SATISFICING AS A DRIVER OF LOW-VALUE CARE AND HEALTHCARE COSTS

A thesis submitted to fulfill requirements for the degree of Doctor of Philosophy for the School of Public Health, Faculty of Medicine, The University of Sydney

Year: 2025

Tet Far Soon

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

Tet Far Soon

Date: 25/9/2024

.

Authorship attribution statement

Chapter 2

Chapter 2 of this thesis will be submitted for consideration for publication in JAMA Internal Medicine as: **Soon J**, Tracy M, Scott I, Elshaug AG, Howard K, Traeger AC. Use of heuristics in testing, screening and treatment decisions: a systematic review. 2024.

I conceived the idea for this systematic review, was one of three authors who screened identified references for inclusion, extracted data and assessed risk of bias for each included article, performed additional analysis of data from the included studies, interpreted the results, wrote the initial draft of the paper and drafted and revised the paper.

Chapter 3

Chapter 3 of this thesis has been published as **Soon J**, Traeger AC, Elshaug AG, Cvejic E, Maher CG, Doust JA, Mathieson S, McCaffery K, Bonner C. Effect of two behavioural 'nudging' interventions on management decisions for low back pain: a randomised vignette-based study in general practitioners. *BMJ Qual Saf*. 2019 Jul;28(7):547-555. doi: 10.1136/bmjqs-2018-008659. Epub 2018 Nov 19. PMID: 30455177.

I conceived the idea for this paper, co-designed the study with all co-authors, obtained ethics approval, planned the research and acquired the data and performed the statistical analysis along with three co-authors, interpreted the results, wrote the initial draft of the paper, drafted and revised the manuscripts for submission to the journal and served as corresponding author.

Chapter 4

Chapter 4 of this thesis will be submitted for consideration for publication in Journal of Medical Imaging and Oncology as: **Soon J**, Andrews M, Badgery-Parker T, Elshaug AG, Howard K. CT and MRI substitution in presence of licencing constraints? 2024

I conceived the idea for this paper, planned the research and acquired the data and performed the statistical analysis, interpreted the results, wrote the initial draft of the paper, and drafted and revised the paper.

Chapter 5

Chapter 5 of this thesis has been published as **Soon J,** Houssami N, Clemson M, Lockie D, Farber R, Barratt A, Elshaug A, Howard K. The financial implications of investigating false-positive and true-positive mammograms in a national breast cancer screening program. *Aust Health Rev.* 2023 Apr;47(2):159-164. doi: 10.1071/AH22120. PMID: 36476540.

I conceived the idea for this paper, co-designed the study with all co-authors, obtained ethics approval, planned the research and acquired the data and performed the cost modelling analysis, interpreted the results, wrote the initial draft of the paper, drafted and revised the manuscripts for submission to the journal and served as corresponding author.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Supervisor Name, Signature, Date

Professor Kirsten Howard

Date: 20/09/2024

Ethical Clearance

Ethics approval for the study described in Chapter 3 was obtained from the Royal Prince Alfred Hospital Ethics Review Committee (Approval number HREC/18/RPAH/138. Ethics approval for the study described in Chapter 5 was obtained from Eastern Health (HREC reference number R19/031).

Abstract

This thesis explores 'satisficing' in clinical decision-making, where clinicians choose ‘good enough’ solutions rather than optimal ones due to cognitive limitations and system constraints. Through four interconnected studies, I investigate this phenomenon and its implications for healthcare delivery and policy, employing diverse methodologies to examine different aspects of satisficing in clinical decision-making.

I began by examining heuristics in Chapter 2, recognising them as a form of clinician satisficing in adapting to high-pressure environments. To understand how heuristics impact medical decisions in actual practice and to assess their statistical significance and magnitudes, I conducted a systematic literature review to synthesise existing evidence on heuristic-driven decision-making by medical practitioners in real-world settings focusing on heuristics prompted by initial patient triage information and prompted by practitioners' prior experiences with medical events. By concentrating on these two heuristics, I hoped to uncover any variations in effect sizes, timing, and persistence of these heuristics.

A systematic search of five databases (Ovid, PsycINFO, Econlit, Scopus, Web of Science) to identify relevant literature from January 1980 to September 2023 was undertaken. Included studies were empirical studies in clinical settings that measured changes in practitioners' behaviour related to testing, screening, or treatment, investigated hypotheses on the heuristics of interest, and were published in English-language peer-reviewed journals. Study selection, data extraction, and quality assessment followed standardized protocols, with quality assessed via the Newcastle-Ottawa Scale (NOS).

This search yielded ten high-quality observational studies (NOS-rated). Two studies examined heuristics shaped by initial patient triage information, showing statistically significant odds ratios (ORs) of 1.38 to 1.7 for heuristic-driven medical decisions (e.g., avoiding pulmonary embolism testing or increasing trauma centre transfers). Eight studies focused on heuristics prompted by previous medical events, with ORs ranging from 1.02 to 1.72 for decisions such as switching obstetric methods, ordering specific tests, or avoiding certain therapeutics. Temporal effects, assessed only for heuristics based on previous medical events, mostly lasted 3-6 months. Studies spanned various specialties including emergency medicine, obstetrics, and primary care. Only two studies reported or estimated measures which provided evidence of a direct link between heuristics and health outcomes.

Despite limited direct evidence linking heuristics to health outcomes, I concluded that heuristic-driven changes in clinical management are likely to result in lower quality of care due to deviations from evidence-based practice.

I considered that while some heuristics may lead to suboptimal care, others could be leveraged to improve decision-making. This led to the question of whether nudges which in substance utilise certain heuristics, can effectively counteract harmful decision-making patterns and promote better care. I was also interested in whether nudges had any longer-term educative effects.

The study in Chapter 3 employed a 2x2 factorial design to examine the effectiveness of different nudge strategies in influencing general practitioners' (GPs') management of low back pain. The experiment involved four clinical vignettes, with GP participants assigned to four groups: one receiving a default nudge (high-value options presented as the default, low-value options presented only after clicking for more), another a partition display nudge (low-value options presented horizontally, high value options listed vertically), a third experiencing both nudges combined, and a control group receiving no nudge. The nudges were integrated into the presentation of treatment options. The primary outcome was the proportion of scenarios where practitioners chose at least one of the low-value care options. The secondary outcome was the likelihood that participants recommended low value options when displayed conventionally (i.e. without nudges) after they had made their initial choice of management.

120 GPs completed the trial (n=480 vignettes). I found that participants exposed to the default option nudge were 44% less likely to choose at least one low-value care option (odds ratio [OR] 0.56, 95% CI 0·37 to 0.85; p=0·006) compared to those not exposed. However, the partition display nudge had no effect on choice of low-value care (OR 1.08, 95% CI 0·72 to 1.64; p=0.7) nor was there any interaction between the nudges (OR 0.94, 95% CI 0·41 to 2.15; p=0·89).

The nudges had no impact on an additional secondary outcome measure of low-value decision-making taken shortly after in a separate format, suggesting that even the default nudge is only effective at the point of decision-making with no long-term educative impact. This variability in effectiveness highlights the importance of specific nudge design in influencing clinical decision-making.

While the first two studies focused on internal cognitive processes, I next turned to the question of whether resource constraints in healthcare systems influence clinical decision-making and potentially lead to satisficing behaviours among medical decision-makers. My research question was: Does limited availability of Medicare Benefits Schedule (MBS) eligible magnetic resonance imaging (MRI) services lead to increased computed tomography (CT) imaging orders as a substitute, indicating satisficing behaviour among medical practitioners? I was also interested in evaluating whether other factors, such as income and availability of different types of medical service providers, influence the utilisation levels of various imaging modalities and whether these relationships differed between GPs and non-GP specialists.

To illuminate these relationships, in Chapter 4 MBS-funded CT and MRI head imaging data for 2018-2019 were analysed across Australian Statistical Area Level 4 regions. Primary models used 2019 data to examine relationships between imaging utilization, MRI density, healthcare provider density, and regional median income. Supplementary fixed-effects models used 2018-2019 data to control for unmeasured regional characteristics.

Primary models revealed that higher MRI density regions had a lower CT share of total imaging. This suggests that if there were a greater availability of MRI machines, there might be some reallocation of existing imaging investigation from CT imaging to MRI. This indicates that medical practitioners may be adapting their imaging choices based on available resources, potentially leading to suboptimal use of CT in regions with limited MRI access.

Primary models also showed that GP-requested CT utilisation and CT share of imaging were negatively associated with regional median income. Supplementary models confirmed the negative relationship between income and GP-requested CT imaging. This was a robust finding across both model types and may also indicate medical practitioner ‘satisficing’ as patients in lower-income regions may be more likely to be referred for CT imaging instead of MRI due to higher out-of-pocket expenses.

The issues considered so far highlight how low-value care can emerge not just from individual clinical decisions, but also from broader systemic issues in healthcare resource allocation and availability. The final study in this thesis in Chapter 5 sought to quantify an example of all these factors coming together in a complex healthcare programme such as breast cancer screening. It asked the questions of: What are the magnitude and share of costs associated with false positive investigations in Australia’s breast cancer screening programme? How do these costs compare internationally, and what do these comparisons reveal about the influence of country-specific factors on false positive rates of testing? The motivation for these questions is not because all false positives are attributable to low-value care but because the different magnitudes of false positive rates and false positive costs in different countries may be reflective of their respective systemic propensities to low-value care. There may therefore be value in formulating approaches to better quantify aspects of healthcare systems such as the costs associated with false positives in their cancer screening programs.

The analysis in Chapter 5 used aggregate-level retrospective cohort data of women screened at a breast-screening clinic. Counts and frequencies of each diagnostic workup-sequence recorded were scaled up to national figures and costed by estimating per-patient costs of procedures using screening clinic cost data. Main outcomes and measures estimated were percentage share of total annual screening and further-investigation costs for the Australian population breast-screening program from investigating false-positive and true-positive mammograms.

Secondary outcomes were average costs of investigating each false-positive and true-positive mammogram. Sensitivity analyses involved recalculating results excluding subgroups of patients below and above the screening age range of 50-74 years.

My analysis estimated that 15.4% (ranging from 13.4%, to 15.4% under different scenarios) of the total annual screening and further-investigation costs of Australia's national breast screening program were likely from investigating false-positive mammograms. This share exceeded the costs from investigating true-positives (13%). The estimated investigation cost per false-positive mammogram in Australia (US$190.65) was estimated to be lower than estimates for the US and Sweden (ranging from US$623 to US$1600). Sensitivity analyses confirmed that excluding patients outside the invited screening age range of 50 to 74 years and below the screening range resulted in a lower false positive rate and reduced cost share of false-positive mammogram investigations although only excluding those above the screening age range did not have much impact.

In conclusion, the studies in this thesis enhance our understanding of clinical satisficing, indicating that both cognitive factors and systemic constraints significantly influence clinical decision-making and resource utilisation. The progression from understanding heuristics, to testing interventions, to examining systemic constraints, and finally to analysing their collective impact offers a multi-faceted view of satisficing in healthcare. My findings suggest that potential improvements in healthcare quality and efficiency can be achieved through enhanced medical education on cognitive biases, careful design of clinical decision support systems (particularly leveraging default option nudges), and more equitable distribution of healthcare resources. The substantial economic impact of satisficing behaviours in complex healthcare programs, as demonstrated in my breast cancer screening study, underscores the importance of addressing these issues at a systemic level. Future research should focus on establishing direct links between satisficing behaviours and health outcomes, exploring long-term intervention effects, and developing more standardised healthcare cost analysis methodologies to facilitate international benchmarking. Interdisciplinary research integrating psychology, health economics, and clinical practice, along with international comparisons, will be crucial in addressing the complex factors influencing healthcare delivery.

Acknowledgements

I gratefully acknowledge the help of my primary supervisors, Professor Kirsten Howard and Professor Adam Elshaug for sticking through the long and arduous process of supervising a part-time PhD student who took the maximum allowable length to complete his thesis (through job changes on both sides) and through the ebbing and flowing of different research ideas and leads, some of which didn't end up panning out and others of which did.

I would also like to thank all my other co-authors on the studies included in this thesis as their contributions have all been crucial. I would like to single out for special thanks Dr Adrian Traeger who was a key collaborator on my first paper on nudges and has been a frequent and enthusiastic collaborator on all the heuristic and nudge associated projects I have been involved in (including those not covered here).

Finally, I thank my family including my parents, and my partner, who has had to put up with more frequent quiet weekends towards the end-phase of this thesis.

I acknowledge that as a research student I have received the RTP Fee Offset.

Contents

Chapter 1 Introduction 15
- **Background – the puzzle of low-value care** 15
- **Economic rationality, supplier induced demand and low-value care** 15
- **Bounded rationality and satisficing** 18
- **Satisficing and its heuristic implications** 19
- **Nudges as heuristic jujitsu** 22
- **Satisficing as adaptation to existing resource settings** 23
- **Questions for investigation** 27
- **References** 28

Chapter 2 Use of heuristics in testing, screening and treatment decisions: a systematic review 34
- **Key points** 35
- **Abstract** 35
- **Introduction** 37
- **Methods** 37
  - **Search strategy and inclusion/exclusion criteria** 37
  - **Study selection** 38
- **Results** 41
  - **Study characteristics** 41
  - **Heuristics shaped by initial patient information** 47
  - **Heuristics shaped by previous medical events** 47
- **Discussion** 53
  - **Main findings and potential implications** 53
  - **Comparison with previous studies** 54
  - **Limitations** 54
- **Conclusions** 55
- **References** 55

Chapter 3 Effect of two behavioural ‘nudging’ interventions on choice of management options for low back pain: A randomised vignette-based study in general practitioners 57
- **Abstract** 58

Introduction 59
Methods 60
Study design and participants 60
Procedures and intervention 60
Randomisation and blinding 61
Primary and secondary outcome measures 61
Other measures and data collected 63
Sample size 63
Statistical plan 64
Results 64
Primary outcome measure 66
Secondary outcome measure 68
Other exploratory hypotheses 68
Discussion 69
Comparison to other research findings 69
Strengths, weaknesses and clinical implications 70
Conclusion 72
References 72
Chapter 4 CT and MRI substitution in presence of licencing constraints? 75
Abstract 76
Introduction and context 77
Methods 78
Data sources 78
Fitting the primary models from the data collected 79
Supplementary analysis 81
Results 81
Primary models 81
Supplementary model 82
Discussion 84
Primary model results – MRI machine density 84
Primary model results – medical practitioner density 85
Primary and supplementary model results on income 85

Conclusions 86
References 88
Chapter 5 The financial implications of investigating false-positive and true-positive mammograms in a national breast cancer screening program 90
Abstract 91
Introduction 92
Methods 93
Setting and study design 93
Population and subgroups 93
Estimating resources and costs 94
Sensitivity analysis 95
Ethics approval 95
Results 95
Discussion 98
Strengths and weaknesses 98
Implications for practice and policy 99
Conclusion 99
References 100
Chapter 6 Discussion and conclusions 102
Summary of findings, specific strengths and weaknesses and additional considerations for each study 102
The impact of heuristics in medical decision-making 102
Use of nudges to reduce low-value care 103
Satisficing in the selection of medical imaging modalities 104
Quantifying the costs of false positive and true positive investigations in a national breast screening program 106
What this research adds 107
Implications for practice and policy 108
Implications for further research 109
Conclusions 112
References 114
Supplement for Chapter 2 115

Database searches ... 116
Instructions for data extraction form ... 125
Methodological notes on sample size ... 127
Data quality ... 128
Supplementary Table 2-1: Summary of data quality ... 128
Converting results into odds ratios ... 129
Additional results omitted from figures and tables ... 130
Results from Simianu et al 2016 ... 130
Supplementary Table 2-2: Full results from Simianu 2016 recalculated as odds ratios ... 130
Results from Keating et al 2017 ... 131
Supplementary Table 2-3: Stratified results from Keating et al 2017 results recalculated as odds ratios ... 131
Supplement for Chapter 3 ... 132
Supplementary File 1: Clinical Vignettes ... 133
Supplementary file 2: Visual appearance of the display of menu options in each of the four experimental conditions ... 136
Supplementary File 3 ... 140
Supplementary Table 3-1: 2x2 factorial design of experiment ... 140
Supplementary File 4: More on the Cognitive Reflection Test ... 141
Supplementary File 5 ... 142
Supplementary Table 3-2: Proportion of vignettes where participants chose at least one low-value option, stratified by vignette type ... 142
Supplement for Chapter 4 ... 143
Methodological notes ... 144
List of MBS item descriptors included in the analysis and the rationale ... 144
Distinction between fully and partially MBS eligible MRI machines ... 146
Supplement for Chapter 5 ... 147
1. Costing of diagnostic open biopsies and its treatment as a cost to the public hospital system ... 148
2. Share of total breast-cancer screening which is private screening ... 148
3. Caveats to inclusions ... 149
4. Additional tables ... 150

Supplementary Table 5-1: Key national indicators from BreastScreen Australia data for 2016 150

Supplementary Table 5-2: Cost components of diagnostic workup sequences 151

Supplementary Table 5-3: Full list of diagnostic workup sequences of BSC service clients recalled to assessment in 2016 153

Supplementary Table 5-4: Recall to assessment rates, false positive rates and positive predictive value by Australian State/Territory 153

5. References 154

List of Tables (except Supplementary Tables which are listed in Table of Contents under subheadings)

Table 2-1 Summary of selected studies 43
Table 2-2 Summary of key outcomes from included studies 45
Table 3-1.Characteristics of participants 66
Table 3-2. Cognitive characteristics of participants selecting low-value options in default option nudge vs no default option nudge conditions 69
Table 4-1: Summary data 80
Table 4-2: Regression model results for 2019 data (n=87) for GP (general practitioner) and non-GP specialist requested imaging. 83
Table 4-3: Regression model results for 2018 data (n=174) for GP (general practitioner) requested CT imaging 84
Table 5-1:Summary statistics of women screened through BreastScreen service in 2016 95
Table 5-2:Diagnostic workup sequences of BreastScreen service clients recalled to assessment in 2016 96
Table 5-3: Total flow-on diagnostic costs in 2016 of representative BreastScreen facility (AUD2020) 96
Table 5-4: Total annual flow-on diagnostic costs for the Australian population from the national breast screening program including costs of breast screen-detected abnormalities in the twelve months following a client's mammogram (AUD2020) 97
Table 5-5: Results of sensitivity analysis 97

List of Figures (except Supplementary Figures which are listed in Table of Contents under subheadings)

Figure 2-1: PRISMA flow diagram of the study selection process 42
Figure 2-2: Odds ratios of medical decisions following exposure to initial patient triage information. 49
Figure 2-3: Odds ratios of obstetric decisions following exposures to previous medical events. 50
Figure 2-4: Odds ratios of testing and investigation decisions following exposures to previous medical events. 51
Figure 2-5: Odds ratios of prescribing decisions following exposures to previous medical events. 52
Figure 3-1: Visual appearance of the partition display nudge (i) and default option nudge (ii) on a computer screen. 62
Figure 3-2: Participant study flow 65
Figure 3-3. Proportion of scenarios where at least one low-value option was selected 67

Chapter 1 Introduction

Background – the puzzle of low-value care

Low-value care in medicine refers to tests, treatments, or procedures that offer little to no benefit to patients, may cause harm, and lead to unnecessary healthcare spending.[1] Research in this area has grown over the years and is often related to and overlaps with research into overuse of medical services and overdiagnosis[2] and has included investigations of the frequency of low value care in different healthcare settings and jurisdictions[3-6] the development of frameworks for measurement and monitoring of low-value care[7] and related research into the various factors contributing to the persistence of low-value care such as defensive medicine,[8] clinician inertia[9] and other healthcare provider characteristics[10] and financial incentives in fee-for-service systems.[11 12] Initiatives for identifying and reducing these practices such as the Choosing Wisely and related campaigns[13 14] have also grown, and this has led to further research into approaches for reducing or de-implementing low-value care.[15-19] Most recently some of this applied research has drawn on insights from other fields, notably economics and psychology.[20 21]

A major theme of this thesis is that the persistence of low-value care can be analysed through the framework of 'satisficing' which originates from behavioural economics. However, it is firstly important to understand and acknowledge the continuing applicability of some more traditional economic concepts to low-value care.

Economic rationality, supplier induced demand and low-value care

Rational choice theory, a cornerstone of traditional economics, assumes that individuals make optimal decisions by considering all available information[22] (this is sometimes known as 'maximising' behaviour). While this assumption may seem far-fetched, economists use it as a working hypothesis to model human behaviour and it has been argued that the theory's value lies not in the literal truth of its assumptions, but in its predictive power and ability to generate useful economic insights.

A classical and memorable restatement of this argument was by Milton Friedman 1966[23] who argued that we could analyse an expert billiards player's actions 'as if' they knew complex mathematical formulae for optimal ball trajectories and could accurately estimate angles by eye, perform instantaneous calculations, and execute shots accordingly even if this is not what they did in reality i.e. that it sufficed if this model yielded no worse predictions than models with more elaborate assumptions about how expert billiard players executed their shots. Similarly, he argued, we can analyse a firm 'as if' they had full knowledge of the cost and demand functions in their business, used this to calculate marginal costs and revenues for all possible actions, and therefore pursued their pricing and production strategies to the point where marginal cost equalled marginal revenue (as per standard economic models of the firm). Thus, he argued that it was sufficient if the resulting modelled behaviour of firms (and expert billiard players) was consistent with the assumptions used to model that behaviour. Extending this argument further, economists would argue in defence of the rational choice model (and the related assumption of maximising behaviour) that the pertinent question is not whether these assumptions are literally true, but to what extent incorporating additional complexities, such as individual cognitive imperfections, into the models yields insights that cannot be gained from existing ones.

It is important to note that rational choice theory does not need to assume perfect information, 'only' perfect rationality. It is feasible for a rational choice approach to encompass imperfect information and asymmetric bargaining but what is retained is the fundamental assumption of perfect rationality in the form of maximisation – for instance in the case of imperfect information, individuals are assumed to gather information until the perceived benefit outweighs the cost (this caveat will become more relevant when the supplier induced demand hypothesis is discussed below).

This framework has been applied to understand aspects of low-value care in healthcare, particularly through the concept of supplier-induced demand (SID). SID describes a situation where healthcare providers influence the demand for their own services due to their superior medical knowledge, patient trust, and the inherent information asymmetry in the provider-patient relationship. This still aligns with rational choice theory insofar as providers are assumed to be maximizing their utility (which may include professional prestige as well as financial gain). It also still aligns with rational choice theory even if under this narrative, patients are assumed to place their trust in their doctors. This is because 'rational ignorance' (a concept originally developed in the economic analysis of political systems[24]) can be an outcome of maximising behaviour by patients in conditions of imperfect information – in this case the argument is that the patient chooses not to acquire all possible information about a (medical) decision because the perceived cost of acquiring that information outweighs the expected benefit and hence delegates their decisions to their doctor.

The SID hypothesis helps explain why fee-for-service payment models are often seen as a notable contributor to low-value care. Under this payment model, providers are paid per visit, test, or procedure, potentially incentivizing the provision of additional services beyond those likely to benefit the patient who trusts in the provider to choose the right level of servicing. This connection between payment models and service provision has motivated research into the development of alternative payment models to mitigate these impacts and their resulting cost inefficiencies.[25,26]

One common methodology underlying investigations of the SID hypothesis is to examine the relationship between doctor-to-population ratios and service utilisation. The rationale is that a higher doctor-to-population ratio (i.e., doctor density) increases market oversaturation or competitiveness, prompting doctors to induce demand to maintain income levels. For instance, Richardson 1981[27] found that a 10% increase in the supply of GPs correlated with a 4.6-5.1% increase in services, while a 10% increase in specialists correlated with a 7.6-11.9% increase. Similarly, Maeda et al 2016[28] discovered that higher hospital physician density significantly correlated with the likelihood of gastrostomy procedures in Japan, while Alinia et al 2021[29] reported that a higher orthopaedic surgeon to population ratio was associated with increased knee replacement surgery rates in Iran. However, this approach has methodological weaknesses, including potential issues with causality – for instance high doctor density might lead to more service use simply because it is more convenient for patients, such as shorter distances to travel to the nearest doctor; and doctors may choose locations based on where they expect higher demand, influenced by factors like income levels, potentially overestimating the elasticity of demand inducement.[30] Patients from other regions may also travel to areas with high doctor density for services (the border crossing issue) so that unless the data strictly includes patients residing in the same region as the doctors, the results might overestimate the elasticity of demand inducement.[30]

Another approach to investigating SID involves examining the impact of changes in remuneration rates. Shigeoka and Fushimi 2014's[31] study on neonatal intensive care units (NICUs) found that when these units were excluded from the introduction of a prospective payment system (making them relatively more profitable than units performing other types of procedures) this led to their increased utilisation including by manipulation of infants' reported birth weights which determine their maximum allowable stay in the NICU. Papanicolas and McGuire 2015 found that the higher financial reimbursement attached to uncemented hip replacement in England compared to the cemented variety led to a faster uptake of the former even though this was contrary to clinical guidelines.[32]

A particularly striking example of SID comes from studies on caesarean section rates. Gruber and Owings 1996[33] found that even a decline in fertility rates (which leads to decreased income for obstetricians) was significantly positively associated with a corresponding increase in caesarean sections. Specifically, they estimated that a 10% drop in fertility was associated with an almost 1% rise in caesarean rates Similarly, studies in other countries have shown higher caesarean rates in private hospitals compared to public ones which can be attributed to the payment gap for this procedure between the two settings (in favour of the private system),[34 35] higher caesarean rates in hospitals which receive higher profits per procedure,[36] and higher caesarean rates among women with private health insurance compared to those in public systems[37] and/or uninsured women.[38] This demonstrates that even for significant surgical procedures with potential risks, economic considerations can play a substantial role in clinical decision-making. This demonstrates the power of the rational choice model in explaining and financial incentives in shaping medical practice, even in contexts where one might expect clinical factors to be the primary determinants of care.

Another area of health services research which has provided evidence of supplier-induced demand are studies of the practice of physicians referring patients to medical facilities in which they have a financial interest (also known as self-referral). These studies have shown that when physicians begin to self-refer (e.g., after acquiring ownership in imaging facilities or leasing such facilities), the rate of imaging studies for their patients increases significantly[39-41] or exceeds the general radiologist rate.[42] For example, Baker 2010[39] studied changes in MRI utilization and overall spending by orthopaedists and neurologists who began billing for MRI scans after acquiring or leasing their own MRI equipment and found a substantial increase in MRI orders once they commenced such billing, (for example orthopaedists had a 38% increase in MRI procedures within thirty days of the first visit after initiating billing). Similarly, Shreibati and Baker 2011[41] found that orthopaedists and primary care physicians who begin self-referral of MRI increased their use of MRI for patients with low back pain. One meta-analysis which included 5 studies analysing almost 77 million total episodes of care estimated a relative frequency of imaging of 2.48 (95% confidence interval, 1.90-3.24) for self-referrers compared with non–self-referrers.[43]

Bounded rationality and satisficing

While the rational choice framework provides valuable insights into how clinicians engaged in economic maximizing behaviour can end up providing low-value care through supplier-induced demand (SID), this framework alone is insufficient for analysing decision-making in healthcare, particularly because not all low-value care can be explained by economic or financial incentives. In this section it is argued that the concept of "satisficing" may add nuance to our understanding of clinical decision-making and how it may lead to low-value care.

Herbert Simon introduced the term "satisficing" in 1956 to describe a decision-making strategy where individuals choose options or seek solutions that are "good enough".[44] This contrasts with the 'maximizing' behaviour of rational choice theory discussed previously, which assumes individuals try to select the globally optimal (as in best possible) option when faced with a decision between alternatives. Satisficing is rooted in the concept of bounded rationality, which recognizes that when individuals make decisions, their rationality is limited by factors such as the complexity of the decision problem, cognitive limitations, and time constraints.[45]

It is important to note that satisficing differs from the concept of 'rational ignorance' discussed previously in the context of SID, which is still part of rational choice theory. While rational ignorance involves weighing up the costs and benefits of gathering additional information before deciding to remain uninformed, satisficing involves decision-makers searching through alternatives sequentially until they find one that meets a predefined minimum criteria or threshold, especially in environments with limited time and information.[45] In other words, rather than continuing an indefinite search through the ‘decision tree’ a predefined stopping process is built into decision-making. In the economic sphere, this may explain behaviours such as brand loyalty and habitual purchasing, as consumers choose products that are "good enough" rather than spending excessive time searching for the perfect option.

The satisficing framework provides an alternative understanding of why low-value care persists, even in systems where financial incentives for overservicing are minimal. While the supplier-induced demand model suggests that low-value care is driven by economic motives, it oversimplifies the issue. Most healthcare providers are motivated by a genuine desire to help patients but operate under time and resource constraints. Bounded rationality and satisficing help explain why doctors may order unnecessary tests or stick to familiar treatment protocols, not for financial gain, but because these decisions are "good enough" under the circumstances. These behaviours, while not optimal in a strict sense, reflect boundedly rational responses to the limitations providers face.

It is also important to recognize that a clinical decision-maker may exhibit different behaviours in different contexts. They may be a 'maximizer' when acting as an economic decision-maker (maximizing utility, which includes income and non-pecuniary aspects like professional reputation or quality of life) while being a 'satisficer' when acting as a clinical decision-maker (focusing on the clinical effectiveness of patient management), or vice versa. This nuance allows for the possibility that some low-value care may result from a combination of economic incentives leading to economically rational maximization behaviour and boundedly rational 'satisficing' behaviour in clinical decision-making.

The following sections will explore how satisficing in two different contexts - high-pressure, low or conflicting information environments and resource-constrained settings - may explain some elements of low-value care. These potential explanations for low-value care are not mutually exclusive and can provide a more comprehensive understanding of the complex factors influencing healthcare decision-making.

Satisficing and its heuristic implications

In high-pressure clinical environments with limited or conflicting information, where physicians must make quick decisions, the concept of "satisficing" becomes especially relevant. For instance, according to one estimate an average clinical encounter is approximately 11 minutes long with less than 2 minutes available to search for reliable information.[46] In such settings, decision-makers often rely on heuristics—mental shortcuts or rules of thumb that simplify decision-making processes. While these heuristics can be practical under pressure, they also come with inherent biases that can lead to low-value care.

There is a duality to the nature of these heuristics, as captured in the debate between Gigerenzer and Goldstein 1996 [47] versus Kahneman and Tversky 1974.[48] Gigerenzer and Goldstein 1996 argue that "fast and frugal" heuristics may lead to more accurate decisions with less effort and time.[47] In contrast, Kahneman and Tversky 1974 proposed a dual-system framework where 'system 1 thinking' is automatic and intuitive, while 'system 2 thinking' is deliberate and effortful.[48] They suggest that poor decision-making may result from overusing system 1 thinking, and techniques enhancing system 2 thinking can improve decision-making. In reality, both perspectives may be valid, and the existing rates of low-value care in the medical system may reflect the trade-offs embodied by these different approaches. Note that sometimes the term ‘heuristics’ is used interchangeably with ‘cognitive biases’ or ‘heuristic biases’ but here we will generally use the neutral term ‘heuristics’ throughout unless we refer to articles which use both terms or the former.

Two common heuristics that illustrate this concept are availability bias and representativeness heuristic. Availability bias leads people to overestimate the likelihood of events based on their ability to recall examples. For instance, physicians in an emergency ward recently exposed to many COVID-19 cases may be more inclined to misdiagnose subsequent patients with respiratory conditions as having COVID-19. The representativeness heuristic leads individuals to judge the likelihood of an event based on how similar it is to a stereotype, which can result in misdiagnosis by aligning symptoms too closely with familiar patterns. In both examples, these heuristics may make sense in terms of increasing performance in terms of an aggregate measure of number of cases cleared per hour but may have negative impacts on the health of the misdiagnosed patients.

It is again important to emphasise that this model of decision-making differs significantly from a rational choice framework, even one modified to account for imperfect information. In a rational choice framework, the decision-maker would systematically evaluate various diagnostic options, considering probability-weighted averages before settling on a specific diagnosis. The heuristic form of thinking driven by 'satisficing' is fundamentally different. **Text Box 1** illustrates this using the example of a physician seeing a patient presenting with chest pain, shortness of breath, and fatigue.

There has been growing evidence of the use of heuristics in healthcare settings and accompanying this, there have been recent reviews of this literature focused on documenting the variety of heuristics identified in these studies, although the scope of these reviews has also extended beyond heuristics as strictly defined here. Examples include Blumenthal-Barby and Krieger 2015,[49] Saposnik et al 2016,[50] Whelahan et al 2020,[51] Thirsk et al 2022[52] and Featherston et al 2020.[53]

Saposnik et al 2016[50] conducted a primary narrative systematic review of cognitive biases and personality traits associated with medical decision-making and included 20 studies comprising 6810 physicians which identified 19 different cognitive biases.

They noted that most studies (60%) looked at cognitive biases in a diagnostic setting followed by treatment or management (35%) and prognosis (10%). Among the common heuristics identified in Saposnik et al 2016's review were the anchoring and availability heuristic. Their review also identified overconfidence and lower risk tolerance as having an impact on medical decision-making although these are best characterised as personality traits rather than heuristics. The recent quasi-systematic review by Whelahan et al 2020[51] identified the availability, anchoring and confirmatory heuristics as the most common heuristics used in medicine.

Recent reviews of the literature on cognitive biases and heuristics in other (non-medical practitioner) healthcare settings are Thirsk et al 2022[52] which looked at nursing and Featherston et al 2020[53] which looked at allied health.

A key feature of the evidence accumulated so far is that a significant majority of studies relied on vignettes, surveys, or recall methods. For instance, 77% of the studies identified by Blumenthal-Barby and Krieger 2015[49] were based on vignettes, as were 70% of the studies identified by Saposnik et al 2016.[50] Featherstone et al 2020's[53] scoping review of cognitive biases in allied health concluded that there was a need for more studies reflecting real clinical settings.

Text Box 1: Maximizing vs. Satisficing in Medical Decision-Making: An Illustration Using Availability Bias

Scenario: A 55-year-old patient presents with chest pain, shortness of breath, and fatigue.

Maximizing Approach (Rational Choice Framework):

The doctor systematically considers all possible causes of these symptoms. Probabilities are mentally assigned to each potential diagnosis based on prevalence data and patient characteristics. The doctor (mentally) creates a decision tree, weighing the costs and benefits of each diagnostic test. Probability-weighted outcomes for each branch of the decision tree. The doctor chooses the diagnostic path that maximizes expected utility (balancing accuracy, cost, and patient outcomes).

Satisficing Approach (Heuristic-Driven):

The doctor immediately recalls recent similar cases or highly publicized medical events (availability bias). If they recently treated a patient with similar symptoms who had a pulmonary embolism, this diagnosis becomes highly salient. The doctor quickly checks for risk factors associated with pulmonary embolism. If the patient has some matching risk factors, the doctor may settle on this diagnosis without extensively considering alternatives.

The doctor orders a test to confirm pulmonary embolism (e.g., D-dimer test or CT pulmonary angiogram). If the test is positive, the doctor proceeds with treatment. If negative, they move to the next most available diagnosis in their mind.

Nudges as heuristic jujitsu

Given the challenges posed by satisficing and the associated use of heuristics in high-risk, low-information clinical environments, the concept of nudging offers a promising strategy to counteract these biases and enhance decision-making. Nudges, first popularized by Thaler and Sunstein 2008,[54] are ways of configuring choices to encourage certain courses of action without taking away the decision-maker's freedom of choice.[55]

Nudges have been shown to be useful in various applications, including shaping patient or consumer behaviour in healthcare settings.[56 57] In the context of medical decision-making, nudges can be used to leverage existing forms of heuristics in a more directed manner, resulting in greater compliance with evidence-based guidelines. For example, Bordeaux et al 2014[58] evaluated the impact of changes to the design of an order set on the delivery of chlorhexidine mouthwash and hydroxyethyl starch (HES) to patients in an intensive care unit. In the study, chlorhexidine mouthwash was added as a default prescription to the prescribing template while HES was removed though both interventions were still available to prescribe manually. Nonetheless the study detected a statistically significant impact in raising prescription of the former and reducing prescription of the latter following the intervention.[58] Other studies which have demonstrated the efficacy of default nudges similar are Mehta et al 2022[59] which investigated the impacts of embedding of Hepatitis C virus screening as a default order in the electronic health record, and Rubins et al 2019 which investigated the impacts of removing preselection of common telemetry orders in the electronic health record.[60]

The efficacy of default nudges can be understood as a mirror image of decision-makers' heuristic-driven thought processes, as decision-makers' thought processes may have a particular inclination towards the default heuristic also known as status quo bias. This cognitive shortcut leads people to often maintain their current or previous decision rather than switch to a new option, even when the alternative might be objectively better. Default nudges exploit this tendency by setting a specific option as the preset choice, thereby increasing the likelihood of its selection. In essence, the nudge aligns with the natural inclination to stick with the status quo, making the desired option the path of least resistance. This synergy between the nudge design and the underlying heuristic explains why such interventions can be remarkably effective in healthcare settings, where decision-makers often face complex choices under time pressure.

Salience-based nudges exploit the tendency to favour recently highlighted or prominently displayed information in decision-making and have also been shown to be effective in healthcare settings where these nudges often take the form of alerts, reminders, or strategically placed information. For instance, Holt et al 2010 [61] demonstrated the effectiveness of screen messages highlighting individuals at raised cardiovascular risk and prompting users to complete risk profiles where necessary.

This intervention increased the proportion of patients with sufficient data who were identifiably at risk. Another example is presenting the cost of medical tests next to their names in ordering systems. Lewis et al 2019 [62] found that displaying individual and annual institutional costs of blood assays in the electronic results system led to a statistically significant reduction in demand for full blood count (FBC) and urea and electrolytes (U&E) tests. Similarly Forgarty et al 2013 found that adding the cost of a commonly used blood test to the reports that clinicians received led to a 32% decrease in demand over 12 months in use of the test,[63] while Gill et al 2020 found that use of a surgeon-targeted surgical receipt cost feedback system reduced supply cost per case in sinonasal surgery.[64] Strikingly these interventions are effective even though the decision makers would not personally incur the cost of ordering these tests or personally financially benefit from any cost savings from reductions in testing (a result that cannot be explained in traditional economic maximising terms). These examples illustrate how making certain information more salient can significantly influence healthcare decisions and resource utilization.

Just as default nudges can be viewed as a mirror image of the default heuristic, salience-based nudges can be viewed as a mirror image of the availability heuristic, a cognitive shortcut where people rely on immediate examples that come to mind when evaluating a specific topic. Just as the availability bias leads individuals to overestimate the likelihood or importance of events that are readily recalled, salience-based nudges leverage this tendency by deliberately making certain information more prominent or memorable. For example, when a clinician sees the cost of a test displayed prominently, it becomes more "available" in their mind, potentially influencing their decision-making process. Similarly, alerts about patients at high cardiovascular risk make this information more cognitively accessible, mirroring how the availability heuristic prioritizes easily remembered information. By strategically enhancing the salience of specific data or options, these nudges essentially create an "artificial availability," directing attention and influencing clinical decision-making.

Thus, the efficacy of nudges in healthcare are effectively a demonstration of heuristic ‘jujitsu’ where existing tendencies in clinical decision-makers’ minds to adopt heuristics can be readily leveraged to help them navigate the complexities of clinical decision-making more effectively, fostering practices that align better with evidence-based guidelines and ultimately improving patient outcomes. This can help transform the inherent limitations of satisficing into opportunities for enhanced healthcare quality.

Satisficing as adaptation to existing resource settings

While the use of heuristics can be seen as a form of satisficing that occurs within a clinician's mind as a cognitive adaptation to complexity and uncertainty, clinicians may also adopt satisficing as an adaptation to external constraints, particularly existing resource settings. For instance, the clinician may choose to use equipment and services that are ‘good enough’ in order to adapt to any external constraints while still providing the best possible care to her patient. Such satisficing in healthcare is analogous to consumer behaviour in economics, where "good enough" choices are made due to resource constraints. The link between the two concepts of heuristics and this scenario is that heuristics are about using rules of thumb that are ‘good enough’ where there are constraints on ‘time to think’ just as under resource constraints or pressures, the clinician may choose to use equipment/services that are ‘good enough’.

Several examples illustrate satisficing as an adaptation to resource constraints in healthcare. When faced with a shortage of MRI machines, a doctor might opt for a CT scan instead, even if an MRI would provide more accurate information. Similarly, medication formularies may limit drug choices due to cost or vendor agreements, forcing clinicians to prescribe less effective or more problematic alternatives. In rural or underserved areas with limited access to allied health care and specialists, clinicians might choose immediate, short-term interventions rather than recommending extended or multidisciplinary care, potentially leading to suboptimal long-term outcomes. Additionally, when coverage limitations dictate the types of treatments or tests available to patients, clinicians might recommend options that are financially feasible for their patients rather than those that are most appropriate or evidence-based, potentially resulting in unnecessary procedures or less effective treatments.

To my knowledge, there has been no literature explicitly studying these cases through the framework of satisficing. However, some instances of supplier-induced demand may also be partially explained by satisficing in the presence of existing resource settings. For example, as previously discussed, studies showing increased imaging rates when physicians begin to self-refer for diagnostic services (e.g., after acquiring ownership in imaging facilities) have typically been interpreted as evidence of supplier-induced demand and the importance of economic incentives in healthcare. However, satisficing theory may provide an alternative explanation for these results.

When physicians have easy access to imaging facilities they own, it may become a default choice to undertake imaging due to convenience and familiarity. An in-house imaging facility provides a readily available option for precautionary investigations. Moreover, the integrated nature of physician-owned practices may foster a sense of quality control and trust, with self-referring physicians believing their own facilities provide superior service or faster results, further reinforcing the tendency to self-refer. For instance, Baker 2010's[39] study on self-referral acknowledged that increased convenience from equipment acquisition might explain the documented increases in imaging referrals following equipment acquisition. However, Baker argued it was not a significant factor in his results because most of the observed increase in MRI use did not occur on the day of the initial patient visit.

Given the potential explanatory power of satisficing in self-referral practices, it may be useful to conduct further investigations that can effectively distinguish between economic motivations and satisficing behaviours associated with these convenience factors even in scenarios which have typically been analysed through a traditional economics lens. Such research could aim to develop methodologies that can isolate the effects of convenience, familiarity, and resource availability from purely financial incentives.

The documented relationship between the 'density' of service providers or medical equipment and the utilisation of associated medical services may also be partially explained by a convenience effect, consistent with the satisficing framework even though these have usually been explained as instances of SID. These include findings such as the positive relationship between increased hospital bed availability and increased hospital bed days[65-67] (sometimes referred to as Roemer's law). Similarly, Baras and Baker 2009[68] who find a positive relationship between availability of MRI units and utilisation of MRI for nonspecific low back pain hypothesise that this may be due to a convenience effect whereby easily available MRI may 'lead physicians to adapt their practice patterns to more routinely incorporate MRI for a wider range of patients'. They speculate that a similar convenience factor may be behind the previously discussed studies of Roemer's law on the relationship between hospital bed availability and hospital bed days. It is plausible that convenience effects may be as much of a driver of this relationship as SID if not more given that these studies have typically been of nonprofit hospitals which are likely to have reduced incentive for SID.

This convenience factor is consistent with the satisficing framework proposed here and could potentially overlap with default or availability heuristics. It is plausible that an established increase in medical facilities such as MRI units or hospital beds in a region may establish itself as a 'default' treatment or management option or become more 'cognitively available' to medical practitioners in that region. This option then becomes the natural 'go to' as decision-makers adapt to these new external settings.

Underlying this theory is the notion that satisficing might manifest as a tendency to make decisions based on immediate resource availability rather than critically assessing whether additional capacity aligns with optimal patient care and overall healthcare efficiency. This perspective offers a nuanced understanding of healthcare decision-making that goes beyond simple economic incentives, highlighting the complex interplay between resource availability, convenience, and clinical judgment.

While this discussion began with the premise that a satisficing-based perspective might constitute an alternative means of explaining the persistence of low-value care, it may be debatable whether some of the examples discussed in this section constitute genuine low-value care, which was previously defined as 'tests, treatments, or procedures that offer little to no benefit to patients, may cause harm, and lead to unnecessary healthcare spending.' For instance, it could be argued that ordering tests or treatments due to a ‘convenience factor’ is a matter of common-sense adaptation to resource constraints and opportunities rather than genuine ‘low-value care’.

One response to this is that the health services research literature on low value care has itself evolved beyond consideration of the more obvious examples of ineffective or obsolete healthcare practices to a more ‘grey zone’ of potential areas of overuse including deviations from evidence-based care. Moreover, the definition of low-value care previously cited does extend to ‘tests and treatments’ that lead to ‘unnecessary healthcare spending’. Therefore these examples of deviations from ‘first best’ or evidence-based care due to patently non-clinical factors can be considered to overlap with low-value care insofar as an alternative configuration of healthcare resources could have led to more evidence-based care and less spending on particular tests or treatments and insofar as such ‘convenience-based factors’ can devolve into the persistence of low-value care in parts of the healthcare sector. This is because low-value care operates on a spectrum, and even seemingly pragmatic adaptations may have downstream effects, such as reinforcing inefficiencies in the healthcare sector. Thus, while satisficing as a heuristic response to resource constraints may appear to reflect common-sense adaptation, it may become part of a general tendency to normalise deviations from evidence-based care.

Many roads to low-value care

The relationship between satisficing, economic maximisation, and low-value care presents a complex irony in healthcare. This irony emerges from the dual role of the clinician as both an economic decision-maker (maximizing utility, including income and non-pecuniary aspects) and a clinical decision-maker (maximizing the effectiveness of clinical services). Each clinician may be a mix of maximiser/satisficer on either of these dimensions. The irony is that on the one hand, satisficing by healthcare providers through use of heuristics in a high pressure/low or conflicting information environment and through adaptation to external settings can lead to low value care (broadly defined to include overuse of second-best clinical practice) even when the provider is attempting to make rational decisions in patients' best interests; but so can maximisation albeit in the economic sphere whereby healthcare providers, acting as profit-maximising agents, may recommend or provide services that are not strictly necessary or of high value to the patient.

This highlights how low-value care can emerge not just from individual clinical decisions, but also from broader systemic issues in healthcare resource allocation and availability, and from existing economic and financial incentives.

Thus, there are many roads to low-value care or more generally to inefficiencies within the healthcare system. Consider the question of false positive rates and their associated costs in screening programmes and how and why these rates may differ in different countries, noting that although false positive tests are inevitable in all healthcare systems the different magnitudes of false positive rates and false positive costs in different countries may be reflective of their respective systemic propensities to low value care.

Studies have proposed several reasons for the phenomenon of the higher false positive rates in US breast cancer screening compared to other countries.[69-71] One factor is the lower frequency of double reading mammograms in the US.[70] There is also less centralization of mammogram reading in the US. This is partly due to a more mobile US population, making prior mammograms less available for comparison when interpreting results.[70] Another factor is the focus of US medical malpractice lawsuits on missed cancer diagnoses.[70 72] Additionally, a higher percentage of younger women are screened in the US compared to European national screening programs. European programs often restrict screening examinations to older age groups.[71] This is significant because younger women have denser breast tissue, resulting in lower screening sensitivity. In addition, it has been noted that mammographic screening in European countries is predominantly undertaken by academic and public screening programs which do not benefit financially from additional diagnostic procedures, compared to screening in private settings which are more common in the US (and which may therefore be influenced by SID-type considerations).[71]

Thus, a whole confluence of internal and external factors including clinical satisficing mixed with economic incentives will leave an imprint in shaping levels of low-value care and therefore in shaping the rates of false positive breast screens in the US versus Europe and other countries.

Questions for investigation

This thesis aims to investigate several crucial questions related to healthcare decision-making and resource allocation, each addressing a unique aspect of the complex interplay between clinical judgment, systemic constraints, and patient outcomes as captured under a satisficing framework.

In **Chapter 2**, we will explore the influence of heuristics on clinical decision-making in real-world settings. Our primary questions are: How do heuristics impact clinical decisions in actual practice, beyond hypothetical scenarios? What is the statistical significance and magnitude of these impacts on testing and prescribing rates? Do different types of heuristics vary in their effect sizes, timing, and persistence? How do heuristics influence diagnostic accuracy and patient outcomes? Is the impact of heuristics on diagnostic accuracy and patient outcomes uniform, or does it vary case by case? By addressing these questions, we aim to contribute to a deeper understanding of how heuristics, as a byproduct of satisficing in high-pressure medical environments, may drive low-value care and influence health outcomes.

Chapter 3 will investigate the efficacy of nudges in clinical decision-making, particularly in aligning practices with evidence-based guidelines. Our key questions include: How effective are different types of nudges, such as default nudges and partition display nudges, in promoting guideline-compliant decision-making? How do these nudges compare when tested simultaneously in a factorial design experiment? Specifically, how do they impact decision-making in low back pain management, a common area of medical consultation? Do these nudges have long-term educative effects beyond the point of decision-making? Through these inquiries, we seek to contribute to a more nuanced understanding of the scope and limitations of nudges as a means of leveraging heuristics in clinical settings.

In **Chapter 4**, we will examine how external constraints in the Australian healthcare system shape medical practitioners' decision-making processes. Our primary research question is: Does limited availability of Medicare Benefits Schedule (MBS) eligible MRI services in a region lead to increased CT imaging orders as a substitute, indicating satisficing behaviour among medical practitioners? Additionally, we will investigate: Do these relationships differ between general practitioners (GPs) and non-GP specialists? What other factors, such as income and availability of different types of medical service providers, influence the utilisation levels of various imaging modalities? By exploring these questions, we aim to provide novel insights into healthcare decision-making under regulatory constraints in the Australian context, contributing to a better understanding of how systemic factors influence clinical choices and resource allocation**.**

Chapter 5 will address an important question arising from the observation that a confluence of internal and external factors shape the rates of low-value care in medical programs, including screening programs. Specifically, we ask: What are the costs of false positive screening rates in different countries, including Australia? How do these costs compare internationally, and what do these comparisons reveal about the influence of country-specific factors on false positive rates of testing? This chapter will develop cost estimates specifically for Australia's breast screening program and a novel methodology for these estimates, contributing to a better understanding of how systemic factors impact screening outcomes and associated cost.

By addressing these four elements, this thesis aims to investigate some of the complex interplay between heuristics, satisficing, and economic factors in healthcare decision-making and resource allocation. Although there is much more that could be the subject of investigation based on the conceptual frameworks discussed here, it is hoped that the findings from these studies will contribute to and suggest a roadmap for further research into the question of how cognitive processes, systemic constraints, and economic incentives shape healthcare delivery and outcomes.

References

1. Scott IA, Duckett SJ. In search of professional consensus in defining and reducing low-value care. *Med J Aust*. 2015;203(4):179-81.

2. Brownlee S, Chalkidou K, Doust J, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390(10090):156-168.
3. Lan T, Chen L, Hu Y, et al. Measuring low-value care in hospital discharge records: evidence from China. *Lancet Reg Health West Pac*. 2023;38:100887.
4. Badgery-Parker T, Feng Y, Pearson SA, et al. Exploring variation in low-value care: a multilevel modelling study. *BMC Health Serv Res*. 2019;19(1):345.
5. Mafi JN, Reid RO, Baseman LH, et al. Trends in Low-Value Health Service Use and Spending in the US Medicare Fee-for-Service Program, 2014-2018. *JAMA Netw Open*. 202;4(2):e2037328.
6. Ellison JE, Kumar S, Steingrimsson JA, et al. Comparison of Low-Value Care Among Commercial and Medicaid Enrollees. *J Gen Intern Med*. 2023;38(4):954-960.
7. Miller G, Rhyan C, Beaudin-Seiler B, et al. A Framework for Measuring Low-Value Care. *Value Health*. 2018;21(4):375-379.
8. Ries NM, Johnston B, Jansen J. A qualitative interview study of Australian physicians on defensive practice and low value care: "it's easier to talk about our fear of lawyers than to talk about our fear of looking bad in front of each other". *BMC Med Ethics*. 2022;23(1):16.
9. Rowe TA, Brown T, Doctor JN, et al. Examining primary care physician rationale for not following geriatric choosing wisely recommendations. *BMC Fam Pract*. 2021;22(1):95.
10. Gerend MA, Bradbury R, Harman JS, et al. Characteristics Associated with Low-Value Cancer Screening Among Office-Based Physician Visits by Older Adults in the USA. *J Gen Intern Med*. 2022;37(10):2475-2481.
11. Park S, Jung J, Burke RE, et al. Trends in Use of Low-Value Care in Traditional Fee-for-Service Medicare and Medicare Advantage. *JAMA Netw Open*. 2021;4(3):e211762.
12. Kini V, Viragh T, Magid D, et al. Trends in High- and Low-Value Cardiovascular Diagnostic Testing in Fee-for-Service Medicare, 2000-2016. *JAMA Netw Open*. 2019;2(10):e1913070.
13. https://www.choosingwisely.org/ [cited 25 September 2024].
14. Soon J, Buchbinder R, Close J, et al. Identifying low-value care: the Royal Australasian College of Physicians' EVOLVE initiative. *Med J Aust*. 2016;204(5):180-1.
15. Kini V, Breathett K, Groeneveld PW, et al. Strategies to Reduce Low-Value Cardiovascular Care: A Scientific Statement From the American Heart Association. *Circ Cardiovasc Qual Outcomes*. 2022;15(3):e000105.
16. Dunsmore J, Duncan E, MacLennan S, et al. Effectiveness of de-implementation strategies for low-value prescribing in secondary care: a systematic review. *Implement Sci Commun*. 2023;4(1):115.
17. Elshaug AG, Rosenthal MB, Lavis JN, et al. Levers for addressing medical underuse and overuse: achieving high-value health care. *Lancet*. 2017;390(10090):191-202.

18. Ingvarsson S, Hasson H, von Thiele Schwarz U, et al. Strategies for de-implementation of low-value care-a scoping review. *Implement Sci*. 2022;17(1):73.
19. Alishahi Tabriz A, Turner K, Clary A, et al. De-implementing low-value care in cancer care delivery: a systematic review. *Implement Sci*. 2022;17(1):24.
20. Brown T, Rowe TA, Lee JY, et al. Design of Behavioral Economic Applications to Geriatrics Leveraging Electronic Health Records (BEAGLE): A pragmatic cluster randomized controlled trial. *Contemp Clin Trials.* 2022;112:106649.
21. Belli HM, Troxel AB, Blecker SB, et al. A Behavioral Economics-Electronic Health Record Module to Promote Appropriate Diabetes Management in Older Adults: Protocol for a Pragmatic Cluster Randomized Controlled Trial. *JMIR Res Protoc*. 2021;10(10):e28723.
22. Blume LE, Easley, D. Rationality. In: Durlauf N, Blume LE, eds. *The New Palgrave Dictionary of Economics*. Palgrave Macmillan; 2008.
23. Friedman, M. The methodology of positive economics. In: Friedman, M. *Essays In Positive Economics*. Univ. of Chicago Press; 1966: 3-43.
24. Downs, A. *An Economic Theory of Democracy*, Harper, NY; 1957.
25. Tobey R, Maxwell J, Turer E, et al. Health Centers and Value-Based Payment: A Framework for Health Center Payment Reform and Early Experiences in Medicaid Value-Based Payment in Seven States. *Milbank Q*. 2022;100(3):879-917.
26. Cattel D, Eijkenaar F. Value-Based Provider Payment Initiatives Combining Global Payments With Explicit Quality Incentives: A Systematic Review. *Med Care Res Rev*. 2020;77(6):511-537.
27. Richardson, J. The inducement hypothesis: that doctors generate demand for their own services. In: van der Gaag J, Perlman M, eds. *Health, Economics and Health Economics*. North-Holland Publishing Co.; 1981:189–214.
28. Maeda T, Babazono A, Nishi T, et al. Investigation of the Existence of Supplier-Induced Demand in use of Gastrostomy Among Older Adults: A Retrospective Cohort Study. *Medicine* (Baltimore). 2016; 95(5):e2519.
29. Alinia C, Takian A, Saravi N, et al. Physician induced demand for knee replacement surgery in Iran. *BMC Health Serv Res*. 2021;21(1):763.
30. Bickerdyke, I., Dolamore, R., Monday, I. et al. *Supplier-Induced Demand for Medical Services*, Productivity Commission Staff Working Paper, Canberra; 2002.
31. Shigeoka H, Fushimi K. Supplier-induced demand for newborn treatment: evidence from Japan. J *Health Econ*. 2014;35:162-78.
32. Papanicolas I, McGuire A. Do financial incentives trump clinical guidance? Hip Replacement in England and Scotland. *J Health Econ*. 2015;44:25-36.
33. Gruber J, Owings M. Physician financial incentives and Caesarean section delivery*, RAND Journal of Economics*. 1996; 1: 99–123.

34. Bhatia M, Dwivedi LK, Banerjee K, et al. An epidemic of avoidable caesarean deliveries in the private sector in India: Is physician-induced demand at play? *Soc Sci Med*. 2020;265:113511.
35. Bhatia M, Banerjee K, Dixit P, et al. Assessment of Variation in Cesarean Delivery Rates Between Public and Private Health Facilities in India From 2005 to 2016. *JAMA Netw Open*. 2020;3(8):e2015022.
36. Sakai-Bizmark R, Ross MG, Estevez D, et al. Evaluation of Hospital Cesarean Delivery-Related Profits and Rates in the United States. *JAMA Netw Open*. 2021;4(3):e212235.
37. Hoxha I, Syrogiannouli L, Braha M, et al. Caesarean sections and private insurance: systematic review and meta-analysis. *BMJ Open*. 2017;7(8):e016600.
38. Hoxha I, Braha M, Syrogiannouli L, et al. Caesarean section in uninsured women in the USA: systematic review and meta-analysis. *BMJ Open*. 2019;9(3):e025356.
39. Baker LC. Acquisition of MRI equipment by doctors drives up imaging use and spending. *Health Aff* (Millwood). 2010;29(12):2252-9.
40. Levin DC, Rao VM. The effect of self-referral on utilization of advanced diagnostic imaging. *AJR Am J Roentgenol.* 2011;196(4):848-52.
41. Shreibati JB, Baker LC. The relationship between low back magnetic resonance imaging, surgery, and spending: impact of physician self-referral status. *Health Serv Res*. 2011;46(5):1362-81.
42. Babiarz LS, Yousem DM, Parker L, et al. Utilization rates of neuroradiology across neuroscience specialties in the private office setting: who owns or leases the scanners on which studies are performed? *AJNR Am J Neuroradiol*. 2012;33(1):43-8.
43. Kilani RK, Paxton BE, Stinnett SS, et al. Self-referral in medical imaging: a meta-analysis of the literature. *J Am Coll Radiol*. 2011;8(7):469-76.
44. Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*. 1956; 63(2): 129-138.
45. Djulbegovic B, Elqayam S, Dale W. Rational decision making in medicine: Implications for overuse and underuse. *J Eval Clin Pract*. 2018;24(3):655-665.
46. Djulbegovic B, Beckstead J, Nash DB. Human judgment and health care policy. *Popul Health Manag.* 2014;17(3):139-140.
47. Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. *Psychol Rev.* 1996;103(4):650-669.
48. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974;185(4157):1124-1131.
49. Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making*. 2015;35(4):539-57.

50. Saposnik G, Redelmeier D, Ruff CC, et al. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak*. 2016;16(1):138.
51. Whelehan DF, Conlon KC, Ridgway PF. Medicine and heuristics: cognitive biases and medical decision-making. *Ir J Med Sci*. 2020;189(4):1477-1484.
52. Thirsk LM, Panchuk JT, Stahlke S, et al. Cognitive and implicit biases in nurses' judgment and decision-making: A scoping review. *Int J Nurs Stud*. 2022;133:104284.
53. Featherston R, Downie LE, Vogel AP, et al. Decision making biases in the allied health professions: A systematic scoping review. *PLoS One*. 2020;15(10):e0240716.
54. Thaler, RH, Sunstein, CR. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press; 2008.
55. Wolf A, Sant'Anna A, Vilhelmsson A. Using nudges to promote clinical decision making of healthcare professionals: A scoping review. *Prev Med*. 2022;164:107320.
56. Milkman KL, Patel MS, Gandhi L, et al. A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment. *Proc Natl Acad Sci* U S A. 2021;118(20):e2101165118.
57. Hansen PG, Larsen EG, Modin A, et al. Nudging hand hygiene compliance: a large-scale field experiment on hospital visitors. *J Hosp Infect*. 2021;118:63-69.
58. Bourdeaux CP, Davies KJ, Thomas MJ, et al. Using 'nudge' principles for order set design: a before and after evaluation of an electronic prescribing template in critical care. *BMJ Qual Saf*. 2014;23(5):382-8.
59. Mehta SJ, Torgersen J, Small DS, et al. Effect of a Default Order vs an Alert in the Electronic Health Record on Hepatitis C Virus Screening Among Hospitalized Patients: A Stepped-Wedge Randomized Clinical Trial. *JAMA Netw Open*. 2022;5(3):e222427.
60. Rubins D, Boxer R, Landman A, et al. Effect of default order set settings on telemetry ordering. *J Am Med Inform Assoc.* 2019;26(12):1488-1492.
61. Holt TA, Thorogood M, Griffiths F, et al. Automated electronic reminders to facilitate primary cardiovascular disease prevention: randomised controlled trial. *Br J Gen Pract*. 2010;60(573):e137-43.
62. Lewis S, Young B, Thurley P, et al. Does cost feedback modify demand for common blood tests in secondary care? A prospective controlled intervention study. *Future Healthc J.* 2019;6(3):204-208.
63. Fogarty AW, Sturrock N, Premji K, et al. Hospital clinicians' responsiveness to assay cost feedback: a prospective blinded controlled intervention study. *JAMA Intern Med*. 2013;173(17):1654-5.
64. Gill AS, Reddy RK, Kulinich AG, et al. Surgeon cost feedback through a surgical receipt program reduces cost in sinonasal surgery. *Int Forum Allergy Rhinol*. 2020; 10(9):1049-1056.

65. Roemer MI. Bed supply and hospital utilization: a natural experiment. *Hospitals*. 1961;35:36-42.
66. Delamater PL, Messina JP, Grady SC, et al. Do more hospital beds lead to higher hospitalization rates? a spatial examination of Roemer's Law. *PLoS One.* 2013;8(2):e54900.
67. Ginsburg PB, Koretz DM. Bed availability and hospital utilization: estimates of the "Roemer effect". *Health Care Financ Rev*. 1983;5(1):87-92.
68. Baras JD, Baker LC. Magnetic resonance imaging and low back pain care for Medicare patients. *Health Aff (Millwood)*. 2009;28(6):w1133-40.
69. Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med*. 1998;338(16):1089-96
70. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United Kingdom. *JAMA*. 2003;290(16):2129-37.
71. Elmore JG, Nakano CY, Koepsell TD, et al. International variation in screening mammography interpretations in community-based programs. J *Natl Cancer Inst*. 2003;95(18):1384-93.
72. Strunk AL, Kenyon S. Medicolegal considerations in the diagnosis of breast cancer. *Obstet Gynecol Clin North Am*. 2002;29(1):43-9.

Chapter 2 Use of heuristics in testing, screening and treatment decisions: a systematic review

Chapter 2 of this thesis will be submitted for consideration for publication in JAMA Internal Medicine as: **Soon J**, Tracy M, Scott I, Elshaug AG, Howard K, Traeger AC. Use of heuristics in testing, screening and treatment decisions: a systematic review. 2024.

Full author list

Jason Soon[1], Marguerite Tracy[2], Ian Scott[3], Adam G Elshaug[4 1], Kirsten Howard[1], Adrian Traeger[5]

1. Menzies Centre for Health Policy and Economics, School of Public Health, Charles Perkins Centre, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia
2. Sydney Medical School, General Practice Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia
3. Centre for Heath Services Research, Faculty of Medicine, University of Queensland, Brisbane, Queensland, Australia.
4. Centre for Health Policy, Melbourne School of Population and Global Health and the Melbourne Medical School, University of Melbourne, Vic., Australia.
5. Institute for Musculoskeletal Health, Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia.

Key points

(as per journal requirements)

Question: How do heuristics prompted by initial patient triage information and clinicians' prior experiences affect medical decision-making in real-world clinical settings?

Findings: This systematic review of ten high-quality observational studies found that heuristic-driven decisions significantly influenced medical decision-making. Studies showed odds ratios of 1.38 to 1.7 for decisions based on initial triage information, and 1.02 to 1.72 for decisions influenced by previous medical events, affecting areas of patient care such as testing and treatment across various specialties.

Meaning: Heuristics substantially impact medical decision-making, highlighting the need for awareness and strategies to mitigate potential negative effects on resource use and healthcare quality.

Abstract

Importance:

Heuristics, or cognitive shortcuts, play a key role in medical decision-making, often unconsciously influencing testing, prescribing, and surgical management. These shortcuts can affect resource use, clinical variation, and patient safety. Understanding their impact across various medical settings is crucial for improving care quality.

Objective:

To synthesise existing evidence evaluating the effects of heuristic-driven decision-making by medical practitioners in real-world clinical settings, focusing on how initial patient triage information and clinicians' prior experiences with medical events influence decisions on testing, screening, and treatment.

Evidence review:

A systematic search of five databases (Ovid, PsycINFO, Econlit, Scopus, Web of Science) was conducted to identify relevant literature from January 1980 to September 2023. Included studies were empirical studies in clinical settings that quantitatively measured changes in practitioners' behaviour related to testing, screening, or treatment, investigated hypotheses on the heuristics of interest, and were published in English-language peer-reviewed journals. Study selection, data extraction, and quality assessment followed standardized protocols, with quality assessed via the Newcastle-Ottawa Scale (NOS).

Findings:

Ten high-quality observational studies (NOS-rated) were included. Two studies examined heuristics prompted by initial patient triage information, showing statistically significant odds ratios (ORs) of 1.38 to 1.7 for heuristic-driven medical decisions impacting patient care (e.g., avoiding pulmonary embolism testing or increasing trauma centre transfers). Eight studies focused on heuristics prompted by previous medical events, with ORs ranging from 1.02 to 1.72 for decisions such as switching obstetric methods, ordering specific tests, or avoiding specific therapeutics. One study found statistically insignificant results. Temporal effects, assessed only for heuristics based on previous medical events, mostly lasted 3-6 months. Studies spanned various specialties including emergency medicine, obstetrics, and primary care.

Conclusions and Relevance:

Heuristics significantly influence medical decision-making, potentially leading to deviations from evidence-based care. Increasing physician awareness about heuristic influences, particularly availability, representativeness, and anchoring heuristics, is crucial. Implementing strategies to mitigate heuristic-induced biases could reduce low-value care and improve patient safety. Further research should directly measure the impact of heuristic-influenced decision-making on health outcomes and develop effective debiasing strategies for clinical practice.

Introduction

Heuristics are mental shortcuts or rules of thumb that influence decision-making, often without the decision-maker's conscious awareness. It has been argued that these cognitive tools are important in various decision-making settings,[1 2] including medical decision-making,[3-5] where practitioners frequently need to make quick decisions based on limited information. If heuristics influence medical decision-making outcomes such as testing, prescribing, and surgical management, these may have flow-on effects to resource usage, clinical variation, and healthcare safety and quality. Therefore, it is important to understand the average magnitude and persistence of heuristic impacts across different medical settings.

However, a key limitation of existing evidence on the use of heuristics by decision-makers in healthcare settings[6 7] including medical practitioners,[8 9] is its reliance on hypothetical scenarios presented through vignettes and surveys. Moreover, of the four most recent literature reviews on heuristics in healthcare settings[6-9] only two[8 9] were specific to routine medical care settings – one was a quasi-review[9] and the other was primarily narrative.[8] Neither looked exclusively at observational data from real clinical settings or attempted to evaluate and compare quantitative data on the impacts of heuristics on medical decisions.

In order to undertake a review which would allow us to evaluate and compare quantitative data on the impacts of heuristics on medical decisions, it was firstly necessary to define some key heuristics as specifically as possible so that only results quantifying the impacts of the same types of heuristics could be pooled together. For this reason, we considered that it would be helpful to focus on two key exposures identified in the literature as prompting heuristic use: i) initial patient triage information e.g. Kulkarni et al 2019[10] and ii) clinicians' prior experience with medical events e.g. Keating et al 2017.[11] The systematic review was therefore formulated to determine the impacts of heuristic use prompted by these exposures on medical decision-making derived from studies from real clinical settings and to compare the results using a common measure. The specific research question was defined as follows:

Among medical practitioners, does (i) exposure to initial information presented about patients in triage records or (ii) prior experience with medical events prompt the use of heuristics that impact on patient care?

Methods

Search strategy and inclusion/exclusion criteria

This systematic review was registered on PROSPERO (ID 447414) and has been reported according to PRISMA guidelines. On 1 October 2023, we conducted a literature search of Ovid (MEDLINE and EMBASE), PsycINFO, Econlit, Scopus and

Web of Science databases from January 1980 to September 1 2023 using a pre-specified search protocol. We were specifically interested in the use of heuristics (e.g. 'anchoring', 'availability', 'premature closure', 'confirmation', 'representativeness') in testing, screening and treatment decisions. A combination of search terms including generic terms such as 'heuristics', 'cognitive bias' and 'heuristic bias' and terms specifying examples of heuristics and biases was used as part of these search strategies. References were also reviewed from previously retrieved articles.

Included studies had to meet all the following criteria:

- Papers published in English in peer reviewed journals;
- Empirical study designs in a real (non-simulated) clinical setting;
- Reported at least one quantitative measure of clinician behaviours relating to testing, screening and treatment decisions;
- Full-text papers investigating hypotheses regarding use of heuristics and/or biases among medical professionals in their medical decision-making;
- Included references to one of the two specific 'exposures' of interest: i) initial patient triage information or ii) previous medical events.

The full search strategy is available in the **Supplement**.

The first exposure of interest represented a narrowing of our criteria from the protocol which define the exposure as 'initial information presented about patients' because selecting studies involving this exposure without specifying the source of the information led to difficulties in reaching agreement between reviewers on study inclusion. Therefore, studies involving this exposure type were restricted to cases where initial patient information was presented in triage records. For instance, this excluded a study such as Fischer et al 2008[12] where patient information could have been gleaned from any of a combination of physical examination, patient complaints and diagnostic procedures and it was unclear whether exposure to such information preceded any decision-making change.

Study selection

Three investigators (JS, AT, and MT) independently screened identified references for inclusion based on title and abstract. Full texts of potentially eligible articles were reviewed by the same three reviewers. Disagreements were resolved by discussion.

Data extraction and risk of bias

Data were extracted using a modified EPOC (Effective Practice and Organisation of Care) data collection checklist embedded in an Excel file. At least two reviewers (JS, AT, MT) independently extracted data from each article. As all studies selected were observational studies, they were assessed for quality using the Newcastle-Ottawa Scale (NOS). NOS thresholds employed in a previous review of medical heuristics[8] were used as the basis for rating the quality of each study. At least two reviewers

(JS, AT, MT) independently assessed risk of bias for each article, and any disagreements were resolved by discussion with the third reviewer.

Analysis and presentation of results

Analysis focused on presenting the influence of the two exposures (initial patient triage information or previous medical events) on rates of various tests, referrals, and/or procedures. To facilitate comparability, where results were reported as percentage differences in care they were converted into mathematically equivalent odds ratios. Results from the two exposures of interest were presented separately.

To compare results with studies where the 'expected direction indicating heuristic use' was positive (e.g. Ly 2019[13]) with studies where the expected direction is negative (e.g. Keating et al 2017[11]), results were transformed into a consistent measure of heuristic use impacting on patient care. Based on this transformation, all results showing an odds ratio (OR) above 1 with confidence intervals not crossing 1 would indicate evidence of heuristics impacting on patient care.

Key outcomes of studies which investigated the same exposure and decision-making 'type' were presented together on the same forest plot. If studies reported more than one outcome, the outcome that study investigators focused on or discussed the most was chosen to be the key outcome. In the case of one study (Simianu et al 2016[14]) which reported four stratified results, the one which could be presented in a compatible format on a forest plot and had a baseline of no testing was presented and the others were discussed separately. If studies measured the same key outcome over multiple time periods, outcomes over these multiple periods were reported. If studies only reported statistically significant results in numerical formats, data for other results were extracted from published figures using plot digitiser software.[15]

Multiple results from the same studies were grouped together from top to bottom based on shortest to longest period for which outcomes were measured after the event.

Text Box 1: Explanations of heuristics

The ***availability heuristic*** is a heuristic resulting in judgments about the likelihood of an event based on how easily an example or instance comes to mind.[1] Such ease of recall may be due to, for instance, an emotionally charged or memorable or very frequent event. In the context of medical decision-making for instance it may result in a clinician who, during flu season and having seen many cases of patients with flu-triggered bronchospastic disorders causing dyspnoea, gives the same diagnosis to another patient presenting with shortness of breath, while ignoring symptoms and risk factors suggestive of pulmonary embolism which can also cause dyspnoea.

As originally defined, the ***anchoring heuristic*** refers to a heuristic whereby initial exposure to a number serves as a reference point and influences subsequent quantitative judgments.[1] However, the concept can also be extended to refer to non-quantitative judgements so that it also refers to a heuristic where the decision maker focuses on a single—often initial—piece of information when making a decision without sufficiently adjusting to later information. In the context of medical decision-making for instance it may result in a clinician who while troubleshooting an alarm on an infusion pump is oblivious to sudden surgical bleeding and hypotension (example from Stiegler et al 2012[2]).

The ***representativeness heuristic*** is a heuristic which results in a decision-maker evaluating the probability that an object or event belongs to a specific class by looking at the degree to which that particular object or event resembles that class. An example from medical decision-making is when a young male enters the emergency room with sudden onset of slurred speech, confusion, and weakness on one side of his body and the clinician's initial diagnosis is that this is due to alcohol or drugs because the patient's young age does not fit the 'representative' image of a stroke patient as being elderly or at least middle-aged.

1. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science.* 1974;185(4157):1124-1131
2. Stiegler MP, Neelankavil JP, Canales C, Dhillon A. Cognitive errors detected in anaesthesiology: a literature review and pilot study. *Br J Anaesth.* 2012 Feb;108(2):229-35.

Results

Study characteristics

Our search identified 4764 records of which 3524 were unique peer reviewed studies. After title/abstract screening and full text assessment, 10 articles met all eligibility criteria. Further details about the screening and selection process are provided in **Figure 2-1**.

Table 2-1 shows the study characteristics. Nine included studies were cohort studies (eight retrospective and one prospective) and one was a cross-sectional study which in substance and for rating purposes we classified as a retrospective cohort study. **Table 2-1** in the **Supplement** summarises the NOS criterion ratings achieved by each study and provides a detailed justification for our classification. The numerical results of studies summarised in forest plots in the subsequent sections are provided in **Table 2-2**.

Two studies investigated heuristics prompted by initial patient triage information and eight studies investigated heuristics prompted by previous medical events. Although not all studies labelled the heuristic under investigation, of those that did, the most investigated heuristic in our included studies was availability followed by representativeness and anchoring (see **Text Box 1** for brief explanations of each heuristic).

In two studies[11 13] the medical decision makers were primary care physicians (20%), three[5 10 16] involved emergency department physicians (30%), three[17-19] involved obstetricians (30%), and one each involved thoracic physicians[20] and colorectal surgeons.[14] Forms of medical decision-making investigated were prescribing[20] (10%), triaging to a trauma centre[10] (10%), obstetric management[17-19] (30%) and testing and investigation (50%).[5 11 13 14 16]

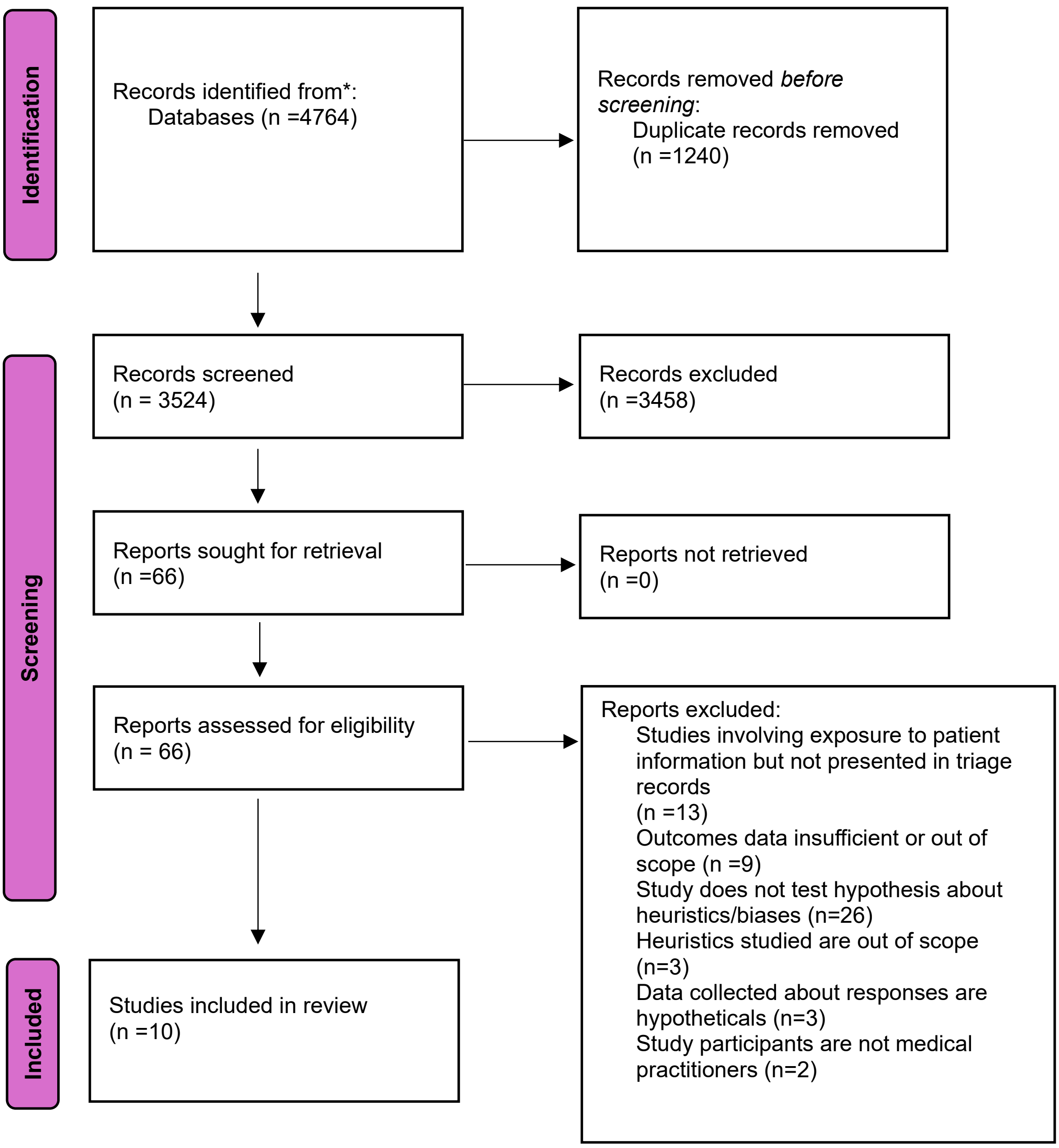

Figure 2-1: PRISMA flow diagram of the study selection process

Table 2-1 Summary of selected studies

Study	Country	n=	Clinical specialty group	Design	Category of exposure	Description of exposure (as described in article)	Authors' label attached to the cognitive bias/heuristic	Outcome measure timepoints following exposure
Choudry et al 2006	Canada	530 physicians, 1060 patients	Thoracic physicians	Retrospective cohort study (matched pair before and after analysis)	Previous medical event	Patients with atrial fibrillation admitted to hospital for adverse bleeding events while taking warfarin	'Availability bias' 'Chagrin factor'	from 0-90 days to 271-360 days
Dan et al 2017	Israel	40 physicians, 4052 patients	Obstetricians	Retrospective chart review	Previous medical event	Full term deliveries resulting in hypoxic–ischemic encephalopathy developed secondary to obstetric mismanagement	'Availability heuristic'	From 0-2 to 4-6 weeks
Keating et al 2017	US	30,704 physicians; 5,360,191 patients	Primary care physicians	Longitudinal study of representative cohort	Previous medical events	Patients who suffered a gastrointestinal bleed or perforation leading to hospitalisation or death within 14 days of colonoscopy	'Availability heuristic'	From Q1 to Q4
Kulkarni et al 2019	US	3199 patients	Emergency physicians	Retrospective cohort study	Initial information	Patients presenting to non-trauma centre noted as having characteristics representative of severe trauma cases on their charts	'Representativeness heuristic'	N/A
Ly 2019	US	20,200 physicians	Primary care physicians	Retrospective analysis of claims data[a]	Previous medical events	Colleague in same practice with recent malpractice injury case	Nil	From Q1 to Q4
Ly et al 2021	US	7370 physicians, 416,720 patient visits	Emergency physicians	Event study [b]	Previous medical events	Patients discharged with a diagnosis of pulmonary embolism	'Availability heuristic'	From 0-10 days to 51-60 days
Ly et al 2023	US	108,019 patients	Emergency physicians	Cross-sectional study [c]	Initial information	Patients who presented to emergency department with shortness of breath who had congestive heart failure mentioned in the patient visit reason section documented in triage	'Anchoring bias'	N/A

Riddell et al 2014	US	2859 hospitals, 1,414,134 deliveries	Obstetricians	Controlled pre-post study	Previous medical events	Uterine ruptures during obstetric management of labour and delivery associated with severe maternal morbidity or stillbirth	‘Availability heuristic’ ‘Regret aversion’ ‘Representativeness heuristic’	From 1 month to 11 months
Simianu et al 2016	US	282 physicians, 4854 operations	Colorectal surgeons	Prospective cohort study	Previous medical events	Patients undergoing colorectal operations who were subsequently found to have radiologically demonstrated anastomotic leak or enterocutaneous fistula, postoperative percutaneous drainage of abscess, or unplanned re-operative intervention requiring colostomy and/or ileostomy, abscess drainage, operative drain placement, or anastomotic revision	‘Recency bias’ ‘Availability heuristic’	6 months after
Singh 2021	US	231 physicians, 86, 345 deliveries	Obstetricians	Retrospective observational study	Previous medical events	Patients who experienced a complication during delivery	“win-stay/lose-shift” heuristic Other heuristics mentioned as possibilities are ‘availability’, ‘salience’, ’recency’ and ‘attribute substitution’	N/A

a) Ly 2019 described his approach as ‘analyses using 2009-2013 claims data for a random 20% sample of Medicare beneficiaries linked by physician to publicly available 2010-2012 injury reports’ and the empirical specification is described as an ‘event study’. For quality assessment purposes we classified this as a retrospective cohort study.
b) Although described as an event study Ly et al 2021 also report that they used electronic health record data which had been used in prior studies. For quality assessment purposes we classified this as a retrospective cohort study.
c) The authors of Ly et al 2023 classified their study as a cross-sectional study but in substance and for purposes of rating we considered this a retrospective cohort study. See the notes to Table 6-1 in the Supplement for a detailed justification of this.

Table 2-2 Summary of key outcomes from included studies

Study	Exposure	Outcome	Outcome measure timepoints following exposure	Odds ratio (95% confidence interval)
Choudry et al 2006	Bleeding in previous patient prescribed warfarin	Decision not to prescribe warfarin[1]	0 to 90 days after exposure	1.21 (1.00-1.38)
			91 to 180 days after exposure	1.4 (1.21-1.54)
			181 to 270 days after exposure	1.39 (1.19-1.54)
			271 to 360 days after exposure	1.28 (1.03-1.46)
Dan et al 2017	Mismanaged full term delivery in a previous patient	Unscheduled caesarean[2]	0 to 2 weeks after exposure	1.72 (1.31-2.21)
			2 to 4 weeks after exposure	1.45 (1.07-1.88)
			4 to 6 weeks after exposure	1.09 (0.77-1.46)
Keating et al 2017	Colonoscopy complication in a previous patient	Decision not to order colonoscopy[3]	Quarter 1 after exposure	1.01 (0.99-1.02)
			Quarter 2 after exposure	1.02 (1.01-1.03)
			Quarter 3 after exposure	1.01 (1.00-1.02)
			Quarter 4 after exposure	1.00 (0.99-1.01)
Kulkarni et al 2019	Representative characteristics of severe trauma mentioned in triage record	Hospital transfer	N/A	1.70 (1.40-2.00)
Ly 2019	Injury report against a practice peer	Decision to order advanced imaging [4]	Quarter 1 after exposure	1.35 (1.06-1.64)
			Quarter 2 after exposure	1.16 (1.05-1.27)
			Quarter 3 after exposure	1.02 (0.93-1.10)
			Quarter 4 after exposure	0.96 (0.88-1.04)
Ly et al 2021	Previous patient pulmonary embolism diagnosis	Pulmonary embolism testing[5]	0 to 10 days after exposure	1.17 (1.05-1.29)
			11 to 20 days after exposure	1.10 (0.96-1.23)
			21 to 30 days after exposure	1.11 (1.00-1.22)
			31 to 40 days after exposure	1.07 (0.96-1.19)
			41 to 50 days after exposure	1.06 (0.95-1.17)
			51 to 60 days after exposure	1.06 (0.94-1.18)
Ly et al 2023	Congestive heart failure mentioned as reason for patient visit in triage record	Avoiding pulmonary embolism testing[6]	N/A	1.38 (1.29-1.45)
Riddell et al 2014	Uterine rupture during labour in previous patient	Failure of trial of labour[7]	1 month after exposure	1.11 (1.03-1.21)
			2 months after exposure	1.01 (0.92-1.08)
			3 months after exposure	1.01 (0.94-1.09)
			4 months after exposure	1.03 (0.95-1.10)
			5 months after exposure	1.01 (0.92-1.10)
			6 months after exposure	1.03 (0.93-1.14)
			7 months after exposure	1.03 (0.93-1.14)

			8 months after exposure	0.99 (0.88-1.09)
			9 months after exposure	1.04 (0.92-1.18)
			10 months after exposure	1.12 (0.96-1.29)
			11 months after exposure	1.01 (0.82-1.24)
Simianu et al 2016	Anastomotic leak in previous patient when no leak testing was done	Leak testing[8]	6 months after exposure	Surgeon volume 5+ cases p.a. + No leak testing during leak episode: 0.60 (0.13-2.28)
Singh 2021	Complication in caesarean delivery of previous patient	Vaginal delivery[9]	N/A	1.10 (1.04-1.18)

1 Original outcome reported in study was the patient's odds of receiving warfarin and the study found a reduction compared to before exposure. This has been recalculated in terms of the doctor's odds of not prescribing warfarin to retain similar directionality with the majority of studies for comparison purposes.

2 Original outcome reported in study was percentage difference in caesarean deliveries as well as in terms of a hazard ratio which is defined as "the rate of change in the probability of unscheduled caesarean delivery." We consider that this is effectively an odds ratio and have interpreted it as such.

3 Original outcome reported in study was percentage change (reduction) in colonoscopies among patients. This has been recalculated in terms of the doctor's decision not to order a colonoscopy to retain similar directionality with the majority of studies for comparison purposes.

4 Original outcome reported in study was change in advanced imaging per 100 evaluation and management visits and this was converted to an odds ratio.

5 Original outcome reported in study was percentage change in pulmonary embolism testing and this was converted to an odds ratio.

6 Original outcome in study was percentage change (reduction) in pulmonary testing. This has been recalculated in terms of the doctor's decision to avoid pulmonary embolism testing to retain similar directionality with the majority of studies for comparison purposes.

7 Original outcome in study was percentage change in trial of labour success and the study found a reduction compared to before exposure. Although another outcome reported was rates of vaginal birth we considered 'trial of labour success' the key outcome because vaginal birth can be defined as a subset of this measure and was the main focus of reporting in the results section of the study. This has been recalculated in terms of a failure of trial of labour to retain similar directionality with the majority of studies for comparison purposes.

8 Original outcome in study was change in rate of leak testing after a postoperative leak occurred so this was recalculated as an odds ratio of leak testing. The study reported 4 different measures – change in rate of leak testing (i) in surgeons who performed under 5 cases a year where leak testing was not performed in the case that resulted in a post-operative leak (ii) in surgeons who performed under 5 cases a year where leak testing was performed in the case that resulted in a post-operative leak (iii) in surgeons who performed 5 cases or more a year where leak testing was not performed in the case that resulted in a post-operative leak (iv) in surgeons who performed 5 cases or more a year where leak testing was performed in the case that resulted in a post-operative leak. All results were found to be statistically insignificant. The table only reports results from scenario (iii) because it was the only one of two results where the comparator was no testing and the other result (scenario (i)) had extremely wide confidence intervals and presenting this result would have obscured presentation of results of all the other studies.

9 Singh reported the percentage change in patient's likelihood of a vaginal delivery following complication in caesarean and this was converted to an odds ratio.

Heuristics shaped by initial patient information

Key decision-making outcomes

Figure 2-2 reports on key outcomes of the two studies which investigated heuristics prompted by initial patient triage information. Both show evidence of use of these heuristics impacting on care indicated by statistically significant ORs.

Ly et al 2023[5] reported an OR of 1.38 (1.29-1.45) representing higher odds of avoiding pulmonary embolism (PE) testing induced by mention of congestive heart failure (CHF) as patient visit reason in triage records. They also reported that the mention of CHF in patient triage records led to longer times to PE testing and increased B-type natriuretic peptide (BNP) testing which is used to diagnose heart failure, while having no impact on overall PE diagnosis rates.

Kulkarni et al 2019[10] reported an OR of 1.7 (1.4-2.0) for transfer to trauma centres following exposure to triage information that a patient had characteristics representative of severe trauma. They also reported a higher OR of 2.4 (1.7-3.3) where the 'representative characteristics' were specifically of trauma associated with motor vehicle collisions and 3.8 (1.7-8.7) where they were penetrating injuries.

Heuristics shaped by previous medical events

Key decision-making outcomes

Figures 2-3, 2-4 and 2-5 report on key outcomes of studies which investigated heuristics prompted by previous medical events. The results from these studies were grouped according to the clinical setting of medical decision-making.

<u>Obstetric decision-making</u>: **Figure 2-3** shows use of heuristics impacting on care as noted by statistically significant ORs ranging from 1.10 (1.04-1.18) to 1.72 (1.31-2.21). These represented higher odds of vaginal delivery following exposure to complications in a previous caesarean delivery (Singh 2021[19]) and higher odds of unscheduled caesarean deliveries or failure of trial of labour following exposure to complications in a previous full-term delivery or labour (Dan et al 2017;[17] Riddel et al 2014[18]).

<u>Testing and investigation decision-making</u>: **Figure 2-4** shows use of heuristics impacting on care as noted by statistically significant ORs ranging from 1.02 (1.01-1.03) to 1.35 (1.06-1.64). These represented higher odds of using three types of investigation: advanced imaging for physicians aware of practice peers who had received adverse injury reports (Ly 2019[13]); PE testing among emergency physicians who had recently diagnosed PE in a previous patient (Ly et al 2021[16]); and in a separate group, higher odds of not ordering colonoscopy were seen among physicians whose previous patients had bleeding complications after colonoscopies (Keating et al 2017[11]).

However, **Figure 2-4** also shows one statistically insignificant result from Simianu et al 2016[14] which examined testing rates for anastomotic leak in colorectal surgery before and six months after surgeons had patients who had experienced a leak. While Simianu et al 2016[14] reports on results from four scenarios, all results were statistically insignificant and only one of these showing an OR of 0.60 (0.13-2.28) is presented in **Figure 2-3** (for reasons discussed in **Methods**). Other results from Simianu et al 2016[14] are presented in the **Supplement**.

Warfarin prescribing: The results of the sole study of prescribing are reported in **Figure 2-5**. Heuristic use impacted on care as noted by statistically significant ORs ranging from 1.21(1.00-1.38) to 1.4 (1.21-1.54) representing higher odds of not prescribing warfarin for physicians whose previous patients had suffered bleeding events after prescribing warfarin.

Age stratified impacts

Keating et al 2017[11] reported that exposure to prior medical events had disparate impacts on medical decision-making based on both physician and patient age. Physicians whose patients previously had complications after colonoscopies but who were less than the median 50.2 years of age of physicians in the study exhibited statistically significant odds of not using colonoscopy in both the subsequent second and third quarters after this adverse event. By contrast, physicians with similar scenarios but who were of median age or older did not exhibit any change in patterns of care during any subsequent periods.

Similarly, patients aged over 75 years faced statistically significant higher odds of not being referred for colonoscopy by physicians who had experienced prior events up to 9 months after the event. By contrast, for patients aged 65-75 years, this avoidance of colonoscopy was only seen up to 6 months following the event (details provided in **Supplement**).

Patient-related outcomes

Singh 2021[19] investigated impacts of heuristics shaped by previous medical events on health outcomes, estimating from her model that physicians who switched obstetric procedures after previous complications reported similar or worse patient outcomes over time, including an absolute increase of 0.11% in maternal/neonatal death and a 0.11% reduction in home discharges.

Although Ly et al 2023[5] did not measure health outcomes, they estimated that mention of CHF in a patient triage record increased time to PE testing by 15.5 minutes.

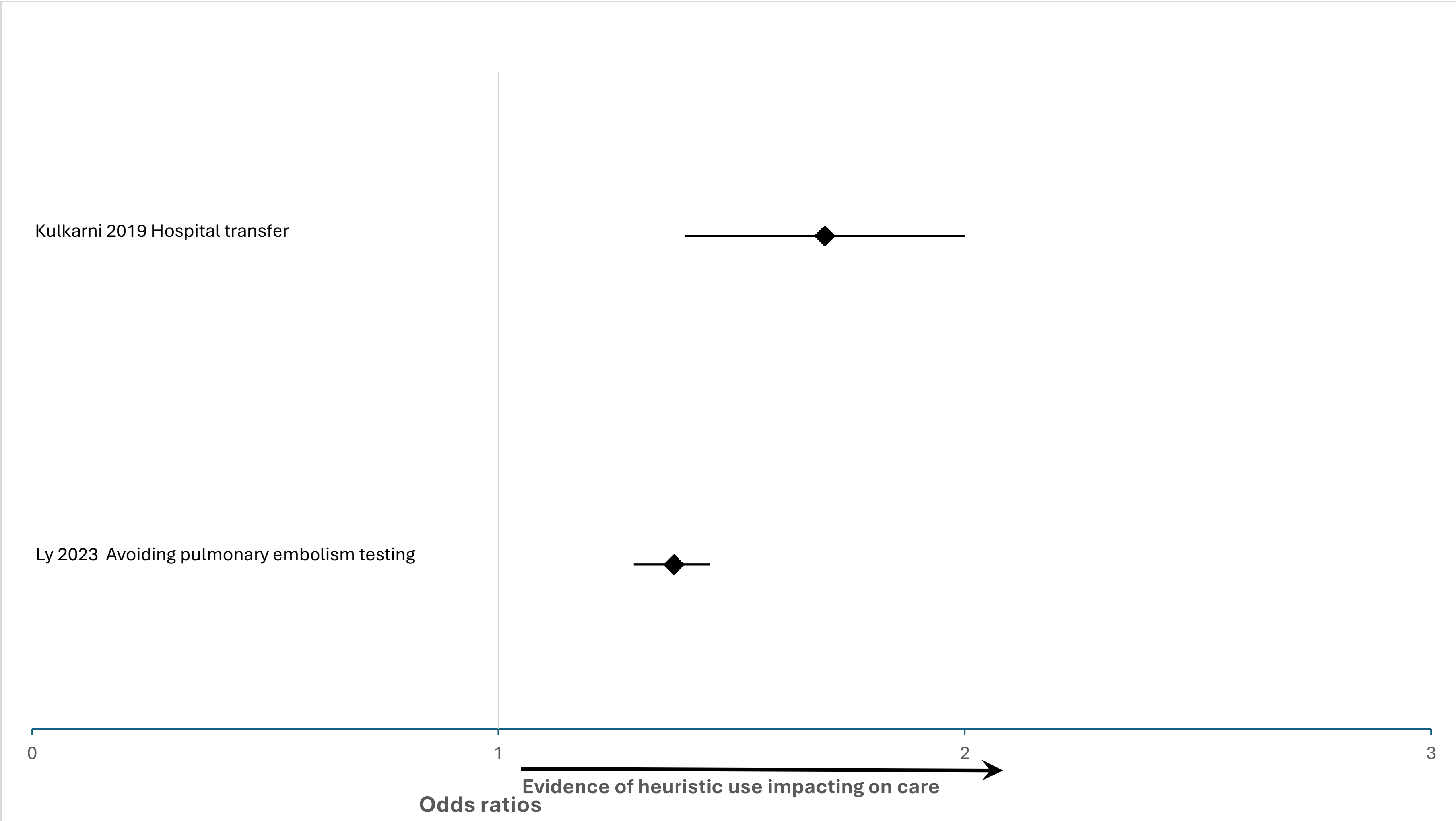

Figure 2-2: Odds ratios of medical decisions following exposure to initial patient triage information.

Initial patient triage information were representative characteristics of severe trauma mentioned (Kulkarni et al 2019) and congestive heart failure mentioned as reason for patient visit (Ly et al 2023).

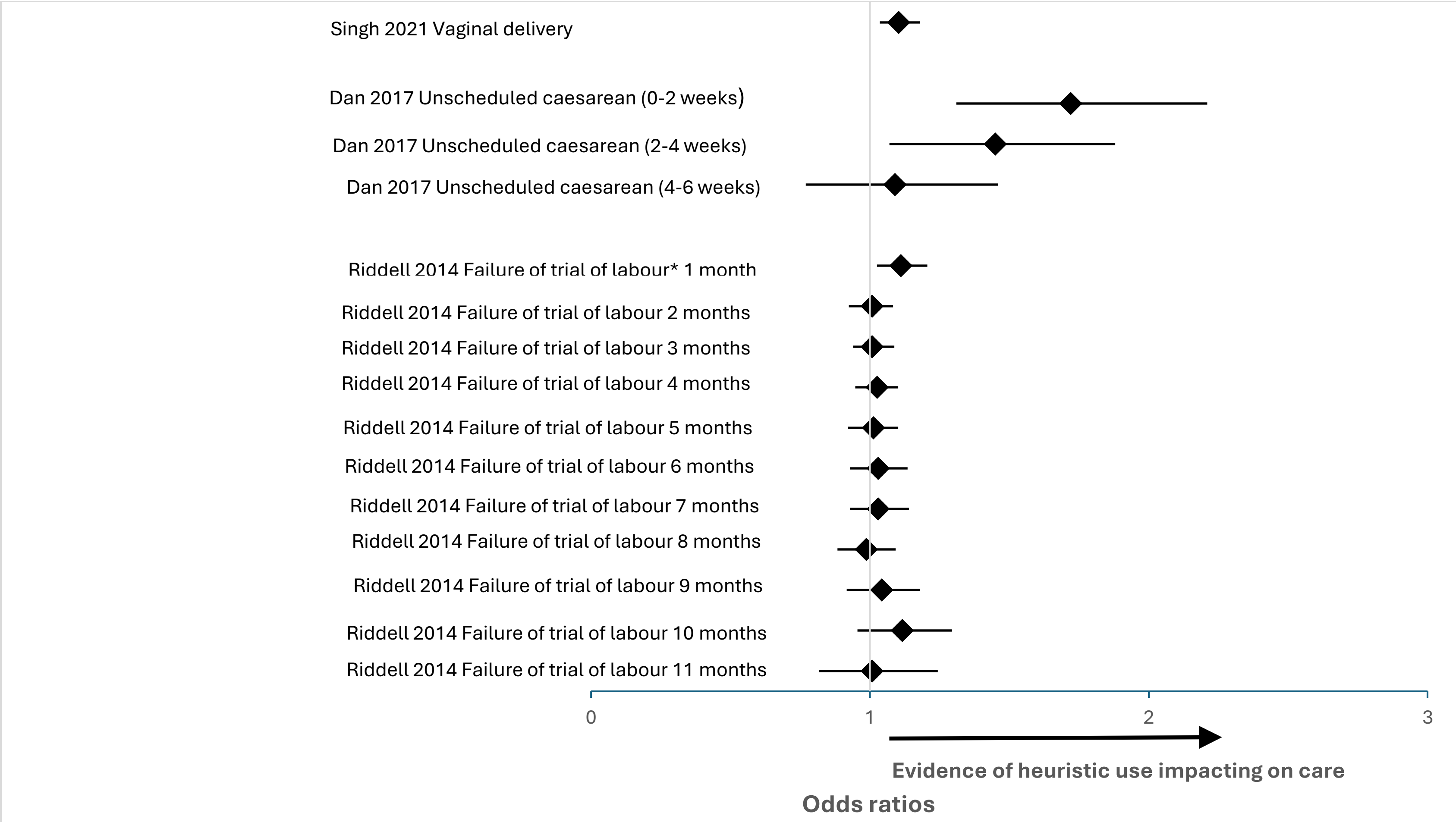

Figure 2-3: Odds ratios of obstetric decisions following exposures to previous medical events.

Time periods refer to period elapsed since the event with the exception of Singh 2021 which did not report this. Previous medical events were complication in previous caesarean delivery (Singh 2021), mismanaged previous full term delivery (Dan et al 2017) and uterine rupture during labour of previous patient (Riddell et al 2014).

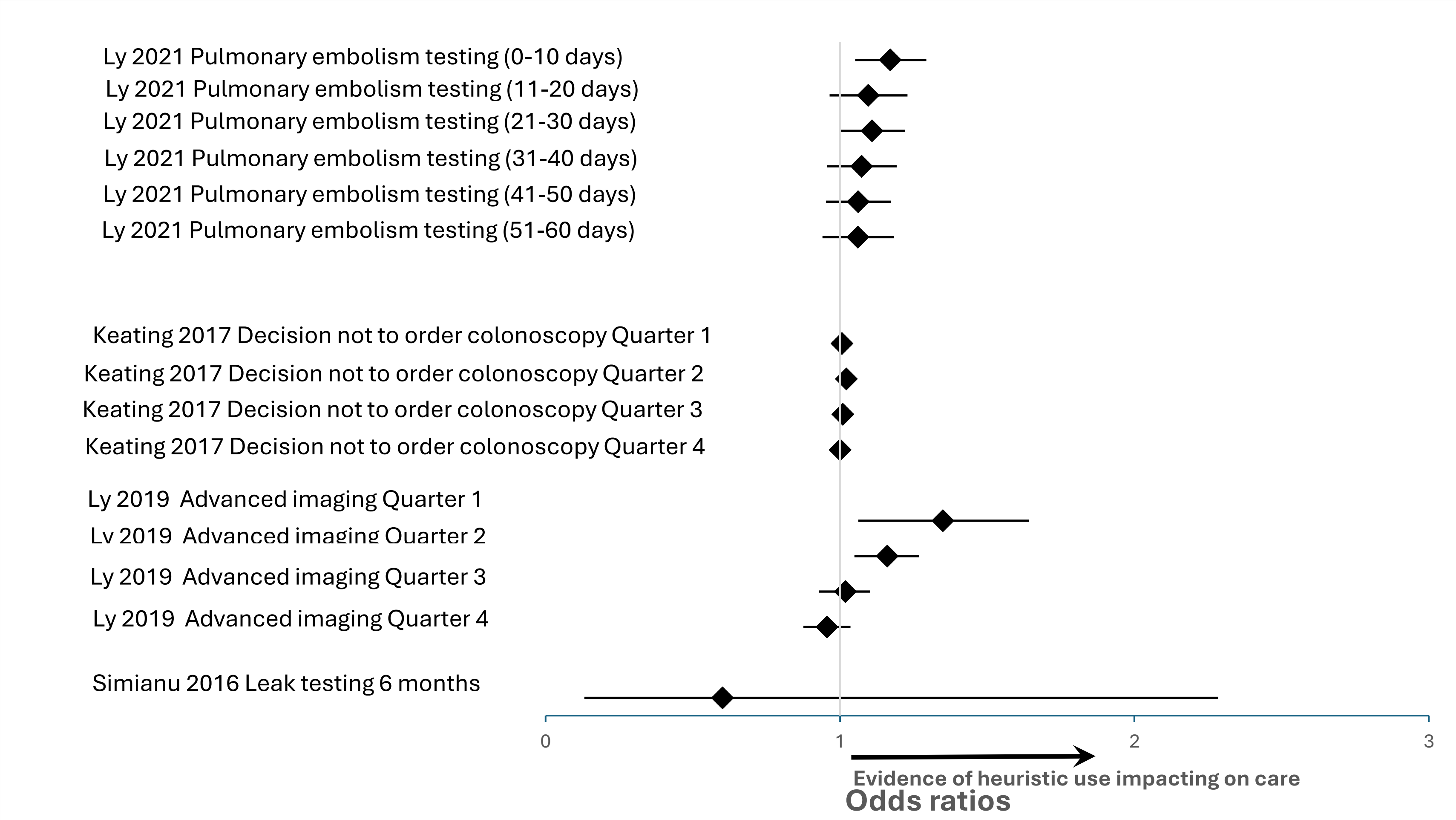

Figure 2-4: Odds ratios of testing and investigation decisions following exposures to previous medical events.

Time periods refer to period elapsed since the event. Previous medical events were previous patient pulmonary embolism diagnosis (Ly et al 2021), colonoscopy complication in previous patient (Keating et al 2017), injury report against a practice peer (Ly 2019) and anastomotic leak in previous patient (Simianu et al 2016). The confidence interval estimate for Simianu et al 2016 is an approximation because of its different method of reporting.

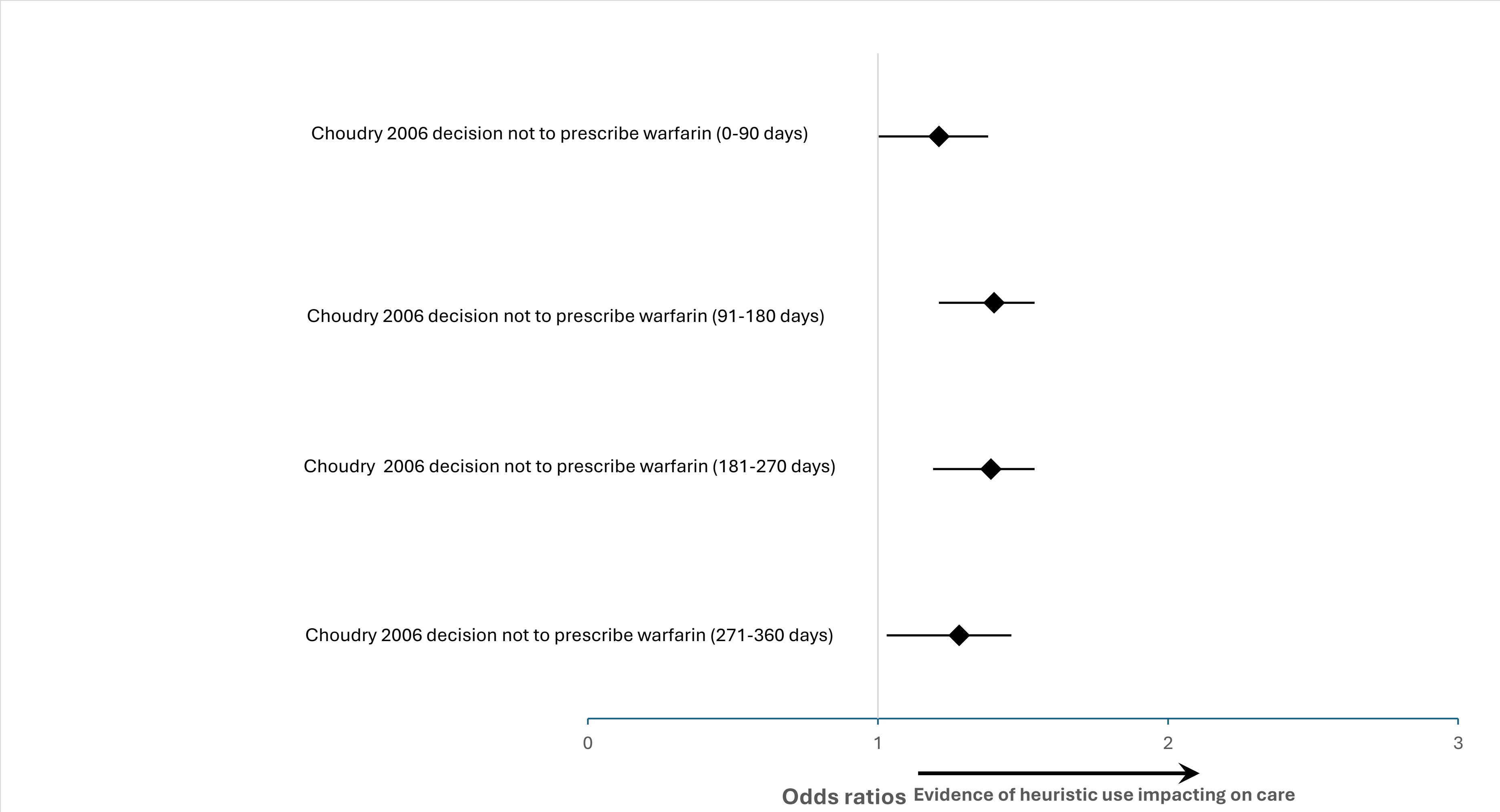

Figure 2-5: Odds ratios of prescribing decisions following exposures to previous medical events.

Time periods refer to period elapsed since the event. Previous medical events were bleeding events in previous patient following prescription of warfarin.

Discussion

Main findings and potential implications

This systematic review analysed data from 10 studies to understand impacts of heuristics prompted by two types of exposures – initial triage information and clinicians' experience of recent prior medical events – on medical decision-making.

With the exceptions of Singh 2021[19] (who found that heuristic induced switching between obstetric procedures worsened health outcomes) and Ly et al 2023[5] (who found that exposure to triage information highlighting CHF as a reason for the patient's visit increased time to diagnosis of PE implying a longer time to correct diagnosis for patients who had PE), none of the included studies reported other measures which provided evidence of a direct link (either positive or negative) between heuristics and patient health outcomes. Nonetheless, these findings suggest potential areas of concern associated with heuristic use.

The findings of high odds ratios (OR 1.38-1.7) from studies of heuristics prompted by initial patient triage information imply that medical decision-makers may disproportionately rely on a selective subset of information in patient records to make clinical management decisions. The specific heuristics postulated by investigators in these studies to have affected decision-making were an anchoring heuristic in the study where mention of CHF increased the odds of avoiding PE testing (Ly et al 2023[5]) and a representativeness heuristic in the case where mention of representative characteristics of severe trauma increased transfers of patients to trauma centres (Kulkarni et al 2019[10]).

The findings from studies of heuristics prompted by previous medical events suggest that prior experience with medical malpractice cases and diagnoses made for previous patients can disproportionately increase imaging or use of specific tests for current patients. Similarly, adverse events previously experienced with caesarean deliveries or normal deliveries can induce physicians to preference full term deliveries or caesarean sections respectively in subsequent patients; and adverse events associated with specific forms of care (colonoscopies, warfarin prescription) in a previous patient can also induce physicians to reduce those forms of care in subsequent patients. While obstetric studies showed larger impacts of previous medical events on decision-making (OR 1.10-1.72) compared to testing-related studies (OR 1.02-1.35), this may reflect small study bias in the former due to the smaller physician sample sizes (see **Supplement** for extended discussion on this issue). The heuristic postulated by all but one of the studies to explain these decision-making changes was the availability heuristic, sometimes known as the recency effect. Only one of the included studies (by Simianu et al 2016[14]) reported all results as statistically insignificant.

Thus, notwithstanding the existence of detailed guidelines on various forms of care,[21 22] our data show that prior events or information exposures can interrupt usual practice and result in deviation from evidence-based care due to selectively processed information in triage notes or recalled experience relating to single episodes of care. Care is unlikely to be evidence-based under such circumstances and these changes are likely to have resulted in a lower quality of care overall, even though by happenstance in some cases it may have led to improvements in care. An example of the latter is from Keating et al 2017[11] where a heuristic-induced reduction in colonoscopies in the elderly may have led to more evidence-based care for this cohort resulting in a reduced incidence of overdiagnosis.

Fortunately, in all but one of the studies of heuristics prompted by prior medical events, the impacts on subsequent decision-making were of limited duration, and did not persist beyond the first 3 to 6 months after the event. The exception was one study of prescribing[20] which found impacts (on warfarin prescription) persisting for up to 12 months following an adverse event.

These results underscore the benefits of increasing physician awareness about the influence of heuristics, particularly availability, representativeness and anchoring heuristics on their decision-making. Further research could examine the best method of improving awareness and mitigating risk, whether through educational materials, decision-making aids, use of ‘nudges’ which leverage other counteracting heuristics, simulation exercises or encouraging regular reflection on clinical practices. If use of heuristics is inevitable, circumstances predisposing to their misuse and debiasing strategies that can be used to counter such misuse, should be made clear. Such measures may reduce low-value care, both underuse and overuse of tests and interventions, and potentially improve patient safety and care quality. Further research is also needed to measure how heuristic-influenced decision-making affects health outcomes directly.

Comparison with previous studies

Previous reviews have included studies investigating a range of heuristics but did not directly compare or quantify the impacts of heuristics on medical decision-making using common metrics such as odds ratios.

Our review also took a different focus by reporting the magnitude and statistical significance of impacts of heuristics specifically prompted by two defined exposures, triage information and clinician past experiences. While resulting in more restricted study inclusion criteria, it allowed a comparison of findings from studies investigating similar heuristic types. Unlike previous reviews our review excluded studies based on responses to hypothetical scenarios which enhances the applicability and relevance of the findings to real world medical practice.

Limitations

The heterogeneity in outcome definitions across studies required the conversion of some study outcomes reported as other measures into odds ratios, which may create potential for errors or oversimplifications in reporting results as would the use of software to extract some results from published figures. As all included studies are observational, they carry an inherent risk of bias, despite being rated high quality on the Newcastle-Ottawa Scale. All these studies may be influenced by uncontrolled confounding factors, although some attempted to account for this through control and falsification analyses.

Only two studies examined heuristics associated with initial patient triage information, due to the need to restrict study inclusion to this narrow criterion for reasons mentioned previously. While our restriction improved clarity in study selection, it may have limited our ability to capture the full spectrum of heuristic impacts in medical decision-making related to initial patient information reported/exhibited clinically by patients or contained in referral letters and past medical records.

Finally we note that although all but one of our included studies reported results that were statistically significant, we cannot rule out the possibility that these outcomes are the result of publication bias i.e. studies which investigated heuristics but did not report any significant results did not proceed to publication.

Conclusions

This systematic review reveals that heuristics shaped by ‘initial patient triage information’ and ‘previous medical events’ significantly influence medical decision-making in testing, prescribing, and choice of obstetric procedures. These findings have implications for resource utilization and potentially for the quality of medical care, and further research is warranted in assessing the direct impacts of heuristically influenced decision-making on patients’ outcomes and the most effective means for countering the misapplication of commonly triggered heuristics.

References

1. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science* 1974;185(4157):1124-1131.
2. Gigerenzer G, Gaissmaier W. Heuristic decision making. *Annu Rev Psychol*. 2011; 62:451-482.
3. Marewski JN, Gigerenzer G. Heuristic decision making in medicine. *Dialogues Clin Neurosci*. 2012;14(1):77-89.
4. Halpern SD, Truog RD, Miller FG. Cognitive Bias and Public Health Policy During the COVID-19 Pandemic. *JAMA*. 2020; 324(4):337-338.
5. Ly DP, Shekelle PG, Song Z. Evidence for Anchoring Bias During Physician Decision-Making. *JAMA Intern Med*. 2023;183(8):818-823.
6. Featherston R, Downie LE, Vogel AP, et al. Decision making biases in the allied health professions: A systematic scoping review. *PLoS One*. 2020;15(10):e0240716.
7. Thirsk LM, Panchuk JT, Stahlke S, et al. Cognitive and implicit biases in nurses’ judgment and decision-making: A scoping review. *Int J Nurs Stud*. 2022;133:104284.

8. Saposnik G, Redelmeier D, Ruff CC, et al. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak*. 2016;16(1):138.
9. Whelehan DF, Conlon KC, Ridgway PF. Medicine and heuristics: cognitive biases and medical decision-making. *Ir J Med Sci*. 2020;189(4):1477-1484.
10. Kulkarni SS, Dewitt B, Fischhoff B, et al. Defining the representativeness heuristic in trauma triage: A retrospective observational cohort study. *PLoS One*. 2019;14(2):e0212201.
11. Keating NL, James O'Malley A, Onnela J-P, et al. Assessing the impact of colonoscopy complications on use of colonoscopy among primary care physicians and other connected physicians: an observational study of older Americans. *BMJ Open 2017*;7:e014239.
12. Fischer T, Fischer S, Himmel W, et al. Family practitioners' diagnostic decision-making processes regarding patients with respiratory tract infections: an observational study. *Med Decis Making*. 2008;28(6):810-8.
13. Ly DP. Rates of Advanced Imaging by Practice Peers After Malpractice Injury Reports in Florida, 2009-2013. *JAMA Intern Med*. 2019;179(8):1140-1141.
14. Simianu VV, Basu A, Alfonso-Cristancho R, et al. Assessing surgeon behavior change after anastomotic leak in colorectal surgery. *J Surg Res*. 2016;205(2):378-383.
15. https://web.eecs.utk.edu/~dcostine/personal/PowerDeviceLib/DigiTest/index.html [cited 25 September 2024].
16. Ly DP. The Influence of the Availability Heuristic on Physicians in the Emergency Department. *Ann Emerg Med*. 2021;78(5):650-657.
17. Dan O, Hochner-Celnikier D, Solnica A, et al. Association of Catastrophic Neonatal Outcomes With Increased Rate of Subsequent Cesarean Deliveries. *Obstet Gynecol*. 2017;129(4):671-675.
18. Riddell CA, Kaufman JS, Hutcheon JA, et al. Effect of uterine rupture on a hospital's future rate of vaginal birth after cesarean delivery. *Obstet Gynecol*. 2014;124(6):1175-1181.
19. Singh M. Heuristics in the delivery room. *Science*. 2021;374(6565):324-329.
20. Choudhry NK, Anderson GM, Laupacis A, et al. Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: matched pair analysis. *BMJ*. 2006;332(7534):141-5.
21. Davidson KW, Barry MJ, Mangione CM, et al. Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2021;325(19):1965-1977.
22. Wigle P, Hein B, Bernheisel CR. Anticoagulation: Updated Guidelines for Outpatient Management. *Am Fam Physician*. 2019;100(7):426-434.

Chapter 3 Effect of two behavioural ‘nudging’ interventions on choice of management options for low back pain: A randomised vignette-based study in general practitioners

Chapter 3 of this thesis has been published as **Soon J**, Traeger AC, Elshaug AG, Cvejic E, Maher CG, Doust JA, Mathieson S, McCaffery K, Bonner C. Effect of two behavioural 'nudging' interventions on management decisions for low back pain: a randomised vignette-based study in general practitioners. *BMJ Qual Saf*. 2019 Jul;28(7):547-555. doi: 10.1136/bmjqs-2018-008659. Epub 2018 Nov 19. PMID: 30455177

References to tables and figures have been amended to fit this document’s labelling conventions

Full author list (institutional affiliations were during time of publication)

Jason Soon[1], Adrian C Traeger[2], Adam G. Elshaug[1], Erin Cvejic[3], Chris G Maher[2], Jenny A Doust[4], Stephanie Mathieson[2], Kirsten McCaffery[3], Carissa Bonner[5]

Correspondence to Jason Soon, Menzies Centre for Health Policy, School of Public Health, Charles Perkins Centre, Faculty of Medicine and Health, University of Sydney, Sydney, NS 2006, Australia; tsoo7799@uni.sydney.edu.au

1. Menzies Centre for Health Policy, School of Public Health, Charles Perkins Centre, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia
2. Institute for Musculoskeletal Health, Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia
3. Sydney Health Literacy Lab, Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia
4. Centre for Research in Evidence-Based Practice (CREBP), Faculty of Health Sciences & Medicine, Bond University, Gold Coast, Queensland, Australia
5. School of Public Health, University of Sydney, Sydney, New South Wales, Australia

Abstract

Objective: ‘Nudges’ are subtle cognitive cues thought to influence behaviour. We investigated whether embedding nudges in a general practitioner (GP) clinical decision support display can reduce choice of low-value management options.

Methods: Australian GPs completed four clinical vignettes of patients with low back pain. Participants chose from three guideline-concordant and three guideline-discordant (low-value) management options for each vignette, on a computer screen. A 2x2 factorial design randomised participants to two possible nudge interventions: ‘partition display’ nudge (low-value options presented horizontally, high value options listed vertically), or ‘default option’ nudge (high-value options presented as the default, low-value options presented only after clicking for more). The primary outcome was the proportion of scenarios where practitioners chose at least one of the low-value care options.

Results: 120 general practitioners (72% male, 28% female) completed the trial (n=480 vignettes). Participants using a conventional menu display without nudges chose at least one low-value care option in 42% of scenarios. Participants exposed to the default option nudge were 44% less likely to choose at least one low-value care option (odds ratio [OR] 0.56, 95% CI 0·37 to 0.85; p=0·006) compared to those not exposed. The partition display nudge had no effect on choice of low-value care (OR 1.08, 95% CI 0·72 to 1.64; p=0.7). There was no interaction between the nudges (OR 0.94, 95% CI 0·41 to 2.15; p=0·89).

Interpretation: A default option nudge reduced the odds of choosing low-value options for low back pain in clinical vignettes. Embedding high value options as defaults in clinical decision support tools could improve quality of care. More research is needed into how nudges impact clinical decision-making in different contexts.

Introduction

Low-value care is healthcare that provides little or no benefit, may cause patient harm, or yields marginal benefits at a disproportionately high cost.[1] This is increasingly recognised as a worldwide problem.[2 3]

Current diagnosis and management of low back pain provides compelling examples of low-value care. Recent research shows that clinical mismanagement of low back pain is an international problem, including in low, middle and high income countries.[4] Imaging of acute low back pain, for example, does not sensibly inform management,[5] and is associated with harms,[6] yet a survey of Australian general practitioners (GP) found low back pain imaging was ordered in 25% of consultations.[7] Trends in analgesic medicine prescribing for low back pain are worsening; one Australian study found that between 2004 and 2014, GPs were prescribing increasing amounts of guideline-discordant complex medicines including opioids and fewer guideline-concordant medicines (e.g. non-steroidal anti-inflammatory drugs).[8]

Traditional knowledge translation tools such as clinical education about current guidelines are one means of changing clinical behaviour to reduce low-value care but are estimated to be effective only 10-20% of the time.[9] For example, one trial found targeted clinical education and reminders reduced low back pain imaging[10] but this has never been replicated. More research is needed into alternative approaches.

One alternative approach is the use of 'nudges' which take advantage of the cognitive biases influencing clinicians' habitual behaviour to make it more difficult to choose low-value options while preserving their freedom to choose .[11] Partition display nudges assume that people are less likely to select items listed horizontally than those listed vertically, for example on a computer screen.[12] One recent randomised study found that embedding a partition display nudge (listing low-value medication options horizontally rather than vertically) in a simulated Electronic Health Record reduced the proportion of respondents choosing aggressive antibiotic medications by 11.2% (43.7% in control group vs. 32.5% in partition display nudge group).[12] Default option nudges exploit a well-documented cognitive bias towards preferring the default choice option over those that require additional thought or action.[13] There is evidence from non-randomised and quasi-randomised experiments that positioning the preferred choices as the default options in Electronic Health Record systems can influence decision making about medicines [14-16] and tests.[17 18]

While there has been growing interest in the use of nudges to reduce low-value care,[19-21] few randomised controlled trials have investigated the effectiveness of default option and partition display nudges in a clinical context. We are aware of only one (previously mentioned) randomised controlled trial[12] of partition display nudges and none involving default option nudges. The study of partition display nudges[12] found that the nudges could influence decisions about antibiotic medicines but did not examine effects on test ordering behaviour e.g. ordering diagnostic imaging.

Thus, it remains unclear whether partition display or default option nudges can influence testing and treatment decisions in other common conditions such as low back pain.

We aimed to determine whether the use of partition display and default option nudges in a simulated clinical decision support display can reduce low-value clinical management choices related to imaging and complex medicines for low back pain.

Methods

Study design and participants

This was a double blind (participant and assessor), 2x2 factorial, parallel group randomised trial. The trial was registered at the Australian New Zealand Clinical Trials Registry (ACTRN12618000769280; trial protocol and statistical analysis plan available in full at https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375033&isReview=true. No amendments to study protocol were made after the enrolment of the first participant.

Participants were qualified general practitioners aged 18 years and over located in Australia. They were recruited by a market research company and informed in the recruitment email that they would receive $40 to complete a 15-minute survey. The company relies on an extensive database of general practitioners who are required to opt-in to their database and therefore have already indicated a willingness to participate in online research. Recruitment began on 21 May 2018 and halted on 4 June 2018.

Procedures and intervention

Participants completed four clinical vignettes of patients with low back pain. The vignettes were developed by experienced clinicians and based on current guidance from international clinical guidelines.[22 23] The first two vignettes described a low back pain scenario paired with six imaging options, while the last two vignettes described a low back pain scenario with six medication options (see **Supplementary File 1** for complete vignettes).

For each vignette, participants had to select at least one, and up to three, of the six options provided. Each participant was randomised to one of four experimental conditions in the presentation of these options across all vignettes:

1. All options displayed conventionally, that is, fully visible and listed vertically (**No nudge**)
2. All options visible but three low-value options listed horizontally rather than vertically (**Partition display nudge**)
3. Three guideline-concordant options visible on the initial view, with message stating ‘click to see more options’ below third option. Upon one click, low-value options fully visible and listed vertically (**Default option nudge**)

4. Three guideline-concordant options visible on initial view, with message stating ‘click to see more options’ below third option. Upon one click, low-value options fully visible but listed horizontally (i.e. a Partition display nudge). (**Double nudge** that is, Default option nudge + Partition display nudge).

Figure 3-1 shows the visual appearance of the partition display nudge and default option nudge on a computer screen. See **Supplementary File 2** for computer screenshots showing how the menu options were displayed to participants in each of the four experimental conditions.

The 2x2 factorial design underlying this experiment is summarised in **Supplementary File 3**.

Randomisation and blinding

The survey was hosted on the Qualtrics survey platform. Randomisation of each participant to one of our four experimental conditions was conducted by the Qualtrics Block Randomizer function set to ‘Evenly present elements’. This meant that the randomiser assigned each new participant to those conditions with the lowest display counts using a random number generator. Participants were blinded to the study hypothesis, and outcome assessors were blinded to group allocation until study completion.

Primary and secondary outcome measures

The primary outcome measure was the proportion of scenarios where participants chose at least one of the low-value options.

The secondary outcome measure was the likelihood participants would recommend each of the six options using a Likert scale based on the following: ‘*How likely would you be to recommend each of the below options for this case? (1 = very unlikely and 5 = very likely)’.* This information was collected after participants had made their initial choice of management option/s. Likert score questions were displayed conventionally (i.e. all options were visible and listed vertically) without nudges.

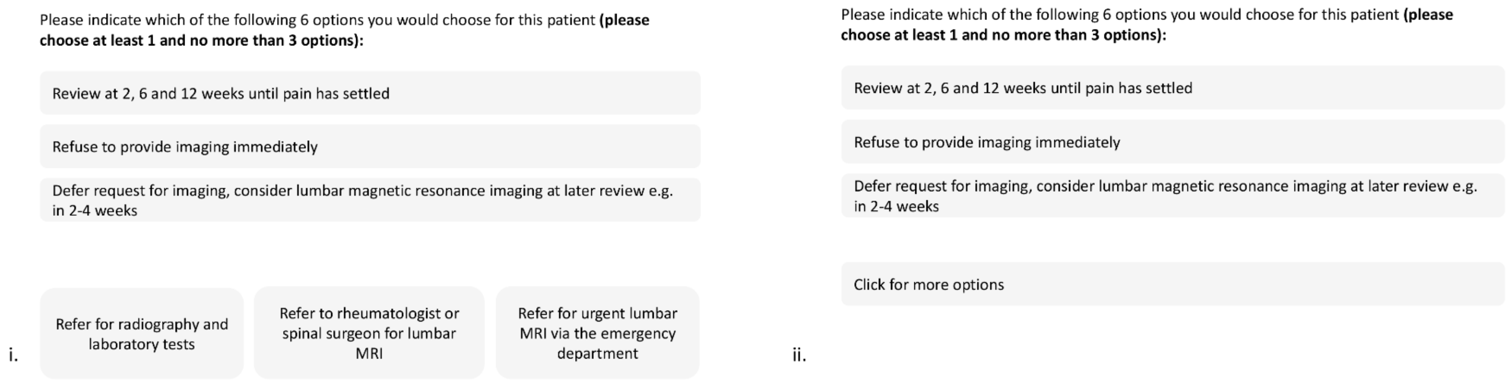

Figure 3-1: Visual appearance of the partition display nudge (i) and default option nudge (ii) on a computer screen.

The partition display nudge (i) displayed the three guideline concordant options first in a vertical list, and then displayed the low value options in a horizontal list. The default option nudge (ii) displayed the three guideline concordant options first in a vertical list, followed by a message stating ‘click to see more options’ below third option. Upon one click, the low-value options were made fully visible in a vertical list.

Other measures and data collected

Information was also collected about participant characteristics, namely sex, practice management software used, years of clinical experience, location of practice and estimated percentage of working hours spent teaching or in academia. This demographic data was collected before participants were allocated to the intervention.

To explore acceptability of the menu formats we planned to examine the intervention on dropout rate; if drop-outs were disproportionately high in one intervention group we felt this could indicate poor acceptability of the menu format.

We also asked participants to complete a Cognitive Reflection Test at the end of the survey. The Cognitive Reflection Test is designed to measure a person's tendency to override their cognitive biases and engage in further reflection to find a correct answer.[24] It is based on three questions (for example, *'A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?'*) and is measured on a 0 to 3 scale (0 = 0 questions answered correctly, 3 = all 3 questions answered correctly). A low test score indicates that a person may be thinking 'fast' (i.e. more susceptible to cognitive biases). The Cognitive Reflection Test has been validated as a measure of 'reflective thinking.' [24 25] We defined a low test score as less than or equal to one (see **Supplementary File 4** for explanation of test and cutoff score).

Sample size

We based our initial sample size estimate on a single vignette per participant. We assumed a conservative baseline rate of 0.45 (i.e., at least one low-value option chosen in 45% of vignettes) in the no nudge group based on a study of general practitioner use of imaging and medicines for low back pain at first consultation.[7] With an allocation proportion of 0.5 to each of the nudges (0.25 to each of the four experimental groups), we estimated a sample of 120 observations would provide 80% power at the 0.025 significance level (to account for multiple comparisons) to detect a difference between no nudge and nudge groups of approximately 25% due to exposure to either nudge. To obtain more reliable estimates and gain additional power to detect smaller effects, each participant was exposed to four vignettes. To account for the likelihood of correlated outcomes from participants completing multiple vignettes, we utilised the sample size formula for clustered binomial outcomes given in Lee and Dubin 1994.[26] Assuming that within-subject responses would follow a beta-binomial distribution (with parameters a=0.5 and b=0.5), with 60 participants per cluster (480 observations in total), we estimated that we could detect between-group differences in the proportion of vignettes where participants choose any low-value option as small as 15% to test the main effect of either of the nudges with more than 80% power and an alpha of 0.025. We undertook a planned, blinded interim analysis after 20 participants had provided complete responses to ensure sample size assumptions were reasonable.

Statistical plan

To determine the main effects of the default option and partition display nudges on the odds of choosing low-value options we fitted a logistic regression model. The binary independent variables were default option nudge (yes, no) and partition display nudge (yes, no). The outcome variable for the primary model was defined as at least one low-value option selected (yes, no). We specified 'at least one low-value option selected' (no) as the reference category.

For our secondary outcome measure (the Likert score assigned to each option), to estimate the effect of the intervention on the likelihood of recommending low-value options we fitted a linear regression model. The binary independent variables were the same as those for the primary outcome analysis above, but the outcome variable was the mean of the Likert scores assigned to low-value options (continuous scale).

For both these models (for the primary and secondary outcomes), we used generalised estimating equations (GEE) model to account for repeated (correlated) observations across vignettes. To assess the robustness of our results we ran these models as a univariable regression (i.e. with the default option nudge as the sole independent variable, omitting the partition display nudge, and vice versa) and as a multivariable regression (with both nudges included in the model). We also checked for the presence of interaction effects between nudges. Our exploratory hypothesis was that if there was an interaction effect present, it would be likely to be positive (i.e. combining the two nudges would be super-additive) rather than negative.

All descriptive statistics for the statistical analysis were undertaken in Excel 2016 and all regression models were undertaken in SPSS Statistics Version 24.

Results

We invited 183 GPs to participate. A total of 120 GPs agreed to participate in the trial and provided complete outcome responses (n=480 scenarios), a response rate of 65.6%. The full participant flow is summarised in **Figure 3-2**. Demographic characteristics of the participants are summarised in **Table 3-1**. Participants were predominantly male, practised in urban areas and spent less than 15% of working hours teaching or in academia. The majority had practised clinically for more than 20 years. None of the participants had fewer than 5 years' experience. Participants mostly reported using MedicalDirector and Best Practice software. The majority had a high Cognitive Reflection Test score (i.e. they were thinking 'slow') meaning they were theoretically more likely to override their cognitive biases

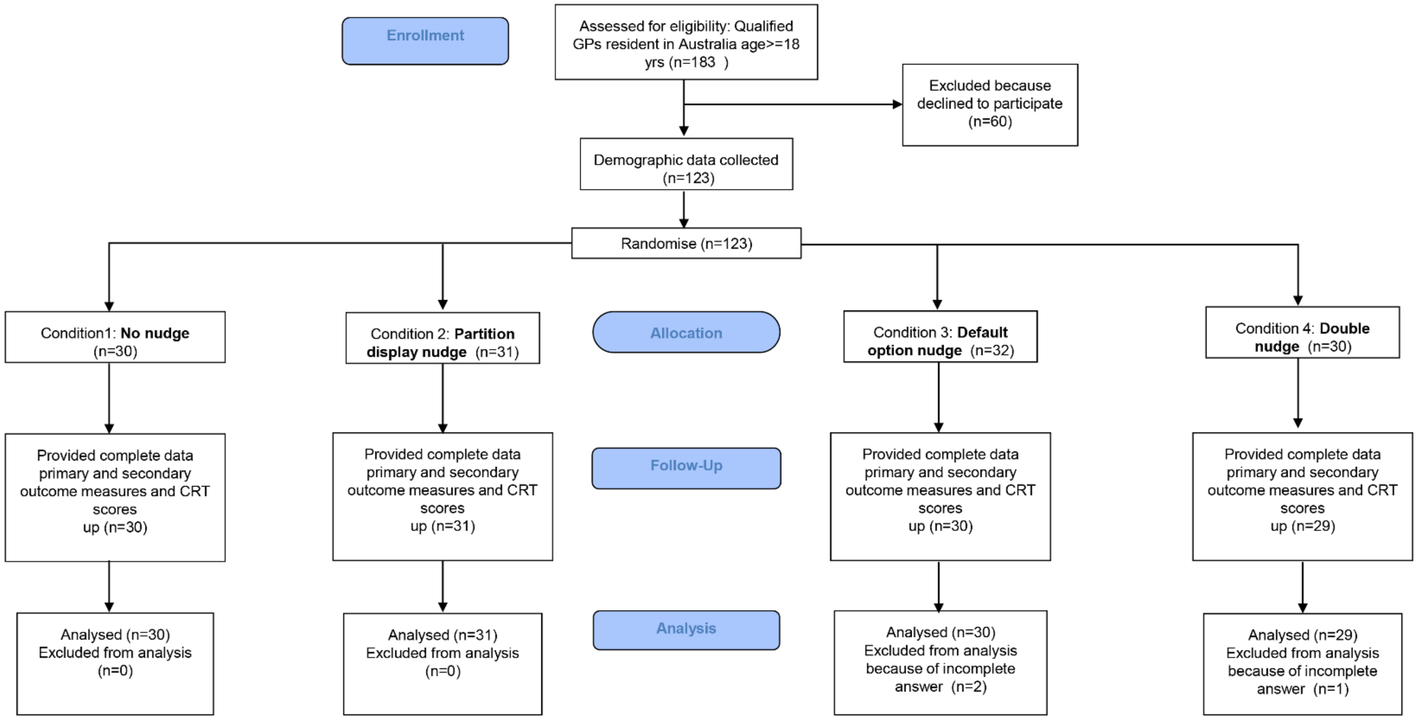

Figure 3-2: Participant study flow

Table 3-1.Characteristics of participants

	All participants (n=120)	No nudge (n=30)	Partition display nudge (n=31)	Default option nudge (n=30))	Double nudge (n=29)
Gender					
Male	86 (71.7%)	21 (70%)	21 (67.7%)	25 (83.3%)	19 (65.5%)
Female	34(28.3%)	9 (30%)	10 (32.3%)	5 (16.7%)	10 (34.5%)
Practice Management Software					
Medical Director	55 (45.8%)	12 (40%)	16 (51.6%)	15 (50%)	12 (41.4%)
Best Practice	52 (43.3%)	15 (50%)	11 (35.5%)	10 (33.3%)	16 (55.2%)
Other	13 (10.8%)	3 (10%)	4 (12.9%)	5 (16.7%)	1 (3.4%)
Years of clinical practice					
<5	0	0	0	0	0
5 to 10	4 (3.3%)	1 (3.3%)	0	1 (3.3%)	2 (6.9%)
11 to 15	22 (18.3%)	11 (36.7%)	3 (9.7%)	5 (16.7%)	3 (10.3%)
16 to 20	16 (13.3%)	3 (10.0%)	5 (16.1%)	4 (13.3%)	4 (13.8%)
>20	78 (65%)	15 (50.0%)	23 (74.2%)	20 (66.7%)	20 (69.0%)
Location of clinical practice					
Urban	107 (89.2%)	28 (93.3%)	28 (90.3%)	27 (90.0%)	24 (82.8%)
Rural	12 (10%)	1 (3.3%)	3 (9.7%)	3 (10.0%)	5 (17.2%)
Remote	1 (0.8%)	1 (3.3%)	0	0	0
% of working hours teaching/academia					
0 to 15%	102 (85%)	26 (86.7%)	27 (87.1%)	28 (93.3%)	21 (72.4%)
16 to 25%	5 (4.2%)	1 (3.3%)	1 (3.2%)	1 (3.3%)	2 (6.9%)
26 to 50%	0	0	0	0	0
51 to 75%	0	0	0	0	0
More than 75%	13 (10.8%)	3 (10.0%)	3 (9.7%)	1 (3.3%)	6 (20.7%)
Cognitive Reflection Test Scores					
Mean Cognitive Reflection Test score (0 to 3)	1.8	1.7	1.7	2.0	1.8
High Cognitive Reflection Test score (%)	76 (63.3%)	16 (53.3%)	18 (58.1%)	23 (76.7%)	19 (65.5%)
Low Cognitive Reflection Test score (%)	44 (36.7%)	14 (46.7%)	13 (41.9%)	7 (23.3%)	10 (34.5%)

Primary outcome measure

Participants in the conventional display (no nudge) condition chose at least one low-value option in 42.5% of scenarios. The proportion of scenarios where participants chose at least one low-value option was lower in the default option nudge and double nudge conditions than for those allocated to the no nudge and partition display nudge conditions (**Figure 3-3**). The proportion of scenarios where participants chose at least one low-value option for the default option and double nudge (which included the default option nudge) condition groups were similar (30% for the default option nudge and 31% for the double nudge). The default option nudge effects appeared to apply to both imaging and medication vignettes (**Supplementary file 5**).

Our multivariable logistic regression model estimated that the partition display nudge had no effect on choice of low-value care (odds ratio [OR] 1.08, 95% CI 0·72 to 1.64; p=0.7). Participants exposed to the default option nudge condition were 44% less likely to choose at least one low-value option (odds ratio [OR] 0.56, 95% CI 0·37 to 0.85; p=0·006) compared to those not exposed to the default option nudge. Results were robust to whether the effects of nudges were tested in univariable (single nudge) or multivariable (account for both nudges) models. There was no statistically significant interaction effect (odds ratio [OR] 0.94, 95% CI 0·41 to 2.15; p=0·89).

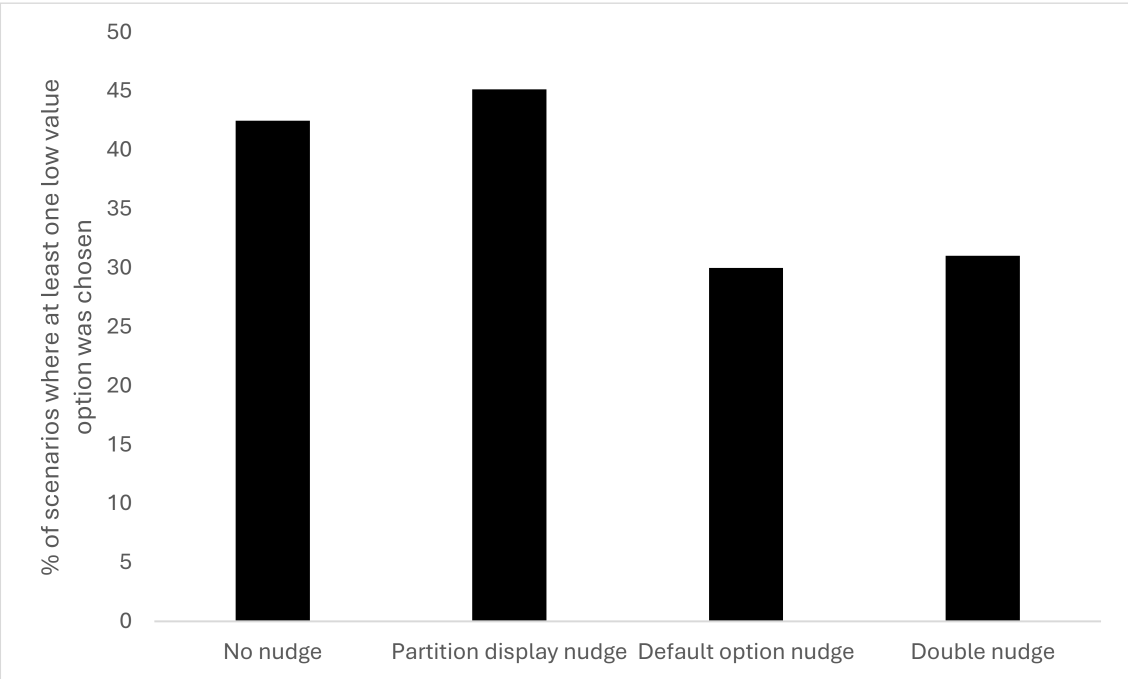

Figure 3-3. Proportion of scenarios where at least one low-value option was selected

Secondary outcome measure

Mean Likert scores (*1 = very unlikely to recommend and 5 = very likely to recommend*) for the three low-value choices were 2.27 (95% CI 2.07 to 2.47) for participants in the partition display nudge condition and 2.17 (95% CI 1.92 to 2.42) for participants in the default option nudge condition compared to 2.34 (95% CI 2.18 to 2.5) for participants in the no nudge condition. Mean Likert scores for the three guideline concordant choices were 2.78 (95% CI 2.56 to 3.01) for participants in the partition display nudge condition and 3.01 (95% 2.84 to 3.18) for participants in the default option nudge condition compared to 2.72 (95% CI 2.56 to 2.87) for participants in the no nudge condition. Our linear regression model did not find any statistically significant effect of the partition display nudge (-0.12, 95% CI -0.36 to 0.12; p=0·33) or the default option nudge (-0.23, 95% CI -0.47 to 0.02; p=0·07) on the secondary outcome measure. Results were robust to whether the effects of nudges were examined in univariable (single nudge) or multivariable (account for both nudges) models. There was no statistically significant interaction effect (-0.11, 95% CI -0.6 to 0.37; p=0·65).

Other exploratory hypotheses

Because post-randomisation drop-outs were negligible (there were a total of three incompletely answered surveys) we did not proceed with the planned analysis of the effect of the interventions on drop-out rate.

We hypothesised that participants with low Cognitive Reflection Test scores (i.e. those who were thinking ‘fast’) would be more susceptible to nudges than participants with high test scores. **Table 3-2** shows that in those allocated to a default option condition, the effect of the default option nudge appeared to be higher for those with low Cognitive Reflection Test scores (in 25% of scenarios at least one low-value option was chosen) compared to those with a high Cognitive Reflection Test score (in 32.7% of scenarios at least one low-value option was chosen). The majority (71.6%) of participants exposed to the default option nudge condition ‘clicked through’ to see the non-default options in all vignettes.

Table 3-2. Cognitive characteristics of participants selecting low-value options in default option nudge vs no default option nudge conditions

	Default option nudge (n=59)	**No default option nudge (n=61)**
	% of scenarios where low-value options selected	% of scenarios where low-value options selected
All participants	30.5	43.9
Participants with low Cognitive Reflection Test score (thinking 'fast')	25	44.4
Participants with high Cognitive Reflection Test score (thinking 'slow')	32.7	43.4

Discussion

Embedding a default option nudge in a simulated electronic decision support display–where high-value options were presented as the default options and low-value options were revealed only after clicking for more–reduced the odds of GPs selecting low-value care by 44%. Embedding a partition display nudge had no effect on the odds of choosing low-value options. Combining default option and partition display nudges in the decision support had no additional effect on the odds of choosing low-value options.

Comparison to other research findings

Three non-randomised studies[14-16] have found that use of a default option nudge in Electronic Health Record systems where generic medications were presented as default options increased the odds these medications would be prescribed in outpatient and primary care clinical settings. One non-randomised study and a quasi-randomised study found that embedding default option nudges in electronic order sets could influence test ordering behaviour. Munigala et al 2017[17] found that removing urine tests other than the default test from the 'frequently ordered' window of the emergency department electronic order set decreased ordering of urine cultures. Leis et al 2017[18] found that whether or not a test was included as part of a standard admission order set strongly influenced whether the test was ordered at all even if there was no explicit prohibition on ordering it.. Our study adds important data to this literature; it is the only randomised controlled trial that we are aware of to study the use of default option nudges in a clinical context, although through the use of vignettes. It provides further robust evidence of the impact of default option nudges on prescribing of medicines and suggests these effects could extend to test ordering behaviour.

Importantly, our study failed to replicate the findings of Tannenbaum et al 2015[12] who found partition display nudges could reduce low value clinical decision making about antibiotics.[12] We were surprised by this finding, given the similarities between our methodology and that of Tannenbaum et al 2015.[12] Both studies used a randomised factorial design, and vignettes where physicians had to decide between aggressive/complex medicines and simple medicines.

However, there were also important differences between our study and Tannenbaum et al 2015's.[12] Their study was focused on decisions about antibiotics whereas ours were about management of low back pain. It remains possible that partition display nudges are more effective for some clinical management decisions than others. For example, these nudges may be more effective when tested in contexts where providers do not hold strong preconceived notions about the importance of a given test or treatment. In Tannenbaum et al 2015's study[12], where the focus was on use of antibiotics, providers may have had fewer preconceived notions regarding specific antibiotic types. Clinicians assessing patients with low back pain, on the other hand, may believe imaging is a valuable test, and be committed to a treatment plan that includes diagnostic workup. This could make them less susceptible to the effect of partition display nudges when provided with vignettes about low back pain. Curiously the physician sample in Tannenbaum et al 2015's study were mostly academic. In our study 85% spent less than 15% of working hours teaching or in academia. It is unclear whether time spent in academia could influence susceptibility to nudges. A final difference between the two studies is that the low-value options grouped by Tannenbaum et al 2015 under a partition display nudge formed clearly discrete clinical categories e.g. broad spectrum (as opposed to narrow spectrum) antibiotics and prescription (as opposed to over the counter) drugs. By contrast, the low-value options grouped under the partition display nudges in our study comprised a more heterogenous set of choices.

Strengths, weaknesses and clinical implications

Key strengths of our study are the randomised, factorial design to provide evidence on the effects of two nudges. We pre-registered our trial protocol and conducted all procedures and analyses as planned. Our analysis was adequately powered and based on 480 patient scenarios. We used planned statistical techniques to ensure that repeated measures from the same participant were accounted for. Our vignettes were developed by experienced clinicians to ensure that our scenarios and management options questions were realistic. The management options provided in the vignettes reflected the current best evidence and clinical practice pathways in our geographical area.[22,23,27] Another strength is the application of nudges to diagnostic imaging decisions. Given the cost of these tests and the prevalence of low back pain, using default option nudges to reduce unnecessary imaging has the potential for substantial cost savings if shown to be effective in a real-world setting.

This study provides insights on the importance of the 'timing' of nudges and their interaction with the psychology of practitioners being nudged. In particular, we found that (1) nudges did not have an effect on measures of low-value decision making collected shortly after exposure; (2) participants with a low Cognitive Reflection Test score (who therefore think 'fast') may be more susceptible to a default option nudge. These findings suggest that nudges may be most effective in situations where thinking 'slow' is difficult or discouraged. They also raise the question of when it is most effective to introduce a nudge during a GP consultation with patients, as there is a possibility that the nudge may wear off after a lengthy conversation with the patient.

This study also has limitations. It is reasonable to ask whether findings from hypothetical vignettes can be fully indicative of clinicians' behaviour in a 'real world' setting. Nonetheless, vignette-based studies are widely used and accepted in research into medical choices.[28 29] There may be important differences between our sample and the broader Australian GP population e.g. our sample had a lower share of early career physicians compared to the Australian sample. [30] However, it is unclear whether this means the nudges would be more or less effective in a more representative GP sample.

An important limitation of the current study is that it does not directly reflect current GP interaction with existing Electronic Health Record systems. For example, current practice management software systems use drop down menus and predictive texts to locate tests and treatments. However, as software systems and decision support tools evolve, there may be scope for our results to inform how management choices are presented on screen. Default option nudges are not necessarily limited to the specific format or clinical management choices we used in this study and could be implemented in new ways as these systems and tools evolve.

Moreover, while it may not necessarily reflect current workflow we believe there are important reasons to consider the results of nudging studies such as this one. First, despite being a potentially powerful tool for enhancing the safety, quality and effectiveness of healthcare without compromising clinician autonomy,[19 21] clinician nudges are still relatively understudied. Second, there is uncertainty around the specific contexts in which these nudges are most effective and the myriad factors which may modify these effects.[31 32] We believe the evidence base for quality improvement interventions would benefit from an inventory of well-researched examples of nudging in different contexts. These would include the target (e.g. medications, pathology tests, imaging tests), desired actions (avoiding something unnecessary versus doing something recommended), the urgency of the clinical situation, the degree to which patients are involved in decision-making, and so on. The insights that such a research programme can provide into nudge principles will benefit the design of future clinical decision aids, such as Electronic Health Record systems, electronic order sets and hand-held apps.

Conclusion

Clinical mismanagement of low back pain is an international problem. Traditional education-based approaches to quality improvement have had limited effectiveness in changing clinician behaviour. Our study suggests that embedding a default option nudge into clinical decision support tools can reduce the choice of low-value care options for low back pain. This has the potential to significantly improve patient care while preserving autonomous clinical decision-making.

References

1. Scott IA , Duckett SJ . In search of professional consensus in defining and reducing low-value care. *Med J Aust* 2015;203:179–81.
2. Levinson W , Kallewaard M , Bhatia RS , et al . 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf* 2015;24:167–74.
3. Mafi JN , Parchman M . Low-value care: an intractable global problem with no quick fix. *BMJ Qual Saf* 2018;27:333–6.
4. Foster NE , Anema JR , Cherkin D , et al . Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet* 2018;391:2368–83.
5. Webster BS , Bauer AZ , Choi Y , et al . Iatrogenic consequences of early magnetic resonance imaging in acute, work-related, disabling low back pain. *Spine* 2013;38:1939–46.
6. Jarvik JG , Hollingworth W , Martin B , et al . Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *JAMA* 2003;289:2810–8.
7. Williams CM , Maher CG , Hancock MJ , et al . Low back pain and best practice care: a survey of general practice physicians. *Arch Intern Med* 2010;170:271–7.
8. Mathieson S , Valenti L , Maher CG , et al . Worsening trends in analgesics recommended for spinal pain in primary care. *Eur Spine J* 2018;27:1136–45.
9. Grimshaw J , Eccles M , Tetroe J . Implementing clinical guidelines: current evidence and future implications. *J Contin Educ Health Prof* 2004;24(Suppl 1):S31–7.
10. Eccles M , Steen N , Grimshaw J , et al . Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet* 2001;357:1406–9.
11. Scott IA , Soon J , Elshaug AG , et al . Countering cognitive biases in minimising low value care. *Med J Aust* 2017;206:407–11.

12. Tannenbaum D , Doctor JN , Persell SD , et al . Nudging physician prescription decisions by partitioning the order set: results of a vignette-based study. *J Gen Intern Med* 2015;30:298–304.
13. Abadie A , Gay S . The impact of presumed consent legislation on cadaveric organ donation: a cross-country study. *J Health Econ* 2006;25:599–620.
14. Patel MS , Day SC , Halpern SD , et al . Generic medication prescription rates after health system-wide redesign of default options within the electronic health record. *JAMA Intern Med* 2016;176:8478.
15. Patel MS , Day S , Small DS , et al . Using default options within the electronic health record to increase the prescribing of generic-equivalent medications: a quasi-experimental study. *Ann Intern Med* 2014;161(10 Suppl):S44–52.
16. Malhotra S , Cheriff AD , Gossey JT , et al . Effects of an e-Prescribing interface redesign on rates of generic drug prescribing: exploiting default options. J *Am Med Inform Assoc* 2016;23:891–8.
17. Munigala S , Jackups RR , Poirier RF , et al . Impact of order set design on urine culturing practices at an academic medical centre emergency department. *BMJ Qual Saf* 2018;27:587–92.
18. Leis B , Frost A , Bryce R , et al . Standard admission order sets promote ordering of unnecessary investigations: a quasi-randomised evaluation in a simulated setting. *BMJ Qual Saf* 2017;26:938–40.
19. Patel MS , Volpp KG , Asch DA . Nudge units to improve the delivery of health care. *N Engl J Med* 2018;378:214–6.
20. Meeker D , Linder JA , Fox CR , et al . Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. *JAMA* 2016;315:562–70.
21. Vaughn VM , Linder JA . Thoughtless design of the electronic health record drives overuse, but purposeful design can nudge improved patient care. *BMJ Qual Saf* 2018;27:583–6.
22. Traeger A , Buchbinder R , Harris I , et al . Diagnosis and management of low-back pain in primary care. *CMAJ* 2017;189:E1386–95.
23. Chou R , Qaseem A , Owens DK , et al . Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. *Ann Intern Med* 2011;154:181–9.
24. Pennycook G , Cheyne JA , Koehler DJ , et al . Is the cognitive reflection test a measure of both reflection and intuition? *Behav Res Methods* 2016;48:341–8.
25. Frederick S . Cognitive reflection and decision making. *Journal of Economic Perspectives* 2005;19:25–42.
26. Lee EW , Dubin N . Estimation and sample size considerations for clustered binary responses. *Stat Med* 1994;13:1241–52.
27. Mid North Coast and Northern NSW Local Health District. Health pathways mid and North Coast of NSW 2018. 2018. Available: https://manc.healthpathways.org.au/index.htm [Accessed 28 July 2018].

28. Jiwa M , Gordon M , Arnet H , et al . Referring patients to specialists: a structured vignette survey of Australian and British GPs. *BMC Fam Pract* 2008;9:2.
29. Peabody JW , Luck J , Glassman P , et al . Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715–22.
30. Britt HMG , Bayram C , Henderson J . A decade of Australian general practice activity 2006–07 to 2015–16. General practice series no. 41. Sydney: Sydney University Press, 2016.
31. Mongin P , Cozic M . Rethinking nudge: not one but three concepts. *Behav Public Policy* 2018;2:107–24.
32. Hertwig R . When to consider boosting: some rules for policy-makers. *Behav Public Policy* 2017;1:143–61.

Footnotes

Contributors JS, AT and CB conceived the idea. All authors contributed to refinement of the study design. JS, AT, CB and EC acquired the data and performed the statistical analysis. All authors interpreted the data. JS wrote the initial draft of the paper. All authors revised the paper critically for important intellectual content and approved the final version.

Funding This work was supported by the National Health and Medical Research Council (NHMRC) ‘Wiser Healthcare’ Research Program Grant (APP1113532). AT was supported by an NHMRC Early Career Fellowship (APP1126082). CB was supported by National Heart Foundation of Australia Vanguard Grant 101326. CM was supported by an NHMRC Principal Research Fellowship (APP 1103022). AGE holds a HCF Research Foundation Professorial Fellowship.

Competing interests None declared.

Patient consent Not required.

Ethics approval Royal Prince Alfred Hospital Ethics Review Committee (Approval number HREC/18/RPAH/138).

Provenance and peer review Not commissioned; externally peer reviewed.

Chapter 4 CT and MRI substitution in presence of licencing constraints?

The material contained in this chapter will be submitted for consideration for publication in Journal of Medical Imaging and Oncology as: **Soon J**, Andrews M, Badgery-Parker T, Elshaug AG, Howard K. CT and MRI substitution in presence of licencing constraints? 2024

Full author list

Jason Soon[1], Matthew Andrews[2], Tim Badgery-Parker[3], Adam G Elshaug[4 1], Kirsten Howard[1]

1. Menzies Centre for Health Policy and Economics, School of Public Health, Charles Perkins Centre, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia.
2. I-MED Radiology Network, Cabrini Hospital, Brighton, Victoria, Australia.
3. Australian Institute of Heath Innovation, Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, New South Wales, Australia.
4. Centre for Health Policy, Melbourne School of Population and Global Health and the Melbourne Medical School, University of Melbourne, Vic., Australia.

Abstract

Introduction: This study investigated whether constraints on magnetic resonance imaging (MRI) supply in Australia may lead to overuse of computed tomography (CT) imaging. The analysis focused on Medicare Benefits Schedule (MBS) funded MRI and CT head scans, examining potential substitution effects as practitioners adapt to resource constraints.

Methods: MBS-funded CT and MRI head scan data for 2018-2019 were analysed across Australian Statistical Area Level 4 regions. Primary models used 2019 data to examine relationships between imaging utilisation, MRI density, healthcare provider density, and regional median income. Supplementary fixed-effects models used 2018-2019 data to control for unmeasured regional characteristics.

Results: Primary models revealed indirect evidence of CT substitution due to limited MRI availability, with higher MRI density regions showing a lower CT share of total imaging. MRI utilization positively correlated with MRI density for both general practitioners (GPs) and non-GP specialists. GP-requested CT utilisation and CT share of imaging negatively associated with regional median income. Supplementary models confirmed the negative relationship between income and GP-requested CT imaging.

Conclusions: The findings suggest potential reallocation of imaging from CT to MRI with increased MRI availability. The most robust finding across both model types was the negative relationship between income and GP-requested CT imaging, even after accounting for fixed effects. This may indicate that patients in lower-income regions are referred for CT imaging instead of MRI due to higher out-of-pocket expenses. These results highlight that more equitable access to imaging facilities can promote appropriate imaging use.

Introduction and context

Medical overuse, when healthcare services are provided in excess of what is necessary or beneficial, can lead to unnecessary costs, and inefficiencies within the healthcare system.[1] One concept invoked to partially explain medical overuse, supplier-induced demand (SID)[2] is when providers are incentivised to suggest or recommend more tests or interventions that they 'supply' in order to increase their revenue, irrespective of their clinical utility. This phenomenon is exacerbated under fee-for-service payment models which reimburse providers based on the volume of services they deliver.

However, there are other incentives in healthcare which cause overservicing including ironically, providers' attempts to adapt to changes in incentives aimed at reducing overservicing. For example, Yip 1998[3] found that cuts in US Medicare benefits for coronary artery bypass grafting (CABG) induced physicians to undertake higher volumes of CABGs to maintain a given income target. Similarly, Shahinian et al 2017[4] found that a decline in reimbursement rates for intensity modulated radiation therapy (IMRT) in the US coincided with increasing rates of use for prostate cancer in physician owned office settings. This is thought likely to occur where providers attempt to sustain revenues through increased service volumes to recoup capital investment costs in the face of declining Medicare reimbursement.

This study aimed to investigate whether a potentially perverse incentive to overuse a specific form of computed tomography (CT) imaging may be induced by constraints on the supply of magnetic resonance imaging (MRI) in Australia. This is a different kind of induced overuse to that reported in Yip 1998[3] and Shahinian et al 2017[4] as this overuse is hypothesised to be the result of adaptation to inadequate clinical resources rather than from any pecuniary maximisation by medical practitioners.

Specifically, the claim being investigated, which has been made by some medical groups,[5] is that CT imaging may be undertaken as a substitute for MRI in regions with limited availability of MRI machines. This is because MRI machines eligible for the purpose of claiming patient rebates under the Australian Medicare Benefits Schedule (MBS) previously operated under a restrictive licensing system that is not in place for CT machines (until recently announced reforms which were rolled out from 1 November 2022, initially in rural, regional and remote areas[6] but ultimately extending to all MRI machines by 2027[7]).

We investigated this hypothesis by testing two relationships: a relationship between above average availability of MBS-eligible MRI machines and below average annual utilisation of CT imaging; and a relationship between above average availability of MBS-eligible MRI machines and below average share of CT imaging as a share of total imaging.

The analysis focussed on different types of MBS funded MRI and CT imaging scans of the head and brain identified as most likely to be substitutable for each other.

These imaging items were therefore amenable to analysis of utilisation by measuring total MBS claims made for each of these modalities (CT imaging and MRI) by location, as requested by two key types of medical practitioners: general practitioners (GPs) and non-GP specialists. Models were developed to explain the relationship between population-adjusted utilisation of each of these modalities, stratified by requester (i.e. non-GP specialist requested MRI, non-GP specialist requested CT imaging, GP requested MRI and GP requested CT imaging) and selected variables to test the hypothesis of substitution between CT imaging and MRI.

Methods

Data sources

Data on the number of CT imaging and MRI head scans reported by location of the rendering provider and stratified by requesting provider (GP or non-GP specialist) for the calendar years 2018-19 was sourced as a customised data request from Services Australia. This analysis covered all the MRI head scan items used by GPs and non-GP specialists but only a selection of CT head scan items which were considered substitutable with these MRI head scan items (see the Supplement for more details) The location of the rendering provider was by Australian Statistical Area Level 4 (SA4), the largest sub-state regions of the Australian Statistical Geography Standards (ASGS). Data specific to 2019 was extracted and adjusted for population based on Australian Bureau of Statistics (ABS) data to calculate the number of head scans per 100,000 people (henceforth the 'population-adjusted ' measure). This was further stratified by each modality and requesting provider.

Data on the location of MBS-eligible MRI machines in each SA4 region in Australia was sourced from an Australian Department of Health website (accessed in 2020)[8] which also identified whether the MRI machine was fully or partially MBS eligible. By consulting the ABS map of statistical boundaries,[9] the location of MRI machines was matched to specific SA4 regions so that the number of MRI machines in each SA4 region (adjusted by population) – henceforth known as 'MRI density'- could be estimated. Only MRI density as of 2019 and not from previous years was estimated as the website data collected had been updated to reflect the latest MRI distributions at the time it was accessed.

Data on the number of full-time equivalent (FTE) GPs and non-GP specialists (defined as the total number of physicians and paediatricians in private practice settings) in each SA4 region was sourced from the National Health Workforce Dataset[10] for 2019 based on total number of hours worked per week in private practice. These figures were also adjusted for population to produce measures of GP and non-GP specialist 'density'.

Data was sourced from the Australian Bureau of Statistics on population[11] and median income[12] of each SA4 region in the years 2018 and 2019. This income data was collected to account for income as a potential explanatory variable because there might be a relationship between income and imaging utilisation (based on out-of-pocket affordability and willingness to pay). Other data on specific clinical characteristics of patients that might be predictive of demand for imaging in each SA4 region was not collected for the model.

Fitting the primary models from the data collected

Primary models of best fit to explain the population-adjusted number of GP requested MRI head scans, GP requested CT imaging head scans, non-GP specialist requested MRI head scans and non-GP specialist requested CT imaging head scans in each region were developed using the data described above. Models of best fit to explain the share of total population-adjusted head scans which are CT scans stratified by provider (non-specialist GP requested and GP requested) were also developed.

Only data for 2019 was used in the models because of the restriction on data on MRI machine density noted previously. Relevant explanatory variables fitted to each of these models were MRI density (as a measure of the accessibility of MRI machines), GP and non-GP specialist density (as a measure of accessibility of requesting practitioners) and regional median income. Each model was based on a total of 87 observations (the number of SA4 regions in the data).

All measures for dependent variables and explanatory variables of interest were converted into ‘z scores’ representing the number of standard deviations above or below the mean (of all SA4 regions) for each data point. This approach also means that the data for dependent variables which is otherwise count data (and therefore non-negative) is converted into something which can take on negative values which means it can be more appropriately modelled using simple linear regression. A simple linear regression model was fitted for each outcome of interest using an iterative process where the final model chosen was the one with the greatest explanatory power (in terms of the highest adjusted R-squared) but only including explanatory variables which had statistically significant coefficients. All analyses were conducted in SPSS (version 29).

Note that under this approach, the coefficients for the explanatory variables will estimate how much above the average imaging utilisation is predicted to be if the explanatory variable is one standard deviation above average.

A summary of the data and related minimum and maximum ‘z scores’ is provided in **Table 4-1**.

Table 4-1: Summary data

	2018				2019			
			z-score				z-score	
Variable	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum
Non-GP specialist MRI scans per 100,00	1183.89	1228.25	-0.96	4.97	1285.49	1214.26	-1.06	4.92
GP MRI scans per 100,000	516.97	279.50	-1.85	2.59	530.96	279.40	-1.90	2.67
Non-GP specialist CT scans per 100,000	408.94	362.65	-1.08	3.34	412.08	366.53	-1.07	3.29
GP CT scans per 100,000	975.77	318.81	-2.35	2.33	978.53	311.33	-2.62	2.43
CT share of total non-GP specialist imaging scans per 100,000	N/A	N/A	N/A	N/A	0.29	0.21	-0.98	3.36
CT share of total GP imaging scans per 100,000	N/A	N/A	N/A	N/A	0.66	0.14	-2.14	2.34
Non-GP specialist density	14.7	16.39	-0.90	5.46	14.64	15.34	-0.95	5.11
GP density	77.18	15.49	-1.75	3.77	76.72	15.36	-2.15	3.84
Radiologist density	4.12	4.42	-0.93	5.74	4.05	4.8	-0.84	6.49
MRI machine density								
Fully licensed	NA	NA	NA	NA	1.01	0.79	-1.28	4.22
Fully and partially licensed	NA	NA	NA	NA	1.48	0.93	-1.58	4.84
Median regional income	49,981	6,033	-1.90	3.08	51,653	6,202	-1.89	3.03

Supplementary analysis

A supplementary model was also developed using an additional year of data from 2018, doubling the number of observations to 174 SA4 regions. This analysis aimed to control for unmeasured variables that may account for regional variations in imaging requirements (e.g., prevalence of specific health risks associated with the head or brain). This was done by fitting dummy variables into the model representing the 'fixed effect' of each region, accounting for unobserved, time-invariant factors that might influence imaging utilization.

This approach required adaptation because data on MRI machines per population prior to 2019 was unavailable (for reasons discussed). Instead, we used data on the number of diagnostic radiologists in private practice for each region in 2018 and 2019, obtained from Health Workforce Australia,[10] as a proxy for general (MBS claimable) imaging capacity and availability. Because of this measure of imaging capacity not being modality specific but equally applicable to both, our supplementary modelling was only undertaken for the four models of CT and MRI utilisation (by requester) but not the two models relating to the CT **share** of imaging.

As in our primary models, all measures were converted to z-scores, and we followed a similar iterative process, incorporating the fixed effects approach to control for region-specific characteristics.

Results

Primary models

Table 4-2 summarises our results from the primary models for GP and non-GP specialist requested imaging, providing information on both the full model (fitted with all explanatory variables) and the final model (as per the process explained in the methodology section). Discussion of the results here refers to the final rather than the full model.

As per the unit of measurement being z-scores, Model 1 (GP requested MRI) estimated that after controlling for GP density, every standard deviation above the mean increase in MRI density is associated with GP requested MRI utilisation being 0.48 standard deviations above the mean; and after controlling for MRI density, every standard deviation above the mean increase in GP density is associated with MRI utilisation being 0.21 standard deviations above the mean.

Model 2 (GP requested CT), unlike the other models found that only median income (not requester density or MRI density) had a statistically significant impact on utilisation. Every standard deviation above the mean increase in median income is associated with CT imaging utilisation being 0.56 standard deviations **below** the mean. None of the other variables are statistically significant (and therefore are not controlled for in the final model).

Model 3 (for non-GP specialist requested MRI) estimated that after controlling for requester density, every standard deviation above the mean increase in MRI machine density is associated with MRI utilisation being 0.32 standard deviations above the mean; and after controlling for MRI density, every one standard deviation above requester density is associated with MRI utilisation being 0.61 standard deviations above the mean.

Model 4 (non-GP specialist requested CT) found that there is only a statistically significant relationship between the level of CT imaging utilisation and requester density.

Models 5 and 6 are different from the other models because the dependent variables are CT utilisation as a share of total imaging utilisation. Model 5 estimated that after controlling for income, every standard deviation above the mean increase in MRI density is associated with the CT share of total imaging requested by GPs being 0.53 standard deviations below the mean; and after controlling for MRI density, every standard deviation above the mean increase in income is associated with this share being 0.20 standard deviations below the mean.

Model 6 estimated that every standard deviation above the mean increase in MRI density is associated with the CT share of total imaging requested by specialists being 0.43 standard deviations below the mean. The impact after controlling for other explanatory variables was not reported because in the final model results none of the other variables are statistically significant.

In all the models, the final results omit at least one of the explanatory variables because it is not significant, and this is usually income. In almost all cases the omission leads to a model of equal or greater explanatory power (as measured by the adjusted R squared), the minor exception being model 2 where the full model has an adjusted R squared of 0.31 compared to 0.3.

Supplementary model

The results of our supplementary model are summarised in **Table 4-3**

Unlike our primary models, our supplementary models did not find statistically significant relationships between any explanatory variables and utilisation of imaging modalities with the exception of GP requested CT imaging – hence only the results of a fitted model for GP requested CT imaging are reported.

This model also estimated that every standard deviation increase in radiologist density was associated with GP requested CT utilisation being 0.12 standard deviations above the mean; and every standard deviation increase in regional median income was associated with GP requested CT utilisation being 0.82 standard deviations below the mean.

Table 4-2: Regression model results for 2019 data (n=87) for GP (general practitioner) and non-GP specialist requested imaging.

	1. GP requested MRI		2. GP requested CT		3. Non-GP specialist requested MRI		4. Non-GP specialist requested CT		5. GP requested CT share		6. Non-GP specialist requested CT share	
Parameter	**Full model**	**Final model**	**Full model**	**Final model**	**Full model**	**Final model**	**Full model**	**Final model**	**Full model**	**Final model**	**Full model**	**Final model**
Adjusted R-squared	0.35	0.36	0.31	0.3	0.65	0.65	0.52	0.53	0.34	0.34	0.17	0.18
GP density	0.23 [0.02,0.44] (p=0.032)	0.21 [0.01, 0.40] (p=0.043)	0.13 [-0.08, 0.35] (p=0.22)	N/A	N/A	N/A	N/A	N/A	-0.1 [-0.31,0.11] (0.361)	N/A	N/A	N/A
Specialist density	N/A	N/A	N/A	N/A	0.65 [0.49,0.81] (p<0.001)	0.61 [0.47, 0.75] (p<0.001)	0.76 [0.57, 0.95] (p<0.001)	0.73 [0.59,0.88] (p<0.001)	N/A	N/A	0.07 [0.17,0.32] (p=0.556)	N/A
MRI density	0.48 [0.28,0.68] (p<0.001)	0.48 [0.28,0.678 (p<0.001)	0.05 [-0.16, 0.25] (0.666)	N/A	0.3 [0.15,0.45] (p<0.001)	0.32 [0.18, 0.46] (p<0.001)	0 [-0.17, 0.17] (p=0.999)	N/A	-0.48 [-0.68,-0.29] (p<0.001)	-0.53 [-0.70,0.36] (p<0.001)	-0.459 [-0.69,-0.23] (p<0.001)	-0.43 [-0.62,-0.24] (p<0.001)
Median income	-0.07 [-0.25.0.11] p=(0.448)	N/A	-0.61 [-0.79,-0.42] (p<0.001)	-0.56 [-0.73,-0.38] (p<0.001)	-0.08 [-0.22,0.07] (p=0.308)	N/A	-0.070 [-0.24,0.1] (p=0.403)	N/A	-0.18 [-0.36,0.00] (p=0.057)	-0.20 [-0.38, -0.03] (p=0.025)	0.07 [-0.15,0.29] (p=0.552)	N/A

*Final model refers to model with highest adjusted R-squared where all included covariates are statistically significant. All figures rounded to 2 decimal places except p-values

Table 4-3: Regression model results for 2018 data (n=174) for GP (general practitioner) requested CT imaging

	GP requested CT	
Parameter	**Full model**	**Final model**
Adjusted R-squared	0.98	0.98
GP density	-0.01 [-0.14,0.13] (p=0.93)	N/A
Radiologist density	0.13 [0.02,0.24] p=(0.03)	0.12 [0.02,0.23] (p=0.027)
Median income	-0.82 [-1.4, -0.25] (p=0.006)	-0.82 [-1.4, -0.25] (p=0.006)

*Final model refers to model with highest adjusted R-squared where all included covariates are statistically significant. Covariates representing regional dummy variables have been omitted. Models results for other GP requested and non-GP specialist requested imaging have not been reported as none of these models reported statistically significant results for the non-dummy variables. All figures rounded to 2 decimal places except p-values

Discussion

Primary model results – MRI machine density

Our primary models reveal indirect evidence of practitioner substitution between MRI and CT imaging in response to MRI availability constraints. While we found no direct relationship between MRI density and CT utilization (in models 2 and 4), the significant negative association between MRI density and CT's share of total imaging for both GPs and non-GP specialists (in models 5 and 6) suggests a substitution effect. This indicates that practitioners may be adapting their imaging choices based on available resources, potentially leading to suboptimal use of CT in regions with limited MRI access.

Two potential explanations exist for this phenomenon. The first is a 'health equity effect' where regions with higher MRI availability are meeting a clinically required level of imaging that is not achieved in areas with lower availability. The second is a 'commission bias' where medical practitioners in regions with greater MRI availability may opt for MR imaging out of preference, regardless of its appropriateness. The extent to which each of these factors contributes to the reallocation of imaging shares remains uncertain.

Similar relationships (and questions) also emerge from models 1 and 3 which show that levels of MRI utilisation are higher in regions with high MRI machine density. These tendencies may also be explained by some possible combination of health equity effects and commission bias. Interestingly this relationship (between MRI utilisation levels and MRI density) is slightly larger for GP than non-GP specialist requesters (0.48 vs 0.32). This is possibly due to non-GP specialists having a narrower range of clinically appropriate imaging options by the time patients reach them e.g. they are likely already referring all the patients they need to for MRIs so that a further increase in MRI availability will have less impact on their decision-making.

Primary model results – medical practitioner density

Non-GP specialist density positively correlates with both CT and MRI utilization in models 3 and 4, as expected because regions with greater access to non-GP specialists would be expected to have greater investigations coverage of their populations' medical issues.

Neither GP nor non-GP specialist density significantly affects the CT share of total imaging, indicating no strong preference for either modality at the population level. However, it is worth noting that some GP-requested imaging may be influenced by specialist recommendations made in preparation for consultations.

In model 1, GP density shows a positive relationship with MRI utilization albeit a weaker one than the relationship between non-GP specialist density and MRI utilisation (coefficient 0.21 vs 0.61 for specialists). This smaller effect likely reflects GPs' limited access to MBS-funded MRI items. Surprisingly, GP density does not significantly impact GP-requested CT imaging, possibly due to other stronger explanatory variables dominating this relationship (discussed in the next sub-section).

As with the case for non-GP specialists, a higher GP density would not be expected to be associated with a statistically significant higher (or lower) CT share of imaging in model 6 unless there were a strong GP preference for one over the other modality. In this case, although there may be a preference it may be swamped by the impact of other determinants (discussed below).

Primary and supplementary model results on income

The relationship between regional median income and imaging utilization emerges as a critical finding across our models. Notably, lower income is associated with higher GP-requested CT imaging (in model 2) and a larger CT share of total imaging (in model 5). This robust relationship persists even in our supplementary fixed-effects model, suggesting a potential socioeconomic driver in imaging choices.

Two key factors may explain this income effect. Firstly, GPs, as primary and frequent patient contacts, may prioritise minimising out-of-pocket expenses when ordering investigations compared to non-GP specialists due to their recognising the burden of (and potentially responding to patients' reluctance to incur) substantial amounts of such costs. This could lead to preferring CT imaging scans over more expensive MRI scans. Secondly, GPs may use imaging in initial patient workups, where CT's lower cost and higher availability make it a more accessible option. By contrast, non-GP specialists manage proportionately more complex patients and by the time a patient has been referred to a non-GP specialist, the most appropriate and accurate form of imaging may already have been determined to be MRI.

The supplementary model, while not replicating most of the primary model relationships, reinforces the income- and GP requested CT utilization link. One reason for the failure to replicate some of the results of the primary models is that the supplementary modelling approach introduces regional dummy variables, potentially controlling for previously omitted variables. This can cause some explanatory variables to lose statistical significance if they were already correlated with these omitted variables (e.g. GP and non-GP specialist density may partly reflect the decisions of these medical service providers to relocate to regions where there is more demand for their services). In addition, private radiologist density was being used as an imperfect proxy for MRI machine density in this modelling approach but may be too crude a proxy since it captures the installed base for all imaging in a region. However, any relationships that remain statistically significant in this model are likely to be robust.

The persistent negative relationship between income and GP-requested CT imaging, even after accounting for fixed effects, appears to be the most robust finding from our investigations across both primary and supplementary models. This relationship may indicate inappropriate substitution, with patients in lower-income regions being referred for CT imaging instead of MRI due to higher out-of-pocket expenses for the latter. Further investigation is warranted, particularly given recent expansions in rural MRI access first announced in the Commonwealth Budget 2022-23.

The supplementary model also reveals a positive association between radiologist density and GP-requested CT imaging, mirroring the MRI density effects in the primary models, supporting the notion that imaging choices are influenced by available capacity. A similar relationship for GP-requested MRI may be too weak to detect in the two-year fixed effects model, possibly due to GPs' limited discretion in requesting MRIs and the model controlling for other region-specific variations in imaging utilization.

Conclusions

This study investigated the hypothesis that constraints on MRI supply may lead to an overuse of CT in imaging investigations by modelling the determinants of MBS claims for MRI and CT head imaging. While we did not find evidence of a statistically significant negative relationship between MRI machine density and CT imaging utilisation by either GPs or non-GP specialists, there is some indirect evidence of this hypothesis in the form of a statistically significant negative relationship between MRI machine density and the CT imaging share of total imaging utilisation. The latter finding suggests that if there were a greater availability of MRI machines, there might be some reallocation of existing imaging investigation from CT imaging to MRI.

Our investigations uncovered other relationships of potential interest.

Firstly, there was a statistically significant positive relationship between MRI utilisation (whether ordered by non-GP specialists or GPs) and MRI density. This relationship may also reflect 'health equity effects' due to medical practitioners in regions with higher MRI density (and therefore availability) being more encouraged and able to order MRIs for investigations.

Secondly, there was a statistically significant negative relationship between GP requested CT utilisation (and share of CT utilisation by GPs but not specialists) and median regional income. This may also indicate inappropriate substitution, with patients in lower-income regions being referred for CT imaging instead of MRI by GPs due to higher out-of-pocket expenses for the latter even though there may also be other factors influencing test choice such as patient age and degree of clinical suspicion.

The study's strengths include its comprehensive analysis of relationships between imaging utilisation, MRI availability, practitioner availability, and income. While one limitation was the lack of control for regional variations in genuine clinical imaging requirements in the primary models, we undertook supplementary modelling to address this.

In supplementary modelling using a fixed effects model to control for any unmeasured differences in clinical characteristics of each region's population and using two years of data, while most of the statistically significant relationships identified in our primary models were not replicated, the finding of a negative relationship between GP requested CT utilisation and median regional income remained and a positive statistically significant relationship between GP requested CT utilisation and private radiologist density was also identified. The latter, because it used an explanatory variable which was a proxy for installed imaging capacity may also reflect the 'health equity effects' postulated previously.

Given recent expansions in rural MRI access announced in the Commonwealth Budget 2022-23 which meant that from 1 November 2022, any MRI equipment located at accredited comprehensive practices in Modified Monash (MM) 2-7 areas (i.e. regional, rural and remote areas) will be able to provide Medicare eligible MRI services, it will be of interest to monitor whether these patterns persist in the coming years.

As this study focused on MBS-funded imaging, it does not account for or test for any possible imaging modality substitutions within the public hospital system.

Another limitation of this study is that our analysis cannot definitively establish the extent to which the annual utilisation of CT imaging analysed in this study represents a "second-best" option because we did not directly quantify the share of referred imaging that could have been appropriately addressed by CT imaging. To do so would have required either collecting detailed clinical records on each of the referred imaging requests in our dataset (which is not plausible for an administrative dataset of this size) or making more precise assumptions about the share of such imaging that could appropriately have been undertaken by CT (which is essentially just a difference of degree from the approach we have used instead which is to assume that some share of CT imaging recorded may be inappropriate).

Instead our analysis was conducted on the presumption that some share of CT imaging captured in our dataset is likely a suboptimal substitute for MRI given (i) clinical opinion that MRI is the preferred imaging modality (over CT imaging) for many neurological conditions due to superior soft tissue contrast and the absence of ionizing radiation and (ii) this is logically borne out by the statistically significant negative relationships observed between CT utilization and both median regional income and MRI density which suggest that resource constraints were a key factor influencing imaging decisions.

References

1. Brownlee S, Chalkidou K, Doust J, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390(10090):156-168.
2. Richardson JR, Peacock SJ. Supplier-induced demand: reconsidering the theories and new Australian evidence. *Appl Health Econ Health Policy*. 2006;5(2):87-98.
3. Yip WC. Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors. *J Health Econ*. 1998;17(6):675-99.
4. Shahinian VB, Kaufman SR, Yan P, et al. Reimbursement and use of intensity-modulated radiation therapy for prostate cancer. *Medicine (Baltimore*). 2017;96(25):e6929.
5. Royal Australian and New Zealand College of Radiologists. *RANZCR Response to Inquiry into Availability and accessibility of diagnostic imaging equipment around Australia*; 2017. Available from https://www.aph.gov.au/DocumentStore.ashx?id=cf7e0cdf-d086-41e7-9812-527e530c4235&subId=560700 [cited 25 September 2024].
6. Department of Health and Aged Care. *Changes to eligibility for magnetic resonance imaging (MRI) equipment from 1 November 2022.* Available from https://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/94D430931943391CCA2588D40001F4B9/$File/PDF%20Changes%20to%20eligibility%20for%20MRI%20equipment%20from%201%20November%202022.pdf

7. Department of Health and Aged Care. *Budget 2024–25: Reforming Diagnostic Imaging*. Available from https://www.health.gov.au/sites/default/files/2024-05/budget-2024-25-reforming-diagnostic-imaging.pdf [cited 25 September 2024].

8. Department of Health and Aged Care. *MRI and PET unit locations in Australia.* Available from https://www.health.gov.au/topics/diagnostic-imaging/MRI-and-PET-locations [cited 25 September 2024].

9. Australian Bureau of Statistics. *ABS Maps*. Available from https://maps.abs.gov.au/ [cited 25 September 2024].

10. Health Workforce Data Tool. Available from https://hwd.health.gov.au/datatool/ [cited 25 September 2024].

11. Australian Bureau of Statistics. Population estimates by SA2 and above, 2001 to 2021 in *Regional population, 2020-21.*

12. Australian Bureau of Statistics. Table 1. Total income, earners and summary statistics by geography, 2015-16 to 2019-20 in *Personal Income in Australia.*

Chapter 5 The financial implications of investigating false-positive and true-positive mammograms in a national breast cancer screening program

Chapter 5 of this thesis has been published as **Soon J,** Houssami N, Clemson M, Lockie D, Farber R, Barratt A, Elshaug A, Howard K. The financial implications of investigating false-positive and true-positive mammograms in a national breast cancer screening program. *Aust Health Rev*. 2023;47(2):159-164. doi: 10.1071/AH22120. PMID: 36476540

References to tables and figures have been amended to fit this document's labelling conventions

Full author list (institutional affiliations were during time of publication

Jason Soon[A], Nehmat Houssami[B],[C], Michelle Clemson[D], Darren Lockie[D], Rachel Farber[B], Alexandra Barratt[B], Adam G Elshaug[A,E], Kirsten Howard[A]

[A]Menzies Centre for Health Policy and Economics, School of Public Health, Faculty of Medicine and Health, University of Sydney, NSW, Australia.

[B] School of Public Health, Faculty of Medicine & Health, University of Sydney, NSW, Australia.

[C]Daffodil Centre, The University of Sydney – joint venture with Cancer Council NSW, NSW, Australia.

[D]Maroondah BreastScreen, Eastern Health, Vic., Australia.

[E] Centre for Health Policy, Melbourne School of Population and Global Health and the Melbourne Medical School, University of Melbourne, Vic., Australia.

Abstract

Objectives To determine the total annual screening and further-investigation costs of investigating false-positive and true-positive mammograms in the Australian population breast-screening program.

Methods This economic analysis used aggregate-level retrospective cohort data of women screened at a breast-screening clinic. Counts and frequencies of each diagnostic workup-sequence recorded were scaled up to national figures and costed by estimating per-patient costs of procedures using screening clinic cost data. Main outcomes and measures estimated were percentage share of total annual screening and further-investigation costs for the Australian population breast-screening program of investigating false-positive and true-positive mammograms. Secondary outcomes determined were average costs of investigating each false-positive and true-positive mammogram. Sensitivity analyses involved recalculating results excluding subgroups of patients below and above the screening age range of 50–74 years.

Results Of 8235 patients, the median age was 60.35 years with interquartile range of 54.17–67.17 years. A total of 15.4% (ranging from 13.4 to 15.4% under different scenarios) of total annual screening and further-investigation costs were from investigating false-positive mammograms. This exceeded the share of costs from investigating true-positives (13%).

Conclusions We have developed a transparent and non-onerous approach for estimating the costs of false-positive and true-positive mammograms associated with the national breast-screening program. While determining an optimal balance between false-positives and true-positive rates must rely primarily on health outcomes, costs are an important consideration. We recommend that future research adopts and refines similar approaches to facilitate better monitoring of these costs, benchmark against estimates from other screening programs, and support optimal policy development.

Keywords: breast cancer, breast neoplasms, BreastScreen Australia, cancer screening, cost of healthcare, digital breast tomosynthesis, false-positive results, health services, mammography, public health, women’s health.

Introduction

Screening mammography aims to identify women more likely to have breast cancer, and therefore requiring diagnostic testing to exclude or confirm malignancy. Its main benefit is early detection to reduce breast cancer mortality through earlier treatment with the additional benefit of reduced treatment intensity for screen-detected cancers.[1,2] However there are also risks associated with screening such as the risk of false-positive mammograms, causing women to be recalled for further investigation. False-positive mammograms may increase anxiety among women with abnormal findings, even after cancer is later ruled out.[3-5] Breast screening policies must strike a balance between reducing the number of women recalled who do not have cancer (false-positives) and the risk of missing (true-positive) cancers. While the goal of breast screening is to extend life and lessen morbidity, program costs and cost-effectiveness remain important considerations. Quantitative estimates of the costs associated with each of these risks therefore can help decision-makers better evaluate these trade-offs and maximise the health value of the program.

Evidence suggests that false-positive rates of screening mammography vary considerably between countries. For instance, a study of European breast-screening programmes estimated false-positive rates of 5.4% for initial screens and a cumulative false-positive risk of 8%-21% after 10 screens[6] compared to recent estimates of a false-positive rate of around 11%[7] and cumulative false-positive risk after 10 screens of almost 50% for US breast-screening.[8]

While there are studies investigating the costs of false-positive mammography investigations, few have reported total national costs of breast-screening programmes or the cost shares of these investigations. One Swedish study[9] estimated the variable costs per-person of false-positive mammography investigations at US$664-$973 (as estimated by Chubak[10] based on prevailing exchange rates). US estimates range from US$527 per-person in a 2010 study[10] to US$852 in a 2015 study in by Ong et al,[11] who also estimated an annual national cost of false-positive mammograms of US$2.8 billion.[11]

This paper addresses these gaps in the Australian literature by estimating the respective shares of the total annual costs of screening and further investigation (hereafter referred to as assessment) within Australia's national breast-screening programme from assessment costs for false-positive and true-positive mammograms. These assessment costs arise from additional tests women undergo after being recalled for a positive mammogram, and may include clinical breast examination with palpation, diagnostic mammography, ultrasound, percutaneous needle biopsy and open biopsy.

Methods

Setting and study design

The Australian national government funded breast-screening program, BreastScreen Australia, invites women aged 50–74 for screening mammograms every two years, and will also screen women aged 40–49 and 75 and over on request. BreastScreen Australia screens approximately 900,000 women per annum. Victoria is one of eight states/territories of Australia, representing approximately 26% of the Australian total population.[12] Aggregate level retrospective cohort de-identified data on clients screened by a BreastScreen Australia screening clinic in Victoria were collected and analysed. This clinic is one of five fixed screening sites located in a metropolitan Reading and Assessment Service. Across the State of Victoria there are eight such Services. The catchment area for this Service undertook approximately one-sixth of the screenings in Victoria in 2016 while this clinic screened over one-fifth of the clients in this service in that same year.

This facility provides mammogram screening using two view, full-field digital mammography (with digital breast tomosynthesis mammography reserved for assessment). All images are double-read independently by radiologists. Each client is screened and then leaves immediately. Results are provided after her mammogram is reported, within 14 days.[13] If the screen is reported as abnormal, the client is 'recalled to assessment' (i.e. further investigation). If normal, she is asked to return for her next screen in 2 years' time (routine recall). Clients recalled to assessment are recorded as either having breast cancer detected, or as routine recall if cancer is not detected.

Services provided through BreastScreen are free to women at the point of service. We only report costs borne by the public health system (i.e. BreastScreen with the exception of diagnostic open biopsies which are assumed to be borne by public hospitals – see **Supplement, Section 1**). We did not consider out-of-pocket costs from private screening which may account for approximately 26% of total breast-screening in Australia (though this may be an overestimate – see **Supplement, Section 2** for methodology and caveats).

Population and subgroups

Participants were all clients screened by the service between 1 January 2016 and 31 December 2016 excluding:

- Women with a history of breast cancer, so that only those undergoing routine screening were included and those returning for follow-up after previously diagnosed cancer were excluded, similar to the exclusion adopted by Ong and Mandl (2015).[11]

- Clients recalled to assessment but recorded as ‘incomplete assessment’ or ‘early review at assessment’ without record of final results, who comprised only 0.04% of the 2016 dataset.

Data were analysed in two client groups - those screened for the first time in 2016, and those undergoing a second or subsequent screen in 2016. This ensured that the different diagnostic frequency patterns of first and subsequent-screen clients were fully captured rather than averaged out.

False-positives were defined as screening results in clients recalled to assessment who ultimately received a ‘routine recall’ recommendation (subject to exceptions detailed in **Supplement, Section 3**). True-positives were defined as screening results in clients recalled to assessment ultimately recorded as having breast cancer detected.

Estimating resources and costs

Counts and frequencies of each diagnostic-workup sequence recorded for clients were quantified separately for those undergoing their first screen and those undergoing a subsequent screen.

Per-patient costs of each screening and assessment procedure (and associated administrative, consent and reporting costs) comprising each recorded sequence were estimated. Cost data were provided by the screening clinic which supplied the client data. Full information on component cost estimates is documented in **Supplementary Table 5-2**.

These cost estimates were applied to the frequency analysis, producing an estimate of the total annual costs of breast cancer screening and assessment in 2016 undertaken in the screening clinic. Cost components were partitioned into false-positive and true-positive results to estimate false-positive and true-positive assessment costs as shares of total screening and assessment costs. These reported costs are incremental costs of screening/assessment and exclude provision for fixed costs such as overheads. Estimates were reported in 2020 Australian dollars, scaled up and converted accordingly using Reserve Bank of Australia data on the Australian Consumer Price Index. All analysis was undertaken using Microsoft Excel.

This diagnostic-workup sequence frequency data was extrapolated to 2016 national figures for all national BreastScreen Australia clients by calculating the percentage of clients recalled to assessment in each 5 year age-group that underwent each distinct diagnostic workup sequence; and then applying these age-based percentages to the number of national BreastScreen Australia clients recalled to assessment in 2016 by first and subsequent screening rounds in each age-group. The per-patient costs of each screening and assessment procedure were then applied to this age-based frequency data to derive national cost estimates for 2016. This approach adjusts the frequency data to take account of differences between national data and our screening clinic data which might affect overall false-positive rates (and costs) such as the share of screens which are first round screens, the recall to assessment (RTA) rate and differences in age distribution.

Sensitivity analysis

Results were recalculated excluding patients (i) outside the screening range of 50-74 years of age (ii) below 50 years of age and (iii) above 74 years.

Ethics approval

Ethics approval for use of the data in this study was obtained from Eastern Health (reference number LR19/031).

Results

Table 5-1 reports summary statistics on the study participants Their median age was 60.35 with an interquartile range of 54.17-67.17 years old. The recall to assessment rate for the sample was 3.7%.

Table 5-1:Summary statistics of women screened through BreastScreen service in 2016

Number of clients	8235
Age range	40-90
Median age	60.35
Interquartile range	54.17-67.17
First round in 2016 (%)	13.3
Subsequent screening round in 2016 (%)	86.7
Recall to assessment (%)	3.7

Table 5-2 summarises the results of the diagnostic-workup sequence frequency analysis derived from client records of the screening clinic. The four most frequent diagnostic work-up sequences in descending order were:

- Assessment mammography and ultrasound following screening (assessment mammography uses digital breast tomosynthesis)
- Assessment mammography only following screening
- Assessment mammography, ultrasound, physical examination and core biopsy following screening
- Assessment mammography, physical examination and core biopsy following screening.

Table 5-2:Diagnostic workup sequences of BreastScreen service clients recalled to assessment in 2016

Diagnostic workup sequence	A. % of clients recalled to assessment after 1st screening round (n=93)	B. % of clients recalled to assessment after subsequent screening round (n=201)
1. Screen + M	23 (24.7%)	48 (23.9%)
2. Screen + M+U	33 (35.5%)	74 (36.8%)
3. Screen + M+PE+CB	8 (8.6%)	14 (7.0%)
4. Screen + M+U+PE+CB	18 (19.4%)	45 (22.4%)
5. Screen + M + U + PE + CB + FN	2 (2.2%)	2(1.0%)
6. Other sequences	9 (9.7%)	18 (9.0%)

M= assessment mammography; U = ultrasound; PE = physical examination; CB = core biopsy; FN = fine needle biopsy

A full listing of all sequences is provided **in Supplementary Table 5-3**.

Table 5-3 presents total costs in 2016 incurred for all diagnostics (including routine screening which did not result in the detection of abnormalities) and resulting cost shares in the screening clinic that we used as the basis for national projections. 13% of total screening and assessment costs were incurred from assessment of false-positive mammograms, more than the share of costs of investigating true-positive mammograms (10%).

Table 5-3: Total flow-on diagnostic costs in 2016 of representative BreastScreen facility (AUD2020)

Total screening and assessment costs	$521 097
Screening costs for all patients	$399 864 (76.7%)
Misc. assessment costs	$1121 (0.2%)
False-positive assessment costs	$68 057 (13.1%)
True-positive assessment costs	$52 247 (10%)

Table 5-4 presents total annual costs in 2016 incurred for all national BreastScreen Australia-related diagnostics (including routine screening which did not result in the detection of abnormalities) based on the extrapolated and age adjusted frequency data.

Table 5-4: Total annual flow-on diagnostic costs for the Australian population from the national breast screening program including costs of breast screen-detected abnormalities in the twelve months following a client's mammogram (AUD2020)

Total screening and assessment costs	$**73 876 961**
Screening costs for all patients	$52 932 160 (71.6%)
False-positive assessment costs	$11 369 043 (15.4%)
True-positive assessment costs	$9 575 758 (13%)

Annual screening and assessment costs totalled $73.88 million with 15.4% due to assessments caused by false-positive mammograms, more than the share of costs of investigating true-positive mammograms (13%). The average assessment cost for each false-positive mammogram was $277.80 compared to $861.99 for each true-positive mammogram.

As would be expected, excluding patients outside the invited screening age range resulted in a lower false positive rate of 2.9% and lower cost share of false-positive mammogram investigations of 13.4% as did excluding only those below the screening range (**Table 5-5**). Results were broadly unchanged if only patients above the screening range were excluded (e.g. cost share of 15.3% for false-positive investigations).

Table 5-5: Results of sensitivity analysis

	All clients	50-74 age range	50+ age range	Age 74 and under
False-positive mammogram assessment costs as % of total annual screening assessment costs	15.4	13.4	13.6	15.3
True-positive mammogram assessment costs as % of total annual screening assessment costs	13	12.4	12.7	12.7
False positive rate (%)	3.8	2.9	3.0	3.7
Average cost of false-positive diagnosis (AUD2020)	277.80	297.24	292.43	281.26
Average cost of true-positive diagnosis (AUD2020)	861.99	902.43	885.29	876.61

Discussion

This is the first study to our knowledge to estimate total annual screening and assessment costs of a national breast-screening programme for women aged ≥40years, partitioned into the costs of investigating false-positive and true-positive mammograms and their respective shares. Total costs of investigating false-positive mammograms are approximately $11.37 million or 15.4% of the annual screening and assessment costs of $73.88 million compared to a share of 13% for investigating true-positive mammograms. Our study is the first to base cost estimates on detailed analysis of complete annual data on diagnostic frequency patterns of women, combined with cost estimates from the same screening clinic and extrapolated nationally. This approach may be widely applicable.

The average investigation cost per false-positive mammogram was $277.80 (US$190.65), much lower than estimates from the US and Sweden which (converted to 2020 USD) range from US$623 to US$1600.[9-11] The estimated national false-positive rate of 3.8% is much lower than the 11% estimated for the US.[7]

Saxby et al (2020)[14] using Australian data reported an average investigation cost of $196 per false-positive mammogram and $676 per true-positive mammogram (in 2020 AUD). They did not report cost shares or total screening programme costs. As our estimates include diagnostic open biopsy costs, when these are removed to facilitate comparability, the differences between our estimates are small ($198.70 per false-positive diagnosis and $787 per true-positive diagnosis compared to $196 and $676 respectively for Saxby et al[14]).

Based on 2000 to 2014 data, Australia's false-positive mammography rate is 3.3% for ages 50-69 and 4.8% for ages 40-49[15] which is not directly comparable with our 3.8% estimate though our estimate nonetheless lies within the appropriate range.

Strengths and weaknesses

While there have been numerous studies on the psychological impact of false-positive mammograms, the healthcare costs of investigating false-positive mammograms is understudied. Our study used a novel and transparent approach which may be widely applicable.

The economy of using data from one screening clinic to construct a frequency analysis of screening and diagnosis tests forming the basis for cost-modelling is a strength. A corresponding weakness is the reliance on data from one clinic. However, while it is true that policies and practices of screening clinics through Australia vary and they work to different service delivery and budgetary models, all clinics operating as part of the national breast-screening program follow standardised practice and policies, and undergo nationally-aligned accreditation and quality assurance processes.

While data from multiple screening clinics could produce a more representative frequency analysis, a cross-check of our estimates against estimates from other Australian studies did not find major disparities in terms of average diagnosis costs or false-positive rates, increasing our confidence in our estimates, including that approximately 15% of diagnostic costs are from false-positive screens. Our sensitivity analysis indicates that excluding patients outside screening age reduces this slightly, as would be expected, but the range of variation produced from our sensitivity analysis is narrow (13.4-15.4%).

While not a weakness per se, the scope of our study excludes private assessment costs (including the costs of false positive investigations) outside the national screening program which may account for 26% of total screening (see caveats to this estimate in **Supplement, Section 2**).

Implications for practice and policy

A key contribution of this study is the transparent and non-onerous (in terms of data requirements) approach described to estimate the costs associated with false-positive and true-positive screens. We encourage future researchers to adopt and refine similar approaches using whatever screening clinic data is available so that more national estimates of false-positive and true-positive costs can be developed and compared. We hope that such estimates can facilitate better informed decision-making about the costs and benefits of screening.

However, in undertaking any benchmarking using these methods, we caution against the implication that if false-positive costs associated with one screening programme are high relative to screening in other jurisdictions that this necessarily means that such costs can or should be further reduced. The approach that we have outlined suggests that the costs associated with a given false-positive recall rate will depend on the relative costs and frequencies of the different kinds of diagnostic-workup sequences to investigate these false-positive screens so there may be multidimensional means of managing these costs (for instance through refining diagnostic algorithms). Trying to reduce false-positive costs simply by reducing false-positive risks may have second-order effects by increasing false-negative risks, reducing the cancer-detection rate, resulting in delayed diagnosis and treatment of cancers that were missed. Importantly, we quantify costs of false-positive and true-positive screens in the context of a national program with ongoing monitoring of these outcomes which recognises that ‘an effective breast cancer screening program will limit any unnecessary investigations by minimising the proportion of women recalled for further assessment without affecting the achievement of high breast cancer detection rates’[13].

Conclusion

To our knowledge we have produced the first estimates of the share of annual national screening and assessment costs of Australia’s national breast-cancer screening from assessment of false-positive mammograms. Our 15.4% estimate is higher than that of the share of assessment costs from true-positive mammograms (13%). Excluding patients outside screening age range did not significantly change these results.

We believe that these estimates and the transparent and non-onerous methodology underlying them may be of interest to future researchers who wish to better monitor these costs and benchmark against estimates from other screening programmes (whether for breast or other forms of cancer).

Future studies should extend the analysis to cover patients after they have exited screening and assessment to include treatment following positive diagnosis (including estimates of overdiagnosis costs) as well as to cover patients through the full screening, assessment, and treatment journey outside the national screening program.

References

1. Roder D, Houssami N, Farshid G, et al. Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia. *Breast Cancer Res Treat* 2008;108:409–416.
2. Department of Health and Ageing. BreastScreen Australia Evaluation. Screening monograph no. 4/2009. Canberra: Department of Health and Ageing; 2009.
3. Lerman C, Track B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. *Ann Intern Med* 1991;114:657–661.
4. Gram I, Lund E, Slenker S. Quality of life following a false positive mammogram. *Br J Cancer* 1990;62:1018–1022.
5. Swanson V, McIntosh IB, Power KG, Dobson H. The psychological effects of breast screening in terms of patients' perceived health anxieties. *Br J Clin Pract* 1996;50:129–135.
6. Hofvind S, Ponti A, Patnick J, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen* 2012;19:57–66.
7. Lehman CD, Arao RF, Sprague BL, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology* 2017; 283:49–58.
8. Ho TH, Bissell MCS, Kerlikowske K, et al. Cumulative Probability of False-Positive Results After 10 Years of Screening With Digital Breast Tomosynthesis vs Digital Mammography. *JAMA Netw Open* 2022;5:e222440.
9. Lidbrink E, Elfving J, Frisell J, Jonsson E. Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the Stockholm trial. *BMJ* 1996;312: 273–276.
10. Chubak J, Boudreau DM, Fishman PA, Elmore JG. Cost of breast-related care in the year following false positive screening mammograms. *Med Care* 2010;48:815–820.

11. Ong MS, Mandl KD. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. *Health Aff* 2015; 34:576–583.
12. Australian Bureau of Statistics. 3101.0 National, state and territory population, September 2020, Table 4. 2022. Available at https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release#data-download [accessed 11 May 2021].
13. BreastScreen Australia. National Accreditation Standards, Information for clinical staff. 2019. https://www.health.gov.au/sites/default/files/documents/2019/09/breastscreen-australia-national-accreditation-standards-nas-breastscreen-australia-national-accreditation-standards.pdf [accessed 14 June 2021].
14. Saxby K, Nickson C, Mann GB, et al. The financial impact of a breast cancer detected within and outside of screening: lessons from the Australian Lifepool cohort. *Aust N Z J Public Health* 2020;44:219–226.
15. Australian Institute of Health and Welfare. Analysis of breast cancer outcomes and screening behaviour for BreastScreen Australia. Cancer series no. 113. Cat. no. CAN 118. Canberra: AIHW; 2018.

Data availability. The data that support this study were obtained from BreastScreen Maroondah by permission/licence. Data will be shared upon reasonable request to the corresponding author with permission from BreastScreen Maroondah.

Conflicts of interest. The authors declare that they have no conflicts of interest.

Declaration of funding. Funding for this research was provided by National Health and Medical Research Council (NHMRC)/Centres of Research Excellence GNT1104136.

Acknowledgements. We gratefully acknowledge the contribution of Karinna Saxby and Dennis Petrie who shared the costing assumptions used in their paper Saxby et al.[14]

Chapter 6 Discussion and conclusions

This thesis explores the concept of 'satisficing' in clinical decision-making, where clinicians choose solutions that are "good enough" rather than optimal. It distinguishes between satisficing resulting in the use of heuristics and satisficing as adaptation to external settings such as healthcare system resource constraints.

Chapters 2 and 3 investigate heuristics in medical decision-making, including how the medical decision-maker's heuristic tendencies can be leveraged constructively. Chapter 4 investigates a scenario in which medical decision-makers may be satisficing in their imaging choices due to constraints on MBS-eligible MRI machines. Chapter 5 presents a method for estimating mammogram costs in Australia's breast-screening program, which could help benchmark the impacts of satisficing and other potential inefficiencies in healthcare systems.

Summary of findings, specific strengths and weaknesses and additional considerations for each study

This section will provide a summary of findings and specific strengths and weaknesses of each of these chapters with reference to the implications these findings have for the significance of clinical satisficing as a driver of low-value care and healthcare costs, as well as some additional considerations which could not be explored in the original studies. Later sections will then discuss the specific contributions of each of these studies to a more comprehensive theory of clinical satisficing, directions for further research and potential policy implications and other applications.

The impact of heuristics in medical decision-making

Chapter 2's systematic literature review included 10 studies and identified two types of heuristics: those induced by initial triage information and those induced by clinicians' recent experiences. The first type has an instantaneous impact at the time of exposure, while the second type's impact persists over time. The review found that both types of heuristics have statistically significant impacts on clinical decision-making.

The studies on heuristics prompted by triage information showed statistically significant impacts of heuristic use impacting on patterns of care with an average magnitude larger than impacts of heuristics prompted by previous medical events. However, because there were only two studies from this first group compared with eight from the second it is unclear if these impact differences are representative of the differences between the two heuristic types.

Studies on heuristics prompted by previous medical events revealed significant impacts of heuristic use impacting on patterns of care with the persistence and magnitude of these impacts after exposure varying based on the type of decision-making (in six out of the eight studies the outcomes were measured at multiple timepoints following exposure and there were some time points following exposure when the impacts were no longer statistically significant). The studies in this group were divided into studies of obstetric decision making, testing and investigations and prescribing. Within this group of eight studies, only one study[1] which was about anastomotic leak testing reported no statistically significant results. Obstetric decision-making showed the largest overall impacts of heuristics, while prescribing decisions showed the most persistent effects. However, these reported differences are again based on a small number of studies (there was only one study in the prescribing setting and three studies in the obstetric decision-making setting).

In addition, taking into account physician sample size issues (see the Supplement for discussion of this) it is also unclear to what extent these differences in decision-making type may be representative of true differences. For instance, the high odds ratios for the obstetric studies were from Dan et al 2017[2] and although the study involved 4052 patients it was based on a sample size of 40 physicians. Similarly, Choudry et a 2006[3] (the prescribing study) was based on a sample size of 530 physicians. By contrast Keating et al 2017[4], one of the testing and investigation studies, which reported relatively small odds ratios of 1.02 (1.01-1.03) for the second quarter results which were statistically significant was based on a sample of 30,704 physicians. On the other hand, there may be some plausible reasons to expect different heuristic impacts for different types of medical decision-making (this point is discussed further in the section on implications for future research).

The review found limited direct evidence linking heuristic use to health outcomes, with only two studies attempting such measures. However, it is plausible that significant changes in management based on once-off events may not be optimal.

The strengths of this review include its focus on real-world clinical settings and rigorous definition of heuristic exposures. The approach of converting results to odds ratios for comparability was unique. However, the conversion to odds ratios may have introduced some inadvertent errors, though the overall directionality and statistical significance of the findings remain consistent with the original studies.

Use of nudges to reduce low-value care

Chapter 3 investigated the efficacy of nudges in clinical decision-making, particularly in aligning practices with evidence-based guidelines. The study employed a 2x2 factorial design to examine the effectiveness of different nudge strategies in influencing general practitioners' management of low back pain.

The experiment involved four clinical vignettes, with participants assigned to four groups: one receiving a default nudge, another a partition display nudge, a third experiencing both nudges combined, and a control group receiving no nudge. The nudges were integrated into the presentation of treatment options.

The study found that the default nudge was highly effective in changing clinical management, reducing the odds of selecting low-value care by 44%. However, the partition display nudge had no impact on decision-making, and combining both nudges showed no additional effect from the partition display nudge. The nudges had no impact on an additional secondary outcome measure of low-value decision-making taken shortly after in a separate (non-nudge) format, suggesting that even the default nudge is only effective at the point of decision-making with no long-term educative impact. This variability in effectiveness highlights the importance of specific nudge design in influencing clinical decision-making.

Further analysis in this study also revealed that participants with low Cognitive Reflection Test scores (indicating 'fast,' non-reflective thinking) were more susceptible to nudges. Together, this and the secondary outcome finding implies that nudges may be most effective in high-pressure environments where careful deliberation is challenging.

Nonetheless, these findings support the proposition that nudging can counteract existing heuristic biases and enhance decision-making. Interestingly, the successful nudge is itself based on a commonly observed heuristic, suggesting a strategy of using one heuristic to counteract others.

The study's strengths included its randomized, factorial design, pre-registered trial protocol, adequate power, and rigorous statistical techniques. The vignettes were developed by experienced clinicians to ensure realism and reflect current best practices. However, the use of hypothetical vignettes limits the study's applicability to real-world settings, and the presentation of default options does not directly reflect current GP interactions with Electronic Health Record systems.

Satisficing in the selection of medical imaging modalities

Chapter 4 investigated a form of satisficing by medical practitioners in response to constraints on Australia’s Medicare Benefits Schedule (MBS)-eligible MRI machines. The hypothesis was that limited MRI availability might lead to increased CT imaging orders as a substitute.

The study used z-score measures for all variables to facilitate comparability and allow the use of simple linear regression models. While no relationship was found between MRI availability and the number of per capita CT imaging claims, a negative relationship was evident when considering the share of imaging claims that were CT-based. This relationship held for both GPs and non-GP specialists.

The findings suggest some indirect evidence for the hypothesis of CT imaging serving as a substitute for MRI insofar as regions with higher MRI density have a lower CT share of imaging, meaning that if there were a greater availability of MRI machines, there might be some reallocation of existing imaging investigation from CT imaging to MRI.

The study also found positive relationships between MRI machine availability and MRI utilisation suggesting that availability sustains its own demand. This could reflect either an equity effect (increasing MRI machine availability may lead to increased MRI utilisation because practitioners in regions with previously low MRI machine availability are getting by with less than the clinically appropriate use of MRI) or a 'commission bias' where practitioners in regions with currently greater MRI availability are requesting MRI even in marginal cases simply because it is widely available. Self-referral effects of the kind discussed in Chapter 1 (whether these are due to supplier induced demand or also attributable to satisficing) were not considered as a factor in this case because we had no data on which MRI machines were in practices owned by physicians who may have incentives to self-refer. This point is addressed further in the section on implications for further research.

An interesting additional finding was a negative relationship between GP-requested CT utilization and median regional income, potentially indicating another form of satisficing where patients in lower-income regions are referred for CT imaging instead of MRI due to GPs recognising the burden of (and potentially responding to patients' reluctance to incur) higher out-of-pocket expenses for the latter.

The study's strengths include its comprehensive testing of relationships between imaging utilisation and MRI availability, controlling for practitioner availability and income. Its main weakness was the lack of control for regional variations in genuine clinical imaging requirements in the primary models, although a supplementary model attempted to address this. It is also worth noting that as this study was focused on MBS funded imaging it does not account for, or test for, any possible imaging modality substitutions within the public hospital system. Another limitation of this study is that it did not directly quantify the share of referred imaging that could have been appropriately addressed by CT imaging. However the assumption that some share of CT imaging captured in the dataset analysed was likely inappropriate was based on plausible clinical opinion given the restricted availability of MRI and the use of imaging for head scans which was the focus of the study given that MRI is the preferred imaging modality for many neurological conditions due to superior soft tissue contrast and the absence of ionizing radiation.

The persistence of the negative relationship between income and GP-requested CT imaging, even after accounting for fixed effects in the supplementary model suggests it is a robust finding indicating potential satisficing based on patients' financial capacity.

A potential conceptual weakness of this study's findings and its implications for the broader significance of clinical satisficing that was previously discussed is that it is debatable whether ordering tests or treatments due to a 'convenience factor' (in this case scarcity of MRI equipment) constitutes genuine low-value care. However also as discussed previously, the fact that an alternative configuration of healthcare resources could have led to more evidence-based care and less spending on a particular form of imaging (in this case CT imaging) does arguably constitute low-value care, especially in this instance where the more overused form of imaging may result in the exposure of patients to more ionising radiation with its attendant higher cancer-related risks for both paediatric[5] and adult patients[6]. This interpretation is of course not inconsistent with the view that the clinician responses observed here represent a pragmatic adaptation to constraints given the conflicting needs of evidence-based care and patient affordability, as the purpose of this study is not about assigning blame, but rather highlighting the extent to which resourcing constraints can shape clinical patterns of care.

Quantifying the costs of false positive and true positive investigations in a national breast screening program

While Chapter 5 does not focus exclusively on either the internal (heuristic-driven) or external (adaptation to resource constraints) determinants of low-value care, it looks at a possible specific manifestation of these determinants, namely the rate of false positive rates for breast cancer, and develops cost estimates associated with these false positives, including in the process developing a novel methodology for these estimates, which it is hoped can contribute to a better understanding of how systemic factors impact screening outcomes and their associated costs. This chapter also discusses how and why these rates may differ in different countries, noting that the different magnitudes of false positive rates and false positive costs in different countries may be reflective of their respective systemic propensities to low value care.

Chapter 5 estimated that 15.4% of the total annual screening and further-investigation costs of Australia's national breast screening program were likely from investigating false-positive mammograms. This share exceeded the costs from investigating true-positives (13%).

It is important to note that these results do not necessarily imply inefficiency in the national breast screening program. In fact, when compared to estimates from other countries, Australia may perform relatively well. While there are no direct comparisons of the share of costs attributable to false positives in other countries' breast screening programs, the chapter reports that the estimated investigation cost per false-positive mammogram in Australia (US$190.65) is lower than estimates for the US and Sweden (ranging from US$623 to US$1600). Some of these overseas studies are older, and therefore costs may not be fully representative of contemporary resource use, however, this demonstrates the potential for results like these to be used as a means of international benchmarking.

The lower costs of false positives in Australian breast screening may reflect a combination of factors including some that reduce clinical satisficing in screening programs. These include a more standardised approach to mammograms, including double reading, and the use of specialised public screening centres rather than private settings.

The study's main strength is its novel and transparent approach to estimating these understudied healthcare costs, which may be widely applicable. A weakness is the reliance on data from a single clinic. However, while data from multiple screening clinics could produce a more representative frequency analysis, a cross-check of their estimates against other Australian studies did not find major disparities in average diagnosis costs or false-positive rates. Additionally, sensitivity analyses were reported for each of the key results to address this limitation.

Also worth noting are that these sensitivity analyses confirmed that excluding patients outside the invited screening age range and below the screening range resulted in a lower false positive rate and reduce the cost share of false-positive mammogram investigations. However, only excluding those above the screening age range did not have much impact on the false positive rate or its cost share. While these sensitivity analyses were used for the purpose of cross-checking our estimates, it is suggested that use of such sensitivity analyses as part of the approach demonstrated in this study can have a role to play in testing the implications of changes to specifications of screening programmes.

What this research adds

The research in Chapter 2, by quantifying the impact of medical heuristics using odds ratios, allows for comparisons of heuristic impacts across different heuristic types and medical decision-making scenarios, as well as tracking the persistence of heuristics like the availability heuristic over time as reported by different studies in different clinical settings. This methodology could serve as a foundation for future research to further explore the effects of specific heuristics in various medical settings in future reviews of the literature.

Chapter 3 contributes to the growing body of evidence on nudge strategy effectiveness, specifically in the context of diagnostic imaging decisions. The study's factorial design allows for testing multiple nudges simultaneously. Another noteworthy feature of the Chapter 3 study is the additional nuance added by the findings on secondary and additional outcomes which found that while default nudges were highly effective at the point of exposure, they had no impact on low-value decision-making measures taken shortly after and that participants with low Cognitive Reflection Test scores (indicating 'fast,' non-reflective thinking) were more susceptible to nudges.

Chapter 4 offers a fresh perspective on the causes of medical overuse. Instead of focusing on supplier-induced demand, it explores the hypothesis that overuse of CT imaging results from clinicians adapting to resource constraints, in this case limited access to MRI machines. This is a different kind of induced overuse to that reported in SID studies as the overuse is hypothesised to be the result of adaptation to inadequate clinical resources by clinicians doing the best that they can to meet patient needs rather than from any pecuniary maximisation by medical practitioners. This study also uncovered an unexpected negative relationship between income and CT imaging, suggesting that clinicians may adapt their practices to provide more affordable care for low-income patients.

Lastly, Chapter 5 presents a transparent and efficient methodology for benchmarking screening investigation costs across different jurisdictions. This approach relies on data from a single screening centre, assuming the screening population is sufficiently representative or that appropriate adjustments can be made. This method offers a practical solution for comparing costs across different healthcare systems.

Implications for practice and policy

Chapter 2's systematic review on heuristics in medical decision-making suggests that cognitive shortcuts significantly influence clinical judgments. This has implications for medical education and ongoing professional development. As discussed in that chapter, policymakers and medical educators should consider incorporating specific training on recognising and mitigating the effects of cognitive biases in clinical decision-making. Additionally, healthcare organizations may also benefit from implementing decision support systems that help counteract these heuristics, particularly in high-stakes or time-pressured situations.

This last point is related to the results from Chapter 3 on the effectiveness of nudges in clinical decision-making which have direct implications for the design of clinical guidelines and electronic health record (EHR) systems. The strong effect of default nudges suggests that careful attention should be paid to how options are presented in EHRs and clinical pathways. Policymakers and healthcare IT developers should consider incorporating evidence-based default options that align with best practices. However, the variability in nudge effectiveness identified in the study also implies that any such interventions should be carefully designed and tested before widespread implementation.

Chapter 4's findings on the relationship between MRI availability and CT usage highlight the complex interplay between resource allocation and clinical practice. Policymakers should consider these findings when making decisions about the distribution of imaging resources. The study suggests that increasing the availability of MBS-eligible MRI machines could potentially reduce reliance on CT imaging, which has implications for both healthcare costs and patient radiation exposure. In other words, these findings suggest there are sound reasons based on efficiency and clinical effectiveness to also pay attention to health equity as the lack of health equity (in the sense of lack of availability to medical infrastructure) may lead to satisficing strategies such as the ones investigated in this chapter, leading to low-value care despite the best efforts of medical practitioners. The observed relationship between income and imaging choices suggests a need for policies that address healthcare equity in terms of ensuring that patients do not bear unacceptably high out of pocket costs (which may then also induce medical practitioners to ‘satisfice’ in terms of their clinical management choices).

The results from Chapter 5 on the costs associated with false-positive mammograms in Australia's breast screening programs underline the importance of transparent cost analyses in screening programs, which could inform both policy decisions and public discussions about the value of such programs.

Implications for further research

Chapter 2's systematic review on heuristics in medical decision-making suggests several avenues for further research. One is the need for more studies that directly link these cognitive shortcuts to health outcomes. Alternatively, the results from more systematic reviews of the empirical literature on heuristic impacts could also be used to feed into simulation models to quantify the impacts of these heuristics on a population wide level. For instance, if previous exposure to a patient with condition X predisposes a medical decision-maker to increase their odds of testing for condition X for a limited period of time after exposure, what might this imply in terms of the average rate of misdiagnoses, net changes in health utilities (e.g. due to increased time to correct diagnosis) and healthcare resource costs over a defined period of time?

Additionally, more work is needed to understand how different types of medical decisions (e.g., diagnostic, treatment, prescribing) are affected by various heuristics, and whether these effects vary across medical specialties or healthcare settings. While some of these differences were detected by the findings in Chapter 2 it was unclear, given the small sample sizes of some of these studies, how truly representative these differences are. However, there may be plausible reasons for differences across some decision-making settings which are worth further investigation. For instance:

- Could the persistence of heuristic impacts on decision-making be partly determined by whether the impact involves shifting to a relatively more time and resource intensive form of management? For instance, of the two obstetric studies which reported broadly similar outcomes involving an increased rate of caesareans (Riddell et al 2014[7] and Dan et al 2017[2]) the persistence of the impacts was also broadly similarly short lived (4 weeks/1 month). This could be explained by the fact that caesarean procedures are a relatively more time and resource intensive activity so if such switching is fundamentally not evidence-based but prompted by a heuristic, the heuristic will ultimately have to contend with pressures for the practitioner to revert to more effective or proportionate approaches.
- The variation in the persistence of heuristic impacts on changes in medical management can also be explained by the strength of incentives and ‘cultural’ supports for previous practices. For instance, the small and short-lived reduction in colonoscopies in the general patient population reported in

Keating et al 2017[4] (following an exposure to a patient suffering an adverse event from a colonoscopy) can potentially be explained by two factors. Firstly, there are guidelines in place for public health screening which it will be very difficult for physicians to ignore. Secondly, if there is a culture of defensive medicine in the US and a litigious environment then this is likely to reinforce the impacts of such guidelines in reducing the incentives for the 'exposed' physicians to change their clinical management to reduce the colonoscopy rate from what it previously was despite the impact of the availability heuristic. In weighing up the risk of sending a patient to undergo a colonoscopy that goes awry versus the risk of not detecting a colon cancer (and the likely legal consequences of that), the latter may still ultimately prove compelling for the 'exposed' physician in Keating et al 2017[4], hence the transience of the impact. The stratified results reported by Keating et al 2017[4] are consistent with this story because there are longer-lasting impacts of the exposure on elderly patients (those aged over 75) lasting up to 9 months after the event because for this patient group, any perceived risk of failure to diagnose is perceived to be significantly lower, especially given the additional consideration of overdiagnosis that arises for that age group.

The results from Chapter 3 on nudges in clinical decision-making suggest a rich area for future research. While the study found default nudges to be effective, more research is needed to understand the long-term effects of such interventions. For instance, although we also found that the default nudge was only effective at the point of decision-making and did not have further impacts on decision-making immediately after the display of options in the order set, our study was not conducted in a real clinical setting. This raises the question of whether if a default nudge was embedded in an EHR in a real clinical setting and decision-making was followed up over a longer time period, it would have led to sustained changes in clinical behaviour even outside decision-making through the EHR order set; or conversely might the effect of the default nudge in the EHR order set itself also diminish over time? The methodology used in the study also points the way to trialling a greater number of different nudges in one study giving a larger number of results for analysis.

It was also noted in our study that our finding that the partition display nudge was ineffective in changing clinical decision-making was not consistent with Tannenbaum et al 2015[8] who found that partition display nudges could reduce low value clinical decision making about antibiotics. This was despite the similarities in study design. It was speculated that these different outcomes might be because partition display nudges also varied in terms of their effectiveness in different settings (for instance Tannenbaum et al 2015's study was on use of antibiotics where providers may have had fewer preconceived notions about the use of some antibiotics compared to others). On the other hand, this example may simply be reflective of the general difficulties associated with replicability in behavioural and psychological research.

Subtle differences in sample characteristics, study settings, and contextual factors may contribute to variability in outcomes, even with similar interventions. These observations point to an important area for future research: exploring how the relative effectiveness of specific nudge types, such as partition display nudges, may be influenced by variations in clinical setting, provider characteristics, or decision-making contexts. Such investigations would not only help clarify the mechanisms driving nudge efficacy but also contribute to addressing concerns about replicability and generalisability in behavioural research.

Based on the additional results stratified by CRT scores, future research to investigate how nudges might be personalised based on individual clinician characteristics or practice patterns may also yield additional insights. In addition, given the links between nudges and heuristics, future nudge-based trials could also take their cue from some of the findings in Chapter 2 such as the finding by Keating et al 2017[4] that older physicians (with age as a proxy for experience) were less susceptible to heuristic impacts than younger ones (an outcome which is consistent with other vignette-based studies of heuristics[9 10]). This might also suggest that more experienced or older physicians will be less susceptible to nudge use and is a promising area of further research.

Chapter 4 suggested that given recent expansions in rural MRI access announced in the Australian Commonwealth Budget 2022-23 it will be of interest to monitor whether the relationships identified in this chapter persist in the coming years. This is the most obvious area for future research. There are also additional enhancements to the modelling that was undertaken that could be made to allow for a more comprehensive investigation of the relationships between utilisation of the different imaging modalities and accessibility of those modalities in future research.

At a basic level, we had originally intended to undertake our analysis using more granular Statistical Area Level 3 data rather than Statistical Area Level 4 data but the data custodian Services Australia was unable to provide us with this more granular data based on the specifications requested. In addition, there are other extensions to the modelling which could have been undertaken with more time and access to more comprehensive data. For instance, the modelling described in Chapter 4 could be enhanced by access to data that also includes the location of CT imaging machines, adding greater specificity to the primary and supplementary models discussed by allowing specific testing of the relationship between CT machine availability and CT imaging utilisation.

As foreshadowed previously, the study in Chapter 4 did not include an investigation of possible self-referral impacts but future research could also undertake additional investigation of the ownership of each diagnostic imaging practice in Australia. Based on this, data could be compiled on the ‘density’ of physician-owned diagnostic imaging practices in each region, potentially allowing for additional modelling comparing imaging utilisation in regions with different densities of physician owned diagnostic imaging to test for the significance of higher utilisation from self-referral.

The results from Chapter 5 on the costs of false-positive mammograms in Australia's breast screening program highlight the need for similar cost analyses in other countries and for other screening programs. Future research could focus on developing standardised methodologies for such analyses to facilitate international comparisons. There is also an opportunity to investigate interventions that could reduce false-positive rates (based on identifying the characteristics of the best performing jurisdictions) without compromising the sensitivity of screening programs.

The findings from these chapters also suggest a need for more interdisciplinary research in healthcare. Future studies could benefit from collaborations between clinicians, cognitive psychologists, health economists, and policy researchers to develop a more comprehensive understanding of the complex factors influencing healthcare delivery and outcomes.

Finally, these chapters highlight the potential value of big data and machine learning approaches in healthcare research, particularly in uncovering patterns in clinical decision-making and resource utilisation that may not be evident in smaller-scale studies. Leveraging large-scale healthcare databases could enable the development of more sophisticated decision-support tools that align with the framework of recognizing and mitigating or leveraging heuristics by clinicians. For instance, Chapter 3 discussed the potential for evolving software systems and decision-support tools to implement nudges, such as default options, in EHR systems by carefully designing how management choices are presented on-screen.

Thinking beyond simple nudges, AI algorithms integrated into EHR systems could improve clinical decision-making by identifying subtle patterns, flagging potential errors, and suggesting evidence-based alternatives. In terms of further research into use of heuristics by clinicians, natural language processing (NLP) provides opportunities to mine clinical records to assess the appropriateness of imaging referrals such as those analysed in this thesis.

However, the integration of AI into healthcare and health services research may also introduce some risks that warrant careful consideration. For example, reliance on AI could perpetuate existing biases if models are trained on non-representative datasets or lead to over-reliance on algorithmic outputs at the expense of clinical judgment. Additionally, the "black box" nature of many machine learning models raises concerns about transparency. As healthcare systems increasingly adopt AI-driven tools, future research should investigate their impact on clinical workflows, patient outcomes, and resource utilisation.

Conclusions

This thesis has explored the concept of 'satisficing' in clinical decision-making, examining two manifestations of satisificing, one through use of heuristics and one through substitution of imaging modalities in the presence of constraints on the supply of one modality. Through a series of interconnected studies, we have shed light on various aspects of this phenomenon and its implications for healthcare delivery and policy.

Our systematic review of heuristics in medical decision-making revealed that cognitive shortcuts significantly influence clinical judgments, with effects varying based on the type of heuristic and medical decision. This finding demonstrates the importance of recognising and addressing cognitive biases in clinical practice. The subsequent study on nudges demonstrated that carefully designed interventions, particularly default nudges, can effectively guide clinicians towards more guideline-compliant decisions. However, the variability in nudge effectiveness highlights the need for thoughtful implementation of such strategies.

In exploring satisficing in the presence of resource constraints, we found evidence suggesting that limited availability of MBS-eligible MRI machines may lead to increased use of CT imaging as a substitute. This finding has important implications for resource allocation in healthcare systems and demonstrates the complex relationship between resource availability and clinical practice and the importance of health equity as a determinant of low-value care. Additionally, our analysis of Australia's breast screening program introduced a novel and transparent methodology for estimating screening and investigation costs. This approach, which efficiently leverages data from a single screening centre, provides a practical tool for benchmarking costs across different healthcare jurisdictions which may reflect the systemic impacts of satisficing in those jurisdictions.

Collectively, these studies contribute to a more comprehensive understanding of clinical satisficing and its impact on healthcare delivery. They suggest that both cognitive factors and systemic constraints play crucial roles in shaping clinical decision-making and resource utilization. This research points to several potential avenues for improving healthcare quality and efficiency, including enhanced medical education on cognitive biases, careful design of clinical decision support systems, and more equitable distribution of healthcare resources.

However, this work also highlights the need for further research. Future studies should focus on directly linking heuristics and satisficing behaviours to health outcomes, exploring the long-term effects of nudge interventions, and developing standardized methodologies for cost analyses in screening programs. Additionally, there is a clear need for more interdisciplinary research and international comparisons to fully understand and address the complex factors influencing healthcare delivery.

In conclusion, this thesis demonstrates that understanding and addressing clinical satisficing is crucial for improving healthcare quality, efficiency, and equity.

References

1. Simianu VV, Basu A, Alfonso-Cristancho R, et al. Assessing surgeon behavior change after anastomotic leak in colorectal surgery. *J Surg Res*. 2016;205(2):378-383.
2. Dan O, Hochner-Celnikier D, Solnica A, et al. Association of Catastrophic Neonatal Outcomes With Increased Rate of Subsequent Cesarean Deliveries. *Obstet Gynecol*. 2017;129(4):671-675.
3. Choudhry NK, Anderson GM, Laupacis A, et al. Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: matched pair analysis. *BMJ*. 2006;332(7534):141-5.
4. Keating NL, James O'Malley A, Onnela J-P, et al. Assessing the impact of colonoscopy complications on use of colonoscopy among primary care physicians and other connected physicians: an observational study of older Americans. *BMJ Open 2017*;7:e014239.
5. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet*. 2012 4;380(9840):499-505.
6. Cao CF, Ma KL, Shan H, et al. CT Scans and Cancer Risks: A Systematic Review and Dose-response Meta-analysis. *BMC Cancer*. 2022; 30;22(1):1238.

7. Riddell CA, Kaufman JS, Hutcheon JA, et al. Effect of uterine rupture on a hospital's future rate of vaginal birth after cesarean delivery. *Obstet Gynecol*. 2014;124(6):1175-1181.
8. Tannenbaum D, Doctor JN, Persell SD, et al. Nudging physician prescription decisions by partitioning the order set: results of a vignette-based study. *J Gen Intern Med.* 2015;30(3):298-304.
9. Monteiro S, Sherbino J, Ilgen JS, et al. The effect of prior experience on diagnostic reasoning: exploration of availability bias. *Diagnosis (Berl).* 2020;7(3):265-272.
10. Richie M, Josephson SA. Quantifying Heuristic Bias: Anchoring, Availability, and Representativeness. *Teach Learn Med*. 2018;30(1):67-75.

Supplement for Chapter 2

Database searches

Database: Ovid MEDLINE(R) ALL <1946 to September 01, 2023>

Search Strategy:

1 exp Physicians/ (178475)

2 Doctor*.mp. (153815)

3 (Medical adj1 prac*).mp. (32378)

4 (Medical adj1 prof*).mp. (20933)

5 1 or 2 or 3 or 4 (350130)

6 exp Clinical Decision-Making/ (15817)

7 delayed diagnosis/ or diagnosis, differential/ or diagnostic errors/ or early diagnosis/ or incidental findings/ or overdiagnosis/ (546211)

8 exp Heuristics/ (1072)

9 exp Bias/ (75812)

10 Heuristic*.mp. (15774)

11 (Cognitive adj2 bias).mp. (2220)

12 (Cognitive adj2 error).mp. (223)

13 (Heuristic adj2 bias).mp. (31)

14 (Anchoring adj2 bias).mp. (144)

15 (Anchoring adj2 error).mp. (4)

16 (Availability adj2 heuristic*).mp. (88)

17 (Availability adj2 bias*).mp. (127)

18 (Anchoring adj2 heuristic*).mp. (29)

19 (Premature adj1 closure).mp. (864)

20 Availability adj3error.mp. (0)

21 (Confirmation adj2 heuristic*).mp. (1)

22 (Confirmation adj2 error*).mp. (20)

23 (Confirmation adj2 bias*).mp. (471)

24 (Representativeness adj2 heuristic*).mp. (54)

25 (Representativeness adj2 bias*).mp. (35)

26 (Representativeness adj2 error*).mp. (8)

27 exp Clinical Reasoning/ (1253)

28 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 (652873)

29 testing.mp. (833358)

30 Prescribing.mp. or Drug Prescriptions/ or Inappropriate Prescribing/ or Drug Utilization/ or Practice Patterns, Physicians'/ (147150)

31 (Treatment adj2 decision*).mp. (30680)

32 (Prescribing adj2 rate*).mp. (1733)

33 (Testing adj2 rate*).mp. (3599)

34 (Screening adj2 rate*).mp. (6758)

35 exp "Early Detection of Cancer"/ (37980)

36 Colonoscopy/ (31453)

37 (Delivery adj3 Obstetric*).mp. (34295)

38 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 (1098978)

39 5 and 28 and 38 (1271)

40 limit 39 to (english language and yr="1980 -Current") (1156)

Database: Embase Classic <1947 to 1973> Part 1 of 2, Embase <1974 to 2023 September 01>

Search Strategy:

1 physician/ or andrologist/ or anesthesiologist/ or cardiologist/ or dermatologist/ or diabetologist/ or emergency physician/ or endocrinologist/ or epileptologist/ or female physician/ or foreign physician/ or gastroenterologist/ or general practitioner/ or geriatrician/ or gerontologist/ or gynecologist/ or hematologist/ or hepatologist/ or hospital physician/ or immunologist/ or infectious disease specialist/ or intensivist/ or internist/ or medical geneticist/ or neonatologist/ or nephrologist/ or neurologist/ or obstetrician/ or occupational physician/ or oncologist/ or ophthalmologist/ or orthopedic specialist/ or osteopathic physician/ or otolaryngologist/ or pathologist/ or pediatrician/ or phlebologist/ or physiatrist/ or podiatrist/ or psychiatrist/ or pulmonologist/ or radiologist/ or rheumatologist/ or surgeon/ or urologist/ or vaccinologist/ or venereologist/ (955509)

2 Doctor*.mp. (334569)

3 (Medical adj1 prac*).mp. (122914)

4 (Medical adj1 prof*).mp. (49726)

5 1 or 2 or 3 or 4 (1308822)

6 clinical decision-making.mp. or clinical decision making/ (85654)

7 heuristics/ or Heuristics.mp. (6007)

8 bias.mp. (304058)

9 Heuristic*.mp. (17507)

10 (Cognitive adj2 bias).mp. (4805)

11 (Cognitive adj2 error).mp. (352)

12 (Heuristic adj2 bias).mp. (38)

13 (Heuristic adj2 error).mp. (14)

14 (Anchoring adj2 bias).mp. (381)

15 (Anchoring adj2 error).mp. (10)

16 (Availability adj2 heuristic*).mp. (119)

17 (Availability adj2 bias*).mp. (243)

18 (Anchoring adj2 heuristic*).mp. (56)

19 (Premature adj1 closure).mp. (1444)

20 (Confirmation adj2 heuristic*).mp. (1)

21 (Confirmation adj2 error*).mp. (25)

22 (Confirmation adj2 bias*).mp. (784)

23 (Representativeness adj2 heuristic*).mp. (64)

24 (Representativeness adj2 bias*).mp. (43)

25 (Representativeness adj2 error*).mp. (11)

26 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 (405132)

27 testing.mp. (1151199)

28 prescribing error/ or unnecessary prescribing/ or prescribing.mp. (100701)

29 practice patterns.mp. (17271)

30 (Treatment adj2 decision*).mp. (50452)

31 (Prescribing adj2 rate*).mp. (2665)

32 (Testing adj2 rate*).mp. (5611)

33 (Screening adj2 rate*).mp. (10746)

34 (Delivery adj3 Obstetric*).mp. (23338)

35 Colonoscopy.mp. (110820)

36 cancer screening.mp. (114734)

37 delayed diagnosis.mp. or delayed diagnosis/ (25838)

38 delayed diagnosis.mp. (25838)

39 differential diagnosis.mp. (463657)

40 early diagnosis.mp. (223555)

41 diagnostic error.mp. (72383)

42 overdiagnosis.mp. (6606)

43 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 (2231332)

44 5 and 26 and 43 (4901)

45 limit 44 to (human and english language and "remove medline records" and yr="1980 -Current") (2125)

Database: APA PsycInfo Search Strategy:

1 Heuristic*.tw. (17111)

2 (Cognitive adj2 bias).tw. (2759)

3 (Heuristic adj2 bias).tw. (51)

4 (Psych* adj1 bias*).tw. (377)

5 (Anchoring adj2 heuristic*).tw. (69)

6 (Availability adj3 bias).tw. (81)

7 (Premature adj1 closure).tw. (137)

8 (Confirmation adj2 heuristic*).tw. (5)

9 (Confirmation adj2 bias).tw. (743)

10 (Representativeness adj2 heuristic*).tw. (191)

11 (Representativeness adj2 bias).tw. (33)

12 (Anchoring adj2 bias).tw. (85)

13 (Availability adj2 heuristic*).tw. (253)

14 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 (21028)

15 Physician*.tw. (71927)

16 Doctor*.tw. (45955)

17 (Medical adj1 prac*).tw. (6597)

18 (Medical adj1 prof*).tw. (6038)

19 15 or 16 or 17 or 18 (119660)

20 (difference-in-differences adj1 design).tw. (92)

21 (retrospective adj2 review).tw. (4625)

22 (case adj1 series).tw. (4996)

23 (event adj1 stud*).tw. (886)

24 (longitudinal adj1 stud*).tw. (63966)

25 (cross-sectional adj1 stud*).tw. (40108)

26 (prospective adj3 stud*).tw. (38161)

27 (retrospective adj3 stud*).tw. (17621)

28 (regression adj1 analys*).tw. (97877)

29 (case-control adj1 stud*).tw. (9078)

30 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 (253809)

31 14 and 30 (367)

32 14 and 19 and 30 (7)

33 limit 32 to (english language and yr="1980 - 2023") (7)

Web of Science

1. Heuristic$ (Abstract)
2. Cognitive NEAR/2 bias (Abstract)
3. Heuristic NEAR/2 bias (Abstract)
4. Psych$ NEAR/1 bias$ (Abstract)
5. Anchoring NEAR/2 bias$ (Abstract)
6. Anchoring NEAR/2 heuristic$ (Abstract)
7. Availability NEAR/2 heuristic$ (Abstract)
8. Availability NEAR/2 bias$ (Abstract)
9. Premature NEAR/2 closure (Abstract)
10. Confirmation NEAR/2 heuristic (Abstract)
11. Confirmation NEAR/2 bias$ (Abstract)
12. Representativeness NEAR/2 bias$ (Abstract)
13. Representativeness NEAR/2 heuristic$ (Abstract)
14. #13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1
15. (((((((AB=(retrospective NEAR2 review)) OR AB=(case NEAR1 series)) OR AB=(event NEAR1 stud$)) OR AB=(longitudinal NEAR1 stud$)) OR AB=(cross-sectional NEAR1 stud$)) OR AB=(prospective NEAR stud$)) OR AB=(retrospective NEAR3 stud$)) OR AB=(regression NEAR1 analys$)) OR AB=(case-control NEAR1 stud$)
16. #15 and #14

Econlit

TX (physicians or doctors or clinicians or healthcare professionals) OR TX medical decision making OR TX (medical practitioners or medical professionals or surgeons)

AB heuristics OR AB cognitive bias OR AB cognitive biases OR AB (heuristic bias or heuristic biases) OR AB (psychological bias or psychological biases) OR AB (anchoring effect or anchoring bias or anchoring heuristic) OR AB (availability heuristic or availability bias) OR AB (confirmation heuristic or confirmation bias) OR AB (representativeness heuristic or representativeness bias)

Scopus

Set search alert

Advanced query

((TITLE-ABS-KEY (case-control W/1 stud*)) OR (TITLE-ABS-KEY (regression W/1 analys*)) OR (TITLE-ABS-KEY (retrospective W/3 stud*)) OR (TITLE-ABS-KEY (prospective W/3 stud*)) OR (TITLE-ABS-KEY (cross-sectional W/1 stud*)) OR (TITLE-ABS-KEY (longitudinal W/1 stud*)) OR (TITLE-ABS-KEY (event W/1 stud*)) OR (TITLE-ABS-KEY (case W/1 series)) OR (TITLE-ABS-KEY (retrospective W/2 review)) OR (TITLE-ABS-KEY (difference-in-differences W/1 design))) AND ((TITLE-ABS-KEY (surgeon*)) OR (TITLE-ABS-KEY (medical W/1 prof*)) OR (TITLE-ABS-KEY (medical W/1 prac*)) OR (TITLE-ABS-KEY (doctor*)) OR (TITLE-ABS-KEY (physician*))) AND ((TITLE-ABS-KEY (representativeness W/2 heuristic*)) OR (TITLE-ABS-KEY (representativeness W/2 bias*)) OR (TITLE-ABS-KEY (confirmation W/2 bias*)) OR (TITLE-ABS-KEY (confirmation W/2 bias*)) OR (TITLE-ABS-KEY (confirmation W/2 heuristic*)) OR (TITLE-ABS-KEY (premature W/2 closure)) OR (TITLE-ABS-KEY (availability AND w2 AND bias*)) OR (TITLE-ABS-KEY (availability W/2 heuristic*)) OR (TITLE-ABS-KEY (anchoring W/2 heuristic*)) OR (TITLE-ABS-KEY (anchoring W/2 bias*)) OR (TITLE-ABS-KEY (psych* W/1 bias*)) OR (TITLE-ABS-KEY (heuristic W/2 bias)) OR (TITLE-ABS-KEY (cognitive W/2 bias)) OR (TITLE-ABS-KEY (heuristic*))) AND (LIMIT-TO (LANGUAGE , "English"))

<u>Supplementary Medline search</u>

Heuristic*.tw.

2.

(Cognitive adj2 bias).tw.

3.

(Heuristic adj2 bias).tw.

4.

(Decision adj2 rule).tw.

5.

(Psych* adj1 bias*).tw.

6.

(Anchoring adj2 heuristic*).tw.

7.

(Availability adj3 bias).tw.

8.

(Premature adj1 closure).tw.

9.

(Confirmation adj2 heuristic*).tw.

10.

(Confirmation adj2 bias).tw.

11.

(Representativeness adj2 heuristic*).tw.

12.

(Representativeness adj2 bias).tw.

13.

(Anchoring adj2 bias).tw.

14.

(Availability adj2 heuristic*).tw.

15.

1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14

16.

(difference-in-differences adj1 design).tw.

17.

(retrospective adj2 review).tw.

18.

(case adj1 series).tw.

19.

(event adj1 stud*).tw.

20.

(longitudinal adj1 stud*).tw.

21.

(cross-sectional adj1 stud*).tw.

22.

(prospective adj3 stud*).tw.

23.

(retrospective adj3 stud*).tw.

24.

(regression adj1 analys*).tw.

25.

(case-control adj1 stud*).tw.

26.

16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25

27.

15 and 26

28.

limit 27 to (english language and yr="1980 -Current")

Supplementary Embase search

Database: Embase Classic <1947 to 1973> Part 1 of 2, Embase <1974 to 2023 September 29>

Search Strategy:

1 Heuristic*.tw. (16262)

2 (Cognitive adj2 bias).tw. (2109)

3 (Heuristic adj2 bias).tw. (36)

4 (Decision adj2 rule).tw. (2518)

5 (Psych* adj1 bias*).tw. (200)

6 (Anchoring adj2 heuristic*).tw. (55)

7 (Availability adj3 bias).tw. (214)

8 (Premature adj1 closure).tw. (1428)

9 (Confirmation adj2 heuristic*).tw. (1)

10 (Confirmation adj2 bias).tw. (513)

11 (Representativeness adj2 heuristic*).tw. (60)

12 (Representativeness adj2 bias).tw. (30)

13 (Anchoring adj2 bias).tw. (373)

14 (Availability adj2 heuristic*).tw. (109)

15 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 (23268)

16 limit 15 to (english language and "remove medline records" and yr="1980 - 2023") (6298)

17 (difference-in-differences adj1 design).tw. (301)

18 (retrospective adj2 review).tw. (214100)

19 (case adj1 series).tw. (148386)

20 (event adj1 stud*).tw. (1180)

21 (longitudinal adj1 stud*).tw. (128953)

22 (cross-sectional adj1 stud*).tw. (347129)

23 (prospective adj3 stud*).tw. (678714)

24 (retrospective adj3 stud*).tw. (678186)

25 (regression adj1 analys*).tw. (537940)

26 (case-control adj1 stud*).tw. (169549)

27 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 (2598675)

28 limit 27 to (english language and "remove medline records" and yr="1980 - 2023") (1073042)

29 16 and 28 (435)

Instructions for data extraction form

Review question: Among medical practitioners, does exposure to previous medical events induce the use of heuristics in testing, screening and treatment decisions as measured in terms of (i) differences in testing, screening and referral rates (ii) differences in odds of recommending or performing therapeutic procedures for patients, in real clinical settings

Main outcome: Medical care decisions as measured by (i) rates of testing or referrals to other tests and treatments or (ii) probability or odds of recommending or performing certain therapeutic procedures for patients.
Clarification: Studies which report on diagnostic error rates (instead of testing or treatment rates) can be included as long as they are reported in a way that allows us to assess the effect of exposure to the hypothesised heuristic-inducing information or event.

How to fill in the form

1. Answer Y or N **to Medical practitioners?** (and you may also briefly state type of medical practitioner)
2. Answer Y or N to **Empirical design in non-simulated setting?**
3. Answer Y or N in first column allocated to **Is objective of the study to test a hypothesis regarding use of heuristics/biases among medical professionals?** In second column allocated to this question you may copy and paste or summarise text from paper which you think outlines the objective of the study to support your conclusion (this is optional not compulsory but can help later in understanding differences in conclusions). If possible put any verbatim quotes in italics.
4. Answer P or I in the first column allocated to **Is the main heuristic under investigation related to previous medical events (P) or initial patient information (I) ?** In second column allocated to this question you can state what the specific previous medical event (P) or initial patient information (I) is (or in ambiguous cases provide further explanation). Again this is optional not compulsory but can help later in understanding differences in conclusions.
5. Answer Y or No in the first column allocated to **Is there at least one relevant quantitative measure reported relating to testing/referrals , recommending or performing treatments (or other clinical decisions) for exposed and unexposed groups or observation period. These could be presented as summary statistics for each group or exposure period (means, proportions), or between group/period differences (e.g. mean difference, odds ratio)** In second column allocated to this question you can state what the measure is (or in ambiguous cases provide further explanation) Again this is optional not compulsory but can help later in understanding differences in conclusions.

If a study fails at one step you don’t have to fill in columns after that step

Examples of Initial information about patient (e.g. preliminary information recorded in a patient's file or observed from an initial consultation):

1. Clinician observes that a patient is recorded as just exceeding the specific age threshold where key clinical trials suggest patients are less likely to benefit from a specific treatment.
2. Clinician observes that a patient visit reason as documented in triage before the patient is seen mentions congestive heart failure.
3. Clinician observes that a patient entering intensive care is noted as having a 'do not resuscitate' order.
4. Clinician observes at the initial consult that the patient is running a high fever.
5. Clinician observes that a patient file reports positive HPV status .

'Initial information' about a patient typically refers to 'first impressions' information extracted by the clinical decision-maker from one of two main settings (i) a patient file/triage record (ii) a test or observation undertaken during an initial consultation/patient visit. Therefore, this will exclude (under scenario (ii)) studies which report on information about a patient provided by a test (or other process) undertaken by the clinical decision-maker after a barrage of other tests or observations have already been undertaken by the **same clinical decision-maker**. On the other hand, where the barrage of tests is undertaken by other clinicians and their results are summarised in a patient file/record read by the clinical decision-maker for the first time this falls under scenario (i) and would not be excluded on those grounds alone.

The 'first impressions' information which leads to different further testing/treatment rates for one group of patients based on that differentiating information vs another should occur in similar settings for the comparisons reported (e.g. the information should be gleaned from similar sources whether this is through information summarised in patient files or information observed from the same type of observation/testing whether it is patient dialogue, visual examination, etc). This will exclude studies which aim to study differences in clinical decision-making due to the intentional introduction of different testing strategies/approaches in the patients being compared (since this is no longer due to first impressions information gleaned in similar settings).

Examples of Previous medical events

1. Clinician has previously referred for colonoscopies patients who subsequently suffered adverse events due to these colonoscopies.
2. Clinician has previously been subject to a medical malpractice claim.
3. Clinician has peers in same clinical practice who had public injury reports filed against them.
4. Clinician has previously diagnosed patients with pulmonary embolism in an ED setting.
5. Clinician has previously undertaken caesarean delivery of patients resulting in severe uterine rupture.

Methodological notes on sample size

As can be seen from Table 2-1, there were a number of inconsistencies in reporting among the 10 studies selected. The first group of studies on heuristics prompted by exposure to triage information did not the report on the number of participating physicians, only the number of patients; while each of the second set of studies on heuristics prompted by previous medical events reported on the number of participating physicians but some did not report on the number of patients. In addition some studies reported on the number of deliveries rather than the number of patients and one reported on the number of hospitals rather than the number of patients. For these reasons we did not report on the total (pooled) number of physicians and patients in our 10 included studies.

In the case of heuristics prompted by exposure to triage information, each 'experiment' is effectively reset every time a physician looks at a new set of patient triage information. Even if it is the same physician each time, the impact on decision-making depends on the extent to which each time that same physician looks at an individual patient's information, the 'direction' of that physician's decision (e.g. to test or not to test for PE as per the Ly et al 2023 study) is consistent with the hypothesised direction given the information provided (e.g. not to test for PE if the triage information mentions CHF as a patient visit reason). Thus the number of physicians is not relevant to the robustness of the result for this heuristic type but the number of patients is (e.g. the inference that a physician is deciding not to test for PE because of mention of CHF based on the physician's decisions from looking at 5 patients will be less robust than if this inference was based on the physician looking at 100 patients).

In the case of heuristics prompted by exposure to previous medical events however, the number of participating physicians will be relevant in considering the robustness of the finding. Using Keating et al 2017 as an example, consider if this study had been based on one doctor having a previous patient with a colonoscopy suffering a bleeding event and how her subsequent decisions to order colonoscopy for later patients compared with 10 other doctors that did not have a similar prior experience. Even if this one doctor went on to see 1000 patients over the time period of the study, a heuristic impact based on this one doctor (vs 10 other doctors, or a total sample size of 11) would still be less robust than if there were more participating doctors. For these reasons, the number of participating physicians is a relevant consideration when assessing the robustness of findings in studies of heuristics prompted by previous medical events.

Data quality

Supplementary Table 2-1: Summary of data quality

Author	Selection*				Comparability	Outcome			Total
	Representativeness	Selection	Ascertainment	Demonstration	Comparability	Assessment	Follow up	Adequacy	
Choudry et al 2006	1	1	1	0.5	2	1	1	1	8.5
Dan et al 2017	0	1	1	1	1	1	1	1	7
Keating 2017	1	1	1	0.5	1	1	1	1	7.5
Kulkarni et al 2019	1	1	1	0	2	1	1	1	8
Ly 2019	1	1	1	0.5	2	1	1	1	8.5
Ly et al 2021	0.75	1	1	0.25	2	1	1	1	8
Ly et al 2023	0.75	1	1	0.25	2	1	1	1	8
Riddell et al 2014	1	1	1	0.5	2	1	1	1	8.5
Simianu et al 2016	0.75	1	1	0.25	1.5	0.5	1	1	7
Singh 2021	0.5	1	1	0.25	2	1	1	1	7.75

Notes for Supplementary Table 2-1

1. The Selection components of the score are an averaged rating across evaluation of these components for both patients and doctors, hence the fractions.
2. As discussed, although Ly et al 2023 classified their study as a cross-sectional study we considered that in substance and for quality assessment purposes it could be treated as a retrospective cohort study and assessed accordingly using the same NOS scale as the other studies (therefore also allowing for a direct comparison between this study and the others). We acknowledge that this study had characteristics of a cross-sectional study such as the fact that it examines associations between presentation and outcomes at a single time point such that exposures and outcomes are measured at the same time, and because each emergency department visit in the study is treated as a discrete event. However it also has significant elements of a retrospective cohort data study like the other studies in this table such as the use of historical data from 2011 to 2018 which is from the US national Veterans' Affairs Electronic Health Record database, and the temporal sequence of medical investigation considered in the study (including non-primary outcomes) from presentation to testing to diagnosis and the fact that in substance the study compares outcomes between patients with different patient visit reasons. Note also that while employing a broadly similar methodology including use of archival patient data to investigate the heuristic impacts of initial patient triage information, the investigators in Kulkarni et al 2019 describe their study as a retrospective observational cohort study.
3. Quality thresholds applied based on Saposnik et al 2016 are:

- 0 to 4: low quality
- 5 to 6: moderate quality
- 7 to 9: high quality.

Converting results into odds ratios

For studies which reported the primary outcome as a percentage change in care (for instance an increase or reduction in testing rates), the results were converted into an odds ratio by determining what the 'baseline' rate was before exposure, the new rate following exposure (with confidence intervals of the change in rates determining the confidence intervals of the new rates), the corresponding odds associated with the baseline and new rates and from this, the respective odds ratios.

For instance, in Ly 2019, the mean testing rate was reported as 2.3% while the change in the quarter following exposure was reported as 0.78% with a 95% confidence interval of 0.14% to 1.42%. This implies an odds of testing of 0.0235 before exposure and 0.0318 after exposure so the point estimate odds ratio is 0.0318/0.0235 which is 1.35. The confidence intervals of the odds ratios around this can then be estimated by estimating the odds for the lower bound testing rate and upper bound testing and also dividing each of these by the odds of testing under the baseline (pre-exposure) rate.

In Simianu et al 2016 this estimate is made more complicated because rather than reporting the change in the baseline testing rate and the 95% confidence interval of the change in testing rate, the previous and new testing rate with confidence intervals for each are reported. For instance in scenario 3 the baseline testing rate was 70.7% (55.6%-85.8%) while the new testing rate following exposure (anastomotic leak in patient) was 59.2% (44.3%-74.1%). The odds ratio was estimated the usual way by dividing the odds of testing after exposure with the odds of testing before exposure. However because there were confidence intervals reported for both the baseline and post-exposure rates the confidence intervals were estimated differently this time. The lower bound of the post-exposure odds ratio was estimated by dividing the odds of testing for the post-exposure lower bound by the odds of testing for the pre-exposure upper bound and the upper bound of the post-exposure odds ratio was estimated by dividing the odds of testing for the post-exposure upper bound by the odds of testing for the pre-exposure lower bound. The rationale for this approach was to account for uncertainties given the confidence intervals provided i.e. the post-exposure lower bound should account for factors leading to the lowest possible bound for the point estimate (and this is captured by using the lowest possible numerator and highest possible denominator suggested by the confidence intervals of the before and after point estimates).

Additional results omitted from figures and tables

Results from Simianu et al 2016

As noted previously, the original outcomes reported by Simianu et al 2016 were change in rate of leak testing after a postoperative leak occurred and this was recalculated as an odds ratio of leak testing. The study reported 4 different scenarios – change in rate of leak testing (i) in surgeons who performed under 5 cases a year where leak testing was not performed in the case that resulted in a post-operative leak (ii) in surgeons who performed under 5 cases a year where leak testing was performed in the case that resulted in a post-operative leak (iii) in surgeons who performed 5 cases or more a year where leak testing was not performed in the case that resulted in a post-operative leak (iv) in surgeons who performed 5 cases or more a year where leak testing was performed in the case that resulted in a post-operative leak. All results were found to be statistically insignificant even in the format originally reported by Simianu et al 2016 which was based on reporting p-values of differences in testing rates rather than odds ratios (the p-values for all scenario were >0.1). The table below reports results from all 4 scenarios. The third scenario was chosen for reporting in the forest plot in Figure 2-4 because (i) it was one of the two scenarios which reported the odds ratio for leak testing where leak testing was not performed during the episode of care that resulted in a post-operative leak (i.e. it had a comparator of no testing) (ii) the other scenario where leak testing was also not performed reported an odds ratio with a very wide confidence interval (0.30-51.06) which if placed in the same figure would have obscured the results from the other studies.

Supplementary Table 2-2: Full results from Simianu 2016 recalculated as odds ratios

Surgeon volume	Leak test	Odds ratio
<5 cases p.a.	No	2.51 (0.30-51.06)
<5 cases p.a.	Yes	1.15 (0.14-13.76)
5+ cases p.a.	No	0.60 (0.13-2.28)
5+ cases p.a.	Yes	0.99 (0.47-2.13)

In addition it is worth noting that Simianu et al 2016 chose to highlight another aspect of their findings which was not directly relevant to the scope of our review and therefore was not discussed in our review which was focused on compiling and synthesising results relating to the magnitude of overall changes in medical decision-making and whether these changes were statistically significant. They noted that out of their total sample of 4,854 elective colorectal operations, there was a leak rate of 2.6% or 124 leaks and that of these 124 leaks, the 40 (32%) in which the anastomosis was not tested occurred across 25 surgeons. They then defined a threshold of an increase in leak testing by 5% as demonstrating the existence of a heuristic (identified as a recency effect) and noted that only 9 (36%) of the 25 surgeons increased their leak testing by 5% or more after leaks in cases where the anastomosis was not tested. Therefore they argued that their findings suggested that a heuristic use was demonstrated by one third of the surgeons in their sample. However as this finding is not based on the statistical significance of the results associated with their primary outcomes and relies on defining an arbitrary threshold of impact of 5% instead we have not reported this in our review.

Results from Keating et al 2017

The odds ratios estimated for 'decision not to order colonoscopy' for Keating et al 2017 stratified by physician and patient age based on their original data are presented below. The values highlighted in red are the statistically significant results.

Supplementary Table 2-3: Stratified results from Keating et al 2017 results recalculated as odds ratios

Keating 2017 model stratified by patient age	
	Odds ratios
Patients 65-75 years	
Quarter 1	1.00 (0.98-1.03)
Quarter 2	**1.04 (1.02-1.07)**
Quarter 3	1.01 (0.99-1.04)
Quarter 4	1.02 (0.99-1.04)
Patients >75 years	
Quarter 1	**1.03 (1.01-1.06)**
Quarter 2	**1.03 (1.00-1.05)**
Quarter 3	**1.04 (1.01-1.06)**
Quarter 4	1.01 (0.98-1.04)
Keating 2017 model stratified by physician experience (age above/below median)	
	Odds ratios
Physicians<50.2 years	
Quarter 1	1.01 (0.99-1.03)
Quarter 2	**1.05 (1.03-1.07)**
Quarter 3	**1.02 (1.00-1.04)**
Quarter 4	1.00 (0.98-1.02)
Physicians≥50.2 years	
Quarter 1	1.01 (0.99-1.02)
Quarter 2	1.00 (0.98-1.02)
Quarter 3	1.00 (0.98-1.02)
Quarter 4	1.00 (0.98-1.02)

Supplement for Chapter 3

Supplementary File 1: Clinical Vignettes

The following clinical vignettes are based on current guidance (Traeger et al. CMAJ. 2017; Chou et al. Ann Intern Med. 2011) [1] [2] combined with local guidance from HealthPathways[3]

The first three options are guideline-concordant, while the second three options are guideline discordant (low value).

A. Imaging scenario 1

A 22 year old female presents with a 1-week history of severe left-sided low back pain and sciatica. Her symptoms are not improving and she reports severe leg pain at night. She is otherwise well with no unusual bladder or bowel disturbance. On physical examination she has a positive left straight leg raise test. Muscle strength is normal, but the patient has a reduced ankle jerk reflex and sensation in left lateral foot. The patient is distressed about her symptoms and her inability to sleep, and requests an imaging test.

Please indicate which of the following 6 options you would choose for this patient (please choose at least 1 and no more than 3 options):

- Refuse to provide imaging immediately
- Review at 2, 6 and 12 weeks until pain has settled
- Defer request lumbar MRI until later review e.g. in 2-4 weeks
- Refer for radiography and laboratory tests
- Refer for urgent lumbar MRI via the emergency department
- Refer to rheumatologist or spinal surgeon for lumbar MRI

B. Imaging scenario 2

A 70 year old female presents with acute low back pain after a fall at home. She has severe pain with movement that ceases with sitting. She takes oral prednisone 5mg/day for management of rheumatoid arthritis. On physical examination she is

---

[1] Traeger A, Buchbinder R, Harris I, et al. Diagnosis and management of low-back pain in primary care. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2017;189(45):E1386-e95. doi: 10.1503/cmaj.170527 [published Online First: 2017/11/15.

[2] Chou R, Qaseem A, Owens DK, et al. Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. Annals of internal medicine 2011;154(3):181-9. doi: 10.7326/0003-4819-154-3-201102010-00008 [published Online First: 2011/02/02.

[3] Mid and North Coast of NSW https://manc.healthpathways.org.au/index.htm

locally tender over L1 and bruising in the upper lumbar region. There are no neurologic signs.

Please indicate which of the following 6 options you would choose for this patient (please choose at least 1 and no more than 3 options):

- Provide no imaging or laboratory tests immediately
- Provide analgesia and defer request for lumbar radiography until later review e.g. in 1-2 weeks
- Provide advice on self-management until later review
- Refer for immediate lumbar radiography and laboratory tests via the emergency department
- Refer for immediate lumbar computed tomography via the emergency department
- Refer for immediate lumbar magnetic resonance imaging via the emergency department

C. Medicine scenario 1

A 55-year old male builder presents with a 2-day history of severe, acute low back pain. He is unable to sit in the consultation room and reports feeling nauseous. He takes Lipitor 10mg/day and has history of GORD. On examination straight leg raise is negative and there are no neurological signs.

Please indicate which of the following 6 options you would choose for this patient (please choose at least 1 and no more than 3 options):

- Prescribe no analgesic medicine; discuss non-pharmacologic options e.g. massage therapy
- Prescribe a muscle relaxant (e.g. diazepam)
- Prescribe a non-steroidal anti-inflammatory drug + gastroprotective agent
- Prescribe an extended release opioid (e.g. oxycodone)
- Prescribe a neuropathic pain medicine (e.g. pregabalin)
- Prescribe an oral corticosteroid (e.g. prednisone)

D. Medicine scenario 2 (Chronic low back pain, weak opioid, NSAID, non-pharm recommended)

A 37-year old mother of two has presented 3 times in 6 months for persistent back pain and sciatica. In the latest visit she reports an exacerbation of leg symptoms and inability to work for the past 3 days due to severe pain. She has a positive straight leg raise but normal reflex, muscle strength and sensation. She is generally well, and the back pain is her primary complaint. She currently takes paracetamol 500mg 3x/day and amitriptyline 10mg/day for depression.

Please indicate which of the following 6 options you would choose for this patient (please choose at least 1 and no more than 3 options):

- Advise to stop paracetamol
- Prescribe an NSAID
- Prescribe a weak opioid or opioid combination (e.g. codeine, codeine + paracetamol)
- Prescribe an extended release opioid (e.g. oxycodone)
- Prescribe a neuropathic pain medicine (e.g. pregabalin)
- Prescribe an oral corticosteroid (e.g. prednisone)

Supplementary file 2: Visual appearance of the display of menu options in each of the four experimental conditions

No nudge

Please indicate which of the following 6 options you would choose for this patient **(please choose at least 1 and no more than 3 options):**

- Review at 2, 6 and 12 weeks until pain has settled
- Refuse to provide imaging immediately
- Defer request for imaging, consider lumbar magnetic resonance imaging at later review e.g. in 2-4 weeks
- Refer to rheumatologist or spinal surgeon for lumbar MRI
- Refer for radiography and laboratory tests
- Refer for urgent lumbar MRI via the emergency department

Partition display nudge

Please indicate which of the following 6 options you would choose for this patient **(please choose at least 1 and no more than 3 options):**

Review at 2, 6 and 12 weeks until pain has settled

Refuse to provide imaging immediately

Defer request for imaging, consider lumbar magnetic resonance imaging at later review e.g. in 2-4 weeks

Refer for radiography and laboratory tests

Refer to rheumatologist or spinal surgeon for lumbar MRI

Refer for urgent lumbar MRI via the emergency department

Default option nudge

Please indicate which of the following 6 options you would choose for this patient **(please choose at least 1 and no more than 3 options):**

Review at 2, 6 and 12 weeks until pain has settled

Refuse to provide imaging immediately

Defer request for imaging, consider lumbar magnetic resonance imaging at later review e.g. in 2-4 weeks

Click for more options

Refer for radiography and laboratory tests

Refer for urgent lumbar MRI via the emergency department

Refer to rheumatologist or spinal surgeon for lumbar MRI

Options revealed

Double nudge

Please indicate which of the following 6 options you would choose for this patient **(please choose at least 1 and no more than 3 options):**

Defer request for imaging, consider lumbar magnetic resonance imaging at later review e.g. in 2-4 weeks

Review at 2, 6 and 12 weeks until pain has settled

Refuse to provide imaging immediately

Click for more options

Refer for urgent lumbar MRI via the emergency department	Refer to rheumatologist or spinal surgeon for lumbar MRI	Refer for radiography and laboratory tests

Options revealed

Supplementary File 3

Supplementary Table 3-1: 2x2 factorial design of experiment

Experimental condition	**Default**	**Partition**
No nudge	No	No
Partition display nudge	No	Yes
Default option nudge	Yes	No
Double nudge	Yes	Yes

Supplementary File 4: More on the Cognitive Reflection Test

For our descriptive statistics we defined a low Cognitive Reflection Test score as less than or equal to one. This cut-off was based on a study reporting that web-based and US college samples typically obtain means between 0.5 and 1 while participants from more ‘elite’ colleges obtained a mean above 1.5.[4]

The three Cognitive Reflection Test questions were:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days

[4] Frederick S. Cognitive reflection and decision making. Journal of Economic perspectives 2005;19(4):25-42.

Supplementary File 5

Supplementary Table 3-2: Proportion of vignettes where participants chose at least one low-value option, stratified by vignette type

	No partition display nudge (n=60)	Partition display nudge (n=60)	No Default option nudge (n=61)	Default option nudge (n=59)
	% of scenarios where low-value options selected	% of scenarios where low-value options selected	% of scenarios where low-value options selected	% of scenarios where low-value options selected
All vignettes	36.3	38.3	43.9	30.5
Imaging vignettes	46.7	45.8	52.5	39.8
Medication vignettes	25.8	30.8	35.2	21.2

Supplement for Chapter 4

Methodological notes

List of MBS item descriptors included in the analysis and the rationale

The CT imaging scans encompassed scans funded by four different MBS imaging item descriptors so that the number of CT head scans in a year were counted by adding up the total number of claims arising from each of these item descriptors in the year. The MBS item descriptors for these CT scans were:

- 56001: Computed tomography—scan of brain without intravenous contrast medium, not being a service to which item 57001 applies (R) (Anaes.)
- 56010: Computed tomography—scan of pituitary fossa with or without intravenous contrast medium and with or without brain scan when performed (R) (Anaes.)
- 56013: Computed tomography - scan of orbits with or without intravenous contrast medium and with or without brain scan when undertaken (R) (Anaes.)
- 56036: Computed tomography—scan of facial bones, para nasal sinuses or both, with scan of brain, with intravenous contrast medium, if:
- (a) a scan without intravenous contrast medium has been performed; and
- (b) the service is required because the result of the scan mentioned in paragraph (a) is abnormal (R) (Anaes.)

Based on the advice of one of the co-authors (MA) and also in order to address the advice of the External Request Evaluation Committee of Services Australia to further narrow down the CT items before requesting data, items 56007, 56028 and 56030 were not included because utilisation of these items is likely to be diminishing (in the case of 56007) and very small (in the case of 56028 and 56030).

The MBS item descriptors for included non-GP specialist requested MRI scans included all of MRI subgroups 1. Scan of head for specified conditions and 2. Scan of head for specified conditions. The MBS item descriptors were:

- 63001: MRI—scan of head (including MRA, if performed) for tumour of the brain or meninges (R) (Anaes.) (Contrast)
- 63004: MRI—scan of head (including MRA, if performed) for inflammation of brain or meninges (R) (Anaes.) (Contrast)
- 63007: MRI—scan of head (including MRA, if performed) for skull base or orbital tumour (R) (Anaes.) (Contrast)
- 63010: MRI—scan of head (including MRA, if performed) for stereotactic scan of brain, with fiducials in place, for the sole purpose of allowing planning for stereotactic neurosurgery (R) (Anaes.) (Contrast)
- 63040: MRI—scan of head (including MRA, if performed) for acoustic neuroma (R) (Anaes.) (Contrast)

- 63043: MRI—scan of head (including MRA, if performed) for pituitary tumour (R) (Anaes.) (Contrast)
- 63046: MRI—scan of head (including MRA, if performed) for toxic or metabolic or ischaemic encephalopathy (R) (Anaes.) (Contrast)
- 63049: MRI—scan of head (including MRA, if performed) for demyelinating disease of the brain (R) (Anaes.) (Contrast)
- 63052: MRI—scan of head (including MRA, if performed) for congenital malformation of the brain or meninges (R) (Anaes.) (Contrast)
- 63055: MRI—scan of head (including MRA, if performed) for venous sinus thrombosis (R) (Anaes.) (Contrast)
- 63058: MRI—scan of head (including MRA, if performed) for head trauma (R) (Anaes.) (Contrast)
- 63061: MRI—scan of head (including MRA, if performed) for epilepsy (R) (Anaes.) (Contrast)
- 63064: MRI—scan of head (including MRA, if performed) for stroke (R) (Anaes.) (Contrast)
- 63067: MRI—scan of head (including MRA, if performed) for carotid or vertebral artery dissection (R) (Anaes.) (Contrast)
- 63070: MRI—scan of head (including MRA, if performed) for intracranial aneurysm (R) (Anaes.) (Contrast)
- 63073: MRI—scan of head (including MRA, if performed) for intracranial arteriovenous malformation (R) (Anaes.) (Contrast)

The MBS item descriptors for included GP requested MRI scans encompassed all the GP-specific items related to MRI head scans namely:

- 63507: MRI—scan of head for a patient under 16 years if the service is for:
    - (a) an unexplained seizure; or
    - (b) an unexplained headache if significant pathology is suspected; or
    - (c) paranasal sinus pathology that has not responded to conservative therapy (R) (Contrast)
- 63551: MRI - scan of head for a patient 16 years or older, after a request by a medical practitioner (other than a specialist or consultant physician), for any of the following:

    (a) unexplained seizure(s);

    (b) unexplained chronic headache with suspected intracranial pathology (R) (Contrast)

Distinction between fully and partially MBS eligible MRI machines

While more MBS item descriptors for MRI head scans are available to non-GP specialists, all the non-GP specialist requested MRI head scans can only be undertaken on fully MBS eligible MRI machines if the patient is to be eligible for the MBS benefit while all MBS-eligible GP requested MRI head scans can be undertaken on partially eligible MRI machines and still qualify the patient for a government rebate.[1] Therefore, in the specifications of our modelling, we assumed GP requesters would treat the total of fully and partially MBS eligible MRI machines as MRI machines available for their use while in the case of non-GP specialists the effective availability of an MRI machine was restricted to the number of fully available MBS eligible MRI machines in that region. However the impact of using these different values in the regressions undertaken is trivial.

1. Refer to relevant tables of imaging items in Australian Government Department of Health and Aged Care. *Diagnostic Imaging Services Table - Full versus Partial Medicare-eligible MRI machines* https://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/3E48097FBDA88685CA258982007BD9A3/$File/PDF%20Version%20Fact%20sheet%20-%20Full%20versus%20Partial%20Medicare%20eligible%20MRI%20machines.pdf

Supplement for Chapter 5

1. Costing of diagnostic open biopsies and its treatment as a cost to the public hospital system

Almost 0.1% of patients in our sample were assessed with atypical ductal hyperplasia (ADH) and had 'diagnostic open biopsy' in their final assessment recommendation field with no further indication of whether they were false positive or true positive. Rather than exclude this group of patients, for costing purposes we assumed that the share of false positives in this group was similar to the share of false positives expressed as a percentage of total assessed clients who were not assessed with ADH and therefore were not assigned to a diagnostic open biopsy, and the share of true positives was similar to the share of true positives expressed as a percentage of total assessed clients who were not assessed with ADH and therefore were not assigned to a diagnostic open biopsy.

Diagnostic open biopsy is a procedure undertaken in hospital rather than in the BreastScreen facility. We do not know what share, if any, of this tiny minority of patients might have sought to expedite their testing by seeking private hospital treatment rather than public hospital treatment. If any of these patients had sought private treatment, then some share of them might also incur expenses 'out of pocket' if their private health insurance does not sufficiently cover the cost of this procedure in a private hospital. Rather than adopt additional complicated assumptions for this small subset of patients, we made the simplifying assumption that all patients recommended for a DOB obtained one in a public hospital (therefore 'free' at point of delivery) so that the cost incurred for this is another public-system incurred cost.

2. Share of total breast-cancer screening which is private screening

The estimate of 26% of total breast screening in Australia being through private screening is based on adding up the number of women screened in 2016 by BreastScreen Australia (from **Supplementary Table 1**) as reported in Table S4.11 of AIHW (2019)[1] and the number of women undergoing Medicare Benefits Schedule (MBS) subsidised breast mammograms under item 59300 (mammography of both breasts) in 2016 and then calculating the number of women undergoing Medicare-subsidised breast mammograms in 2016 as a percentage of this total. The MBS is a list of the medical services for which the Australian Government will pay a Medicare rebate, to provide patients with financial assistance towards the costs of private medical services. Thus, if a patient chooses to use a private provider and then claims a MBS rebate to recoup the costs of the private service, data on the usage of the private service can be captured. We note a number of caveats associated with this estimate:

- The private mammograms undertaken in this context may be for the purpose of investigating breast symptoms, for surveillance of women at high risk of developing breast cancer or for surveillance of women who have a personal

history of breast cancers. It is unclear what percentage of women undergoing these private mammograms would be doing so as an alternative to and in complete substitution of undergoing screening through a BreastCancer Australian service.
- At the same time, MBS data are not able to capture all mammography that occurs outside BreastScreen Australia because some women may choose to access private screening mammography on a full user-pays basis, for which a MBS rebate cannot be claimed.

While the second caveat may lead to an underestimate of the use of private screening services, the first caveat may be more significant in leading to an overestimate as it is unclear how many women are claiming MBS rebates for private screening in cases where these are supplementary to rather than in substitution of of services provided by BreastScreen Australia. It is only in the latter case that these services would count towards the share of total screening which is undertaken through private screening. We note that our approach is also likely to be an overestimate (and therefore in that respect resulting in a more conservative estimate of the share of breast screening undertaken through BreastScreen Australia) compared to the approach taken in the 2009 BreastScreen Australia Evaluation Report[5]which after a series of exclusions (which we were not able to replicate because we did not request more customised MBS data), estimated that 45% of bilateral mammograms in women aged 40+ in 2006 were diagnostic (therefore not for screening purposes), leaving an estimated 23% which were non diagnostic and 32% 'unknown'. Assuming for simplicity that the bilateral mammograms for 'unknown' purposes are also for 'non-diagnostic' purposes, and therefore assuming that 55% of MBS-funded bilateral mammograms under item 59300 were for screening purposes in 2016 would give us an estimate of approximately 14.5% of breast screening being through private services (if we included only services provided to women aged 35 years and over).

3. Caveats to inclusions

Within the included sample 0.05% were recalled to assessment and then classified under 'routine recall' as a final recommendation in 2016 but their discharge fields were subsequently updated to account for breast cancer being detected after. Thus technically these clients fell under our working definition of a 'false-positive' (women recalled for assessment who ultimately received a 'routine recall' recommendation). Possible alternatives for classifying these cases would have been to (i) classify them as false-positive cases as they fall within the definition; (ii). exclude them from consideration altogether or (iii) treat them as neither false positives nor false negatives but classify them as 'miscellaneous' because they were found to be positive after 2016. For simplicity we have adopted the first approach.

[5] Evaluation of the BreastScreen Australia Program – Evaluation Final Report – June 2009.

4. Additional tables

Supplementary Table 5-1 summarises some key national indicators from 2016 data that we used as the basis for extrapolating the results of our sample analysis to a national scale.

Supplementary Table 5-1: Key national indicators from BreastScreen Australia data for 2016[6]

No of women (aged 40+) recalled to assessment, first screening round in 2016	15,626
Crude rate for women (aged 40+) recalled to assessment, first screening round in 2016	11.1%
No of women screened (aged 40+), first screening round in 2016	140,835
No of women (aged 40+), diagnosed with cancer, first screening round in 2016	1257
No of women (aged 40+) recalled to assessment, subsequent screening rounds in 2016	36,769
Crude rate for women (aged 40+) recalled to assessment, subsequent screening rounds in 2016	3.9%
No of women screened (aged 40+), subsequent screening rounds in 2016	952,331
No of women (aged 40+), diagnosed with cancer, subsequent screening round in 2016	6,570
Crude rate for women (aged 40+) recalled to assessment, all rounds in 2016	4.8%

[6] Australian Institute of Health and Welfare, *BreastScreen Australia monitoring report 2018: supplementary data tables, Table S3.1*

Supplementary Table 5-2 outlines the cost assumptions used to estimate the costs per patient of each of the diagnostic workup sequences. The figures are presented in Australian dollars as per the original source material. We gratefully acknowledge the work of Saxby (2020)[2] who shared the cost assumptions underlying the analysis of Saxby et al (2020) [3] with us which we then adapted to our analysis. Unless a different or additional primary source is identified these figures are derived directly from their cost assumptions. While Saxby (2020)[2] also utilise Lockie et al (2018)[4] they sometimes utilise different tables in that report.

Supplementary Table 5-2: Cost components of diagnostic workup sequences

Activity	**Cost (A$2017)**	**Source and other comments**
1. Screening		
Invitation to screen	0.56	Saxby (2020)
ITS and other coordination	2.50	Saxby (2020)
Administration Coordination	3.17	Saxby (2020)
Mammo staff time	31.70	Lockie et al (2018) Table 18
Mammo consumables	3.50	Lockie et al (2018) Table 19
Consent	5	Lockie et al (2018) Table 18
Reading screen	15.04	Saxby (2020)
2. Assessment		
Mammo DBT staff time	30.9	Lockie et al (2018) Table 18
Mammo DBT consumables	3.50	Lockie et al (2018) Table 19
Ultrasound staff time	20	Lockie et al (2018) Table 18
Physical examination staff time	44	Assuming 10 mins of surgeon and nurse time per patient based on wage estimates in Lockie et al (2018) Table 17
Ultrasound consumables	6.50	Lockie et al (2018) Table 19
Core biopsy staff time	138	Lockie et al (2018) Table 18
CB stereo consumables	600	Lockie et al (2018) Table 19
CB stereo with clip consumables	695	Lockie et al (2018) Table 19
CB ultrasound guided consumables	86	Lockie et al (2018) Table 19
CB ultrasound guided with clip consumables	201	Lockie et al (2018) Table 19
Fine needle biopsy staff time	23	Lockie et al (2018) Table 18
Fine needle consumables	10	Lockie et al (2018) Table 19
Consent	5	Lockie et al (2018) Table 18
Open biopsy	2816	Lockie et al (2018) Table 20 – includes both consumables and labour, incurred by hospital system
Multidisciplinary team meeting	19.58	Lockie et al (2018) Table 17 Saxby (2020) cost assumptions - based on wage costs of radiologist, breast surgeon and pathologist assuming 1 to 1.25 hours are spent for every 40 cases. These costs are assumed to be incurred for every diagnostic workup involving a

		core biopsy and/or fine needle biopsy
Pathology	28.58	Saxby (2020). These costs are assumed to be incurred for every diagnostic workup involving a core biopsy and/or fine needle biopsy
Administration Coordination	3.17	Saxby (2020)
Delivery of results (benign)	5.83	Saxby (2020)
Delivery of results (cancer)	18.63	Saxby (2020)

Supplementary Table 5-3: Full list of diagnostic workup sequences of BSC service clients recalled to assessment in 2016

Diagnostic workup sequence	A. % of clients RTA after 1st screening round	B. % of clients RTA after subs. screening round
1. Screen + M	24.7%	23.9%
2. Screen + M+U	35.5%	36.8%
3. Screen + M+PE	1.1%	0.0%
4. Screen + M+CB	0.0%	1.0%
5. Screen + U+CB	0.0%	0.5%
6. Screen + M+U+PE	1.1%	0.5%
7. Screen + M+U+CB	0.0%	1.0%
8. Screen + M+PE+CB	8.6%	7.0%
9. Screen + M+U+PE+CB	19.4%	22.4%
10. Screen + M+U+PE+CB+FN	2.2%	1.0%
11. screenx2+M+U	0.0%	0.5%
12.screenx2+M+PE+CB	1.1%	0.0%
13.screenx2+2M+U+PE+2CB	0.0%	0.5%
14. Screen + 2M	0.0%	0.5%
15. Screen + 2M+U+PE+CB	0.0%	0.5%
16 Screen + 2M+U+CB+FN	1.1%	0.0%
17. Screen + 2M+2U+PE+CB	0.0%	0.5%
18. Screen + 2M+U+3CB	1.1%	0.0%
19. Screen + M+2U+PE+CB	0.0%	0.5%
20. Screen + M+2U+PE+CB+FN	0.0%	0.5%
21. Screen + M+2U+2PE+CB+FN	1.1%	0.0%
22. Screen + M+3PE+3CB	1.1%	0.0%
23. Screen + M+PE+2CB	1.1%	0.5%
24. Screen + M+U+PE+2CB	1.1%	1.5%
25. Screen + M+PE+3CB	0.0%	0.5%

M= assessment mammography, U = ultrasound, PE = physical examination, CB = core biopsy, FN = fine needle biopsy

Supplementary Table 5-4: Recall to assessment rates, false positive rates and positive predictive value by Australian State/Territory

	Recall rate (%)	False positive rate (%)	Positive predictive value (%)
NSW	4.7	4.0	15.3
Victoria	4.7	4.0	15.2
Queensland	5.4	4.6	13.6
Western Australia	3.6	2.9	19.3
South Australia	5.6	4.9	12.8
Tasmania	3.6	3.0	16.6
ACT	4.4	3.6	17.5
NT	14.6	13.5	7.4

Given that our extrapolated national results were so heavily reliant on data from one facility in the state of Victoria, we estimated and compared the recall to assessment rates, false positive rates and positive predictive values across the different States and Territories. These results are presented in **Supplementary Table 4**. We note from this table that false positive rates were broadly similar across the most populous States, with the exception of Western Australia (while Tasmania, ACT and the NT had significant differences in these three measures from Victoria, with the NT being the most significant outlier, these are also the State and Territories with the smallest populations.)

5. References

1. *Cancer in Australia 2019. Cancer series no.119. Cat. no. CAN 123.* Canberra: Australian Institute of Health and Welfare 2019.
2. Saxby K. Unpublished cost assumptions for Saxby K, Nickson C, Mann GB, et al. The financial impact of a breast cancer detected within and outside of screening: lessons from the Australian Lifepool cohort. *Aust N Z J Public Health*. 2020;44(3):219-226.
3. Saxby K, Nickson C, Mann GB, et al. The financial impact of a breast cancer detected within and outside of screening: lessons from the Australian Lifepool cohort. *Aust N Z J Public Health.* 2020;44(3):219-226.
4. Lockie D, Clemson M, Nickson C, Petrie D, Tew M, Munro J, et al. Assessment of Return on Investment from using Tomosynthesis in BreastScreen Assessment. BreastScreen; 2018.