable improvements in all indicators at either the feedback reporting or the facilitating level. It would also be of interest to compare indicators where different directional movement is desired. In this scenario, the desire is for both PSM and sexual bother rates to be reduced.

5. How does surgeon experience alter the effectiveness of robot-assisted surgery in localised prostate cancer?

5.1. Introduction

Prostate cancer is the most commonly diagnosed cancer in Australian men, with an estimated 1 in 7 men diagnosed with prostate cancer by their 85th birthday (Australian Institute of Health and Welfare, 2016). Initial treatment for prostate cancer is often to remove cancerous tissue surgically via radical prostatectomy. Laparoscopic radical prostatectomy (LRP) was introduced as an alternative to traditional open surgery for prostate cancer in 1991 in the USA (Schuessler, Schulam, Clayman, & Kavoussi, 1997). Due to the learning curve and limited perceived benefits, the technique never substantially replaced open surgery (Shuford, 2007). However, since the early 2000s in the USA and Australia, open surgery has been progressively replaced with RARP. The change in practice has been driven by the expectation that robotic surgery may provide benefits over open surgery, both in terms of surgery efficiency (e.g., reduced blood use, shorter surgery times), and patient outcomes (e.g. shorter recovery times). Financial incentives may also encourage surgeons to use RARP: surgeries conducted in private practice may charge more for RARP and access to new technologies may encourage patients to use their services due to expectations associated with a new technology, particularly where patients can see immediate benefits such as shorter recovery times. In Australia, robot-assisted surgery has been adopted in addition to traditional open surgery. However, the extent of the potential benefits of RARP over open surgery is likely influenced by surgeon experience and the frequency of operating. Therefore, a comparison of RARP to open surgery should consider how experience with RARP may influence surgery outcomes.

Previous studies into the comparative effectiveness of robot-assisted surgery compared to open surgery (open radical prostatectomy, ORP) have shown minimal differences in surgical and patient-reported outcomes (Cao et al., 2019; Huang et al., 2019; Ilic et al., 2018; Pan et al., 2015). Evidence suggests that RARP surgeries result in a shorter length of stay, lower blood loss, fewer transfusions and a reduction in

remaining tumour left behind after surgery (positive surgical margin, PSMs), supported primarily by non-randomised longitudinal studies and one Australian RCT (Barry, Gallagher, Skinner, & Fowler, 2012; Basto et al., 2016; Coughlin et al., 2018; Fode, Sønksen, & Jakobsen, 2014; Gardiner et al., 2012; Herlemann, Cowan, Carroll, & Cooperberg, 2018; Huang et al., 2019; Ilic et al., 2017; Leow et al., 2016; Medical Advisory, 2010; Pan et al., 2015; Ramsay et al., 2012; Robertson et al., 2013; Vora et al., 2013; Yaxley et al., 2016). Nearly all comparative effectiveness studies have included small numbers of surgeons or hospitals, and studies have been limited by the size and scope of the data collected. The non-randomised studies do not appear to control for selection bias. Results are confounded by attributing outcome improvements from RARP to the procedure, rather than the clinical skills of surgeons who switched to RARP or their choice of patients (e.g. patients with fewer comorbidities or lower risk of complications). Some international cost-effectiveness studies considered differences in comparative effectiveness of RARP, e.g., higher rates of preserved urinary continence, erectile function, no residual cancer (Hohwu et al., 2011) as important drivers of cost-effectiveness of RARP.

Studies suggest that a surgeon's experience with RARP may alter how effective RARP is versus ORP. Firstly, there is the learning curve with RARP. Earlier estimates varied, suggesting surgeons should perform between 20-1000 RARP to be proficient in RARP, with recent studies suggesting surgeons should perform at least 100 RARPs (Gumus, Boylu, Turan, & Onol, 2011; Jaulim et al., 2018; Sivaraman et al., 2017). These studies have been conducted on small numbers of surgeons (maximum 9). Though they have limited external validity to other surgeons, the studies suggest that the number of RARP performed prior may influence the outcomes of the current surgery.

Comparative effectiveness studies (including RCTs) often compare surgeons with high levels of experience. The one RCT of RARP versus ORP the two surgeons involved were highly experienced in their respective surgeries: the RARP surgeon had completed a two-year robotic fellowship, and at the start of the trial had performed 200 robotic prostatectomies (over 1000 by the end); the ORP surgeon had 15 years post-fellowship experience and had performed 1,500 operations at the start of the trial. Rarely do studies consider scenarios where surgeons have low levels of experiences

or conduct both RARP and ORP. Studies that assess the effect of RARP uptake on ORP outcomes have not been identified. There is some evidence that regularity of surgeries may also alter the effectiveness of radical prostatectomies, suggesting that 20 cases per year on average is the minimum cut off where the quality of surgery improves (Eastham et al., 2003; Vesey, McCabe, Hounsome, & Fowler, 2012). No studies were identified where the effect of the regularity of surgeries on patient outcomes had been explicitly explored for RARP.

Cost of RARP have been shown to be driven by the high cost of purchasing, maintain and operating equipment both in Australia (Basto et al., 2016) and internationally (Bolenz et al., 2014; Close et al., 2013; Liberman et al., 2012; Ramsay et al., 2012). In economic analyses, reduced blood use, surgery time, length of hospital stay can offset the high cost of RARP if a large number of RARP are conducted annually. In Victoria, a minimum of 140 RARPs annually per hospital was recommended to be cost-equivalent to ORP (Basto et al., 2016), but Basto et al. only considered short term cost offsets (e.g. blood use and length of hospital stay). Economic analyses of RARP versus ORP also suffered from low-quality comparative evidence as described above, particularly in terms of outcomes that indicate long term costs or benefits, and did not consider the effect of surgeon experience with RARP. RARP appears likely to remain a popular surgery choice among patients and surgeons despite the additional cost. It is therefore important to ascertain how to use RARP most efficiently, particularly whether there are circumstances where long term outcomes may be improved compared to ORP.

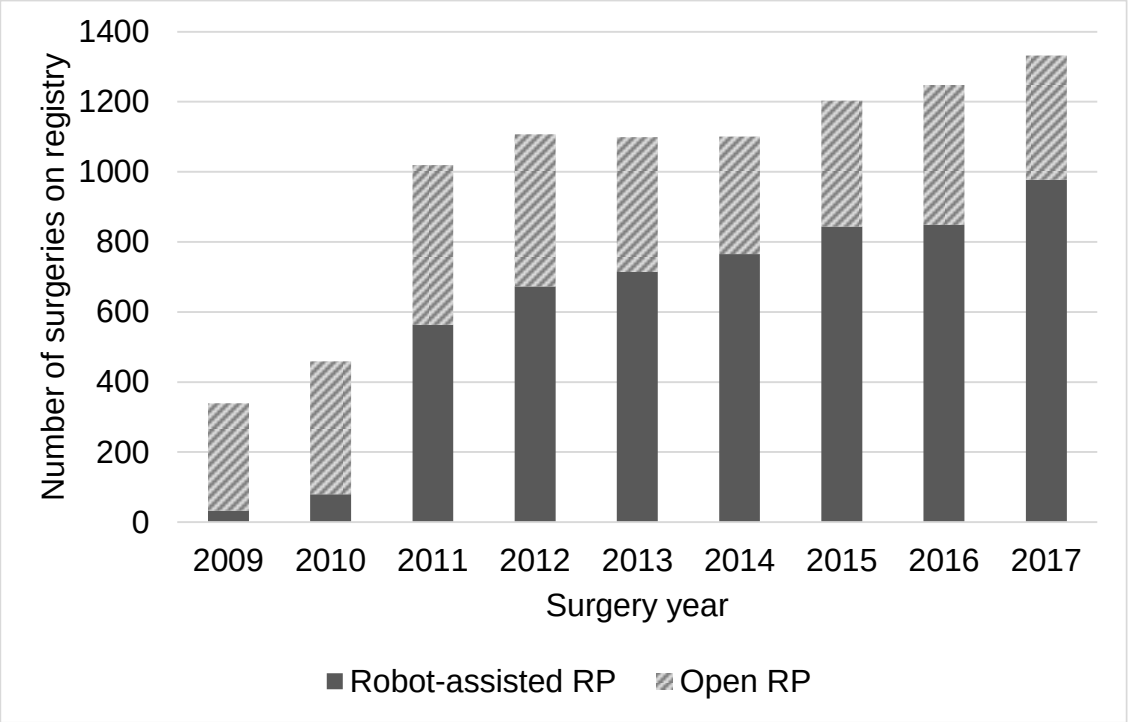
This study aimed to address these potential issues in the published studies to date, by taking advantage of a long-term registry, with 8 years of data on patients with clinical and quality of life measured within one year of surgery. The effect of surgeon RARP experience on these outcomes is explored by comparing ORP patients prior to surgeons commencing RARP, to 1) ORP patients whose surgeon had conducted at least one RARP, 2) RARP patients, grouped by surgeon experience of RARP (the annual volumes of RARP and numbers of RARP prior to their surgery).

5.2. Methods and Data

RARP is currently a popular method of surgery for prostate cancer in Australia. Between 2009 and 2018, 3,534 ORP (35.2% of all radical prostatectomies) and 5,745 RARP (57.3% of all radical prostatectomies) were recorded on the PCOR-Vic, with a minority (7.5%) of patients receiving other types of surgery. The PCOR-Vic collects information on >90% of newly diagnosed prostate cancer cases and therefore is likely indicative of all patients (Sampurno F and Evans SM (eds) for the Victorian Prostate Cancer Clinical Registry Steering Committee, 2015). This study focused on the two most common surgical modalities: RARP and ORP and excluded LRP as it only made up 6.5% of all surgeries from 2009-2018 (4.2% in 2017).

The PCOR-Vic also showed an increase in RARPs over time (Figure 19), as hospitals with robotic equipment joined the registry, and the procedure increased in popularity. For surgeons who joined before 2011 and perform both ORP and RARP, RARP accounted for less than 25% of all surgeries before 2011 and over 60% of surgeries from 2011, increasing up to ~80% of surgeries in 2015.

Figure 19 Number of radical prostatectomies on PCOR-Vic by surgery type



Notes: Increase in surgeries (and RARP in particular) seen in 2011 likely driven both by new hospitals joining the registry and increased access to robotic equipment. In 2012 and 2014 there were peaks of hospitals joining the registry (>6 in the year), but these were mostly hospitals with low numbers of surgeries that did not greatly increase the overall number of radical prostatectomies. The probability of RARP in hospitals that joined prior to 2011 is 55% over the whole period (7.5% prior to 2011), the probability of RARP in hospitals that joined from 2011 is 78%.

Outcome data available from PCOR-Vic

As well as data on surgery types, surgeon identifiers, and patient and hospital characteristics, the PCOR-Vic collects patient and surgery outcome data. Outcomes include the presence of a positive surgical margin (PSM), additional management and patient-reported outcomes at 12-month follow-up after surgery (median time to follow up of included patients who completed follow up was 11.1 months and 95% of patients received follow up within 13.6 months of surgery). Patient-reported outcomes take the form of SF-12 summary scores. The SF-12® (Short Form 12 item survey) (John E Ware, 2002; J. E. Ware, Kosinski, & Keller, 1996) is a quicker to administer subset of questions from the SF-36® (John E Ware, 2008; Ware Jr, 1999), which measures health status as reported by patient response across eight health domains. These are used to calculate aggregate summary scores: physical component summary and mental component summary (Turner-Bowker & Hogue, 2014). A higher score

indicates a better health state, and the mean population score in physical and mental summary scores is 50, with a standard deviation of 10 (Turner-Bowker & Hogue, 2014).

Probability of PSM and additional management are reported or recoded into binary variables, and the SF-12 summary scores are continuous.

A patient inclusion flow diagram is given in Figure 2. Patients were included if they received RARP or ORP at least two years prior to 1st Jan 2019. Brachytherapy and radiotherapy treatment information was restricted to 2 years following surgery. Patients are rarely followed beyond 2 years after initial treatment, and treatment beyond this period is less likely to be influenced by the initial surgery. Patients of surgeons who had less than a total of 5 patients of each surgery type recorded on the registry were excluded. Surgeons are restricted to those who entered the registry performing ORP and then conducted at least 5 ORPs and 5 RARPs to create a more homogenous group of surgeons based on their experience and reduce confounding on selection into treatment by surgeon characteristics.

For each patient, data were available for their age, risk status (including PSA level, Gleason score, clinical T stage and the combination NCCN risk level), surgeon, hospital (including whether it was public or private, metropolitan or regional), date of surgery, surgery type, date of other treatments, follow up information. SEIFA decile was also estimated based on the reported postcode for each patient and patients were excluded if this could not be calculated (this affected five patients)

Outcomes that relate to surgery effectiveness

Ultimately, it should be ascertained whether RARP offers value for money in the long-term, once surgeons have sufficient experience. Six surgery groups by surgery type and surgeon experience with RARP are identified, which give five treatment groups vs a control comparison group for the analysis:

- Control group: patients who received ORP prior to their surgeon commencing RARPs (ORP before RARP);

- Treatment group 1: patients who received ORP after their surgeon began performing RARPs (ORP after RARP);

- Treatment group 2: patients who received RARP from low volume RARP surgeons (<20 per year) and were one of the first 50 recorded RARPs for their surgeon (RARP, low vol., 1st 50);

- Treatment group 3: patients who received RARP from low volume RARP surgeons (<20 per year) after the first 50 recorded RARPs for their surgeon (RARP, low vol., after 1st 50);

- Treatment group 4: patients who received RARP from high volume RARP surgeons (20+ per year) and were one of the first 50 recorded RARPs for their surgeon (RARP, high vol., 1st 50);

- Treatment group 5: patients who received RARP from high volume RARP surgeons (20+ per year) after the first 50 recorded RARPs for their surgeon (RARP, high vol., after 1st 50).

ORPs that occur prior to surgeons commencing RARP are chosen as the control group as patient outcomes are unaffected by a surgeon's experience with RARP. As some surgeons switch to majority RARP once they begin, ORPs conducted after they have commenced RARP could decline in quality due to the fewer ORPs performed. A comparison of ORPs prior to and after the surgeons commence RARP assesses whether there is a significant change in ORP quality after RARP commenced. 50 surgeries were chosen as a large experience threshold that resulted in a substantial number of surgeries for each treatment group (minimum 99) for which matching could occur. Only 2 surgeons on PCOR-Vic had performed more than 100 RARPs. Annual volume was based upon the literature (Eastham et al., 2003; Vesey et al., 2012).

Potential indicators of treatment quality include survival and patient quality of life. However, operable localised prostate cancer is associated with long survival for patients, and the registry does not yet collect long-term survival data. Instead, outcome measures associated with surgical quality and some shorter-term prognostic and patient outcomes are chosen. The most widely accepted measure of surgical quality is the extent of positive surgical margins (PSMs). The presence of tumour tissue

remaining after surgery is associated with higher rates of biochemical recurrence and poorer patient prognosis (Boorjian et al., 2010; Oh et al., 2013; Ploussard et al., 2011; Silberstein & Eastham, 2014; Stephenson et al., 2009; Swindle et al., 2005; Wright et al., 2010). Patient quality of life is measured through summary scores from a widely used generic quality of life measure (SF-12) (John E Ware, 2002; J. E. Ware et al., 1996), recorded during patient 1-year follow-up appointments.

There is a concern that cancer will recur in the longer term, and treatment is more effective if it reduces the rate of recurrence. An additional outcome, the likelihood of a patient receiving additional curative treatment after surgery (i.e. brachytherapy or radiotherapy), is used as a proxy measure of a surgeon's expectation of cancer recurrence. Additional management may also be associated with worse quality of life for patients (Alsinnawi et al., 2019) and increased total cost of treatment. The decision to provide additional management is not based solely on the initial surgery received by a patient. Interpretations of any differences in additional management as an outcome of surgery type alone must be made cautiously. Surgeon and patient preference, and patient circumstance play a role in treatment decisions, which are not observed in the data. Furthermore, because treatment after surgery is not reported to the surgeon (unlike PSM), there is some concern that this information may not be as accurately recorded in the registry, and may systematically underestimate the numbers of patients receiving additional management.

5.2.1. Effectiveness of robotic surgery in prostate cancer surgery using real-world data

5.2.1.1. Methods: adjusting for confounding

As PCOR-Vic data is observational, patient selection into treatment (the type of surgery) and surgery outcomes will confound the relationship between treatment type and outcomes. For example, older patients are less likely to be offered robotic surgery because of a poorer prognosis (Fode et al., 2014; Florian Rudolf Schroeck et al., 2017; Tang et al., 2017). This link between prognosis and surgery type will bias the treatment effect in favour of robotic surgery. Inverse-probability weighting (IPW) is used to adjust for observed confounding by estimating a population in which the treatment is

independent of the measured confounders. As such, IPW improves the balance within each of the two treatment allocation groups (Mansournia & Altman, 2016; Robins, Hernán, & Brumback, 2000). The IPW approach has been used previously for comparing radical prostatectomy outcomes across different institutions (Pooli et al., 2020).

Comparing those who got robotic surgery with those received an open surgery IPW allows additional weight to those patients who received an open surgery who were more similar to those that received robotic surgery.

To estimate the required inverse probability weights, the confounders that affect treatment choice and outcomes are identified and included within a logistic regression of the form:

$$log\left(\frac{P(RARP)}{1-P(RARP)}\right) = log\left(\frac{P(RARP)}{P(ORP)}\right) = \mathbf{X_i B} + \epsilon_i \qquad (5.1)$$

Where $\mathbf{X_i}$ is the group of confounders that may affect surgery choice and ϵ_i the error term for patient *i*.

Inverse probability weights for each patient are estimated such that

$$w_i = \begin{cases} 1, & \text{if patient } i \text{ received RARP} \\ \dfrac{P(RARP|X_i)}{1-P(RARP|X_i)}, & \text{if patient } i \text{ received ORP} \end{cases} \qquad (5.2)$$

Identifying confounders

Many confounders affect both surgery choice and patient outcomes, and patients may be more likely to receive RARP over ORP and have better surgical and patient outcomes if patients are:

- Younger (Fode et al., 2014; Florian Rudolf Schroeck et al., 2017; Song, Lee, Lee, & Hong, 2019; Tang et al., 2017)
- Lower risk at diagnosis (S. M. Evans et al., 2014; Fode et al., 2014)
- Treated in a private hospital [PCOR-Vic data](S. M. Evans et al., 2014)

Evidence from the PCOR-Vic also suggests that patients who live in wealthier areas, who attend metropolitan (vs regional) hospitals, and who have later surgery dates are associated with a higher likelihood of RARP over ORP.[PCOR-Vic data](S. M. Evans et al., 2014)

There are other biases for surgery choice for which the PCOR-Vic does not provide data. For example, receiving RARP is associated with healthier patients (e.g., lower BMI, fewer comorbidities) (Florian Rudolf Schroeck et al., 2017). For this reason, ORPs that occur prior to RARP is the chosen control group, as these patients will be unaffected by these biases. However, there may also be temporal changes in comorbidities and BMI which lead to bias. The PCOR-Vic does not report BMI information and only reports comorbidity status for 147 patients (or which 9 report a comorbidity) and therefore SEIFA, hospital status (e.g. private) and age are hoped to also capture possible BMI and comorbidity biases. In the RCT of RARP versus ORP, similar rates of comorbidities were recorded in the two surgery arms after patients were stratified by age group (Yaxley et al., 2016), suggesting the inclusion of age as a confounder may control for some of the comorbidity biases.

Surgery date is not controlled for in either the surgery choice or outcome models. Surgery date is an almost perfect predictor of whether a patient receives RARP. Only 13 RARPs were conducted prior to 2011, and matching these outliers would give them unreasonably large weights in the outcome model. Individual surgeon effects are not included because of limited data on individual surgeon characteristics.

PCOR-Vic does collect data on both regional and metropolitan hospitals. However, most regional hospitals do not conduct RARP (8.9% of patients at regional hospitals receive RARP), and a large proportion of regional patients attend one hospital. Therefore, it would not be appropriate to match RARP metropolitan patients to ORP regional patients. The base case restricts to metropolitan hospitals. This restriction does result in few ORPs prior to RARP after 2011 (4 between 2012 and 2014).

A complete list of included confounders for the surgery choice and outcome models is given in Table 11.

Table 11 Confounders for surgery choice and outcome models

Confounder	Grouping in model	Reason confounder is in surgery choice model	Reason confounder is in outcome model
Age	<55, 55-74 years old	Younger patients more likely to receive RARP (Fode et al., 2014; Florian Rudolf Schroeck et al., 2017; Tang et al., 2017)	Younger patients have better outcomes (particularly functional) (Song et al., 2019)
Patient risk status	NCCN intermediate, high risk	Lower risk patients more likely to receive RARP (Fode et al., 2014)	Lower risk patients have better outcomes (particularly PSM) (S. M. Evans et al., 2014)
Hospital public or private	public, private	Patients of private hospitals more likely to receive RARP [PCOR-Vic data on file]	Private hospitals associated with better patient outcomes (particularly PSM) (S. M. Evans et al., 2014)
SEIFA decile	<40%, 40-60%, >60%, unknown	Patients from wealthier areas more likely to receive RARP [PCOR-Vic data on file]	Higher SEIFA decile associated with better patient outcomes (particularly PSM) [PCOR-Vic data on file]

Notes: NCCN= National Comprehensive Cancer Network; PCOR-Vic= Prostate Cancer Outcomes Registry-Victoria; PSM= positive surgical margin, RARP= robot-assisted radical prostatectomy; SEIFA= Socio-Economic Indexes for Areas

The IPW for ORP prior to the availability of RARP versus each treatment was estimated for each of the five comparisons using equation (5.3). Patients were grouped within each variable where similarities were expected to control for extreme outliers and avoid overidentification. No RARPs of high-volume surgeons were conducted in public hospitals, and therefore hospital status (public or private) was excluded as a variable in the IPW. Covariates pre and post weighting for each surgery type are compared to assess how well the model balances the covariates.

$$log\left(\frac{P(RARP)}{1-P(RARP)}\right) = \beta_0 + \beta_1 \mathrm{NCCN}_i + \beta_2 \mathrm{Age}_i + \beta_3 (\mathrm{SEIFA}\ 41-60\%)_i + \beta_4 (\mathrm{SEIFA}\ 61-80\%) + \beta_5 (\mathrm{SEIFA}\ 81-100\%)_i + \beta_6 \mathrm{Private}_k + \epsilon_i \quad (5.3)$$

Where NCCN=NCCN risk (intermediate, high), SEIFA= SEIFA decile (<40%, 41-60%, >60%), Age= age group (<55, 55-74), Private=hospital is private (vs public).

Estimating the average effect of surgery type

Following the identification of the IP weights, a linear regression was fit using the covariates included in the IPW model for each of the 5 comparisons versus ORP prior to surgery

To estimate weighted linear regressions using all covariates included in the IPW model. Robust standard errors are calculated. Comparisons were made based on how the outcomes of RARP varied by the average annual volume of RARPs and the number of prior RARPs (first 50 RARPs, more than 50 RARPs). To explore the indirect effect of surgeon experience of RARP on outcomes for ORP, an indicator of if the patient received ORP after each surgeon began performing RARP is included. The outcome model for the probability of PSM following surgery is therefore built as follows, in equation (11):

$$P(\mathrm{PSM})_{\mathrm{i}} = \gamma_0 + \gamma_1 S_i + \gamma_2 \mathrm{NCCN}_i + \gamma_3 \mathrm{Age}_i + \gamma_4 (\mathrm{SEIFA}\ 41-60\%)_i + \gamma_5 (\mathrm{SEIFA}\ 61-80\%) + \gamma_6 (\mathrm{SEIFA}\ 81-100\%)_i + \gamma_7 \mathrm{Private}_k + \varepsilon_i \quad (5.4)$$

Where S_i= surgery type and surgeon experience (patient *i* receives surgery after RARP introduction vs. patient receives ORP prior to RARP begins for surgeon [base]. NCCN=NCCN risk (high vs intermediate [base],), Age= age group (55-74 years vs <55 [base]), SEIFA= SEIFA decile (40-60%, >60% vs <40% [base]), Private=hospital *k* is private (vs public). As with the IPW, for the comparison of high-volume surgeons only private hospitals are included.

Separate IPW and linear regressions are estimated for each pairwise comparison of RARP experience versus the ORP before RARP control group, and therefore equations (5.1-5.4) are repeated for each comparison.

The IPW and linear regression were estimated using the '*teffects ipwra*' command in Stata/SE 16.0 (StataCorp, 2019). As the primary interest is what would have happened to treated patients had they not received RARP (or ORP after RARP commenced), the focus is on the average treatment effect for the treated (ATET). A covariate balance check was performed after weighting for each analysis.

Additional outcomes

IPW and weighted linear regressions were also calculated for the other outcomes: the probability of additional management and the SF-12 summary scores. As no SF-12

data was recorded for public hospital patients who received RARP from low volume surgeons (when their surgeon has conducted >50 RARPs) or RARP from high volume surgeons, the IPW and regression analyses were restricted to private hospitals for these comparisons.

5.3. Results

5.3.1. Patient characteristics

Full patient characteristics are given in Appendix 7.9, p160. Of the 1,816 included patients, 1,099 received RARP. RARP patients tended to live in wealthier postcodes, be of lower risk, were more frequently treated in metropolitan and private hospitals, with later surgery dates. Age at surgery was similar in both groups.

PSM status was recorded for 1,786 patients, and SF-12 summary scores were recorded for 887 patients.

5.3.2. Base case results

The full IPW and regression models for PSM are given in Appendix 7.10, p166. In general, lower NCCN risk and attendance at a private hospital are associated with a higher likelihood of RARP. For RARP patients of high-volume surgeons, living in a wealthier postcode was associated with an increase in the likelihood of PSM. For lower volume surgeons, there was a negative association between wealthier postcodes and likelihood of receiving RARP, which may be offset by whether the patient attended a private hospital (which gave a probability of receiving RARP above 100%). In all outcome models, higher NCCN risk was associated with a statistically significant increase in PSM probability at 5% threshold. Attending a private hospital was associated with a reduction in PSM for patients of surgeons with RARP experience compared to attending a public hospital, especially for ORP patients after RARP commenced (-16.0%, $p<0.001$, 95%CI -26.0, -5.9%), but no statistically significant difference in PSM was seen for private versus public hospitals for ORP patients prior to RARP commencing ($p>0.10$).

The average treatment effects for the treated (i.e. patients who received surgery after RARP commenced for their surgeon) for each of the five comparisons are given in Table 12. Counterfactual estimates differed across the comparisons (generally reducing as experience with RARP increased). Covariate balance was tested in each analysis, and none were found that rejected the null hypothesis that the covariates are balanced.

RARP was associated with a reduction in PSM compared to ORP prior to RARP for all experience levels. However, the magnitude and significance of this reduction were larger when RARP experience was greater, such that patients of low volume surgeons saw a statistically significant reduction in PSM at the 10% level compared to ORP prior to RARP when surgeons had conducted more than 50 RARPs (-10.5%, p=0.066, 95% CI -21.4, 0.7%). Patients of high-volume surgeons saw a statistically significant reduction in the probability of PSM as a 5% threshold, even for patients who were in the first 50 RARPs for their surgeon (-16.2%, p=0.016, 95%CI -29.3, -3.0%)

ORP after RARP commenced, showed a statistically non-significant reduction in the probability of PSM compared to ORP prior to RARP (-1.8%, p=0.703, 95%CI -11.0, 7.5%)

Table 12 Probability of PSM by surgery type and surgeon experience of RARP

Comparison	N (weighted)	P(PSM) (%)	Robust SE (%)	p- value	Lower 95% CI (%)	Upper 95% CI
ORP comparison (before and after RARP begins)	516					
Counterfactual (ORP before RARP)	144 (258.0)	37.3***	4.1	<0.001	29.3	45.3
ATET ORP after RARP vs ORP before RARP	372 (258.0)	-1.8	4.7	0.703	-11.0	7.5

Comparison	N (weighted)	P(PSM) (%)	Robust SE (%)	p- value	Lower 95% CI (%)	Upper 95% CI
ORP vs early low volume RARP experience	700					
Counterfactual (ORP before RARP)	144 (352.9)	37.0***	4.9	<0.001	27.3	46.6
ATET RARP low vol 1st 50 pts vs ORP before RARP	556 (347.1)	-6.9	5.3	0.187	-17.2	3.4
ORP vs established RARP low volume	352					
Counterfactual (ORP before RARP)	144 (179.1)	36.4***	4.9	<0.001	26.9	46.0
ATET RARP low vol >50 pts vs ORP before RARP	208 (172.9)	-10.5*	5.7	0.066	-21.6	0.7
ORP vs early RARP high volume	170					
Counterfactual (ORP before RARP)	71 (85.1)	34.4***	5.6	<0.001	23.3	45.4
ATET RARP high vol 1st 50 pts vs ORP before RARP	99 (84.9)	-16.2**	6.7	0.016	-29.3	-3.0
ORP vs established high vol RARP	280					
Counterfactual (ORP before RARP)	71 (140.0)	32.0***	5.9%	<0.001	20.5	43.6
ATET RARP high vol >50 pts vs ORP before RARP	209 (140.0)	-13.8**	6.4	0.031	-26.4	-1.2

Notes: * p<0.10, ** p<0.05, ***p<0.01; counterfactual refers to the probable outcomes estimated for the treated if they were untreated; ATET=average treatment effect on the treated; counterfactual= ORP prior to surgeon commencing RARP (ORP before RARP), treated= level of RARP experience.

5.3.2.1. Additional patient and surgical outcomes

Additional outcome results are presented in Table 13. RARP experience was associated with a non-statistically significant reduction (numerically worse) in SF-12 physical summary scores compared to ORP prior to RARP, except for patients who

received RARP from high volume surgeons who completed at least 50 surgeries where there was non-significant improvement. Low volume RARP, particularly for later patients, had the most considerable reduction in SF-12 physical summary scores than ORP patients prior to RARP, (score change of -5.42, p=0.111, 95% CI -12.09, 1.24), but these results were not statistically significant. There is some evidence that RARP experience was associated with an increase in the SF-12 mental summary score. In particular, for the initial 50 RARPs, there was an increase in SF-12 mental summary score is statistically significant at a 5% threshold: for patients of low volume surgeons there is an increase in score of 2.18 (p=0.034; 95%CI 0.17, 4.18) and for high volume surgeons there is an increase in score of 2.38 (p=0.035; 95%CI 0.17, 4.58)

There is evidence that RARP experience is associated with a reduction in additional management, but this is not statistically significant for any level of RARP experience compared to ORP prior to RARP. Additional management rates are estimated to be low for all patients (<10% in the counterfactual arm of all comparisons) and in the comparison of RARP patients of high-volume surgeons who have conducted >50 surgeries versus ORP prior to RARP patients, the probability of additional management for ORP prior to RARP patients is not statistically significant at a 5% level, suggesting the absolute probability of additional management for these patients is not proven to be different from 0%.

Table 13 Additional outcomes by surgery type and surgeon experience of RARP

Outcome	Comparison (Treatment vs ORP before RARP)	Raw N	Estimated counterfactual (95%CI)	ATET (95% CI)
Additional management	ORP after RARP	532 (378 treated)	8.9%*** (4.6%,13.2%)	-2.3% (-7.3%,2.7%)
	RARP low vol 1st 50 pts	717 (563 treated)	6.6%*** (2.0%,11.2%)	-2.5% (-7.3%,2.3%)
	RARP low vol >50 pts	364 (210 treated)	6.5%*** (2.1%,10.9%)	-2.7% (-7.8%,2.4%)
	RARP high vol 1st 50 pts	178 (99 treated)	4.8%** (0.2%,9.4%)	-3.8% (-8.8%,1.2%)
	RARP high vol >50 pts	288 (209 treated)	3.6%* (-0.4%,7.6%)	-1.7% (-6.1%,2.7%)

Outcome	Comparison (Treatment vs ORP before RARP)	Raw N	Estimated counterfactual (95%CI)	ATET (95% CI)
SF-12 Physical summary score	ORP after RARP	381 (229 treated)	51.66*** (50.39,52.93)	-1.41 (-3.11, 0.29)
	RARP low vol 1st 50 pts	413 (261 treated)	52.77*** (51.49,54.06)	-0.58 (-2.15,0.98)
	RARP low vol >50 pts	87 (9 treated)	53.44*** (51.39,55.50)	-5.42 (-12.09,1.24)
	RARP high vol 1st 50 pts	173 (95 treated)	53.11*** (51.64,54.59)	-0.94 (-3.11,1.22)
	RARP high vol >50 pts	124 (46 treated)	53.38*** (51.97,54.78)	0.85 (-1.44,3.13)
SF-12 Mental summary score	ORP after RARP	381 (229 treated)	51.19*** (49.73,52.65)	1.57 (-0.34,3.48)
	RARP low vol 1st 50 pts	413 (261 treated)	51.46*** (49.79,53.13)	2.18** (0.17,4.18)
	RARP low vol >50 pts	87 (9 treated)	51.75*** (49.63,53.86)	1.50 (-4.85,7.84)
	RARP high vol 1st 50 pts	173 (95 treated)	51.88*** (50.08,53.69)	2.38** (0.17,4.58)
	RARP high vol >50 pts	124 (46 treated)	51.56*** (49.55,53.57)	2.93* (-0.44,6.29)

Notes: *p<0.10, **p<0.05, *** p<0.01. Patients reweighted for SF-12 analyses. ORP before RARP= ORP occurs prior to surgeon commencing RARP

5.4. Discussion

5.4.1. Summary

This analysis is the first to examine the impact of surgeon experience of RARP on RARP outcomes and ORP outcomes, using a large dataset of typical prostate cancer surgeons and patients.

Like previous research, this study suggests that the uptake of RARP in high volume surgeons was associated with some benefits to patients compared to ORP (reduced PSM, higher SF-12 mental health scores), even when controlling for surgery choice factors. Patients who receive RARP from surgeons with less experience (fewer total

number of RARP, lower annual volumes of RARP) also see a non-significant decrease in PSM for their first 50 patients compared to ORPs prior to RARP. For patients receiving ORP, there is limited evidence that outcomes for ORP may improve after RARP is introduced (e.g. decrease in PSM compared to ORP prior to first RARP for each surgeon).

There is some evidence that RARP outcomes improve with more RARPs performed, especially for low volume surgeons: the probability of PSM reduces above a 50-patient threshold, compared to the first 50 RARPs, although this is not significant. In the SF-12 analyses, no patients were identified who had surgeons with over 100 RARPs, and 9 patients were identified as receiving RARP by low volume surgeons who had previously performed more than 50 RARP, so the significance of these results should be interpreted with caution.

The analysis found some evidence that experience with RARP is associated with a reduction in the likelihood that a patient is recorded as receiving additional radiotherapy or brachytherapy within 2 years of surgery. However, there were many factors that the study could not account for (e.g., input of the multidisciplinary team for each patient, hospital procedures, financial incentives), which may influence surgery choice as well as later treatment decisions. Therefore, reporting of these additional managements and these results should be interpreted with caution. Though there are financial incentives for subsequent brachytherapy or radiotherapy in private practice, the raw numbers from the PCOR-Vic suggest that patients were more likely to receive additional management if they received surgery in a public hospital (6.6% public hospital versus 3.7% private hospital), suggesting those incentives may not drive the additional management rates.

This analysis has also provided insight into surgeons' behaviour in a healthcare system with multiple surgery options. Of the 21 included surgeons who performed both RARP and ORP, 9 had performed more than 50 RARPs in the dataset's timeframe and 4 had an average annual volume of at least 20 patients. Therefore, most surgeons are not conducting as many RARPs as recommended in the effectiveness and cost-effectiveness literature (upwards of 100 RARPs per year) (Basto et al., 2016; Gumus et al., 2011; Leow et al., 2016).

One concern with the analysis was the restriction to metropolitan hospitals because very few RARPs were conducted in regional hospitals. While this reduces the effect of location on surgery choice and outcome, it results in very few ORPs prior to the introduction of RARP after 2011. Thus, adjustment cannot be made for surgery date in the analysis and may overestimate the effect of RARP experience if surgeries, in general, improve over time.

Outcomes from the comparison of ORP prior to and after RARP commences were not statistically significant. However, the change for ORP patients after RARP began for each outcome was similar in magnitude and direction to patients receiving RARP. It is difficult to separate whether this resulted from the changes over time (and an increase in experience of radical prostatectomies in general) or the introduction of RARP.

5.4.2. Strengths and limitations

One limitation of the study is the lack of information on surgeon experience prior to joining the registry. Therefore, the quantity of training and experience surgeons previously had for each surgery type is unknown. The effect of RARP for the first 50 patients may be exaggerated as the surgeons have had more experience of RARP than expected. The difference in patients receiving ORP before and after RARP commences could also be underestimated. This study tried to limit to surgeons who were developing RARP experience, by restricting to surgeons who perform both types of surgery and ensuring that ORP was the first surgery type for each surgeon recorded by the registry.

Surgeons with varying numbers of surgeries were included to reflect the healthcare system as a whole. Some surgeons had small numbers of patients, so the models were not adjusted for individual surgeon fixed effects. Therefore, the results applied across surgeons, rather than the effect of RARP experience for individual surgeons.

The majority of surgeries included in the analysis were conducted in private hospitals, which are highly correlated with the likelihood of receiving RARP and better outcomes. While this may reflect the Australian system, the extrapolation of these results to other healthcare settings should be conducted with caution.

There may be significant confounders that are not included in the IPW. The registry has limited information on comorbidity, BMI and family history that could influence surgery choice and outcomes. It is therefore possible that unintentional imbalances in these risk factors may have occurred through the matching process. Similarly, there are only limited short-run outcome data reported by the registry. Longer-term outcomes, such as biochemical recurrence or survival; or cost relevant outcomes such as blood use, length of surgery, length of hospital stay; could indicate other costs and benefits of each surgery type.

One of this study technique's advantages is that IPW provides the ability to adjust longitudinal data such that it mimics the random selection process of an RCT under certain assumptions (e.g. no selection on unobservables). Furthermore, IPW is a doubly robust method, meaning that either the treatment or outcome model may be mis-specified, and the estimator would remain unbiased (Funk et al., 2011). Furthermore, unlike other matching methods, IPW does not require the removal of patients from the analysis.

However, IPW regressions tend to be less precise than correctly specified maximum likelihood estimators using correctly specified models. Individuals with extreme characteristics may lead to unstable estimates and large standard errors (Funk et al., 2011). To minimise the effect of extreme characteristics in the analysis, they were grouped by similarity.

Linear regressions were used for outcome models, as these are easier to interpret. However, these are unbounded, which may produce erroneous predictions for probability outcomes near 0 or 1. One such instance is the additional management outcome model where 3-6% of patients receive additional management depending upon surgery type, and confidence intervals could extend below 0.

5.4.3. Conclusion

Overall this analysis supports the conclusion that RARP has similar outcomes to ORP (and introduction of RARP has no significant effect on ORP outcomes). It also suggests that RARP effectiveness improves with higher volume. As RARP is an expensive technology, the findings do not provide evidence of the cost-effectiveness

of RARP compared to ORP. However, it does not capture short term costs and outcomes associated with a less invasive procedure. In previous studies, reduction in length of stay and blood transfusions associated with RARP resulted in similar total costs for RARP and ORP for high volume hospitals (Basto et al., 2016).

Given the fast uptake of RARP on the PCOR-Vic (and the subsequent reduction in ORPs) with no immediate significant observable changes in surgery outcomes, potential policy implications surround the uptake of these kinds of costly technologies in health care. One potential suggestion from this analysis is that consideration of surgeons' training needs may be required to gain the full benefits of the technology if, as found here, lower volume surgeons may be expected to have longer learning curves.

6. Conclusion

This thesis aimed to examine whether the quality of care can be improved through changes in provider behaviour, notably through interventions that reduce unwarranted variation or through experience of new technologies. Prostate cancer was chosen as a costly disease to the healthcare system, where there may be systematic variation, making the quality of care less efficient. The thesis focused on guidelines for prostate cancer diagnosis, feedback on management decisions, feedback on surgical and patient outcomes, and surgeon experience of robot-assisted surgery on surgical and patient outcomes.

Chapter 2 considered the effect of GP guidelines to reduce overdetection of prostate cancer. Data included prostate cancer incidence from the New South Wales, Queensland, Western Australia and Victorian cancer registries, PSA testing incidence (as recorded by Medicare) and population numbers reported by the Australian Bureau of Statistics. All data were recorded between 2002 and 2015. ITSA were used to compare incidence rates, testing rates and cases per test between Victoria and an average of the other states both pre- and post-2009. All states saw an increase in PSA testing and prostate cancer cases detected prior to 2009, with Victoria showing a similar increase and absolute rate of cancer detected, but with a faster increase in PSA testing rate than the other states, and a faster reduction in cases detected per 10,000 PSA tests than the other states. From 2009, Victoria saw a faster reduction in prostate cancer cases detected and PSA tests per 100,000 men than the other states. However, neither Victoria nor the other states saw a significant increase in cases detected per 10,000 PSA tests from 2009. Victoria saw the most significant change in cases per test, particularly for men 75 and over, where the rate of case per test remained near-constant from 2009 compared to decreasing prior to 2009. The results provide some evidence that all states responded to guideline changes by reducing the rate of testing and cases detected (for men 45 and above), and that Victoria has different testing practices to the other states. Because there are differences in PSA testing rates between Victoria and the other states prior to 2009, it is unclear if Victoria responded to a change other than the guidelines after 2009 (e.g., introducing the Prostate Cancer Outcomes Registry-Victoria). Though there is evidence of a reduction PSA testing, there is only some evidence that the guidelines have improved the

efficiency of testing for men aged 45 years and over, suggesting that patient selection for testing could be better directed. Similarity cross states of downstream testing could not be assessed and may have affected the rate of cancers detected. Information on the family history, risk or staging stratification of patients across states was also not available. This information may drive testing (e.g. men with higher-risk prostate cancer more likely to present with symptoms, or men with family history more likely to be tested) and demonstrate improved efficiency of testing (e.g. PSA testing should result in cancers diagnosed earlier). The expansion of the PCOR to across Australia may be an opportunity to collect this data and further explore the testing rates.

One of the other ways that evidence-based recommendations have attempted to reduce overdetection is by also reducing the number of men who receive curative management (e.g., surgery or radiotherapy) for those diagnosed with low-risk disease. Chapter 3 explored how peer comparative feedback to urologists reinforced these recommendations. It also explored the effect of feedback concerning the recommendation that high-risk patients should receive active treatment. Time to curative treatment (surgery, radiotherapy, brachytherapy) is explored in the year following diagnosis, before and after feedback begins. Individual patient data from the PCOR-Vic identified time to curative treatment, managing clinician, patient and hospital characteristics. Cox-regression survival analyses were performed on time to curative treatment and the effect of feedback. Once patient and hospital characteristics and clinician fixed effects are accounted for, feedback to clinicians was associated with a statistically significant lower likelihood of receiving treatment within the first year following diagnosis for low and intermediate-risk men. There was also a small reduction in treatment for high-risk men, but this was not significant. These results suggest there is some evidence that feedback reinforced recommendations to reduce treatment in low-risk men. The analysis did not find evidence that feedback reduced time to curative treatment for high-risk men. Potentially, treatment in high-risk men was already at capacity (e.g., treatment rates for high-risk were already optimal). Alternatively, in the last ten years, international guidelines have supported recommendations to reduce overtreatment in low-risk men, rather than the treatment decision for high-risk men.

The analysis did not investigate the effect of a time since diagnosis greater than one year. Advice for not treating low-risk men extends beyond one year, but identifying the managing clinician becomes more complex as time continues. Also, the identity of each patient's surveillance clinician is not well-documented in the registry. For men without a recorded managing clinician, the diagnosing clinician was assumed to be the managing clinician. This disproportionately affected patients receiving surveillance. This analysis also looked at a process measure of quality (decision to treat) and therefore, did not explore how feedback may affect clinical outcomes or patient experience.

Chapter 4 explored the effect of peer comparative feedback on surgical and patient outcomes for men undergoing radical prostatectomy, using individual patient data from the PCOR-Vic. The outcomes of interest were positive surgical margins (a well-documented measure of surgical quality associated with patient prognosis) and reporting of sexual bother (as reported at the patient's 1-year follow up). Linear regressions at the patient level controlled for patient and hospital characteristics and estimated the average effect of feedback for each clinician. Trends in PSM and sexual bother rates were compared across clinicians before and after they received feedback. There was no evidence of an average effect of feedback across all clinicians for PSM or sexual bother rates, but some evidence of a negative relationship between changes in the two (a decrease in sexual bother was associated with an increase in PSM), at the urologist level. There was some evidence that this may be partly due to the initial rates of each outcome; urologists were more likely to improve the outcome with worse performance prior to feedback. The mechanism for the negative relationship was not entirely clear but may be driven by additional treatment or changes in the aggressiveness of the surgery. Sexual bother reporting can also be affected by patient expectations. Therefore, improvements in sexual bother may result from urologists improving how they prepare patients prior to surgery rather than a change to the surgery itself. Sexual bother status was reported for less than half of included patients. However, there was little evidence of a deliberate or systematic lack of reporting by the clinicians. PSM and patient reported sexual bother status are recorded only in the short-term following surgery, and therefore, feedback has not been shown to affect long-term patient outcomes. As RARP became more commonly used in Victoria over

the last 10 years, this analysis controlled for surgery type. However, it did not capture how experience with surgery type may affect patient and surgery outcomes.

Chapter 5 considered how experience with RARP may improve surgical outcomes (reduce PSM), management outcomes (reduce additional treatment), and patient-reported outcomes (increase SF-12 summary scores). Several studies have compared surgeons who perform RARP to those who perform ORP, but few have considered the impact on the healthcare system when surgeons perform both. Experience with RARP is explored through average annual numbers of RARP, and the total number of RARP recorded on the registry; the comparison group was ORP that occurred prior to the first recorded RARP for each surgeon. IPW adjusted for patient and hospital characteristics that might affect surgery selection and outcomes. There is some evidence that PSM rates reduce and SF-12 mental summary scores increase for surgeons with experience of RARP compared to ORP prior to RARP, with PSM decreasing significantly for surgeons with higher annual throughput, and a greater number of total RARPs. There was a non-significant reduction in additional treatment and SF-12 physical summary scores when surgeons experienced RARP. These results support previous research that suggests a lead-in time for RARP experience, and recommendations that RARP is more likely to be cost-effective when throughput is higher. However, this analysis also highlighted that few surgeons have yet conducted high numbers of RARP (e.g. >50 patients). There is also the possibility that some of the effect is related to time, as ORPs tend to happen earlier in the data collection period and RARPs later. However, the comparison of ORPs before and after the commencement of RARPs showed similar outcomes after matching. This analysis also highlighted that surgery choice is not equal across patients: most RARPs were received in private institutions (particularly for higher throughput surgeons who are likely to have better outcomes) and metropolitan locations.

Across analyses, hospital characteristics were associated with different outcomes: patients managed in regional hospitals were less likely to receive treatment in the year following diagnosis (particularly if they were low-risk) or received RARP over ORP. These patients were also more likely to have surgeries that resulted in a PSM and were more likely to report sexual bother. Similarly, patients managed in public hospitals were associated with lower treatment rates in the year following diagnosis;

were less likely to receive RARP than ORP; were more likely to have surgeries that resulted in a PSM (particularly if receiving RARP) and reported higher rates of sexual bother. While not the focus of this thesis, further research could consider the acceptability of this variation in care according to hospital characteristics.

The results of this thesis correspond with existing literature in other clinical areas such that there is more evidence of interventions for changing process measures of quality of care (e.g., test and treatment rates) than outcome measures (e.g. efficiency of PSA testing, surgical and patient outcomes) (Hohn, 2012; Hoque et al., 2017; Modi et al., 2018; Sohn et al., 2016; The Australian Commission on Safety and Quality in Health Care, 2016), even though improvements in outcome measures are more likely to be preferred by patients and clinicians (Pross et al., 2017). The thesis also highlighted that outcome measures are also not as well-reported as process measures. One improvement for the registry could be to increase collection of data relevant to patients (e.g., short term care outcomes such a hospital stay and longer-term outcomes such as ongoing sexual bother or quality of life detriment).

There is also some evidence that clinicians may require more guidance in their behaviour changes. Efficiency of PSA testing has not greatly improved in any state following guidelines to reduce testing, suggesting more could be done to guide the identification of patients who should receive asymptomatic PSA testing. In terms of feedback, there was potential for indicators to affect populations other than their target: the reduction in curative treatment for intermediate-risk patients following feedback may indicate an application of low-risk guidelines to intermediate-risk patients; and for surgical outcomes such as PSM and patient reported sexual bother, surgeons appeared to improve the outcome they performed worse in prior to feedback to the detriment of the other outcome. Clinicians may therefore need shorter or interactive summary feedback, where they can compare their performance in multiple quality indicators at the same time. It may also be beneficial involve patients in the feedback development, such that important outcomes to patients and clinicians can be highlighted, perhaps through a change in the order of feedback reports. One other potential area for improvement is the introduction of benchmarks. In this thesis, clinicians received feedback which only compared then to the average of their peers for each quality indicator. The provides little incentive for improvement for above

average clinicians and has the potential for worsening quality indicators if the average decreases with successive feedback reports.

The thesis also presented an opportunity to compare the effect of structural quality measures (e.g. RARP experience) on patient and surgical outcomes. It highlighted that data on both structure and outcome measures are not well-documented within the registry (e.g. surgeon skill level and potential RARP benefits such as length of hospitalisation are not reported). Additional registry data would need to be collected to allow for exploration of other structural quality measures, e.g. use of MRI or hormone therapies in localised prostate cancer care. The uptake of a new technology (RARP) has occurred in spite of little evidence of long-term benefits to patients. This suggests that either more patient-relevant measures should be collected to understand the perceived benefit of RARP to patients, or the current measures of surgical quality should be made available to patients to help inform their choice of surgeon and surgery type.

One of the advantages of this thesis is the use of large datasets that represent clinical practice. In healthcare where variation in care occurs, a large dataset is likely more representative of all patients with the disease and is robust to statistical approaches to control for unwanted variation. The PCOR-Vic provides individual patient data on a range of patient demographics, treatments and outcome data, and identifiers for clinicians and hospitals. Hospital and clinician data, including the date they first received individualised peer comparative feedback, were also available. However, there are limitations to the datasets, including unobserved characteristics and missing data. In particular, patients on PCOR-Vic were not followed beyond two years after treatment and outcomes such as biological recurrence, long term quality of life changes and survival were not routinely collected. There were also limitations on the data collected at diagnosis. Sexual, urinary and bowel bother were not recorded at diagnosis, and therefore analyses were not adjusted for each patient's baseline levels. In PCOR-Vic data, patient-reported outcomes were not collected for all men, which may result from patients choosing not to answer or the registry failing to collect the data. No evidence was identified that this was occurring systematically. For clinicians, specialism was not reported, and often a clinician ID was not reported for men

undergoing surveillance. Clinician details would help identify further what characteristics may be associated with clinicians changing behaviours.

Cancer registries provide a less costly alternative to randomised control trials that can collect data regarding health care interventions that also capture clinical practice. Randomised control trials remain the gold standard for testing health care interventions. This thesis demonstrated one way of testing for changes in provider care through data collected by pre-established registries. The thesis attempted to get closer to showing a causal relationship between interventions, changes in provider behaviour, and clinical and patient outcomes by applying statistical methods that adjusted for confounding variables. Using non-randomised data to ascertain causality can be quite data-intensive, and there are limitations to the datasets used in the analyses. Some of this could be addressed through additional data collection or conducting randomised experiments within the registry. However, the cost of these measures should be weighed against the additional information they could provide.

Overall, the thesis finds that there is some evidence that quality of care can be improved through changes to provider behaviour, particularly in terms of processes of care. However, there is less evidence that changes to provider behaviour result in improved patient outcomes, particularly if ways to improve these behaviours are less clear. Potential policy implications/areas for further research that may help the effectiveness of interventions to change provider behaviour and improve quality of care in prostate cancer have been identified:

- There is potential for improving the efficiency of PSA testing in asymptomatic men. Guidelines have so far brought the number of tests down, but not restricted testing to men with prostate cancer, and therefore there is potential that an increasing number of cases missed. Guidelines should consider how best to guide men to PSA testing, and further research may need to be done to identify risk factors in men.

- There is potential for improving peer comparative feedback in several ways. There may be scope for providing clinicians with more directions, to help navigate responding to multiple elements of feedback or encourage improvement in

outcomes for all patients, including those of clinicians who are already high-performing.

- There is also potential to direct appropriate technology uptake. While there has been an increase in robotic surgery uptake, one potential suggestion is that consideration of the training needs of surgeons may be required to gain the full benefits of the technology, and lower volume surgeons may be expected to have longer learning curves.

7. Appendices

7.1. Chapter 2 Timeline of events in prostate cancer diagnosis

Table 14 Timeline of possible modifier to prostate cancer diagnosis

Year	Possible modifiers of prostate cancer diagnosis
1998	PSA tests available on Medicare (66656, 66659)
	SA-PCCOC established
2001	PSA test item 66655 introduced (diagnosis only, 1 per man per year)
	5th Guidelines for preventive activities in general practice released
2002	PSA test 66656 description updated for prostate cancer follow up (including follow up for 66655)
	Localised prostate cancer Australia guidelines released (NHMRC)
2005	6th edition Guidelines for preventive activities in general practice released
2008	PCOR-Vic established (Aug 2008)
	NICE guidelines for diagnosis and treatment of prostate cancer
	US Preventative Services Task Force publishes updated prostate cancer guidelines
2009	PSA test introduced on Medicare (66660) for follow up high previous PSA test (max 4/year)
	PSA test 66659 updated for follow up of high PSA test (maximum 1 every year)
	7th edition Guidelines for preventive activities in general practice released
2010	Australia Government Cancer Screening statement on prostate cancer surveillance in asymptomatic men
2012	8th edition Guidelines for preventive activities in general practice released
	US Preventative Services Task Force publishes updated prostate cancer guidelines (May 2012)
	European Association of Urology publishes prostate cancer guidelines
2013	PCOR-ANZ officially established
	Technical report of PSA testing in asymptomatic men (NHMRC) [guidelines released 2014]
	Cochrane review of prostate cancer screening

Year	Possible modifiers of prostate cancer diagnosis
2014	PCOR-QLD established
	Digital/online system for data entry to registry introduced
	NHMRC PSA Testing for Prostate Cancer in Asymptomatic Men Information for Health Practitioners
2016	PCOR-NSW begins
	PSA testing guidelines announced by Cancer Council Australia
	9th edition Guidelines for preventive activities in general practice released

Sources: ("Medicare Benefits Schedule Online," 2020; National Health and Medical Research Council, 2014; National Institute for Health and Care Excellence, 2019; Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel, 2015; SA-PCCOC, 2020; US Preventative Services Task Force, 2008, 2012)

7.2. Chapter 2 Interrupted time series analyses with additional time cut-offs

Figure 20 ITSA prostate cancer cases detected per 100,000 men, time trends between multiple guidelines

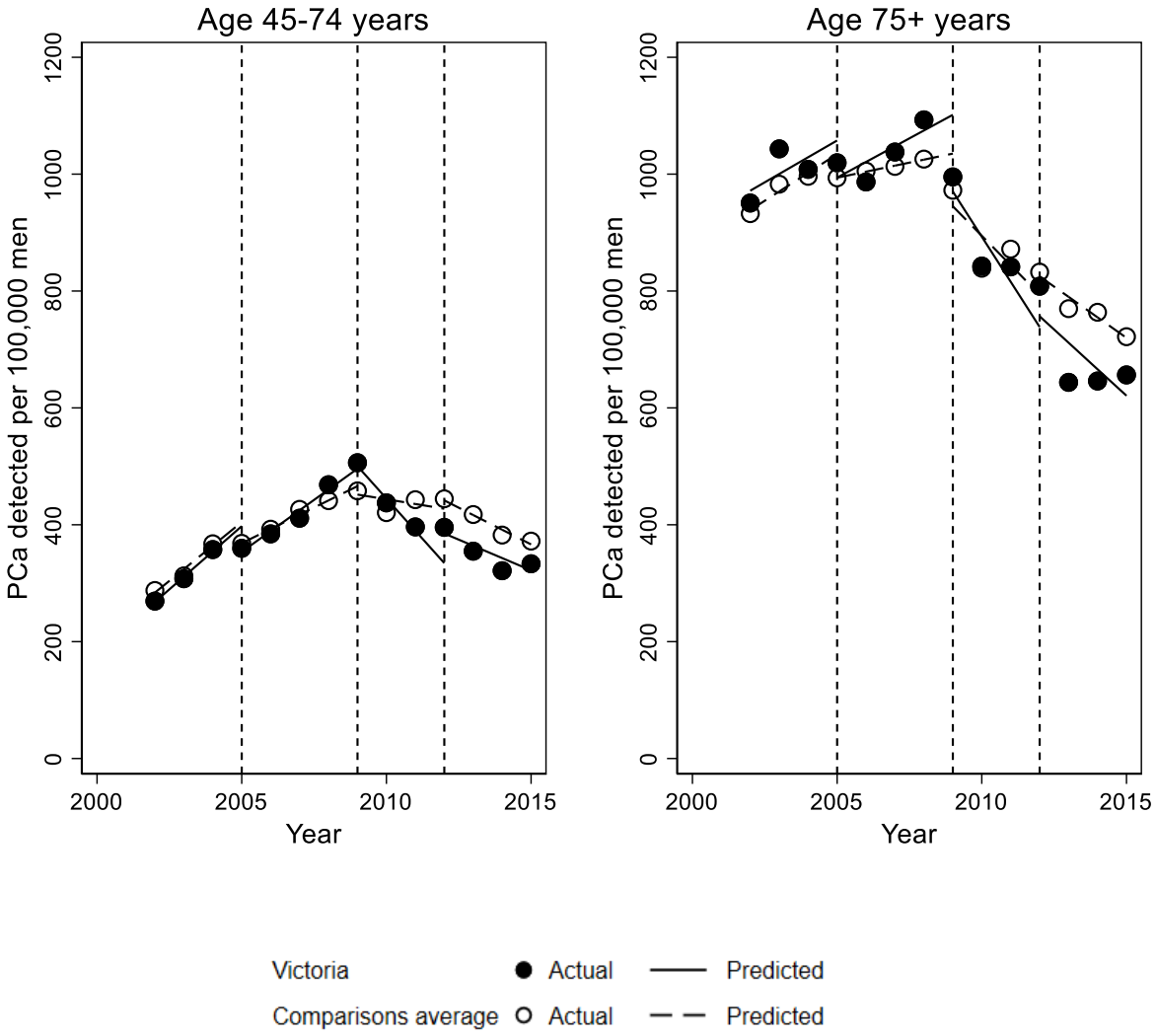


Figure 21 ITSA PSA testing per 100,000 men, time trends between multiple guidelines

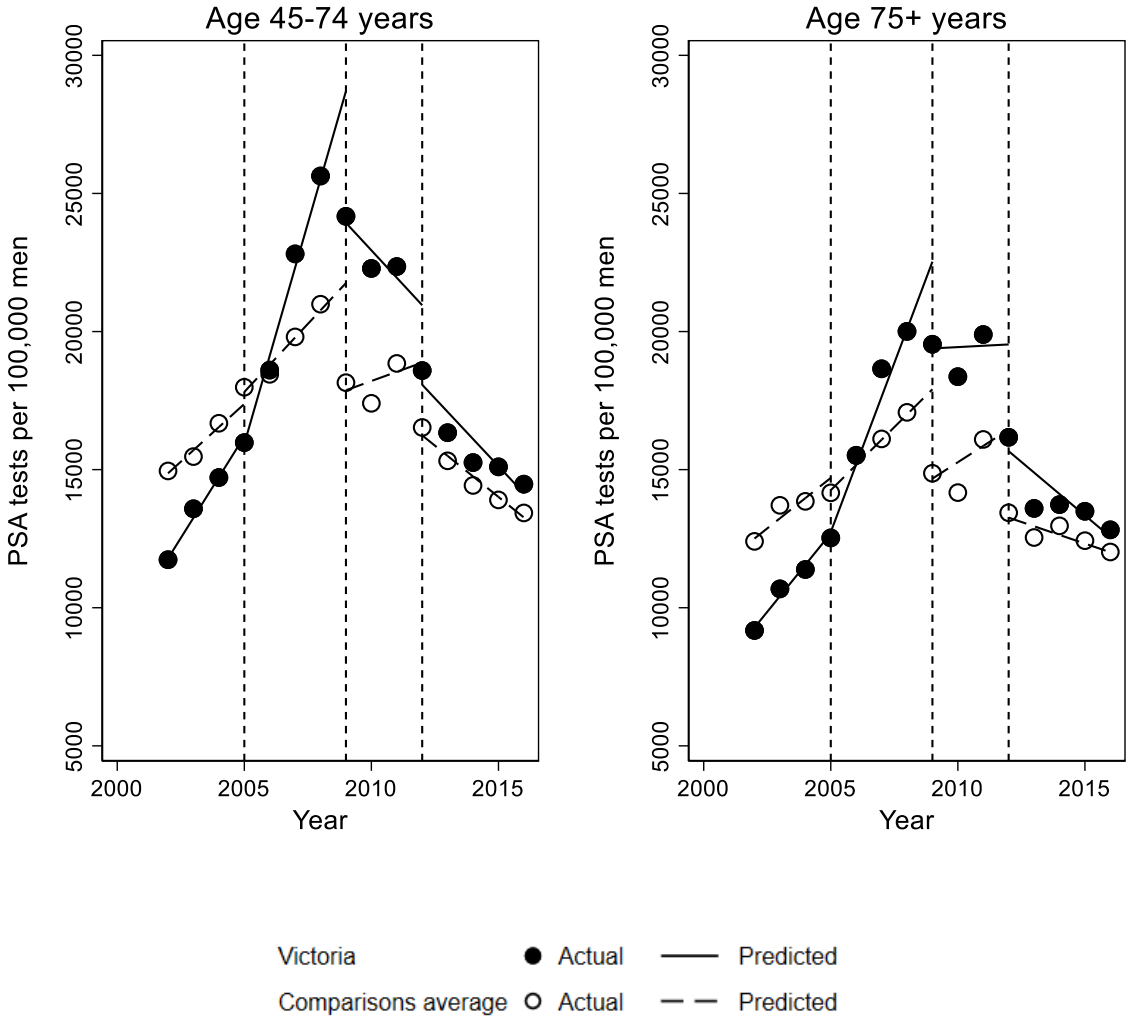
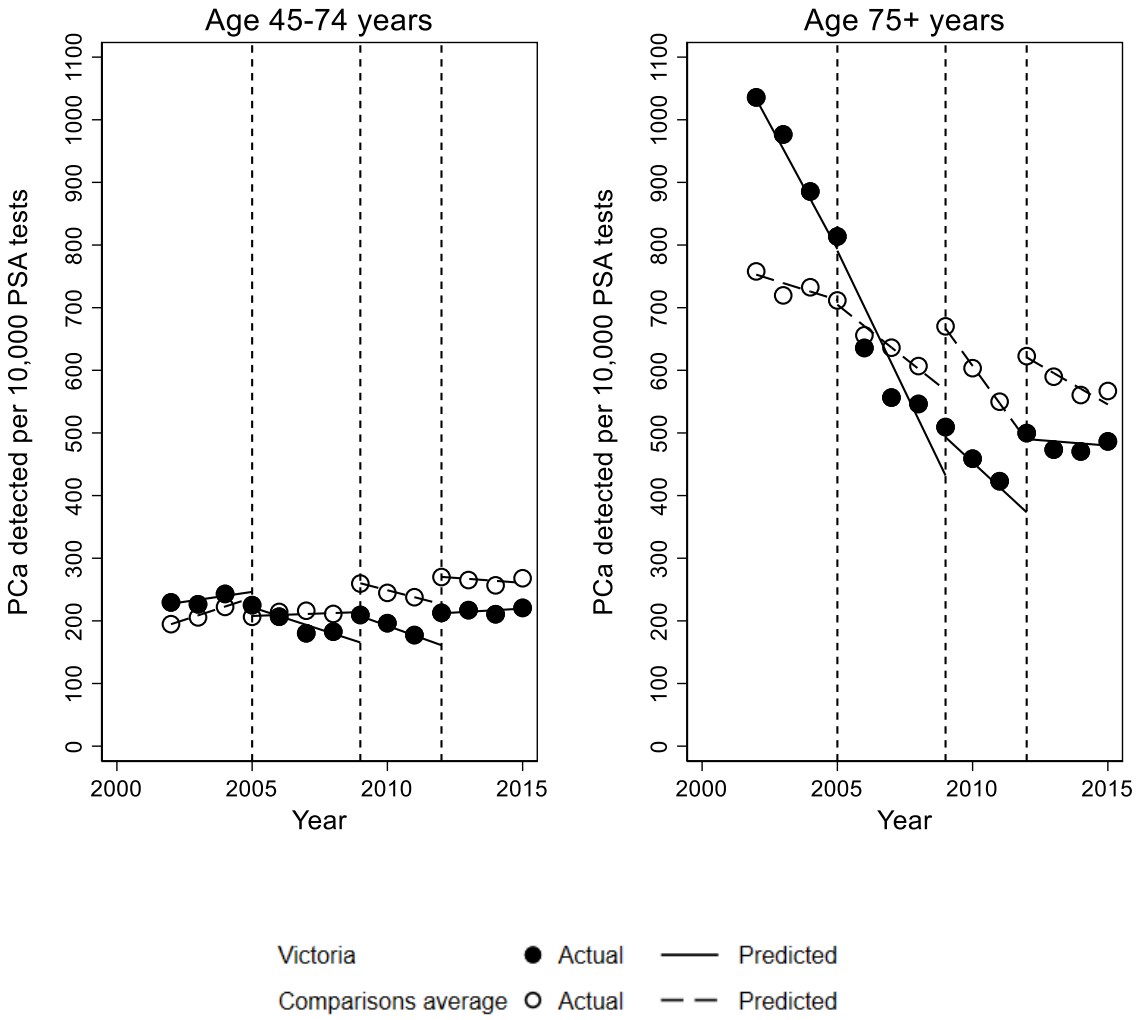


Figure 22 ITSA prostate cancer cases detected per 10,000 PSA tests, trends between multiple guidelines



7.3. Chapter 2 Interrupted time series analyses for men under 45 years old

Figure 23 Interrupted times series analyses for men under 45 years old

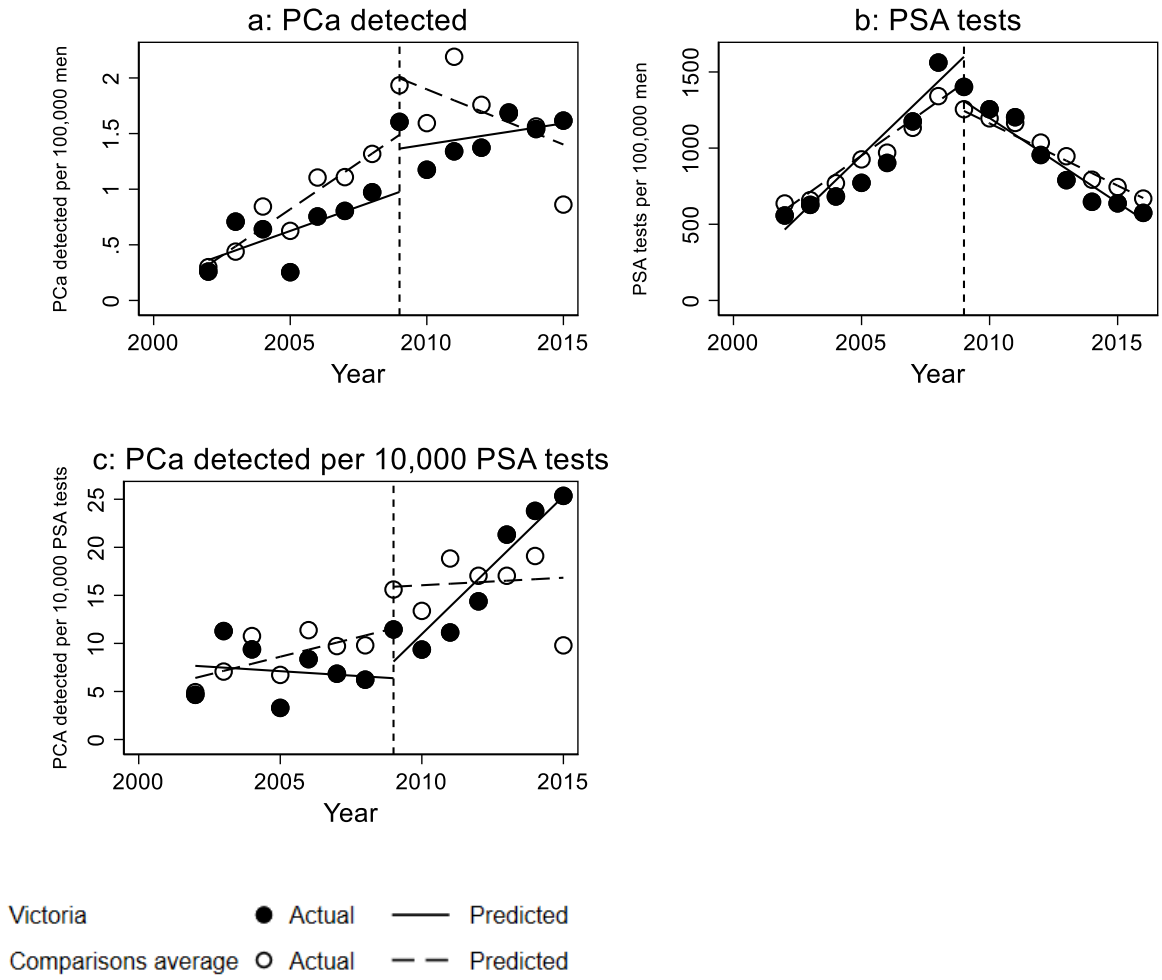


Table 15. Full results for men <45 years old

Analysis: VIC versus states with no registry (NSW, QLD, WA)	Measure of interest	Point estimate [95% CI]
Prostate cancer detected per 100,000 men	Pre-2009 trend: NSW, QLD, WA	0.17*** [0.11, 0.22]
	Pre-2009 trend: VIC	0.08*** [0.03, 0.14]
	Difference in trends pre-2009: VIC vs NSW, QLD, WA	-0.08** [-0.16, -0.002]
	Post-2009 trend: NSW, QLD, WA	-0.10* [-0.21, 0.007]
	Post-2009 trend: VIC	0.04 [-0.03, 11.01]
	Difference in trends post-2009 VIC vs NSW, QLD, WA	0.14** [0.01, 0.27]
	Difference in trends pre and post-2009: NSW, QLD, WA	-0.27*** [-0.39, -0.15]
	Difference in trends pre and post-2009: VIC	-0.05 [--0.14, 0.04]
	Difference between VIC and NSW, QLD, WA trends pre- and post- 2009	0.22*** [0.07, 0.37]
PSA tests per 100,000 men	Pre-2009 trend: NSW, QLD, WA	119.2*** [84.9, 153.4]
	Pre-2009 trend: VIC	162.1*** [91.0, 233.1]
	Difference in trends pre-2009: VIC vs NSW, QLD, WA	42.9 [-36.0, 121.8]
	Post-2009 trend: NSW, QLD, WA	-81.7*** [-101.8, -61.6]
	Post-2009 trend: VIC	-111.2*** [-150.4, -71.9]
	Difference in trends post-2009 VIC vs NSW, QLD, WA	-29.5 [-73.6, 14.6]
	Difference in trends pre and post-2009: NSW, QLD, WA	-200.9*** [-244.3, -157.5]
	Difference in trends pre and post-2009: VIC	-273.2*** [-375.7, -170.7]
	Difference between VIC and NSW, QLD, WA trends pre- and post- 2009	-72.3 [-184.7, 39.0]

Analysis: VIC versus states with no registry (NSW, QLD, WA)	Measure of interest	Point estimate [95% CI]
Prostate cancer cases detected per 10,000 PSA tests	Pre-2009 trend: NSW, QLD, WA	0.74** [0.16, 1.32]
	Pre-2009 trend: VIC	-0.18 [-1.16, 0.79]
	Difference in trends pre-2009: VIC vs NSW, QLD, WA	-0.92 [-2.05, 0.21]
	Post-2009 trend: NSW, QLD, WA	0.16 [-1.19, 1.50]
	Post-2009 trend: VIC	2.86*** [1.94, 3.78]
	Difference in trends post-2009 VIC vs NSW, QLD, WA	2.71*** [1.08, 4.34]
	Difference in trends pre and post-2009: NSW, QLD, WA	-0.581 [-2.05, 0.89]
	Difference in trends pre and post-2009: VIC	3.05*** [1.71, 4.39]
	Difference between VIC and NSW, QLD, WA trends pre- and post- 2009	3.63*** [1.64, 5.62]

Notes: *p<0.10, **p<0.05, ***p<0.01

7.4. Chapter 3 Patient characteristics by risk status

Table 16 Low-risk patient characteristics by management strategy

Factor	Non-curative management (surveillance, ADT, chemo)	Curative management (prostatectomy, radiotherapy, brachytherapy)	p-value	Test
N	1871	1152		
Feedback in 1st year following diagnosis			<0.001	Pearson's chi-squared
No feedback	564 (30.1%)	617 (53.6%)		
Feedback	1307 (69.9%)	535 (46.4%)		

Factor	Non-curative management (surveillance, ADT, chemo)	Curative management (prostatectomy, radiotherapy, brachytherapy)	p-value	Test
Age (years)			<0.001	Pearson's chi-squared
<55	255 (13.6%)	289 (25.1%)		
55-74	1479 (79.0%)	861 (74.7%)		
75+	137 (7.3%)	2 (0.2%)		
PSA level at diagnosis, median (IQR)	5.0 (3.4, 6.5)	5.0 (3.8, 6.3)	0.098	Wilcoxon's rank sum
Clinical T stage			<0.001	Pearson's chi-squared
T1	1330 (71.1%)	760 (66.0%)		
T2	133 (7.1%)	189 (16.4%)		
Unknown	408 (21.8%)	203 (17.6%)		
SEIFA decile			<0.001	Pearson's chi-squared
Lowest 20%	187 (10.0%)	115 (10.0%)		
Lowest 21-40%	197 (10.5%)	145 (12.6%)		
Lowest 41-60%	195 (10.4%)	163 (14.1%)		
Highest 61-80%	394 (21.1%)	290 (25.2%)		
Highest 81-100%	894 (47.8%)	437 (37.9%)		
Unknown	4 (0.2%)	2 (0.2%)		
Managing hospital location			0.069	Pearson's chi-squared
Metro	1426 (76.2%)	1006 (87.3%)		
Regional	165 (8.8%)	145 (12.6%)		
Unknown	280 (15.0%)	1 (0.1%)		

Factor	Non-curative management (surveillance, ADT, chemo)	Curative management (prostatectomy, radiotherapy, brachytherapy)	p-value	Test
Managing hospital			0.033	Pearson's chi-squared
Public	295 (15.8%)	216 (18.8%)		
Private	1576 (84.2%)	935 (81.2%)		
Unknown	0 (0.0%)	1 (0.1%)		
Diagnosis date (years since 1st December 2012), median (IQR)	1.0 (-1.0, 3.1)	-0.3 (-1.5, 1.8)	<0.001	Wilcoxon's rank sum

Notes: SEIFA data extracted on 06 Jun 2018 03:30 UTC (GMT) from ABS.Stat © Commonwealth of Australia

Table 17 Intermediate-risk patient characteristics by management strategy

Factor	Non-curative management (surveillance, ADT, chemo)	Curative management (prostatectomy, radiotherapy, brachytherapy)	p-value	Test
N	1028	4964		
Feedback in 1st year following diagnosis			<0.001	Pearson's chi-squared
No feedback	303 (29.5%)	1767 (35.6%)		
Feedback	725 (70.5%)	3197 (64.4%)		
Age (years)			<0.001	Pearson's chi-squared
<55	65 (6.3%)	628 (12.7%)		
55-74	668 (65.0%)	4205 (84.7%)		
75+	295 (28.7%)	131 (2.6%)		
PSA level at diagnosis, median (IQR)	7.8 (4.9, 11.4)	6.1 (4.7, 8.3)	<0.001	Wilcoxon's rank sum
Gleason score			<0.001	Pearson's chi-squared
<7	320 (31.1%)	272 (5.5%)		
7	704 (68.5%)	4688 (94.4%)		
	4 (0.4%)	4 (0.1%)		
Clinical T stage			<0.001	Pearson's chi-squared
T1	558 (54.3%)	2184 (44.0%)		
T2	235 (22.9%)	1638 (33.0%)		
Unknown	235 (22.9%)	1142 (23.0%)		

Factor	Non-curative management (surveillance, ADT, chemo)	Curative management (prostatectomy, radiotherapy, brachytherapy)	p-value	Test
SEIFA decile			<0.001	Pearson's chi-squared
Lowest 20%	120 (11.7%)	539 (10.9%)		
Lowest 21-40%	125 (12.2%)	569 (11.5%)		
Lowest 41-60%	105 (10.2%)	686 (13.8%)		
Highest 61-80%	191 (18.6%)	1106 (22.3%)		
Highest 81-100%	484 (47.1%)	2046 (41.2%)		
Unknown	3 (0.3%)	18 (0.4%)		
Managing hospital location			<0.001	Pearson's chi-squared
Metro	776 (75.5%)	4493 (90.5%)		
Regional	148 (14.4%)	471 (9.5%)		
Unknown	104 (10.1%)	0 (0.0%)		
Managing hospital			<0.001	Pearson's chi-squared
Public	250 (24.3%)	976 (19.7%)		
Private	777 (75.6%)	3982 (80.2%)		
Unknown	1 (0.1%)	6 (0.1%)		
Diagnosis date (years since 1st December 2012), median (IQR)	1.0 (-0.8, 3.0)	1.4 (-0.7, 3.4)	<0.001	Wilcoxon's rank sum

Table 18 High-risk patient characteristics by management strategy

Factor	Non-curative management (surveillance, ADT, chemo)	Curative management (prostatectomy, radiotherapy, brachytherapy)	p-value	Test
N	455	1655		
Feedback in the year following diagnosis			0.33	Pearson's chi-squared
No feedback	173 (38.0%)	588 (35.5%)		
Feedback	282 (62.0%)	1067 (64.5%)		
Age (years)			<0.001	Pearson's chi-squared
<55	7 (1.5%)	110 (6.6%)		
55-74	112 (24.6%)	1443 (87.2%)		
75+	336 (73.8%)	102 (6.2%)		
PSA Level at diagnosis (ng/mL), median (IQR)	23.9 (12.3, 44.2)	8.3 (5.8, 13.8)	<0.001	Wilcoxon's rank sum
Gleason stage			<0.001	Pearson's chi-squared
<8	151 (33.2%)	309 (18.7%)		
8+	292 (64.2%)	1341 (81.0%)		
Unknown	12 (2.6%)	5 (0.3%)		
cT stage			0.005	Pearson's chi-squared
T1 or T2	226 (49.7%)	1028 (62.1%)		
T3 or T4	102 (22.4%)	319 (19.3%)		
Unknown	127 (27.9%)	308 (18.6%)		

Factor	Non-curative management (surveillance, ADT, chemo)	Curative management (prostatectomy, radiotherapy, brachytherapy)	p-value	Test
SEIFA decile			0.12	Pearson's chi-squared
Lowest 20%	61 (13.4%)	215 (13.0%)		
Lowest 21-40%	59 (13.0%)	194 (11.7%)		
Lowest 41-60%	41 (9.0%)	228 (13.8%)		
Highest 61-80%	95 (20.9%)	345 (20.8%)		
Highest 81-100%	197 (43.3%)	659 (39.8%)		
Unknown	2 (0.4%)	14 (0.8%)		
Managing hospital location			0.018	Pearson's chi-squared
Metropolitan	343 (75.4%)	1456 (88.0%)		
Regional	67 (14.7%)	198 (12.0%)		
Unknown	45 (9.9%)	1 (0.1%)		
Managing hospital			<0.001	Pearson's chi-squared
Public	139 (30.5%)	378 (22.8%)		
Private	316 (69.5%)	1275 (77.0%)		
Unknown	0 (0.0%)	2 (0.1%)		
Diagnosis date (years since 1st December 2012), median (IQR)	0.2 (-1.6, 2.4)	1.3 (-0.8, 3.0)	<0.001	Wilcoxon's rank sum

7.5. Chapter 3 Proportional hazard testing base model

Results here are presented for the intermediate-risk model, before PH testing

Table 19 Cox-regression on time to curative treatment for intermediate risk patients, before proportional hazard testing

Variable	HR	SE	p-value	Lower 95%CI	Upper 95% CI
Clinician received feedback	0.47	0.027	0.000	0.42	0.53
Annual time trend	1.12	0.014	0.000	1.10	1.15
Age group (base <55)					
55-74	0.86	0.043	0.002	0.78	0.95
75+	0.17	0.020	0.000	0.13	0.21
PSA level at diagnosis (Ng/mL)	1.01	0.005	0.291	1.00	1.01
Clinical staging at diagnosis at least T2 (versus cT1)	1.36	0.047	0.000	1.27	1.46
Gleason score at diagnosis at least 7 (vs Gleason score <7)	3.51	0.243	0.000	3.06	4.02
SEIFA decile (versus lowest 20%)					
Lowest 21-40%	1.05	0.074	0.532	0.91	1.20
Lowest 41-60%	1.17	0.081	0.024	1.02	1.34
Highest 61-80%	1.07	0.068	0.321	0.94	1.21
Highest 81-100%	0.98	0.059	0.707	0.87	1.10
Unknown	1.18	0.313	0.529	0.70	1.98
Treating hospital					
Regional (vs metropolitan)	0.85	0.133	0.285	0.62	1.15
Private (vs public)	1.93	0.095	0.000	1.75	2.13

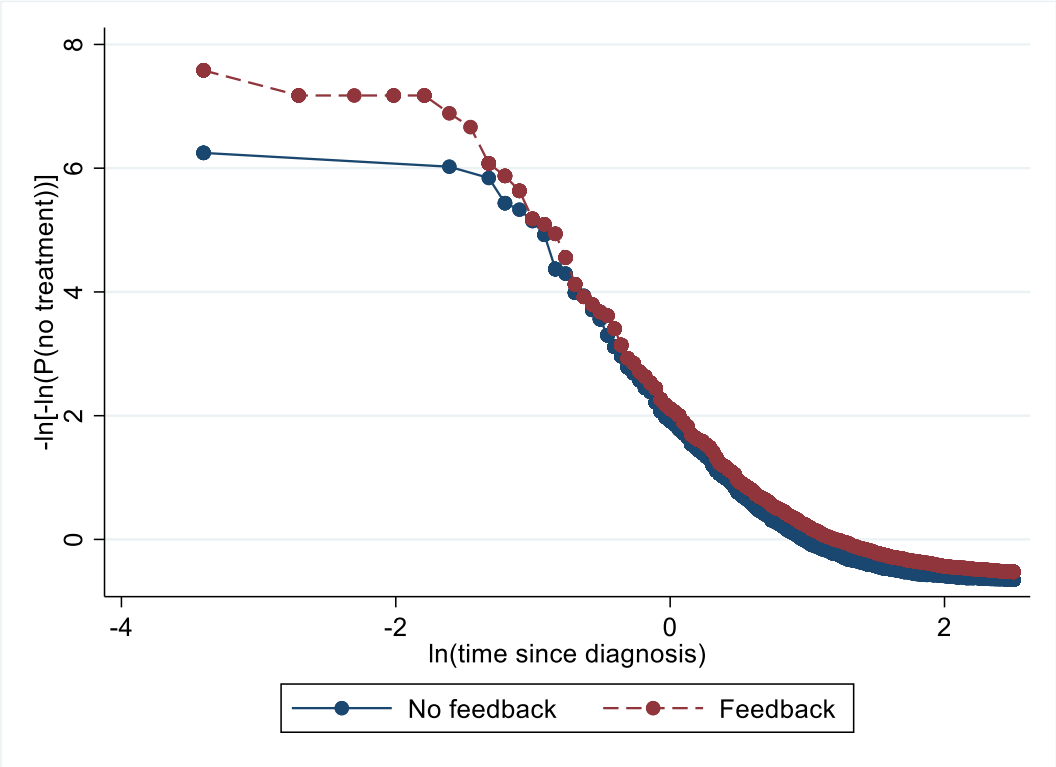
Proportional hazard assumption testing using the Schoenfeld residuals (identifies some variables for which the assumption does not hold: highest age group, private

hospitals and several individual clinician effects (not reported for brevity). The model stratified by these variables rather than including as covariates.

Table 20 Proportional hazard testing for non-stratified model

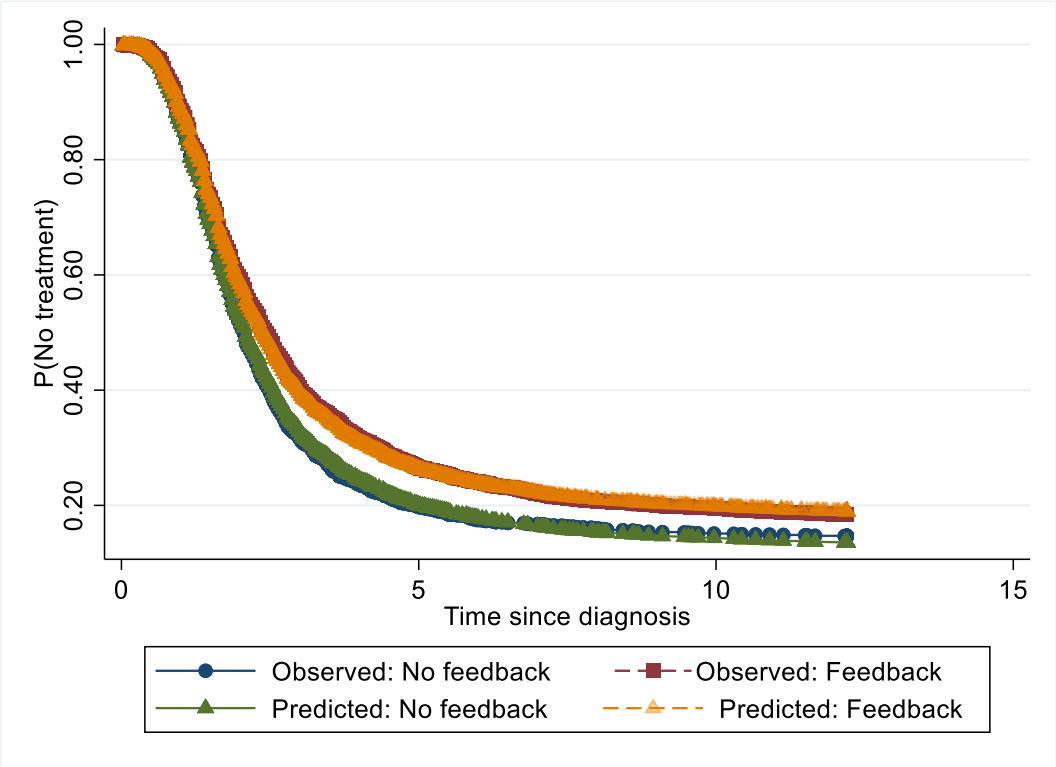
Variable	rho	chi2	df	P>chi2
No feedback	.	.	1	.
Feedback	-0.0081	0.24	1	0.621
Annual time trend	0.02357	2.03	1	0.154
Age group (vs <55)	.	.	1	.
55-74	-0.00361	0.05	1	0.8213
75+	-0.09558	36	1	<0.001
Diagnostic PSA level	-0.00371	0.05	1	0.8234
Clinical T stage (vs <cT2)	.	.	1	.
cT2	0.00756	0.22	1	0.6387
Gleason score (vs score <7)	.	.	1	.
Score=7	-0.03033	3.63	1	0.0568
SEIFA decile (vs lowest 20%)	.	.	1	.
Lowest 21-40%	-0.00947	0.35	1	0.5545
Lowest 41-60%	0.00059	0	1	0.9704
Highest 61-80%	-0.01115	0.49	1	0.4848
Highest 81-100%	-0.02301	2.08	1	0.1493
Unknown	-0.01352	0.7	1	0.4024
Treating hospital	.	.	1	.
Metropolitan (vs regional)	-0.00229	0.02	1	0.884
Private (vs public)	-0.24169	214.76	1	<0.001

Figure 24 Log-log plot of stcox analysis of the effect of feedback on intermediate-risk patients



Log-log plots are parallel across most of the period, parallel at the beginning, merging within the first month and remaining relatively parallel after this point.

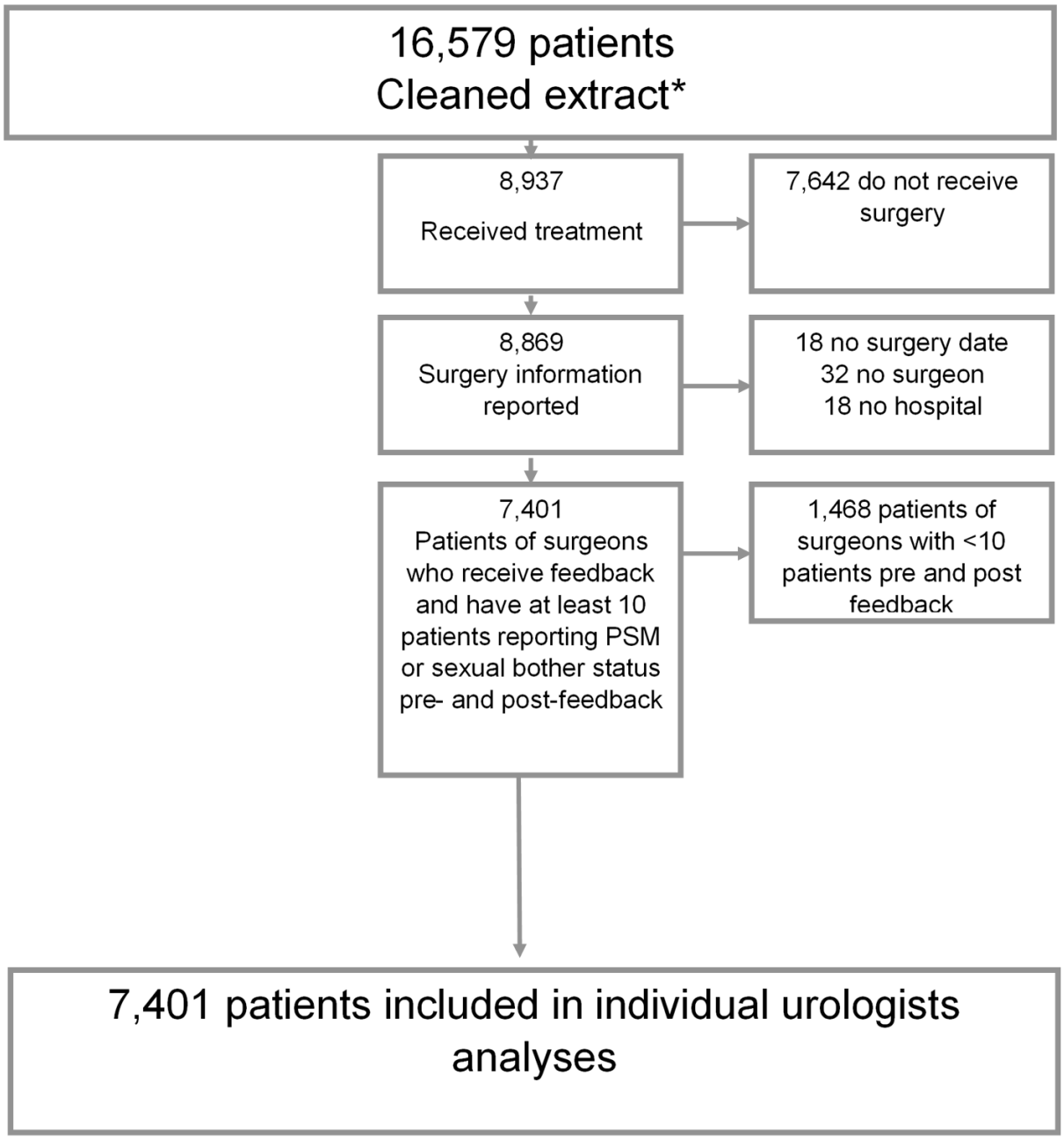
Figure 25 Predicted vs observed Kaplan Meier estimates for intermediate-risk patients, by clinician feedback status



The Kaplan-Meier and predicted survival plot were similar for the feedback and no feedback arms, suggesting the proportional hazard assumption was not violated for the feedback variable.

7.6. Chapter 4 Patient characteristics

Figure 26 Patient inclusion criteria



*Data cleaning information reported in Figure 7, p66

Table 21 Patient characteristics, by response to sexual bother question at follow up

Characteristic	Response to sexual bother question?		
	Not reported	Reported	p-value (Pearson's chi-squared)
N	4802	3508	
Age group			<0.001
<55	521 (10.8%)	528 (15.1%)	
55-74	4081 (85.0%)	2910 (83.0%)	
75+	200 (4.2%)	70 (2.0%)	
NCCN risk at surgery			<0.001
Low Risk	604 (12.6%)	637 (18.2%)	
Intermediate Risk	3139 (65.4%)	2143 (61.1%)	
High Risk	1059 (22.1%)	728 (20.8%)	
PSA level at surgery			<0.001
<=10.0	3808 (79.3%)	2912 (83.0%)	
10.01-20.0	735 (15.3%)	437 (12.5%)	
>20.0	189 (3.9%)	127 (3.6%)	
NR	70 (1.5%)	32 (0.9%)	
Surgery type			<0.001
Less invasive surgery	3408 (71.0%)	2021 (57.6%)	
Open surgery	1336 (27.8%)	1472 (42.0%)	
NR	58 (1.2%)	15 (0.4%)	
Surgery Year			<0.001
2009	64 (1.3%)	281 (8.0%)	
2010	130 (2.7%)	290 (8.3%)	
2011	226 (4.7%)	751 (21.4%)	
2012	288 (6.0%)	750 (21.4%)	
2013	317 (6.6%)	723 (20.6%)	
2014	371 (7.7%)	642 (18.3%)	
2015	1030 (21.4%)	58 (1.7%)	
2016	1056 (22.0%)	11 (0.3%)	
2017	1104 (23.0%)	1 (<1%)	
2018	216 (4.5%)	1 (<1%)	

Characteristic	Response to sexual bother question?		
	Not reported	Reported	p-value (Pearson's chi-squared)
Surgical margins			0.74
No PSM	3479 (72.4%)	2551 (72.7%)	
PSM	1189 (24.8%)	887 (25.3%)	
NR	134 (2.8%)	70 (2.0%)	
Urologist level feedback			<0.001
No feedback	926 (19.3%)	2102 (59.9%)	
Feedback	3876 (80.7%)	1406 (40.1%)	

7.7. Chapter 4 Relationship between PSM, additional treatment and sexual bother rates

Table 22 Linear probability model for additional treatment, urologist and feedback included in fixed effects

Variable N=7,284 (34 urologists)	Change in probability of additional treatment	SE.	P-value	95% CI Low	95% CI high
Constant probability of additional treatment	4.5%	3.8%	0.237	-3.0%	12.1%
NCCN risk (vs low risk)					
Intermediate risk	5.1%***	1.0%	<0.001	3.1%	7.1%
High risk	19.3%***	1.2%	<0.001	16.9%	21.6%
Annual time trend (surgery date)	-0.9%***	0.3%	0.001	-1.4%	-0.4%
PSA level	0.03%***	0.01%	0.001	0.01%	0.05%
Age group (base 55-74)					
<55	0.7%	1.1%	0.543	-1.4%	2.7%
75+	-6.6%***	2.0%	0.001	-10.5%	-2.7%
Open surgery	4.9%***	1.4%	<0.001	2.2%	7.6%
Hospital Characteristics					
Metropolitan (vs regional)	3.0%	4.1%	0.463	-5.0%	11.0%
Private (vs public)	-5.6%***	1.1%	<0.001	-7.8%	-3.4%

Notes: * p<0.10, **p<0.05, ***p<0.01. Urologist fixed effects interacted with feedback status.

Figure 27 Relationship between PSM and additional treatment, by urologist (adjusted to full sample)

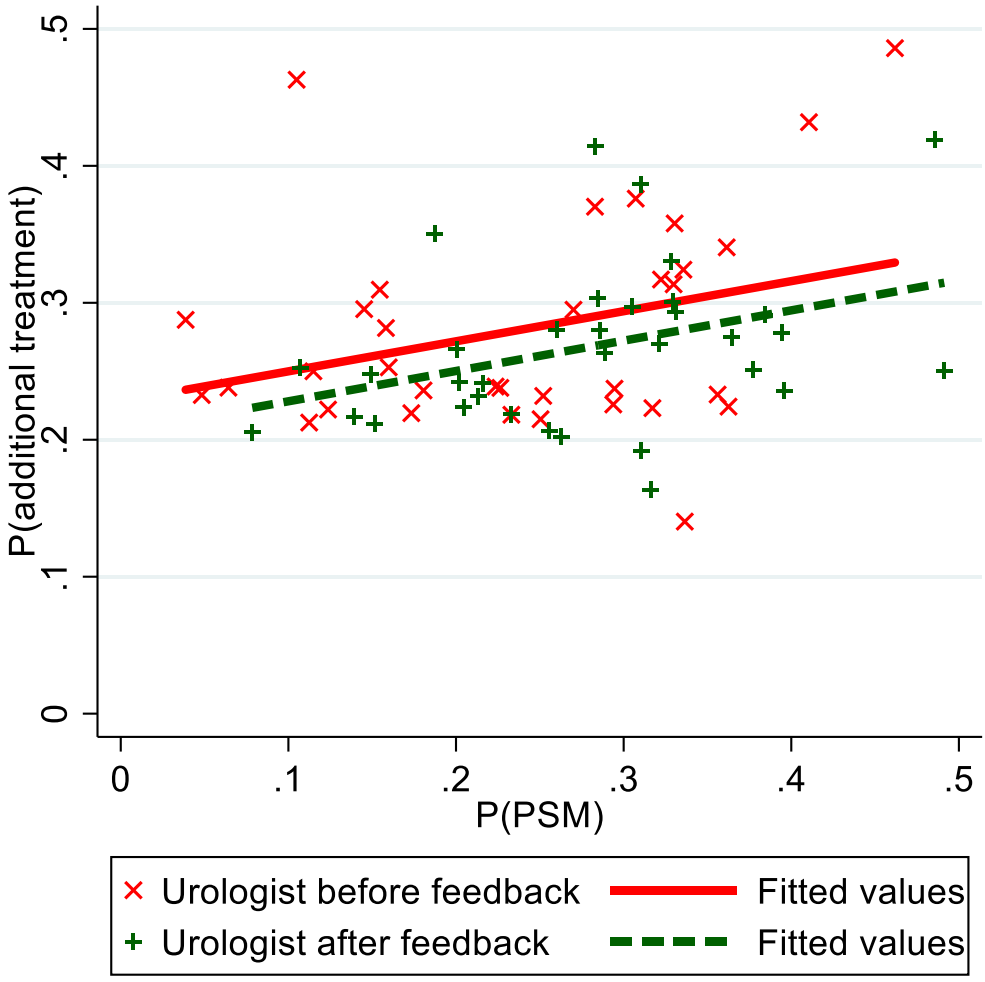


Figure 28 Relationship between sexual bother and additional treatment, by urologist (adjusted to full sample)

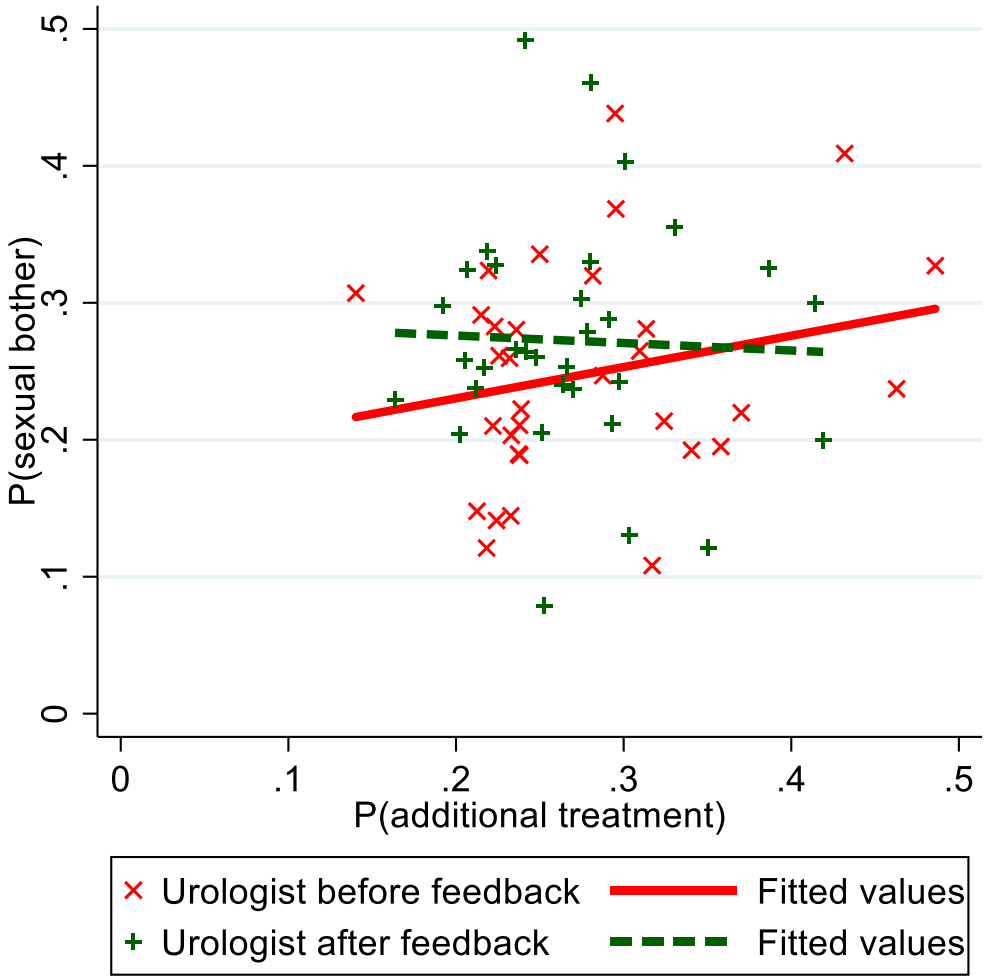
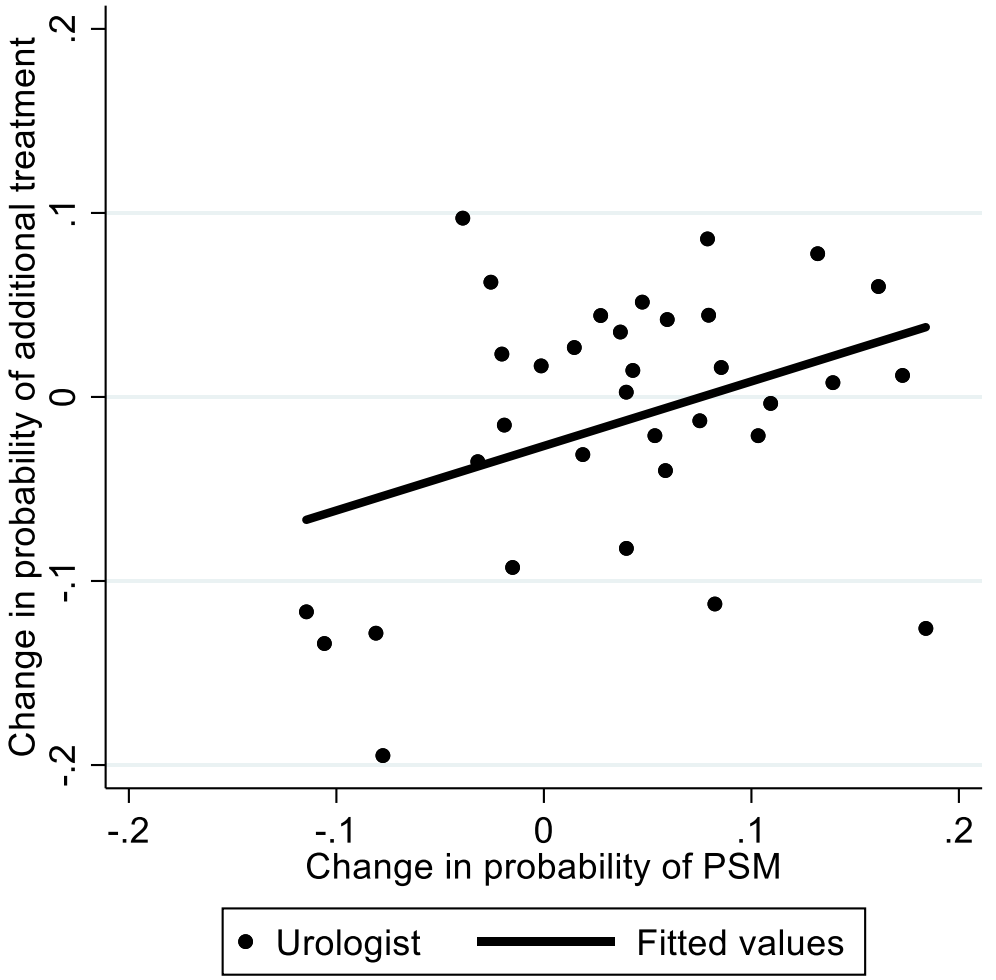
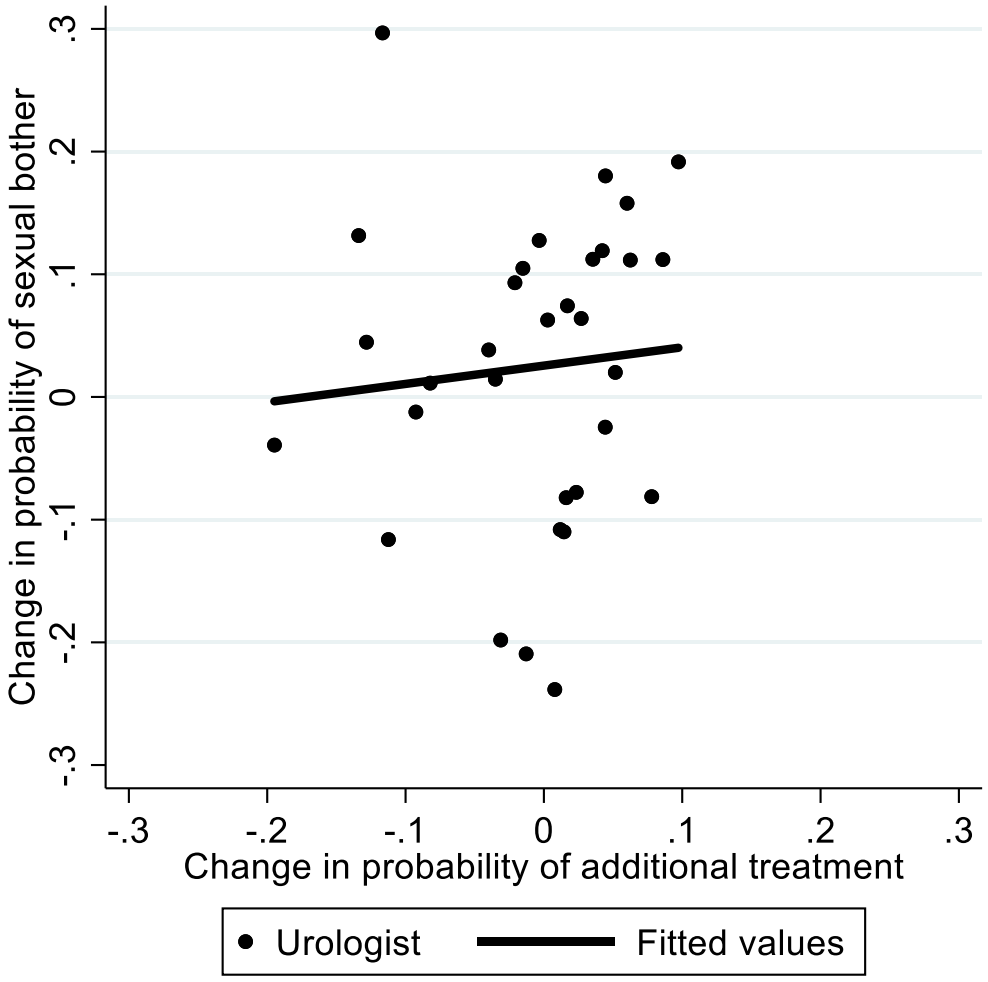


Figure 29 Relationship between change in PSM and additional treatment, by urologist



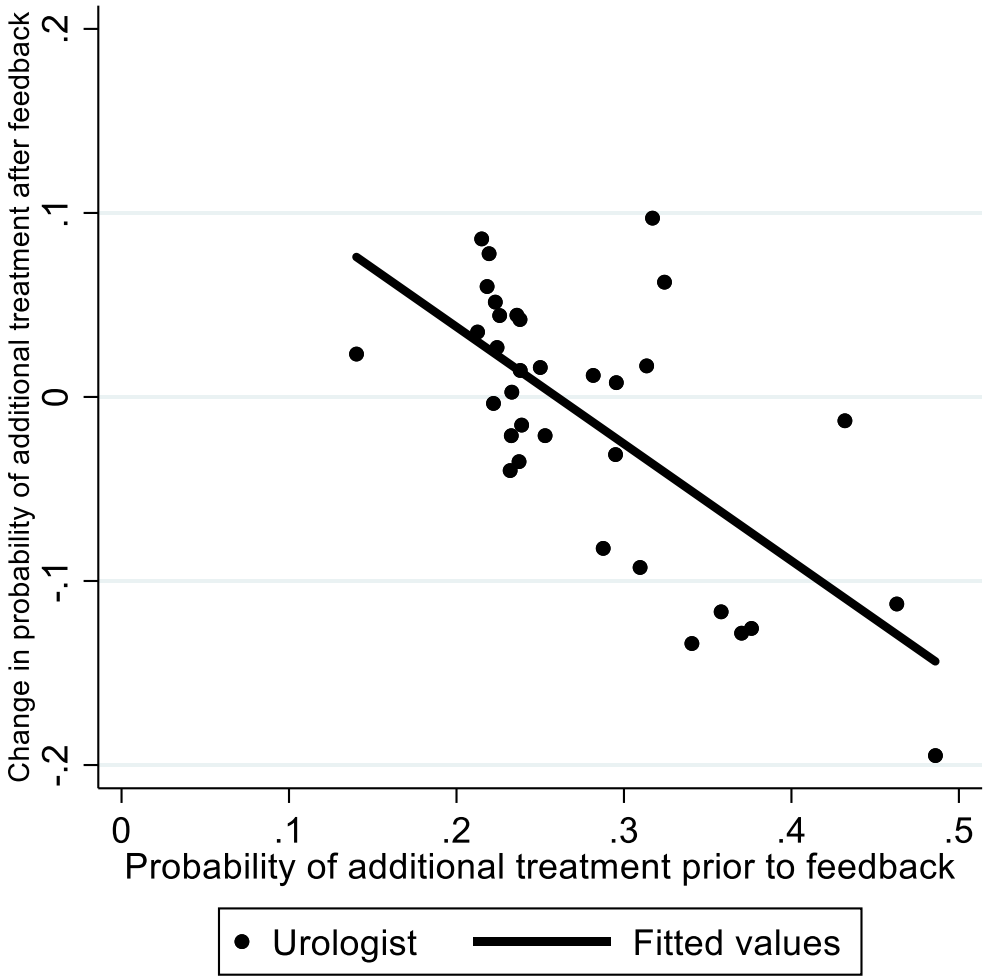
Increase 1% PSM= 0.35% increase additional treatment

Figure 30 Relationship between change in sexual bother and additional treatment, by urologist



Notes: 1% increase in additional treatment results in 0.15% increase in sexual bother

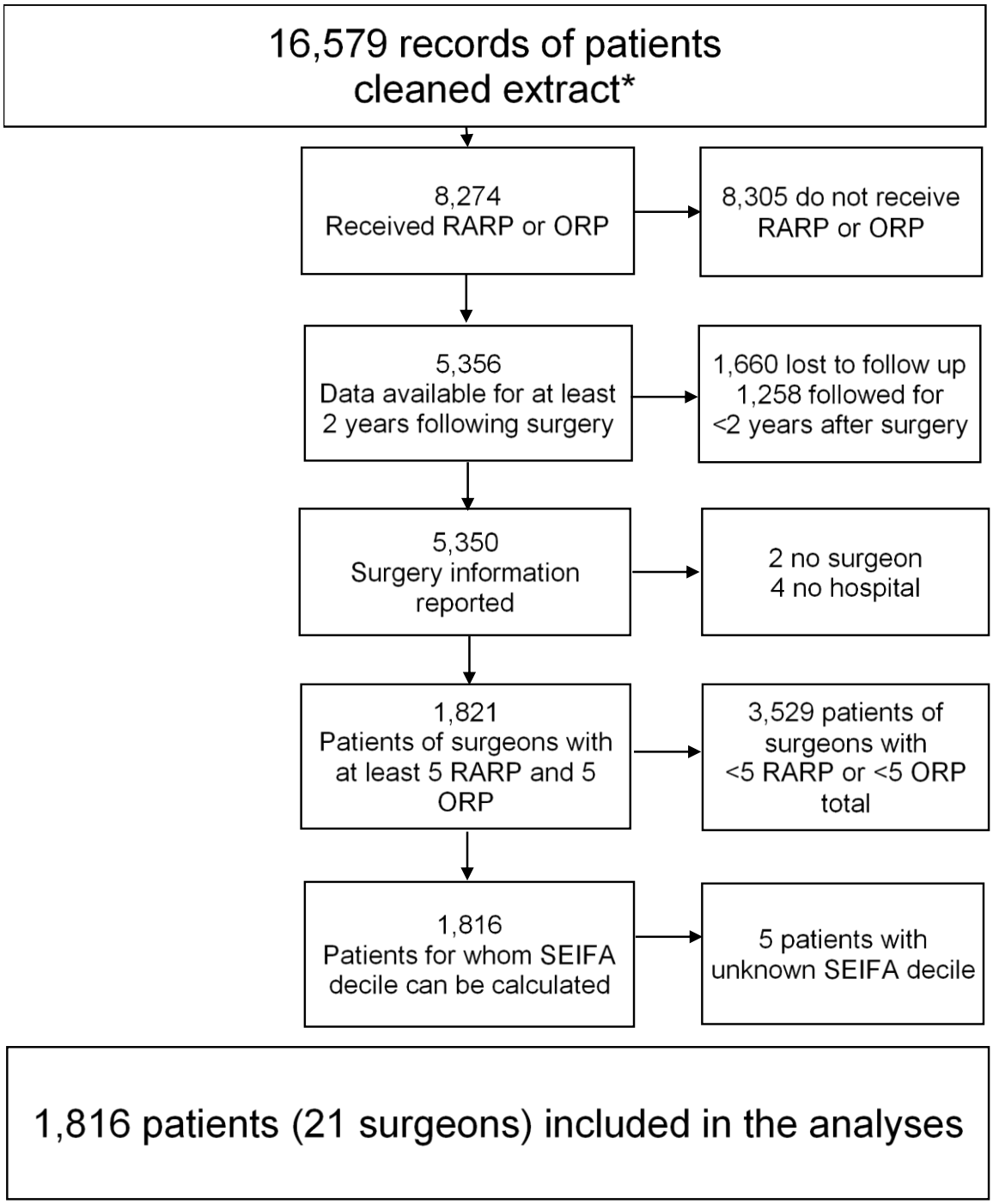
Figure 31 Change in additional treatment compared to additional treatment rates pre-feedback



Notes: For every 1% more likely to have add treat at baseline, 0.64% reduction in additional treatment following feedback. By examining where the linear fit crosses the x-axis, urologists with <27% of patients receiving additional treatment prior to feedback are estimated to increase their additional treatment rates following feedback.

7.8. Chapter 5 Patient inclusion criteria

Figure 32 RARP and ORP patient inclusion criteria



*Data cleaning information reported in Figure 7, p66.

7.9. Chapter 5 Patient characteristics by surgery type

Table 23 Patient characteristics by surgery type and surgeon experience with RARP

Factor	ORP before RARP	ORP after RARP	RARP low vol 1st 50 pts	RARP low vol >50 pts	RARP high vol 1st 50 pts	RARP high vol >50 pts	p-value	Test
N	307	410	565	210	115	209		
Outcomes								
No PSM	206 (67.1%)	262 (63.9%)	389 (68.8%)	154 (73.3%)	91 (79.1%)	171 (81.8%)	<0.001	Pearson's chi-squared
PSM	87 (28.3%)	141 (34.4%)	169 (29.9%)	54 (25.7%)	24 (20.9%)	38 (18.2%)		
NR	14 (4.6%)	7 (1.7%)	7 (1.2%)	2 (1.0%)	0 (0.0%)	0 (0.0%)		
No additional management	291 (94.8%)	384 (93.7%)	542 (95.9%)	202 (96.2%)	114 (99.1%)	205 (98.1%)	0.048	Pearson's chi-squared
Additional management	16 (5.2%)	26 (6.3%)	23 (4.1%)	8 (3.8%)	1 (0.9%)	4 (1.9%)		
SF-12 physical summary score, median (IQR)	54.697 (49.161, 56.723)	53.532 (45.728, 56.705)	55.090 (49.160, 56.705)	49.000 (42.231, 56.416)	55.312 (48.586, 57.203)	56.148 (52.622, 57.761)	0.025	Kruskal-Wallis
SF-12 mental summary score, median (IQR)	54.531 (47.335, 59.429)	54.800 (49.314, 58.892)	57.061 (51.409, 59.530)	57.740 (46.466, 59.429)	56.420 (50.841, 58.154)	57.061 (51.833, 59.429)	0.40	Kruskal-Wallis

Factor	ORP before RARP	ORP after RARP	RARP low vol 1st 50 pts	RARP low vol >50 pts	RARP high vol 1st 50 pts	RARP high vol >50 pts	p-value	Test
Patient characteristics prior to surgery								
Age group (years)							0.088	Pearson's chi-squared
<55	42 (13.7%)	42 (10.2%)	69 (12.2%)	21 (10.0%)	18 (15.7%)	14 (6.7%)		
55-74	265 (86.3%)	368 (89.8%)	496 (87.8%)	189 (90.0%)	97 (84.3%)	195 (93.3%)		
SEIFA decile							<0.001	Pearson's chi-squared
Lowest 40%	98 (31.9%)	100 (24.4%)	151 (26.7%)	57 (27.1%)	21 (18.3%)	13 (6.2%)		
Lowest 41-60%	59 (19.2%)	58 (14.1%)	72 (12.7%)	23 (11.0%)	20 (17.4%)	35 (16.7%)		
Highest 61-100%	150 (48.9%)	252 (61.5%)	342 (60.5%)	130 (61.9%)	74 (64.3%)	161 (77.0%)		
NCCN risk prior to surgery							0.001	Pearson's chi-squared
Intermediate Risk	206 (67.1%)	277 (67.6%)	427 (75.6%)	165 (78.6%)	82 (71.3%)	164 (78.5%)		
High Risk	101 (32.9%)	133 (32.4%)	138 (24.4%)	45 (21.4%)	33 (28.7%)	45 (21.5%)		
Hospital characteristics								
Location							<0.001	Pearson's chi-squared
Metro	154 (50.2%)	378 (92.2%)	563 (99.6%)	210 (100.0%)	99 (86.1%)	209 (100.0%)		
Regional	153 (49.8%)	32 (7.8%)	2 (0.4%)	0 (0.0%)	16 (13.9%)	0 (0.0%)		
							<0.001	

Factor	ORP before RARP	ORP after RARP	RARP low vol 1st 50 pts	RARP low vol >50 pts	RARP high vol 1st 50 pts	RARP high vol >50 pts	p-value	Test
Public	102 (33.2%)	195 (47.6%)	63 (11.2%)	32 (15.2%)	0 (0.0%)	0 (0.0%)		Pearson's chi-squared
Private	205 (66.8%)	215 (52.4%)	502 (88.8%)	178 (84.8%)	115 (100.0%)	209 (100.0%)		
Surgery year							<0.001	Pearson's chi-squared
2008	3 (1.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
2009	76 (24.8%)	6 (1.5%)	2 (0.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
2010	48 (15.6%)	28 (6.8%)	9 (1.6%)	0 (0.0%)	2 (1.7%)	0 (0.0%)		
2011	53 (17.3%)	65 (15.9%)	67 (11.9%)	0 (0.0%)	54 (47.0%)	0 (0.0%)		
2012	33 (10.7%)	70 (17.1%)	93 (16.5%)	0 (0.0%)	17 (14.8%)	17 (8.1%)		
2013	25 (8.1%)	73 (17.8%)	86 (15.2%)	8 (3.8%)	23 (20.0%)	24 (11.5%)		
2014	25 (8.1%)	48 (11.7%)	138 (24.4%)	10 (4.8%)	3 (2.6%)	55 (26.3%)		
2015	27 (8.8%)	47 (11.5%)	108 (19.1%)	78 (37.1%)	0 (0.0%)	67 (32.1%)		
2016	17 (5.5%)	73 (17.8%)	62 (11.0%)	114 (54.3%)	16 (13.9%)	46 (22.0%)		

Table 24 Patient characteristics for surgeries conducted in metropolitan hospitals only

Factor	ORP before RARP	ORP after RARP	RARP low vol 1st 50 pts	RARP low vol >50 pts	RARP high vol 1st 50 pts	RARP high vol >50 pts	p-value	Test
N	154	378	563	210	99	209		
Outcomes								
No PSM	88 (57.1%)	240 (63.5%)	389 (69.1%)	154 (73.3%)	81 (81.8%)	171 (81.8%)	<0.001	Pearson's chi-squared
PSM	56 (36.4%)	132 (34.9%)	167 (29.7%)	54 (25.7%)	18 (18.2%)	38 (18.2%)		
NR	10 (6.5%)	6 (1.6%)	7 (1.2%)	2 (1.0%)	0 (0.0%)	0 (0.0%)		
No additional management	138 (89.6%)	353 (93.4%)	540 (95.9%)	202 (96.2%)	98 (99.0%)	205 (98.1%)	<0.001	Pearson's chi-squared
Additional management	16 (10.4%)	25 (6.6%)	23 (4.1%)	8 (3.8%)	1 (1.0%)	4 (1.9%)		
SF-12 physical summary score, median (IQR)	54.224 (47.774, 56.705)	53.460 (45.627, 56.513)	55.090 (49.160, 56.705)	49.000 (42.231, 56.416)	55.312 (48.586, 57.203)	56.148 (52.622, 57.761)	0.013	Kruskal-Wallis
SF-12 mental summary score, median (IQR)	53.083 (43.771, 57.890)	54.964 (49.440, 59.036)	57.061 (51.409, 59.530)	57.740 (46.466, 59.429)	56.420 (50.841, 58.154)	57.061 (51.833, 59.429)	0.024	Kruskal-Wallis
Patient characteristics prior to surgery								
Age group (years)							0.031	Pearson's chi-squared
<55	25 (16.2%)	40 (10.6%)	69 (12.3%)	21 (10.0%)	17 (17.2%)	14 (6.7%)		
55-74	129 (83.8%)	338 (89.4%)	494 (87.7%)	189 (90.0%)	82 (82.8%)	195 (93.3%)		

Factor	ORP before RARP	ORP after RARP	RARP low vol 1st 50 pts	RARP low vol >50 pts	RARP high vol 1st 50 pts	RARP high vol >50 pts	p-value	Test
SEIFA decile							<0.001	Pearson's chi-squared
Lowest 40%	43 (27.9%)	79 (20.9%)	150 (26.6%)	57 (27.1%)	18 (18.2%)	13 (6.2%)		
Lowest 41-60%	26 (16.9%)	52 (13.8%)	72 (12.8%)	23 (11.0%)	14 (14.1%)	35 (16.7%)		
Highest 61-100%	85 (55.2%)	247 (65.3%)	341 (60.6%)	130 (61.9%)	67 (67.7%)	161 (77.0%)		
NCCN risk prior to surgery							0.017	Pearson's chi-squared
Intermediate Risk	107 (69.5%)	257 (68.0%)	425 (75.5%)	165 (78.6%)	72 (72.7%)	164 (78.5%)		
High Risk	47 (30.5%)	121 (32.0%)	138 (24.5%)	45 (21.4%)	27 (27.3%)	45 (21.5%)		
Hospital characteristics							<0.001	Pearson's chi-squared
Public	75 (48.7%)	164 (43.4%)	63 (11.2%)	32 (15.2%)	0 (0.0%)	0 (0.0%)		
Private	79 (51.3%)	214 (56.6%)	500 (88.8%)	178 (84.8%)	99 (100.0%)	209 (100.0%)		
Surgery year							<0.001	Pearson's chi-squared
2008	3 (1.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
2009	76 (49.4%)	6 (1.6%)	2 (0.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
2010	47 (30.5%)	28 (7.4%)	9 (1.6%)	0 (0.0%)	2 (2.0%)	0 (0.0%)		
2011	24 (15.6%)	62 (16.4%)	67 (11.9%)	0 (0.0%)	54 (54.5%)	0 (0.0%)		
2012	2 (1.3%)	67 (17.7%)	93 (16.5%)	0 (0.0%)	17 (17.2%)	17 (8.1%)		
2013	1 (0.6%)	67 (17.7%)	86 (15.3%)	8 (3.8%)	23 (23.2%)	24 (11.5%)		
2014	1 (0.6%)	44 (11.6%)	137 (24.3%)	10 (4.8%)	3 (3.0%)	55 (26.3%)		
2015	0 (0.0%)	44 (11.6%)	107 (19.0%)	78 (37.1%)	0 (0.0%)	67 (32.1%)		
2016	0 (0.0%)	60 (15.9%)	62 (11.0%)	114 (54.3%)	0 (0.0%)	46 (22.0%)		

7.10. Chapter 5 Full inverse probability weighted regression adjustment models for PSM

Table 25 Full equations for treatment selection and outcome models for estimating the effect of RARP experience on the probability of PSM

		Comparison 1 ORP after RARP vs ORP prior to RARP		Comparison 2 RARP low vol, 1st 50 pts vs ORP prior to RARP		Comparison 3 RARP low vol, >50 pts vs ORP prior to RARP		Comparison 4 RARP high vol, 1st 50 pts vs ORP prior to RARP		Comparison 5 RARP high vol, >50 pts vs ORP prior to RARP	
Model	**Variable**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**
Control outcome model (ORP prior to RARP)	High vs intermediate NCCN risk	23.4	8.8	24.9	10.4	27.1	10.0	20.8	11.5	17.8	13.8
	SEIFA decile vs lowest 40%										
	Lowest 41-60%	-4.8	12.6	-12.1	17.0	-10.7	16.3	-8.2	20.7	-20.9	20.7
	Highest 61-100%	-19.1	9.1	-25.9	12.3	-24.3	12.2	-29.6	12.9	-28.1	14.5
	Age group 55-74 years vs <55 years	2.9	9.7	16.9	10.7	15.4	10.4	21.9	12.5	14.1	15.6

		Comparison 1 ORP after RARP vs ORP prior to RARP		Comparison 2 RARP low vol, 1st 50 pts vs ORP prior to RARP		Comparison 3 RARP low vol, >50 pts vs ORP prior to RARP		Comparison 4 RARP high vol, 1st 50 pts vs ORP prior to RARP		Comparison 5 RARP high vol, >50 pts vs ORP prior to RARP	
Model	**Variable**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**
	Private vs public hospital	0.5	8.3	3.1	8.6	5.3	9.0	N/A	N/A	N/A	N/A
	Constant	40.1	11.7	30.4	12.9	28.4	12.5	31.8	15.9	40.1	19.6
Treated outcome model (RARP experience)	High vs intermediate NCCN risk	14.1	5.4	12.3	4.7	21.5	8.0	9.2	9.5	22.0	7.5
	SEIFA decile vs lowest 40%										
	Lowest 41-60%	-17.5	8.9	-12.9	6.1	4.4	9.8	12.0	14.6	-3.1	11.8
	Highest 61-100%	-14.9	6.3	-1.3	4.7	11.1	7.0	1.3	10.4	3.5	11.1
	Age group 55-74 years vs <55 years	6.2	7.8	9.2	5.3	-1.1	10.1	6.0	9.4	11.0	8.7
	Private vs public hospital	-16.0	5.1	-7.0	6.5	-18.0	9.5	N/A	N/A	N/A	N/A
	Constant	46.8	8.7	27.6	7.9	30.2	12.7	8.1	12.9	1.0	13.8

		Comparison 1 ORP after RARP vs ORP prior to RARP		Comparison 2 RARP low vol, 1st 50 pts vs ORP prior to RARP		Comparison 3 RARP low vol, >50 pts vs ORP prior to RARP		Comparison 4 RARP high vol, 1st 50 pts vs ORP prior to RARP		Comparison 5 RARP high vol, >50 pts vs ORP prior to RARP	
Model	**Variable**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**	**Change in prob (%)**	**SE (%)**
Surgery selection model	High vs intermediate NCCN risk	2.1	21.4	-22.3	22.4	-37.7	26.7	15.2	35.8	-29.2	32.8
	SEIFA decile vs lowest 40%										
	Lowest 41-60%	6.3	31.2	-40.7	32.0	-70.7	40.0	23.3	57.8	160.1	54.5
	Highest 61-100%	39.8	24.1	-35.2	25.5	-54.3	31.6	-1.0	41.4	132.5	41.9
	Age group 55-74 years vs <55 years	46.0	28.4	24.0	26.9	29.5	35.3	-51.0	47.7	70.1	47.4
	Private vs public hospital	18.7	20.5	218.3	22.5	183.3	27.8	N/A	N/A	N/A	N/A
	Constant	19.6	31.0	-10.4	31.7	-63.6	40.8	70.8	59.4	-69.6	59.1

8. References

Alghamdi, M., Taggar, A., Tilley, D., Kerba, M., Kostaras, X., Gotto, G., & Sia, M. (2016). An audit of referral and treatment patterns of high-risk prostate cancer patients in Alberta. *Canadian Urological Association journal = Journal de l'Association des urologues du Canada, 10*(11-12), 410-415. doi:10.5489/cuaj.3910

Alsinnawi, M., Cullen, J., Hurwitz, L. M., Levie, K. E., Burns, J. F., Rosner, I. L., . . . Porter, C. R. (2019). A prospective study of health-related quality of life outcomes among men treated for intermediate- and high-risk prostate cancer: the impact of primary and secondary therapies. *Can J Urol, 26*(4), 9809-9820.

Ansari, F., Gray, K., Nathwani, D., Phillips, G., Ogston, S., Ramsay, C., & Davey, P. (2003). Outcomes of an intervention to improve hospital antibiotic prescribing: interrupted time series with segmented regression analysis. *Journal of Antimicrobial Chemotherapy, 52*(5), 842-848. doi:10.1093/jac/dkg459

Audet, A.-M. J., Doty, M. M., Shamasdin, J., & Schoenbaum, S. C. (2005). Measure, learn, and improve: physicians' involvement in quality improvement. *Health affairs (Project Hope), 24*(3), 843-853. doi:10.1377/hlthaff.24.3.843

Australian Bureau of Statistics. (2017). *National, state and territory population*. 3101.0 Australian Demographic Statistics. Retrieved from: https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release

Australian Bureau of Statistics. (2018, 23rd March 2018). Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2033.0.55.001Main+Features12016?OpenDocument

Australian Cancer Network Management of Metastatic Prostate Cancer Working Party. (2010). *Clinical Practice Guidelines for the Management of Locally Advanced and Metastatic Prostate Cancer*. Retrieved from Sydney:

Australian Commission on Safety and Quality in Health Care. (2019). Australian Register of Clinical Registries. Retrieved from https://www.safetyandquality.gov.au/australian-register-clinical-registries

Australian Government Services Australia. (2017). Medicare Group Reports. Retrieved from http://medicarestatistics.humanservices.gov.au/statistics/mbs_group.jsp

Australian Institute of Health and Welfare. (2016). Prostate Cancer. *Authoritative information and statistics to promote better health and wellbeing*. Retrieved from http://www.aihw.gov.au/cancer/prostate/

Australian Institute of Health and Welfare. (2020). *Cancer data in Australia*. Retrieved from https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/summary

Avery, A. J., Rodgers, S., Cantrill, J. A., Armstrong, S., Boyd, M., Cresswell, K., . . . Sheikh, A. (2010). *PINCER trial: a cluster randomised trial comparing the effectiveness and cost-effectiveness of a pharmacist-led ITbased intervention with simple feedback in reducing rates of clinically important errors in medicines management in general practices. A report for the Department of Health Patient Safety Research Portfolio*. Retrieved from

Barry, M. J., Gallagher, P. M., Skinner, J. S., & Fowler, F. J. (2012). Adverse Effects of Robotic-Assisted Laparoscopic Versus Open Retropubic Radical Prostatectomy Among a Nationwide Random Sample of Medicare-Age Men. *J Clin Oncol, 30*(5), 513-518. doi:10.1200/jco.2011.36.8621

Basto, M., Sathianathen, N., Te Marvelde, L., Ryan, S., Goad, J., Lawrentschuk, N., . . . Murphy, D. G. (2016). Patterns-of-care and health economic analysis of robot-assisted radical prostatectomy in the Australian public health system. *BJU Int, 117*(6), 930-939. doi:10.1111/bju.13317

Bevan, G., & Hamblin, R. (2009). Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance measurement within the UK. *Journal of the Royal Statistical Society. Series A, (Statistics in Society), 172*(1), 161-190. doi:10.1111/j.1467-985X.2008.00557.x

Bighelli, I., Ostuzzi, G., Girlanda, F., Cipriani, A., Becker, T., Koesters, M., & Barbui, C. (2016). Implementation of treatment guidelines for specialist mental health care. *Cochrane Database of Systematic Reviews*(12). doi:10.1002/14651858.CD009780.pub3

Bird, S. M., Sir David, C., Farewell, V. T., Harvey, G., Tim, H., & Peter C., S. (2005). Performance indicators: good, bad, and ugly. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 168*(1), 1-27. doi:10.1111/j.1467-985X.2004.00333.x

Bolenz, C., Freedland, S. J., Hollenbeck, B. K., Lotan, Y., Lowrance, W. T., Nelson, J. B., & Hu, J. C. (2014). Costs of radical prostatectomy for prostate cancer: a systematic review. *Eur Urol, 65*(2), 316-324. doi:10.1016/j.eururo.2012.08.059

Boorjian, S. A., Karnes, R. J., Crispen, P. L., Carlson, R. E., Rangel, L. J., Bergstralh, E. J., & Blute, M. L. (2010). The impact of positive surgical margins on mortality following radical prostatectomy during the prostate specific antigen era. *J Urol, 183*(3), 1003-1009. doi:10.1016/j.juro.2009.11.039

Burt, L. M., Shrieve, D. C., & Tward, J. D. (2018). Factors influencing prostate cancer patterns of care: An analysis of treatment variation using the SEER database. *Adv Radiat Oncol, 3*(2), 170-180. doi:10.1016/j.adro.2017.12.008

Calopedos, R. J. S., Bang, A., Baade, P., Yu, X. Q., Ruthven, S., Patel, M. I., & Smith, D. P. (2019). Patterns of prostate-specific antigen testing by remoteness of residence and socio-economic status: An Australian

population-based study. *Aust J Rural Health, 27*(3), 216-223. doi:10.1111/ajr.12504

Cancer Institute NSW. (2020). NSW Clinical Cancer Registry (2008–2012). Retrieved from https://www.cancer.nsw.gov.au/research-and-data/cancer-data-and-statistics/request-unlinked-unit-record-data-for-research/nsw-clinical-cancer-registry-2008-2012

Cao, L., Yang, Z., Qi, L., & Chen, M. (2019). Robot-assisted and laparoscopic vs open radical prostatectomy in clinically localized prostate cancer: perioperative, functional, and oncological outcomes: A Systematic review and meta-analysis. *Medicine (Baltimore), 98*(22). doi:10.1097/md.0000000000015770

Cary, C., Odisho, A. Y., & Cooperberg, M. R. (2016). Variation in prostate cancer treatment associated with population density of the county of residence. *Prostate Cancer Prostatic Dis, 19*(2), 174-179. doi:10.1038/pcan.2015.65

Chamie, K., Williams, S. B., Hershman, D. L., Wright, J. D., Nguyen, P. L., & Hu, J. C. (2015). Population-based assessment of determining predictors for quality of prostate cancer surveillance. *Cancer, 121*(23), 4150-4157. doi:10.1002/cncr.29574

Chhapola, V., Tiwari, S., Brar, R., & Kanwal, S. K. (2016). An interrupted time series analysis showed suboptimal improvement in reporting quality of trial abstract. *Journal of Clinical Epidemiology, 71*, 11-17. doi:10.1016/j.jclinepi.2015.10.013

Close, A., Robertson, C., Rushton, S., Shirley, M., Vale, L., Ramsay, C., & Pickard, R. (2013). Comparative cost-effectiveness of robot-assisted and standard laparoscopic prostatectomy as alternatives to open radical prostatectomy for treatment of men with localised prostate cancer: a health technology assessment from the perspective of the UK National Health Service. *Eur Urol, 64*(3), 361-369. doi:10.1016/j.eururo.2013.02.040

Coughlin, G. D., Yaxley, J. W., Chambers, S. K., Occhipinti, S., Samaratunga, H., Zajdlewicz, L., . . . Gardiner, R. A. (2018). Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: 24-month outcomes from a randomised controlled study. *The Lancet Oncology, 19*(8), 1051-1060. doi:https://doi.org/10.1016/S1470-2045(18)30357-7

Curtis, H. J., Walker, A. J., & Goldacre, B. (2018). Impact of NICE guidance on tamoxifen prescribing in England 2011-2017: an interrupted time series analysis. *Br J Cancer, 118*(9), 1268-1275. doi:10.1038/s41416-018-0065-2

Damani, A., Van Hemelrijck, M., Wulaningsih, W., Crawley, D., & Cahill, D. (2017). Are you now a good surgeon? T2 positive margin status as a quality outcome measure following radical prostatectomy. *World J Urol, 35*(1), 35-43. doi:10.1007/s00345-016-1836-0

Dasgupta, P., Baade, P. D., Aitken, J. F., Ralph, N., Chambers, S. K., & Dunn, J. (2019). Geographical Variations in Prostate Cancer Outcomes: A Systematic Review of International Evidence. *Front Oncol, 9*, 238. doi:10.3389/fonc.2019.00238

Dickson, C., Taljaard, M., Friedman, D. S., Metz, G., Wong, T., & Grimshaw, J. M. (2017). The antibiotic management of gonorrhoea in Ontario, Canada

following multiple changes in guidelines: an interrupted time-series analysis. *Sexually Transmitted Infections, 93*(8), 561-565. doi:10.1136/sextrans-2017-053224

Dinh, M. M., Bein, K. J., Hendrie, D., Gabbe, B., Byrne, C. M., & Ivers, R. (2015). Incremental cost-effectiveness of trauma service improvements for road trauma casualties: experience of an Australian major trauma centre. *Aust Health Rev*. doi:10.1071/ah14205

Donabedian, A. (2005). Evaluating the quality of medical care. 1966. *The Milbank quarterly, 83*(4), 691-729. doi:10.1111/j.1468-0009.2005.00397.x

Drazer, M. W., Huo, D., & Eggener, S. E. (2015). National Prostate Cancer Screening Rates After the 2012 US Preventive Services Task Force Recommendation Discouraging Prostate-Specific Antigen-Based Screening. *J Clin Oncol, 33*(22), 2416-2423. doi:10.1200/JCO.2015.61.6532

Eastham, J. A., Kattan, M. W., Riedel, E., Begg, C. B., Wheeler, T. M., Gerigk, C., . . . Scardino, P. T. (2003). Variations among individual surgeons in the rate of positive surgical margins in radical prostatectomy specimens. *J Urol, 170*(6 Pt 1), 2292-2295. doi:10.1097/01.ju.0000091100.83725.51

Emery, J. D., Gray, V., Walter, F. M., Cheetham, S., Croager, E. J., Slevin, T., . . . Holman, C. D. (2014). The Improving Rural Cancer Outcomes (IRCO) Trial: a factorial cluster-randomised controlled trial of a complex intervention to reduce time to diagnosis in rural patients with cancer in Western Australia: a study protocol. *BMJ Open, 4*(9), e006156. doi:10.1136/bmjopen-2014-006156

Eredics, K., Dorfinger, K., Kramer, G., Ponholzer, A., & Madersbacher, S. (2017). Active surveillance for low-risk prostate cancer in Austria: the online registry of the Qualitatspartnerschaft Urologie (QuapU). *Wien Klin Wochenschr, 129*(11-12), 375-379. doi:10.1007/s00508-016-1149-8

Evans, M. A., Millar, J. L., Earnest, A., Frydenberg, M., Davis, I. D., Murphy, D. G., . . . Evans, S. M. (2018). Active surveillance of men with low risk prostate cancer: evidence from the Prostate Cancer Outcomes Registry-Victoria. *Med J Aust, 208*(10), 439-443. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29793403

Evans, S. M., Millar, J. L., Frydenberg, M., Murphy, D. G., Davis, I. D., Spelman, T., . . . McNeill, J. J. (2014). Positive surgical margins: rate, contributing factors and impact on further treatment: findings from the Prostate Cancer Registry. *BJU Int, 114*(5), 680-690. doi:10.1111/bju.12509

Feldman-Stewart, D., Brundage, M. D., Hayter, C., Groome, P., Nickel, J. C., Downes, H., & Mackillop, W. J. (1998). What prostate cancer patients should know: variation in professionals' opinions. *Radiother Oncol, 49*(2), 111-123. doi:10.1016/s0167-8140(98)00091-7

Feletto, E., Bang, A., Cole-Clark, D., Chalasani, V., Rasiah, K., & Smith, D. P. (2015). An examination of prostate cancer trends in Australia, England, Canada and USA: Is the Australian death rate too high? *World J Urol, 33*(11), 1677-1687. doi:10.1007/s00345-015-1514-7

Filson, C. P., Schroeck, F. R., Ye, Z., Wei, J. T., Hollenbeck, B. K., & Miller, D. C. (2014). Variation in use of active surveillance among men undergoing

expectant treatment for early stage prostate cancer. *J Urol, 192*(1), 75-80. doi:10.1016/j.juro.2014.01.105

Flodgren, G., Conterno, L. O., Mayhew, A., Omar, O., Pereira, C. R., & Shepperd, S. (2013). Interventions to improve professional adherence to guidelines for prevention of device-related infections. *Cochrane Database of Systematic Reviews*(3). doi:10.1002/14651858.CD006559.pub2

Flodgren, G., Eccles, M. P., Shepperd, S., Scott, A., Parmelli, E., & Beyer, F. R. (2011). An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database of Systematic Reviews*(7). doi:10.1002/14651858.CD009255

Flodgren, G., Gonçalves-Bradley, D. C., & Pomey, M. P. (2016). External inspection of compliance with standards for improved healthcare outcomes. *Cochrane Database of Systematic Reviews*(12). doi:10.1002/14651858.CD008992.pub3

Fode, M., Sønksen, J., & Jakobsen, H. (2014). Radical prostatectomy: initial experience with robot-assisted laparoscopic procedures at a large university hospital. *Scand J Urol, 48*(3), 252-258. doi:10.3109/21681805.2013.868514

Forsmark, A., Gehrman, J., Angenete, E., Bjartell, A., Björholt, I., Carlsson, S., . . . Haglind, E. (2018). Health Economic Analysis of Open and Robot-assisted Laparoscopic Surgery for Prostate Cancer Within the Prospective Multicentre LAPPRO Trial. *Eur Urol, 74*(6), 816-824. doi:10.1016/j.eururo.2018.07.038

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly Robust Estimation of Causal Effects. *American Journal of Epidemiology, 173*(7), 761-767. doi:10.1093/aje/kwq439

Gardiner, R. A., Yaxley, J., Coughlin, G., Dunglison, N., Occhipinti, S., Younie, S., . . . Chambers, S. K. (2012). A randomised trial of robotic and open prostatectomy in men with localised prostate cancer. *BMC Cancer, 12*, 189. doi:10.1186/1471-2407-12-189

Goldsbury, D. E., Yap, S., Weber, M. F., Veerman, L., Rankin, N., Banks, E., . . . O'Connell, D. L. (2018). Health services costs for cancer care in Australia: Estimates from the 45 and Up Study. *PLoS One, 13*(7), e0201552. doi:10.1371/journal.pone.0201552

Gordon, L. G., Tuffaha, H. W., James, R., Keller, A. T., Lowe, A., Scuffham, P. A., & Gardiner, R. A. (2018). Estimating the healthcare costs of treating prostate cancer in Australia: A Markov modelling analysis. *Urol Oncol, 36*(3), 91 e97-91 e15. doi:10.1016/j.urolonc.2017.10.024

Gore, J. L., Wright, J. L., Daratha, K. B., Roberts, K. P., Lin, D. W., Wessells, H., & Porter, M. (2012). Hospital-level variation in the quality of urologic cancer surgery. *Cancer, 118*(4), 987-996. doi:10.1002/cncr.26373

Gude, W. T., Roos-Blom, M.-J., van der Veer, S. N., Dongelmans, D. A., de Jonge, E., Francis, J. J., . . . de Keizer, N. F. (2018). Health professionals' perceptions about their clinical performance and the influence of audit and feedback on their intentions to improve practice: a theory-based

study in Dutch intensive care units. *Implement Sci, 13*(1), 33-33. doi:10.1186/s13012-018-0727-8

Gumus, E., Boylu, U., Turan, T., & Onol, F. F. (2011). The learning curve of robot-assisted radical prostatectomy. *J Endourol, 25*(10), 1633-1637. doi:10.1089/end.2011.0071

Hayen, A., Smith, D. P., Patel, M. I., & O'Connell, D. L. (2008). Patterns of surgical care for prostate cancer in NSW, 1993-2002: rural/urban and socio-economic variation. *Aust N Z J Public Health, 32*(5), 417-420. doi:10.1111/j.1753-6405.2008.00272.x

Hayes, J. H., Ollendorf, D. A., Pearson, S. D., Barry, M. J., Kantoff, P. W., Lee, P. A., & McMahon, P. M. (2013). Observation versus initial treatment for men with localized, low-risk prostate cancer: a cost-effectiveness analysis. *Annals of internal medicine, 158*(12), 853-860. doi:10.7326/0003-4819-158-12-201306180-00002

Hefermehl, L. J., Disteldorf, D., & Lehmann, K. (2016). Acknowledging unreported problems with active surveillance for prostate cancer: a prospective single-centre observational study. *BMJ Open, 6*(2), e010191. doi:10.1136/bmjopen-2015-010191

Herlemann, A., Cowan, J. E., Carroll, P. R., & Cooperberg, M. R. (2018). Community-based Outcomes of Open versus Robot-assisted Radical Prostatectomy. *European Urology, 73*(2), 215-223. doi:10.1016/j.eururo.2017.04.027

Hjertholm, P., Fenger-Grøn, M., Vestergaard, M., Christensen, M. B., Borre, M., Møller, H., & Vedsted, P. (2015). Variation in general practice prostate-specific antigen testing and prostate cancer outcomes: an ecological study. *Int J Cancer, 136*(2), 435-442. doi:10.1002/ijc.29008

Hohn, D. C. (2012). Health care reform and its implications for the academic cancer center. *J Natl Compr Canc Netw, 10*(1), 130-132. doi:10.6004/jnccn.2012.0012

Hohwu, L., Borre, M., Ehlers, L., & Venborg Pedersen, K. (2011). A short-term cost-effectiveness study comparing robot-assisted laparoscopic and open retropubic radical prostatectomy. *J Med Econ, 14*(4), 403-409. doi:10.3111/13696998.2011.586621

Hoque, D. M. E., Kumari, V., Hoque, M., Ruseckaite, R., Romero, L., & Evans, S. M. (2017). Impact of clinical registries on quality of patient care and clinical outcomes: A systematic review. *PLoS One, 12*(9), e0183667. doi:10.1371/journal.pone.0183667

Huang, W., Zhang, Y., Shen, B. H., Wang, S., Meng, H. Z., & Jin, X. D. (2019). Outcomes of health-related quality of life after open, laparoscopic, or robot-assisted radical prostatectomy in China. *Cancer Manag Res, 11*, 899-907. doi:10.2147/cmar.S189893

Ilic, D., Evans, S. M., Allan, C. A., Jung, J. H., Murphy, D., & Frydenberg, M. (2017). Laparoscopic and robotic-assisted versus open radical prostatectomy for the treatment of localised prostate cancer. *Cochrane Database Syst Rev, 2017*(9). doi:10.1002/14651858.CD009625.pub2

Ilic, D., Evans, S. M., Allan, C. A., Jung, J. H., Murphy, D., & Frydenberg, M. (2018). Laparoscopic and robot-assisted vs open radical prostatectomy for the

treatment of localized prostate cancer: a Cochrane systematic review. *BJU Int, 121*(6), 845-853. doi:10.1111/bju.14062

International Agency for Research on Cancer. (2018). Cancer Today- IARC. Retrieved from https://gco.iarc.fr/today/online-analysis-table?v=2018&mode=population&mode_population=countries&population=900&populations=909_900_36&key=asr&sex=1&cancer=27&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1#collapse-group-0-5

Inwald, E. C., Klinkhammer-Schalke, M., Koller, M., & Ortmann, O. (2014). Quality Assurance for Patients with Breast Cancer - the Impact of Clinical Cancer Registries. *Geburtshilfe und Frauenheilkunde, 74*(9), 868-874. doi:10.1055/s-0034-1383052

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., . . . Oxman, A. D. (2012). Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*(6), Cd000259. doi:10.1002/14651858.CD000259.pub3

Jamtvedt, G., Young, J. M., Kristoffersen, D. T., O'Brien, M. A., & Oxman, A. D. (2006). Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*(2), Cd000259. doi:10.1002/14651858.CD000259.pub2

Jaulim, A., Srinivasan, A., Hori, S., Kumar, N., Warren, A. Y., Shah, N. C., & Gnanapragasam, V. J. (2018). A comparison of operative and margin outcomes from surgeon learning curves in robot assisted radical prostatectomy in a changing referral practice. *Annals of the Royal College of Surgeons of England, 100*(3), 226-229. doi:10.1308/rcsann.2018.0001

Jayadevappa, R., Chhatre, S., Johnson, J. C., & Malkowicz, S. B. (2011). Variation in quality of care among older men with localized prostate cancer. *Cancer, 117*(11), 2520-2529. doi:10.1002/cncr.25812

Keegan, K. A., Dall'Era, M. A., Durbin-Johnson, B., & Evans, C. P. (2012). Active surveillance for prostate cancer compared with immediate treatment: an economic analysis. *Cancer, 118*(14), 3512-3518. doi:10.1002/cncr.26688

Kiefe, C. I., Allison, J. J., Williams, O. D., Person, S. D., Weaver, M. T., & Weissman, N. W. (2001). Improving quality improvement using achievable benchmarks for physician feedback - A randomized controlled trial. *Jama-Journal of the American Medical Association, 285*(22), 2871-2879. doi:DOI 10.1001/jama.285.22.2871

Koerber, F., Waidelich, R., Stollenwerk, B., & Rogowski, W. (2014). The cost-utility of open prostatectomy compared with active surveillance in early localised prostate cancer. *Bmc Health Services Research, 14*, 163. doi:10.1186/1472-6963-14-163

Koh, H. (2017). *Understanding Data Interpretation: A Mixed Method Approach*. Monash University, (25687352)

Kötter, T., Blozik, E., & Scherer, M. (2012). Methods for the guideline-based development of quality indicators--a systematic review. *Implement Sci, 7*, 21-21. doi:10.1186/1748-5908-7-21

Kunz, J., Staub, K., & Winkelmann, R. (2017). Estimating Fixed Effects: Perfect Prediction and Bias in Binary Response Panel Models, with an Application to the Hospital Readmissions Reduction Program. *IZA Institute of Labor Economics, Discussion Paper Series*(11182).

Lacey, K., Bishop, J. F., Cross, H. L., Chondros, P., Lyratzopoulos, G., & Emery, J. D. (2016). Presentations to general practice before a cancer diagnosis in Victoria: a cross-sectional survey. *Med J Aust, 205*(2), 66-71. doi:10.5694/mja15.01169

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics, 95*(2), 391-413. doi:Doi 10.1016/S0304-4076(99)00044-5

Lang, M. F., Tyson, M. D., Alvarez, J. R., Koyama, T., Hoffman, K. E., Resnick, M. J., . . . Barocas, D. A. (2017). The Influence of Psychosocial Constructs on the Adherence to Active Surveillance for Localized Prostate Cancer in a Prospective, Population-based Cohort. *Urology, 103*, 173-178. doi:10.1016/j.urology.2016.12.063

Laviana, A. A., Ilg, A. M., Veruttipong, D., Tan, H. J., Burke, M. A., Niedzwiecki, D. R., . . . Saigal, C. S. (2016). Utilizing time-driven activity-based costing to understand the short- and long-term costs of treating localized, low-risk prostate cancer. *Cancer, 122*(3), 447-455. doi:10.1002/cncr.29743

Lawson, K. A., Daignault, K., Abouassaly, R., Khanna, A., Martin, L., Goldenberg, M., . . . Finelli, A. (2020). Hospital-level Effects Contribute to Variations in Prostate Cancer Quality of Care. *Eur Urol Oncol*. doi:10.1016/j.euo.2020.08.010

Lenderink, A. F., Spreeuwers, D., van der Klink, J. J. L., & van Dijk, F. J. H. (2010). Information and feedback to improve occupational physicians' reporting of occupational diseases: a randomised controlled trial. *International Archives of Occupational and Environmental Health, 83*(4), 381-388. doi:10.1007/s00420-009-0468-8

Leow, J. J., Chang, S. L., Meyer, C. P., Wang, Y., Hanske, J., Sammon, J. D., . . . Trinh, Q. D. (2016). Robot-assisted Versus Open Radical Prostatectomy: A Contemporary Analysis of an All-payer Discharge Database. *Eur Urol, 70*(5), 837-845. doi:10.1016/j.eururo.2016.01.044

Liberman, D., Trinh, Q. D., Jeldres, C., & Zorn, K. C. (2012). Is robotic surgery cost-effective: yes. *Curr Opin Urol, 22*(1), 61-65. doi:10.1097/MOU.0b013e32834d543f

Linden, A. (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. *Stata Journal, 15*(2), 480-500. Retrieved from http://www.stata-journal.com/article.html?article=st0389

Linden, A. (2017). A comprehensive set of postestimation measures to enrich interrupted time-series analysis. *Stata Journal, 17*(1), 73-88. Retrieved from http://www.stata-journal.com/article.html?article=st0389_3

Lipitz-Snyderman, A., Sima, C. S., Atoria, C. L., Elkin, E. B., Anderson, C., Blinder, V., . . . Bach, P. B. (2016). Physician-Driven Variation in Nonrecommended

Services Among Older Adults Diagnosed With Cancer. *JAMA Intern Med, 176*(10), 1541-1548. doi:10.1001/jamainternmed.2016.4426

Loeb, S., Folkvaljon, Y., Makarov, D. V., Bratt, O., Bill-Axelson, A., & Stattin, P. (2015). Five-year nationwide follow-up study of active surveillance for prostate cancer. *Eur Urol, 67*(2), 233-238. doi:10.1016/j.eururo.2014.06.010

Löppenberg, B., Sood, A., Dalela, D., Karabon, P., Sammon, J. D., Vetterlein, M. W., . . . Abdollah, F. (2017). Variation in Locoregional Prostate Cancer Care and Treatment Trends at Commission on Cancer Designated Facilities: A National Cancer Data Base Analysis 2004 to 2013. *Clin Genitourin Cancer, 15*(6), e955-e968. doi:10.1016/j.clgc.2017.04.014

Lowrance, W. T., Elkin, E. B., Jacks, L. M., Yee, D. S., Jang, T. L., Laudone, V. P., . . . Eastham, J. A. (2010). Comparative Effectiveness of Surgical Treatments for Prostate Cancer: A Population-Based Analysis of Postoperative Outcomes. *J Urol, 183*(4), 1366-1372. doi:10.1016/j.juro.2009.12.021

Mansournia, M. A., & Altman, D. G. (2016). Inverse probability weighting. *BMJ, 352*, i189. doi:10.1136/bmj.i189

Martin, E., Persaud, S., Corr, J., Casey, R., & Pillai, R. (2018). Nurse-led active surveillance for prostate cancer is safe, effective and associated with high rates of patient satisfaction-results of an audit in the East of England. *Ecancermedicalscience, 12*, 854-854. doi:10.3332/ecancer.2018.854

McCredie, M., Coates, M., Churches, T., & Taylor, R. (1991). Cancer incidence in New South Wales, Australia. *European Journal of Cancer and Clinical Oncology, 27*(7), 928-931. doi:https://doi.org/10.1016/0277-5379(91)90149-8

Medical Advisory, S. (2010). Robotic-assisted minimally invasive surgery for gynecologic and urologic oncology: an evidence-based analysis. *Ont Health Technol Assess Ser, 10*(27), 1-118. Retrieved from https://pubmed.ncbi.nlm.nih.gov/23074405

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3382308/

Medicare Benefits Schedule Online. (2020). Retrieved from http://www.mbsonline.gov.au

Michigan Urological Surgery Improvement Collaborative. (2020). MUSIC Program Overview. Retrieved from https://musicurology.com/

Miller, D. C., Spencer, B. A., Ritchey, J., Stewart, A. K., Dunn, R. L., Sandler, H. M., . . . Litwin, M. S. (2007). Treatment choice and quality of care for men with localized prostate cancer. *Medical care, 45*(5), 401-409. doi:10.1097/01.mlr.0000255261.81220.29

Modi, P. K., Kaufman, S. R., Borza, T., Yan, P., Miller, D. C., Skolarus, T. A., . . . Hollenbeck, B. K. (2018). Variation in prostate cancer treatment and spending among Medicare shared savings program accountable care organizations. *Cancer, 124*(16), 3364-3371. doi:10.1002/cncr.31573

Mohler, J. L., & Antonarakis, E. S. (2019). NCCN Guidelines Updates: Management of Prostate Cancer. *J Natl Compr Canc Netw, 17*(5.5), 583-586. doi:10.6004/jnccn.2019.5011

Mohler, J. L., Armstrong, A. J., Bahnson, R. R., Boston, B., Busby, J. E., D'Amico, A. V., . . . Shead, D. A. (2012). Prostate Cancer, Version 3.2012 Featured Updates to the NCCN Guidelines. *Journal of the National Comprehensive Cancer Network, 10*(9), 1081-1087. doi:DOI 10.6004/jnccn.2012.0114

Morias, S., Buckley, E., Beckmann, K., O'Callaghan, M., Borg, M., & Short, M. (2020). Variation in radiotherapy patterns of care in the radical treatment of South Australian men with non-metastatic prostate cancer between 2005-2015. *Radiother Oncol, 145*, 138-145. doi:10.1016/j.radonc.2019.12.022

Mouraviev, V., Nosnik, I., Sun, L., Robertson, C. N., Walther, P., Albala, D., . . . Polascik, T. J. (2007). Financial comparative analysis of minimally invasive surgery to open surgery for localized prostate cancer: a single-institution experience. *Urology, 69*(2), 311-314. doi:10.1016/j.urology.2006.10.025

National Health and Medical Research Council. (2014). *NHMRC PSA Testing for Prostate Cancer in Asymptomatic Men, Information for Health Practitioners*. Retrieved from https://www.nhmrc.gov.au/sites/default/files/documents/reports/clinical%20guidelines/men4d-psa-testing-asymptomatic.pdf

National Institute for Health and Care Excellence. (2019). *NICE guideline NG131 Prostate cancer: diagnosis and management*. Retrieved from https://www.nice.org.uk/guidance/ng131

Naughton, C., Feely, J., & Bennett, K. (2009). A RCT evaluating the effectiveness and cost-effectiveness of academic detailing versus postal prescribing feedback in changing GP antibiotic prescribing. *Journal of Evaluation in Clinical Practice, 15*(5), 807-812. doi:10.1111/j.1365-2753.2008.01099.x

Nguyen, P. L., Gu, X., Lipsitz, S. R., Choueiri, T. K., Choi, W. W., Lei, Y., . . . Hu, J. C. (2011). Cost Implications of the Rapid Adoption of Newer Technologies for Treating Prostate Cancer. *J Clin Oncol, 29*(12), 1517-1524. doi:10.1200/jco.2010.31.1217

Oh, J. J., Hong, S. K., Byun, S. S., Choe, G., & Lee, S. E. (2013). Prognostic significance of positive surgical margins after radical prostatectomy among pT2 and pT3a prostate cancer. *Urol Oncol, 31*(5), 595-600. doi:10.1016/j.urolonc.2011.05.003

Ong, W. L., Evans, S. M., Evans, M., Tacey, M., Dodds, L., Kearns, P., . . . Millar, J. (2019). Trends in Conservative Management for Low-risk Prostate Cancer in a Population-based Cohort of Australian Men Diagnosed Between 2009 and 2016. *Eur Urol Oncol*. doi:10.1016/j.euo.2019.04.006

Pan, X. W., Cui, X. M., Teng, J. F., Zhang, D. X., Wang, Z. J., Qu, F. J., . . . Xu, D. F. (2015). Robot-Assisted Radical Prostatectomy vs. Open Retropubic Radical Prostatectomy for Prostate Cancer: A Systematic Review and Meta-analysis. *Indian J Surg, 77*(Suppl 3), 1326-1333. doi:10.1007/s12262-014-1170-y

Pickles, K., Carter, S. M., & Rychetnik, L. (2015). Doctors' approaches to PSA testing and overdiagnosis in primary healthcare: a qualitative study. *BMJ Open, 5*(3), e006367. doi:10.1136/bmjopen-2014-006367

Ploussard, G., Agamy, M. A., Alenda, O., Allory, Y., Mouracade, P., Vordos, D., . . . Salomon, L. (2011). Impact of positive surgical margins on prostate-specific antigen failure after radical prostatectomy in adjuvant treatment-naive patients. *BJU Int, 107*(11), 1748-1754. doi:10.1111/j.1464-410X.2010.09728.x

Pollack, C. E., Weissman, G., Bekelman, J., Liao, K., & Armstrong, K. (2012). Physician social networks and variation in prostate cancer treatment in three cities. *Health Serv Res, 47*(1 Pt 2), 380-403. doi:10.1111/j.1475-6773.2011.01331.x

Pooli, A., Salmasi, A., Faiena, I., Lenis, A. T., Johnson, D. C., Lebacle, C., . . . Chamie, K. (2019). Variation in surgical treatment patterns for patients with prostate cancer in the United States: Do patients in academic hospitals fare better? *Urol Oncol, 37*(1), 63-70. doi:10.1016/j.urolonc.2018.10.018

Pooli, A., Salmasi, A., Johnson, D. C., Lenis, A. T., Faiena, I., Lebacle, C., . . . Chamie, K. (2020). Positive surgical margins at radical prostatectomy in the United States: Institutional variations and predictive factors. *Urol Oncol, 38*(1), 1.e17-11.e23. doi:10.1016/j.urolonc.2019.08.016

Preisser, F., Mazzone, E., Knipper, S., Nazzani, S., Bandini, M., Shariat, S. F., . . . Karakiewicz, P. I. (2019). Rates of Positive Surgical Margins and Their Effect on Cancer-specific Mortality at Radical Prostatectomy for Patients With Clinically Localized Prostate Cancer. *Clin Genitourin Cancer, 17*(1), e130-e139. doi:10.1016/j.clgc.2018.09.024

Prochaska, J. O., & DiClemente, C. C. (1984). *The Transtheoretical Approach: Crossing Traditional Boundaries of Therapy.*: Dow Jones Irwin, Homewood, IL.

Pross, C., Geissler, A., & Busse, R. (2017). Measuring, Reporting, and Rewarding Quality of Care in 5 Nations: 5 Policy Levers to Enhance Hospital Quality Accountability. *The Milbank quarterly, 95*(1), 136-183. doi:10.1111/1468-0009.12248

Prostate Cancer Foundation of Australia, & Griffith University. (2016). *Economic modelling of healthcare services for prostate cancer*. Retrieved from Monographs in prostate cancer:

Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel. (2015, 14th January 2016). Clinical practice guidelines PSA Testing and Early Management of Test-Detected Prostate Cancer. Retrieved from http://wiki.cancer.org.au/australiawiki/index.php?oldid=152048

Prostate Cancer Foundation of Australia and Cancer Council Australia PSA Testing Guidelines Expert Advisory Panel. (2016). *Draft clinical practice guidelines for PSA testing and early management of test-detected prostate cancer.* . Retrieved from http://www.prostate.org.au/media/612113/PSA-Testing-Guidelines.pdf

Ramsay, C., Pickard, R., Robertson, C., Close, A., Vale, L., Armstrong, N., . . . Soomro, N. (2012). Systematic review and economic modelling of the relative clinical benefit and cost-effectiveness of laparoscopic surgery and robotic surgery for removal of the prostate in men with localised

prostate cancer. *Health Technol Assess, 16*(41), 1-313. doi:10.3310/hta16410

Riedinger, C. B., Womble, P. R., Linsell, S. M., Ye, Z., Montie, J. E., Miller, D. C., & Lane, B. R. (2014). Variation in prostate cancer detection rates in a statewide quality improvement collaborative. *J Urol, 192*(2), 373-378. doi:10.1016/j.juro.2014.02.088

Robertson, C., Close, A., Fraser, C., Gurung, T., Jia, X., Sharma, P., . . . Pickard, R. (2013). Relative effectiveness of robot-assisted and standard laparoscopic prostatectomy as alternatives to open radical prostatectomy for treatment of localised prostate cancer: a systematic review and mixed treatment comparison meta-analysis. *BJU Int, 112*(6), 798-812. doi:10.1111/bju.12247

Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology, 11*(5), 550-560. Retrieved from https://journals.lww.com/epidem/Fulltext/2000/09000/Marginal_Structural_Models_and_Causal_Inference_in.11.aspx

Robotic Surgical System for Radical Prostatectomy: A Health Technology Assessment. (2017). *Ont Health Technol Assess Ser, 17*(11), 1-172.

Roehrborn, C. G., & Black, L. K. (2011). The economic burden of prostate cancer. *BJU Int, 108*(6), 806-813. doi:10.1111/j.1464-410X.2011.10365.x

Royal Australian College of General Practitioners. (2005). Guidelines for preventative activities in general practice. In R. A. C. o. G. Practitioners (Ed.), *The Red Book* (6th edn ed., pp. 56-57). South Melbourne, Vic: RACGP.

Royal Australian College of General Practitioners. (2009). Guidelines for preventative activities in general practice. In R. A. C. o. G. Practitioners (Ed.), *The Red Book* (7th edn ed., pp. 56-57). South Melbourne, Vic: RACGP.

Royal Australian College of General Practitioners. (2012). Guidelines for preventative activities in general practice. In R. A. C. o. G. Practitioners (Ed.), *The Red Book* (8th edn ed., pp. 71-72). East Melbourne, Vic: RACGP.

Royal Australian College of General Practitioners. (2018). Guidelines for preventative activities in general practice. In R. A. C. o. G. Practitioners (Ed.), *The Red Book* (9th edn, updated ed., pp. 104). East Melbourne, Vic: RACGP.

Ruseckaite, R., Beckmann, K., O'Callaghan, M., Roder, D., Moretti, K., Millar, J., & Evans, S. (2016). A retrospective analysis of Victorian and South Australian clinical registries for prostate cancer: trends in clinical presentation and management of the disease. *BMC Cancer, 16*, 607. doi:10.1186/s12885-016-2655-9

SA-PCCOC. (2020). South Australian Prostate Cancer Clinical Outcome Collaborative Retrieved from https://www.prostatehealth.org.au/

Sampurno F and Evans SM (eds) for the Victorian Prostate Cancer Clinical Registry Steering Committee. (2015). *Victorian Prostate Cancer Clinical*

Registry - Five Year Report [Internet]. Retrieved from Melbourne, Victoria: http://pcr.registry.org.au

Sampurno, F., Earnest, A., Kumari, P. B., Millar, J. L., Davis, I. D., Murphy, D. G., . . . Evans, S. M. (2016). Quality of care achievements of the Prostate Cancer Outcomes Registry-Victoria. *Med J Aust, 204*(8), 319. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27125808

Satkunasivam, R., Lo, M., Stern, M., Gill, I. S., Fleming, S., Wu, X.-C., . . . Hamilton, A. S. (2018). The Role of Provider Characteristics in the Selection of Surgery or Radiation for Localized Prostate Cancer and Association With Quality of Care Indicators. *American journal of clinical oncology, 41*(11), 1076-1082. doi:10.1097/COC.0000000000000442

Schneider, A., Wensing, M., Biessecker, K., Quinzler, R., Kaufmann-Kolle, P., & Szecsenyi, J. (2008). Impact of quality circles for improvement of asthma care: results of a randomized controlled trial. *J Eval Clin Pract, 14*(2), 185-190. doi:10.1111/j.1365-2753.2007.00827.x

Schroeck, F. R., Jacobs, B. L., Bhayani, S. B., Nguyen, P. L., Penson, D., & Hu, J. (2017). Cost of New Technologies in Prostate Cancer Treatment: Systematic Review of Costs and Cost Effectiveness of Robotic-assisted Laparoscopic Prostatectomy, Intensity-modulated Radiotherapy, and Proton Beam Therapy. *European Urology, 72*(5), 712-735. doi:10.1016/j.eururo.2017.03.028

Schroeck, F. R., Jacobs, B. L., & Hollenbeck, B. K. (2013). Understanding variation in the quality of the surgical treatment of prostate cancer. *Am Soc Clin Oncol Educ Book*, 278-283. doi:10.14694/EdBook_AM.2013.33.278

Schuessler, W. W., Schulam, P. G., Clayman, R. V., & Kavoussi, L. R. (1997). Laparoscopic radical prostatectomy: initial short-term experience. *Urology, 50*(6), 854-857. doi:10.1016/s0090-4295(97)00543-8

Seidenwurm, D., & Logsdon, M. (2014). Localized prostate cancer: regional variation and national policy. *Medical care, 52*(8), 677-679. doi:10.1097/mlr.0000000000000192

Share, D. A., Campbell, D. A., Birkmeyer, N., Prager, R. L., Gurm, H. S., Moscucci, M., . . . Birkmeyer, J. D. (2011). How a regional collaborative of hospitals and physicians in Michigan cut costs and improved the quality of care. *Health Aff (Millwood), 30*(4), 636-645. doi:10.1377/hlthaff.2010.0526

Shaw, A. B. (1968). The oldest medical societies in Great Britain. *Medical history, 12*(3), 232-244. doi:10.1017/s0025727300013272

Shuford, M. D. (2007). Robotically assisted laparoscopic radical prostatectomy: a brief review of outcomes. *Proceedings (Baylor University. Medical Center), 20*(4), 354-356. doi:10.1080/08998280.2007.11928322

Silberstein, J. L., & Eastham, J. A. (2014). Significance and management of positive surgical margins at the time of radical prostatectomy. *Indian Journal of Urology : IJU : Journal of the Urological Society of India, 30*(4), 423-428. doi:10.4103/0970-1591.134240

Simonton, D. K. (1977). Cross-Sectional Time-Series Experiments - Some Suggested Statistical-Analyses. *Psychological Bulletin, 84*(3), 489-502. doi:Doi 10.1037/0033-2909.84.3.489

Simonton, D. K. (1979). Alternatives to Simonton Analyses of the Interrupted and Multiple-Group Time-Series Designs - Reply. *Psychological Bulletin, 86*(5), 927-928. Retrieved from <Go to ISI>://WOS:A1979JU03500003

Sivaraman, A., Sanchez-Salas, R., Prapotnich, D., Yu, K., Olivier, F., Secin, F. P., . . . Cathelineau, X. (2017). Learning curve of minimally invasive radical prostatectomy: Comprehensive evaluation and cumulative summation analysis of oncological outcomes. *Urol Oncol, 35*(4), 149.e141-149.e146. doi:10.1016/j.urolonc.2016.10.015

Smith, D. P., & Armstrong, B. K. (1998). Prostate-specific antigen testing in Australia and association with prostate cancer incidence in New South Wales. *Med J Aust, 169*(1), 17-20. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9695696

Smith, D. P., Supramaniam, R., Marshall, V. R., & Armstrong, B. K. (2008). Prostate cancer and prostate-specific antigen testing in New South Wales. *Med J Aust, 189*(6), 315-318. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18803534

Sohn, W., Resnick, M. J., Greenfield, S., Kaplan, S. H., Phillips, S., Koyama, T., . . . Barocas, D. A. (2016). Impact of Adherence to Quality Measures for Localized Prostate Cancer on Patient-reported Health-related Quality of Life Outcomes, Patient Satisfaction, and Treatment-related Complications. *Medical care, 54*(8), 738-744. doi:10.1097/MLR.0000000000000562

Soleymani, F., Rashidian, A., Dinarvand, R., Kebriaeezade, A., Hosseini, M., & Abdollahi, M. (2012). Assessing the effectiveness and cost-effectiveness of audit and feedback on physician's prescribing indicators: study protocol of a randomized controlled trial with economic evaluation. *Daru, 20*(1), 88. doi:10.1186/2008-2231-20-88

Sondergaard, J., Andersen, M., Vach, K., Kragstrup, J., Maclure, M., & Gram, L. F. (2002). Detailed postal feedback about prescribing to asthma patients combined with a guideline statement showed no impact: a randomised controlled trial. *Eur J Clin Pharmacol, 58*(2), 127-132. doi:10.1007/s00228-002-0454-5

Song, B., Lee, H., Lee, M. S., & Hong, S. K. (2019). Outcomes of men aged ≤50 years treated with radical prostatectomy: a retrospective analysis. *Asian journal of andrology, 21*(2), 150-155. doi:10.4103/aja.aja_92_18

Spiegelhalter, D. J. (2005). Funnel plots for comparing institutional performance. *Stat Med, 24*(8), 1185-1202. doi:10.1002/sim.1970

StataCorp. (2019). Stata: Release 16. *Statistical Software*.

Steele, M. A., Bess, D. T., Franse, V. L., & Graber, S. E. (1989). Cost effectiveness of two interventions for reducing outpatient prescribing costs. *Dicp, 23*(6), 497-500.

Stephenson, A. J., Wood, D. P., Kattan, M. W., Klein, E. A., Scardino, P. T., Eastham, J. A., & Carver, B. S. (2009). Location, extent and number of positive surgical margins do not improve accuracy of predicting prostate cancer recurrence after radical prostatectomy. *J Urol, 182*(4), 1357-1363. doi:10.1016/j.juro.2009.06.046

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., . . . Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, d4002. doi:10.1136/bmj.d4002

Stey, A. M., Russell, M. M., Ko, C. Y., Sacks, G. D., Dawes, A. J., & Gibbons, M. M. (2015). Clinical registries and quality measurement in surgery: a systematic review. *Surgery, 157*(2), 381-395. doi:10.1016/j.surg.2014.08.097

Swindle, P., Eastham, J. A., Ohori, M., Kattan, M. W., Wheeler, T., Maru, N., . . . Scardino, P. T. (2005). Do margins matter? The prognostic significance of positive surgical margins in radical prostatectomy specimens. *J Urol, 174*(3), 903-907. doi:10.1097/01.ju.0000169475.00949.78

Tang, K., Jiang, K., Chen, H., Chen, Z., Xu, H., & Ye, Z. (2017). Robotic vs. Retropubic radical prostatectomy in prostate cancer: A systematic review and an meta-analysis update. *Oncotarget, 8*(19), 32237-32257. doi:10.18632/oncotarget.13332

The Australian Commission on Safety and Quality in Health Care. (2016). *Economic evaluation of clinical quality registries: Final report*. Retrieved from

Tran, K., Sandoval, C., Rahal, R., Porter, G., Siemens, R., Hernandez, J., . . . Technical Working, G. (2015). Wait times for prostate cancer treatment and patient perceptions of care in Canada: a mixed-methods report. *Current oncology (Toronto, Ont.), 22*(5), 361-364. doi:10.3747/co.22.2795

Turner-Bowker, D., & Hogue, S. J. (2014). Short Form 12 Health Survey (SF-12). In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 5954-5957). Dordrecht: Springer Netherlands.

Tzortziou Brown, V., Underwood, M., Mohamed, N., Westwood, O., & Morrissey, D. (2016). Professional interventions for general practitioners on the management of musculoskeletal conditions. *Cochrane Database of Systematic Reviews*(5). doi:10.1002/14651858.CD007495.pub2

US Preventative Services Task Force. (2008). Prostate Cancer Screening statements (Archived) Retrieved from https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/prostate-cancer-screening-2008

US Preventative Services Task Force. (2012). Prostate Cancer Screening statements (Archived) Retrieved from https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/prostate-cancer-screening-2012

Vesey, S. G., McCabe, J. E., Hounsome, L., & Fowler, S. (2012). UK radical prostatectomy outcomes and surgeon case volume: based on an analysis of the British Association of Urological Surgeons Complex Operations Database. *BJU Int, 109*(3), 346-354. doi:10.1111/j.1464-410X.2011.10334.x

Vora, A. A., Marchalik, D., Kowalczyk, K. J., Nissim, H., Bandi, G., McGeagh, K. G., . . . Hwang, J. J. (2013). Robotic-assisted prostatectomy and open radical retropubic prostatectomy for locally-advanced prostate

cancer: multi-institution comparison of oncologic outcomes. *Prostate Int, 1*(1), 31-36. doi:10.12954/pi.12001

Vratsistas-Curto, A., McCluskey, A., & Schurr, K. (2017). Use of audit, feedback and education increased guideline implementation in a multidisciplinary stroke unit. *BMJ Open Quality, 6*(2), e000212. doi:10.1136/bmjoq-2017-000212

Wang, S. Y., Wang, R., Yu, J. B., Ma, X., Xu, X., Kim, S. P., . . . Gross, C. P. (2014). Understanding regional variation in Medicare expenditures for initial episodes of prostate cancer care. *Medical care, 52*(8), 680-687. doi:10.1097/mlr.0000000000000158

Ware, J. E. (2002). *The SF-12v2TM how to score version 2 of the SF-12® health survey:(with a supplement documenting version 1)*: Quality metric.

Ware, J. E. (2008). *User's manual for the SF-36v2 health survey*: Quality Metric.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey - Construction of scales and preliminary tests of reliability and validity. *Medical care, 34*(3), 220-233. doi:Doi 10.1097/00005650-199603000-00003

Ware Jr, J. E. (1999). SF-36 health survey.

Wennberg, J., & Gittelsohn, A. (1973). Small Area Variations in Health-Care Delivery. *Science, 182*(4117), 1102-1107. doi:DOI 10.1126/science.182.4117.1102

Wright, J. L., Dalkin, B. L., True, L. D., Ellis, W. J., Stanford, J. L., Lange, P. H., & Lin, D. W. (2010). Positive surgical margins at radical prostatectomy predict prostate cancer-specific mortality: support for optimizing surgical technique and pathological evaluation at radical prostatectomy. *J Urol, 183*(6), 2213-2218. doi:10.1016/j.juro.2010.02.017

Wu, R. C., Prebay, Z. J., Patel, P., Kim, T., Qi, J., Telang, J., . . . Johnston, W. K., 3rd. (2020). Using video review to understand the technical variation of robot-assisted radical prostatectomy in a statewide surgical collaborative. *World J Urol, 38*(7), 1607-1613. doi:10.1007/s00345-019-02906-4

Yaxley, J. W., Coughlin, G. D., Chambers, S. K., Occhipinti, S., Samaratunga, H., Zajdlewicz, L., . . . Gardiner, R. A. (2016). Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. *Lancet, 388*(10049), 1057-1066. doi:10.1016/s0140-6736(16)30592-x

Zhang, L., Wu, B., Zha, Z., Zhao, H., Yuan, J., Jiang, Y., & Yang, W. (2018). Surgical margin status and its impact on prostate cancer prognosis after radical prostatectomy: a meta-analysis. *World J Urol, 36*(11), 1803-1815. doi:10.1007/s00345-018-2333-4