



Applications of behavioural economics to health: three studies in health decision

making and behaviour

By

Robert P. Murphy

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

University of Stirling, Department of Economics

June 2023

Table of Contents

1	Introduction	1
1.1	The Influence of Neoclassical Economics on Public Policy and Health Policy	1
1.2	The Rise in Importance of Behavioural Economics for Public and Health Policy	5
1.3	Overview of the Thesis.....	8
2	Valuing the Q in QALYs: Does Providing Patients' Ratings Affect Population Values?.....	19
2.1	Abstract	20
2.2	Chapter Introduction.....	21
2.3	Methods	25
2.4	Results	31
2.5	Discussion.....	37
3	Do Misconceptions about Health-Related Quality of Life Affect General Population Valuations of Health States?	41
3.1	Abstract	42
3.2	Highlights	43
3.3	Introduction	44
3.4	Method	48
3.5	Results	53
3.6	Discussion.....	62
4	Nudging to improve patient engagement with management of hospital waiting lists: a Randomised Controlled Trial	65
4.1	Abstract	66
4.2	Introduction	67
4.3	Method	70
4.4	Results	76
4.5	Discussion.....	79
5	Discussion	85
5.1	Summary of Findings	85
5.2	Strengths and Limitations.....	89
5.3	Directions for Future Research.....	92
5.4	Applications.....	98
5.5	Conclusion.....	100
	References	104
	Supplementary Files.....	119

List of Figures

Figure 2.1 CONSORT Participant Flow Diagram.....	26
Figure 2.2 Mean valuation (EQ VAS) of conditions by prior belief of patients' mean rating and group.	35
Figure 3.1 CONSORT Participant Flow Diagram.....	49
Figure 3.2 Actual Difference in Well-being Associated with Each Health State Compared to the Control Group's Prior Belief of Difference	56
Figure 3.3 Proportion of Respondents Selecting Health State as 1st, 2nd or 3rd Preference by Group	58
Figure 4.1 CONSORT Participant Flow Diagram.....	72
Figure 4.2 Did not return (DNR) rate by letter type.	77
Figure 4.3 Did not return (DNR) rate by participant characteristic.....	78
Figure 4.4 Did not return (DNR) rate by healthcare characteristic.....	79

List of Tables

Table 2.1 Participant Characteristics	33
Table 2.2 Prior Belief by Group of Patients' Mean Rating of Health Conditions.....	34
Table 2.3 Results of Mixed ANOVA	37
Table 3.1 Participant and Population Characteristics	54
Table 3.2 Ordered Logistic Model of Preferences for "Moderate Anxiety / Depression"	60
Table 3.3 Marginal Effects of Information Provision on Preferences for "Moderate Anxiety / Depression"	61
Table 4.1 Nudge Design Elements Used in the Intervention Letter	75
Table 4.2 Participant Characteristics	76

Acknowledgements

I would like to thank my supervisors during my work on this thesis. This includes Liam Delaney, Conny Wollbrant, and Mirko Miro who stepped in during times of staff changes at the University of Stirling. Thank you to Christopher Boyce and Alex Wood for their first-class supervision from the very beginning to the point of submission of this thesis. Both Christopher and Alex have always made themselves available to give me great guidance and advice. A special thank you to Alex for our regular meetings, for his continued enthusiasm to support my academic development, and for his sharing of knowledge.

I would like to thank SIRE for a funding award (Early Career Engagement Grants) which supported the data collection for the studies reported in Chapters 2 and 3. I would also like to thank Alex Wood for encouraging me to apply for this funding and shaping my application.

I would like to thank my co-authors for their input into Chapters 2 to 4. For their invaluable feedback I would also like to thank the editors and anonymous peer reviewers of the journals in which Chapters 2 and 3 are published. In addition, I would like to thank the staff working in health policy and service management who facilitated the field trial reported in Chapter 4.

Finally, I would like to thank my family for their love. Unfortunately, during the course of my work on this thesis my father, father-in law, and mother-in-law passed away. We will always remember your love, and Mum thank you for your continued love. I want to thank my wife, Leona, and my children Alannah, Rob and Seán for their love, patience, and support during my work on this thesis. A very special thank you to my wife who made many a sacrifice as I worked on this thesis. Leona, even after 30 plus years together, your smile continues to give purpose and joy to my life.

Declaration

The research reported in this thesis is my own work unless otherwise stated. No part of this thesis has been submitted for a degree at another institution.

Chapter 2 was written in collaboration with Christopher J. Boyce, Paul Dolan, and Alex M. Wood. Their input was provided as follows: advice on data collection by Christopher and Paul, advice on data analysis by Christopher and Alex, and advice on write-up by all co-authors. I was wholly responsible for the idea and hypothesis generation, and I wholly undertook the data collection, data analysis, and write-up of the analysis.

Chapter 3 was written in collaboration with Christopher J. Boyce, Paul Dolan, Gordon D. A. Brown, and Alex M. Wood. Their input was provided as follows: advice on idea and hypothesis generation by Alex, advice on data collection by Christopher and Paul, advice on data analysis by Christopher, Alex and Gordon, and advice on write-up by all co-authors. I wholly undertook the data collection, data analysis, and write-up of the analysis.

Chapter 4 was written in collaboration with Carol Taaffe, Molly Byrne, Liam Delaney, Peter D. Lunn, Deirdre A. Robertson, Helen Ryan, and Alex M. Wood. All co-authors advised on the intervention design and commented on drafts of the write-up of the analysis. I was wholly responsible for the idea and hypothesis generation, and I wholly undertook the data analysis and write-up of the analysis.

Robert P. Murphy

Note on Inclusion of Published Work

Certain chapters have been previously published during the period of the PhD registration. Copyright of these papers resides with the publishers, but under the terms of the copyright agreements these papers are reproduced as chapters in this thesis. These papers are as follows.

Chapter 2: Murphy, R. P., C. J. Boyce, P. Dolan, and Alex M. Wood. 2020. “Valuing the Q in QALYs: Does Providing Patients’ Ratings Affect Population Values?” *Health Psychology* 39(1):37–45. doi: 10.1037/hea0000806.

Chapter 3: Murphy, RP, CJ Boyce, P. Dolan, GDA Brown, and AM Wood. 2023. “Do Misconceptions About Health-Related Quality of Life Affect General Population Valuations of Health States?” *VALUE IN HEALTH* 26(5):750–59. doi: 10.1016/j.jval.2022.10.009.

Chapter 4 is currently under review.

List of Abbreviations

BLI	Better Letter Initiative
CEA	Cost-effectiveness Analysis
DA	Day Affect
DNA	Did Not Attend
DNR	Do Not Return
HRQOL	Health-related Quality of Life
HSE	Health Service Executive
HTA	Health Technology Assessment
LS	Life Satisfaction
MTurk	Amazon Mechanical Turk
NTPF	National Treatment Purchase Fund
OECD	Organisation for Economic Cooperation and Development
QALYs	Quality-Adjusted Life Years
QOL	Quality of Life
RCT	Randomised Controlled Trial
SMS	Short Message Service
SWB	Subjective Well-being
TTO	Time Trade-off
VAS	Visual Analogue Scale
WHO	World Health Organization

Abstract

To maximise the health of society within a limited budget, decision makers in public health systems need to decide which health treatments to fund and how best to support engagement by patients with treatments. Current practice is heavily influenced by rational choice theory. In this thesis we apply an alternative behavioural economics perspective to inform decisions on which treatments to fund and how to support engagement with services.

Decisions on which treatments to fund are often informed by the expected gains in patients' quality-adjusted life years (QALYs). QALYs are derived from objective mortality data weighted by appraisals made by members of the general population of the likely impact on well-being associated with health states (the Q in QALYs). Concerns have been raised about the way in which the quality component of QALYs is calculated, leading to calls for ways to obtain experience informed general population appraisals.

In Chapter 2 we test the effect on general population preferences ($N = 155$) of being informed of patients' mean ratings of their health state and whether the direction of an effect depends on people's prior beliefs of patients' mean rating of the health states. We find that when the mean ratings given by patients are higher (lower) than expected, participants in the intervention group provide significantly higher (lower) valuations than participants in the control group. In Chapter 3 we examine whether people ($N = 1259$) selectively underestimate the well-being consequences of moderate anxiety / depression as compared to other dimensions of health, and we test if being informed of actual changes in well-being associated with health states changes appraisals of their relative undesirability. We find that people provided with information on the consequences of health states for life satisfaction or for day affect report a higher preference for avoiding living with moderate anxiety / depression. Both Chapters show that informing people of these summary measures before they appraise health states is a feasible way to obtain experience informed preferences and that experience informed preferences differ to those obtained using the current method.

Non-attendance for hospital appointments is a problem. One way to increase attendance is to improve the accuracy of waiting lists by writing to patients to check if a procedure is still required. However, the did not return (DNR) rate to such letters is substantial. In Chapter 4 we test ($N = 2855$) whether the DNR rate is reduced by introducing nudges to validation letters. We find that the redesigned validation letter reduced DNRs, by 4.73 percentage points or 19.73%.

Taken together these studies show the importance of applying a behavioural economics perspective to inform decisions on how to maximise the health of society.

1 Introduction

A fundamental objective of public health systems is to maximise the health of society given a limited budget. To achieve this objective, health policy decision makers need to decide on the upstream issue of which health treatments to fund and on the downstream issue of how to support patients' engagement with funded health treatments. This thesis applies a behavioural economics perspective to inform approaches to these two key issues. This introduction starts by providing a brief description of the neoclassical economics model of rational choice theory and how it has influenced the current approach to determining which health treatments should be funded in public health systems and the approach taken to communication with patients in relation to health services. It then summarises how behavioural economics challenged many of the assumptions and predictions of rational choice theory. It describes the rise in applications of findings from behavioural economics to public policy and outlines some key criticisms of the dominant approach taken to behavioural public policy. An overview of the thesis, which addresses some of the criticism directed to behavioural public policy, is then provided.

1.1 The Influence of Neoclassical Economics on Public Policy and Health Policy

1.1.1 Neoclassical Economics and Public Policy

Economics is a social science concerned chiefly with description and analysis of the production, distribution, and consumption of goods and services. Economics has a long history of policy application and has often been described as the most politically influential social science (Hirschman and Berman 2014). Neoclassical economics is an approach in which the production, consumption, and valuation (pricing) of goods and services are driven by demand and supply models. In such models, individuals demand goods and services to maximise their utility (expected satisfaction) subject to an income-constraint, and firms supply goods and services to maximise profits subject to available factors of production.

According to neoclassical economics, the choices that individuals make should be consistent with the predictions of rational choice theory.

Rational choice theory assumes that individuals make self-interested decisions to maximise their utility. It also assumes that individuals do this effectively by comparing all options (implementing mathematical rules of cost benefit analysis) to identify the option that is expected to provide the highest utility given their preferences and income constraints (Baddeley, Castells, and Guiora 2019). The theory makes two technical axioms about individuals' preferences over alternatives. First, completeness or that alternatives can be compared with each other – for any two alternatives X and Y in the set, either X is preferred to Y, or Y is preferred to X, or the individual is indifferent between X and Y. Second, transitivity – if alternative X is preferred to Y, and alternative Y is preferred to Z, then X is preferred to Z. These two axioms imply that, given a set of exhaustive and exclusive actions to choose from, an individual can rank the elements of such a set by his/her preferences in an internally consistent way. When faced with choice under uncertainty (i.e., when it is uncertain how actions lead to eventual outcomes) the additional axiom of independence of irrelevant alternatives provided the basis for expected utility theory (Von Neumann and Morgenstern 1944). Expected utility theory predicts that individuals will choose the best action according to their personal preferences and the constraints facing them. It has been argued that expected utility theory is the most influential model of decision making in economics (Kahneman 2011, p. 270). The influence it has had on how to determine which health treatments to fund and how to support patients' use of services is now discussed.

1.1.2 Neoclassical Economics and Determining Which Health Treatments to Fund

Expected utility theory has greatly influenced the approach adopted in many countries to determine which health treatments to fund in a public health system. Many economics guides contain recommendations that decision makers in public health systems should compare the costs and benefits of alternative treatments using the approach of cost-effectiveness analysis (CEA) and when doing so should capture the benefits of alternative treatments using a single measure called a quality-adjusted life year (QALY). A QALY captures both the increase in longevity arising from a treatment being provided and patients' quality of life (QOL) during this time (NICE 2013). QALYs for a specific health state are obtained by multiplying expected average life-years gained from treatment by a utility value which measures the strength of people's preferences for a specific health state. By combining the cost per QALY gained for alternative treatments with appropriate decision rules such as a league table, ranking treatments in ascending order of their cost per QALY gained (Briggs and Gray 2000) or a threshold approach, only funding a new treatment if its cost per QALY gained is less than a threshold level (Birch and Gafni 2004), decision makers can decide which treatments to fund in a way that will maximise the health of society within a given budget.

This approach has been widely adopted. Many countries have developed national health technology assessment guidelines requiring QALYs to be used as the reference measure of benefits in CEA.¹ Some countries have made it a legal requirement to provide treatments that pass a pre-determined cost per QALY gained threshold. The National Health Service (NHS) in England and Wales is legally required to provide funding and resources for treatments recommended by the National Institute for Health and Care Excellence (NICE) in the UK (Anderson et al. 2022). NICE recommends all treatments for which health technology

¹ This includes Australia, Canada, England and Wales, the Netherlands, Scotland, Ireland, Italy, and Hungary (RW.ERROR - Unable to find reference:4688).

assessments indicate a cost per additional QALY of less than about £30,000. More broadly it is reported that QALYs are used by a range of Government authorities across the world to inform decisions on coverage and pricing of treatments (Neumann and Greenberg 2009; Zhang and Garau 2020).

1.1.3 Neoclassical Economics and the Approach to Corresponding with Patients

Expected utility theory has also informed the widespread approach taken when corresponding with patients to encourage engagement with funded health treatments. In economics, demand for health treatments is conceptualised as a derived demand. From a demand standpoint, people want better health because of the desire to enjoy life, in all its consumption and production aspects, to a fuller extent than would be the case with less health. So, to improve their health status, they demand healthcare as a way of achieving this. For this reason, the demand for healthcare is said to be a derived demand. This is because the real demand is for health and the demand for healthcare is to help to achieve this desired demand. Applying an expected utility theory model leads to the prediction that individuals who have expressed a desire for health services will subsequently engage with communications from such services providers (e.g., return forms or attend appointments). This is in line with their personal preferences (they have already expressed a preference for a service) and the constraints facing them are low as many public health services are free at the point of use. Existing health service management practice is consistent with this prediction. Correspondence from health service providers to patients emphasise the factual aspects of health services (such as the time and date of an appointment, or the deadline to return a form by) but do not incorporate insights from psychology or behavioural economics as to what motivates engagement with such requests.

However, when patients are faced with a choice of responding today to a letter (where the benefit is in the future) or engaging in activity that provides immediate benefit, some may ignore a letter in favour of immediate gratification due to present bias. That is they may

succumb to the tendency to settle for a smaller present reward rather than wait for a larger future reward, and there is some evidence of present bias in some areas of health (Wang and Sloan 2018).

1.2 The Rise in Importance of Behavioural Economics for Public and Health Policy

1.2.1 Challenges of Behavioural Economics to Neoclassical Economics

In the latter half of the 20th century behavioural economics, which applies the inductive scientific method and psychological insights to the study of economic behaviour (Simon 1955), challenged both the assumptions and the behavioural predictions of rational choice theory (Baddeley et al. 2019; Kahneman 2011; Schimmelpfennig and Muthukrishna 2023). A fundamental challenge to a core assumption of rational choice theory was the concept of “bounded rationality” (Simon 1955). It was argued that individuals were bounded by constraints in cognitive processing ability and in information available. Bounded rationality argued therefore it was more realistic to assume that when comparing options people would engage in satisficing behaviour rather than optimising their preferences.

Important challenges to the predictions of rational choice theory were the empirically observed phenomena of preference reversals in the face of risky options (Grether and Plott 1979; Lichtenstein and Slovic 1971), the use of heuristics in decision making (Tversky and Kahneman 1974), utilities of risky prospects being non-linear (Kahneman and Tversky 1979), and mental accounting in consumer choice (Thaler 1985). Also important is the phenomenon of errors in affective forecasting (Wilson and Gilbert 2003) or, using the terminology of Kahneman (2011), there are often widespread discrepancies between decision utility (“wantability” of future consumption or of an event) and experienced utility (the actual hedonic experience of consumption or of an even).

1.2.2 Behavioural Economics and Public Policy

During the first 20 years of the 21st century there was a rapid rise in interest in the relevance of behavioural economics for public policy (Straßheim 2020). The popular book “Nudge: Improving Decisions About Health, Wealth and Happiness” (Thaler & Sunstein, 2008) provided practical solutions to important policy issues and gained a following among many politicians and policymakers (Schimmelpfennig and Muthukrishna 2023). Indeed, its authors respectively went on to advise the policy making teams of the then UK Prime Minister David Cameron and the US President Barack Obama (Baddeley et al. 2019). In 2010, the UK government set up its Behavioural Insights Team (Hallsworth and Kirkman 2020) which became an example for the translation of behavioural insights into public policy (Oliver 2017). In 2011, the popular book “Thinking, Fast and Slow” was published and increased awareness among policymakers of dual-process theories of behaviour. Dual-process theories propose (in contrast to rational choice theory) that rapid, intuitive and non-conscious cognitive processes sit alongside deliberative, reflective and self-aware ones (Hallsworth 2023).

In 2015, President Barack Obama signed an executive order for the incorporation of behavioural science insight into public policy. In addition, major international organisations – such as the European Commission (Sousa Lourenco J et al. 2016), the Organisation for Economic Development and Cooperation (Lunn 2014), the World Bank (World Bank 2015), and the World Health Organization (WHO Director-General 2022) – have issued flagship publications endorsing the use of behavioural sciences to support policy-making and implementation. It is reported that within ten years of the publication of Nudge there were over 200 behavioural insights teams around the world (Hallsworth 2023). It has been observed that the application of behavioural insights to public policy has “taken root in many ways across many countries around the world and across a wide range of sectors and policy areas” (OECD 2017).

The term “behavioural public policy” is now part of the lexicon. It refers to the “application of insights from behavioural economics specifically and behavioural science more broadly to public policy design” (Oliver 2019). A common approach to the application of findings from behavioural economics to public policy and public services is “to analyse and address practical issues in real-world settings, usually coupled with a rigorous evaluation of the effects of any interventions” (Hallsworth 2023). This typically involves redesigning an existing intervention based on insights from behavioural economics / science, testing the redesigned intervention against a pre-existing intervention using an experiment, and comparing performance on a pre-determined outcome. An advantage of this empirical approach is that it has demonstrated the impact that behavioural economics can have on public policy objectives. This has been illustrated by examples documented in literature reviews and meta-analysis of field experiments (DellaVigna and Linos 2022; Hummel and Maedche 2019).

However, applications of behavioural economics to public policy have also been criticised (for an overview see Hallsworth 2023). Two particularly important criticisms from a public policy perspective are now summarised. First, while many applications of findings from behavioural science to public services have been shown in experiments to improve outcomes, not many of these redesigns have been “scaled up” or adopted as normal practice (Mažar and Soman 2022). Second, it has been argued that applications of findings from behavioural science has focused mostly on easy to influence and easy to measure changes rather than the fundamental problems facing public policy (Chater and Loewenstein 2022; Marteau et al. 2011; Schmidt and Stenger 2021).

1.2.3 Behavioural Economics and Health Policy

For this thesis, the author undertook a review of international policy orientated documents from leading international organisations. These include the European Commission (Sousa Lourenco J et al. 2016), the OECD (Lunn 2014; OECD 2017), the World Bank

(World Bank 2015), the Institute for Government (Dolan et al. 2012), and the World Health Organization (WHO Director-General 2022). Case studies described within these documents show that applications of findings from behavioural economics have been used in many areas of healthcare to improve health decision making and health behaviours.

Applications of findings from behavioural science to improve public health through prevention have taken place in the areas of physical activity and exercise, healthy eating, smoking, healthy lifestyles, sexual health, and public safety, while health protection applications include the area of disease transmission.² Applications in the area of healthcare have taken place for the topics of patient engagement (improving attendance for health appointments and for screening, and adherence to medication), patient safety (better detection of patient deterioration), and healthcare resources (reducing unnecessary prescribing or testing). This thesis reports on three new applications of findings from behavioural economics to the health sector. These applications address some of the criticism directed to previous applications of behavioural economics to public policy.

1.3 Overview of the Thesis

A fundamental objective of public health systems is to maximise the health of society given a limited budget. This thesis applies a behavioural economics perspective to two key issues facing decision makers in health policy and healthcare when trying to achieve this objective: (a) the upstream issue of deciding which health treatments to fund and (b) the downstream issue of how to communicate with patients to maximise their engagement with funded health treatments. This thesis overcomes two of the key criticisms of behavioural public policy. First, this thesis considers a fundamental problem facing health policy. Second, the changes to current practice tested in the experiments involve relatively simple changes

² Applications of behavioural economics to health protection during the COVID-19 pandemic were also widespread, although such applications were typically not tested through RCTs (OECD 2020a).

that could be “scaled up” or adopted as normal practice with relative ease (indeed, the redesigns reported in Chapter 4 have already been adopted as national practice in Ireland).

1.3.1 Applying a Behavioural Economics Perspective to Which Health Treatments to Fund

We saw in Section 1.1 of this thesis that the use of the quality-adjusted life year (QALY) as a measure of benefit in cost-effectiveness analysis is widely used to inform decisions in public health systems on which health treatments to fund in order to maximise, within a limited budget, the health of society. The quality component of the QALY or utility value is based on people’s preferences for hypothetical health states obtained in surveys of members of the general population. When participating in such surveys, members of the general population are asked to imagine living in a range of potential health states and to express preferences for the health states. Participants are provided with functional definitions of the health states, and information on patients’ experience of living in the health states is never provided (Stiggelbout and De Vogel-Voogt 2008). Viewing through a behavioural economics perspective or lens the way in which such appraisals are obtained in health valuation surveys raises some potential concerns. The concept of “bounded rationality” (Simon 1955) suggests that survey participants will have limited information on what different forms of ill-health are like. The phenomenon of errors in affective forecasting (Wilson and Gilbert 2003) / differences between decision and experienced utility (Kahneman 2011) suggests that participants may have difficulty accurately judging what life would be like living in different forms of ill-health. Evidence of heuristics and biases in decision making (Tversky and Kahneman 1974) raise concerns about biases that could influence appraisals of hypothetical health states.

Examining the health valuation literature, we find that concerns have indeed been raised about the way in which the “quality” component of the QALY is calculated. It has been argued that the valuations of members of the general population produced using the

current approach are largely uninformed by what it is like to be in the health states so described (Karimi, Brazier, and Paisley 2017a). This is because many members of the general population have no direct experience and only limited knowledge of the health states being judged. These criticisms have been further fuelled by two empirical “puzzles” in health economics research.

One “puzzle” is that health valuation studies seem to suggest that the mean valuations or preferences for the same health states differ between nonpatients and patients, and that patients’ valuations appear to be higher than nonpatients. This is the conclusion of the influential Washington Panel on Cost-effectiveness in Health and Medicine (Gold et al. 1996). This conclusion is also supported by more recent research: notably a review (which found that patients provided higher values than the public in 22 or 58% of studies, De Wit, Busschbach, and De Charro, 2000), by a meta-analysis of prostate cancer (Bremner et al. 2007), and by a meta-analysis of multiple health states (Peeters and Stiggelbout 2010). These findings suggest that using valuations obtained from the general population results in a higher valuation for services and medicines that restore health than if patients’ valuations were used. Indeed, it has been observed that the results of cost-effectiveness analysis, upon which funding decisions are made, would change depending on whose valuations are used (Gold et al. 1996).

A second “puzzle” is that the weights assigned to different dimensions of health (e.g., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) differ depending on whether they are based on preferences obtained using current methods or based on the impact that health dimensions have on the actual well-being of people living in those health states. This holds true regardless of whether well-being is measured by life satisfaction (Dolan and Metcalfe 2012), day affect (Dolan and Metcalfe 2012) or happiness (Mukuria and Brazier 2013). A common finding across these studies (Dolan and Metcalfe 2012; Mukuria and Brazier 2013) is that overall the impact of anxiety / depression is significantly worse than

the impact of other health domains on people's well-being, but anxiety / depression is not ranked as the most undesirable health domain according to people's preferences. Depending on whether the quality component of the QALY is based on preferences obtained using the current method or empirically derived impacts of health states on people's well-being would have substantial practical implications for the allocation of health resources which is informed by QALYs.

Many reasons have been put forth for the differences in valuations of ill-health by members of the general population and patients, and for the discord between the ranking of domains of health based on the preferences of members of the general population and the actual impact of different domains of health on people's well-being (Menzel et al. 2002; Stiggelbout and De Vogel-Voogt 2008; Ubel, Loewenstein, and Jepson 2003; Versteegh and Brouwer 2016). One possibility is that there is a measurement bias or error by patients when rating their ill-health or reporting their well-being. For example, patients' self-reports could be influenced by scale recalibration, where patients norm their scale relative to other patients or relative to levels lower than their pre-ill-health experience. It could also be the case that patients may misrepresent their well-being, for instance, if overall evaluative measures of well-being do not capture moment-to-moment lower levels of well-being. However, findings from empirical studies are not consistent with scale recalibration (Baron et al. 2003; Lacey et al. 2008) nor with patients misrepresenting their well-being (Riis et al. 2005). Alternatively, the difference between the valuations of patients and nonpatients may arise due to nonpatients overestimating the negative impacts of ill-health. When undertaking appraisals of hypothetical health states, members of the general population may overestimate the negative impact as they may experience focusing illusions, focusing disproportionately on what will change (Wilson et al. 2000), and may fail to anticipate adaptation, experiencing immune neglect thereby overlooking their ability to rid themselves of negative emotions (Gilbert et al. 1998). Such biases in appraisals of hypothetical health states may occur as health judgement

exercises draw attention to the transition from one health state to another; health state descriptions typically cover only a limited number of dimensions and draw attention to those dimensions that will be adversely affected; and responses may reflect immediate affective reactions to the health state being considered (Dolan and Kahneman 2008). However, empirical studies testing for focusing illusions or failure to anticipate adaptation show mixed results (Damschroder, Zikmund-Fisher, and Ubel 2005, 2008; Ubel, Loewenstein, and Jepson 2005). Therefore, our knowledge of the reasons for the differences identified in these two empirical puzzles remains incomplete.

Furthermore, our level of understanding of the factors considered and the cognitive mechanisms at play when members of the general population appraise hypothetical health states is relatively limited. Previous studies have shown that a variety of factors are considered by members of the general population when valuing health states (Baker and Robinson 2004; Brouwer, Grootenboer, and Sendi 2009; Krol, Brouwer, and Sendi 2006; Myers, McCabe, and Gohmann 2007; van Osch and Stiggelbout 2005; van der Pol and Shiell 2007). However, there is limited evidence on the relationship between descriptions of health states, the individual's underlying beliefs, and the individual's preferences. The one study that investigates the thought processes of participants in health valuation studies (Karimi, Brazier, and Paisley 2017b) uses semi-structured interviews and a think-aloud protocol, but it only involves 21 participants. It found that participants used their imagination or other information to form a view on what living in the health states would be like. While there is a clear need for a better understanding of the cognitive mechanisms that influence such appraisals by members of the general population (Menzel et al. 2002; Stiggelbout and De Vogel-Voogt 2008; Ubel et al. 2003; Versteegh and Brouwer 2016), it is likely to be some time before this is achieved.

In summary, viewing through a behavioural economics perspective or lens raises concerns about the way in which the quality component in the QALY is currently obtained in

health valuation surveys. Rational choice theory of neoclassical economics predicts that when members of the general population are asked to appraise hypothetical health states, that the values they place on health states will be consistent with the utility they will obtain from each health state. In contrast, behavioural economics suggests that when imagining their future self in hypothetical health states, people may not have accurate expectations of what the negative consequences of ill-health on important domains of their life would be and on how their overall well-being would be impacted. This could affect the strength of people's preferences for hypothetical health states, and there may be a divergence between decision utility and experienced utility. Indeed, there is some empirical evidence that appears to contradict the predictions of rational choice theory. People with knowledge of what it is like to live in a particular health state appear to value that state very differently to people without knowledge of it. The reasons for these differences have not been empirically established.

These issues have led to calls for the use of “informed general population values” as the basis for forming the quality component in QALYs (Drummond et al. 2009; Fryback 2003). The suggestion is that respondents in health valuations studies should be provided with information on what living with specific health states is like before appraising those health states for CEA. Providing patient information would ensure people from the general population obtain information from people with experience of the health state being judged. Providing such information could also reduce biases in appraisals, such as failure to anticipate adaptation or the occurrence of focusing illusions. It has been pointed out that such an approach is consistent with the deliberations of the influential Washington Panel on Cost-effectiveness in Health and Medicine (Gold et al. 1996). The Panel recommended that fully informed members of the community would provide the best articulation of society's preferences for particular health states. But the Panel also recognised that there is a need to create a better understanding in the general population of the experiences of patients living in different health states (Stiggelbout and De Vogel-Voogt 2008). Developing ways of briefing

people from the general population on the experiences of those in particular health states is identified as one of thirteen research priorities for the future of the QALY by the Consensus Development Group (Drummond et al. 2009).

Despite the importance of exploring experience-informed general population values, little effort has gone into examining the effect on health valuations of making people from the general population more informed (Brazier et al. 2005). The one notable exception (prior to the two papers in this thesis) was McTaggart-Cowan, Tsuchiya, O'Cathain, and Brazier (2011) which examined the effect of providing the general population with audio recordings of three patients discussing adapting to life with rheumatoid arthritis. They found that provision of this information positively influences the valuation of this health condition. Audio recordings of patients discussing their experiences provide one way of briefing members of the general population on the experiences of those in particular health states, but this method is time and resource intensive (Dolan, Kavetsos, and Tsuchiya 2013). Furthermore, this approach cannot practically be used to provide experiences across a representative sample of patients. Therefore, it could not be adopted as part of normal practice in health valuation studies to inform CEA. This thesis includes two studies that overcome these limitations by providing a summary measure of patients' experience of living in ill-health and that examine the effect of obtaining experience-informed general population preferences. The approach to briefing participants used in Chapters 2 and 3 could be incorporated into existing representative surveys of people from the general population (i.e., could be incorporated into normal practice) as it is simple for participants to understand, does not add substantial labour or processing costs, and can be applied to a wide range of health conditions and generic health states.

In Chapter 2 a novel way is explored of obtaining informed preferences by briefing people from the general population on the experiences of people with a health condition. This is done by providing patients' mean ratings of their condition. Mean ratings were previously

computed based on a representative sample of patients in each condition. The study includes the three health conditions of arthritis, colostomy, and tinnitus. The study establishes whether the beliefs of members of the general population of patients' mean ratings of health conditions are accurate or not. It then tests, for these health conditions, whether there is an effect on the preferences of members of the general population of being informed of patients' mean ratings of the conditions. Furthermore, it tests whether the direction of this effect depends on whether people have overly negative or overly positive prior beliefs of patients' mean rating of the conditions.

Chapter 3 expands on the approach of Chapter 2. Rather than examining health conditions it looks at generic health states (i.e., the type of descriptions of ill-health recommended for use in HTAs). It informs participants using a summary measure that relates directly to the underlying construct that the quality component of the QALY is intended to capture, namely health-related quality of life (HRQOL). HRQOL is the consequences of ill-health for well-being. Chapter 3 explores a novel way of obtaining informed preferences by briefing people from the general population on the consequences of each health state for differences in either life satisfaction or day affect associated with living in each health state. The summary measures were previously derived in a statistical study that controls for a range of demographic and other variables.

Previous empirical findings (Dolan and Metcalfe 2012; Mukuria and Brazier 2013) are consistent with members of the general population underestimating the impact of anxiety / depression on well-being relative to other health domains. If this is the case, it may affect the relative preferences of members of the general population for hypothetical health states. Chapter 3 examines if members of the general population selectively underestimate the well-being consequences of “anxiety / depression” as compared to the well-being consequences of “self-care” or of “pain / discomfort”. In addition, Chapter 3 tests if informing people of the actual changes in well-being associated with health states changes appraisals of their relative

undesirability. Specifically, it tests if being more informed about the association between health states and differences in well-being results in living with “anxiety / depression” being judged as less preferred relative to moderate problems with “self-care” or “pain / discomfort”, and so of increased priority for being addressed by health treatments.

1.3.2 Applying a Behavioural Economics Approach to Correspondence with Patients

We saw in Section 1.1 of this thesis that the expected utility theory model leads to the prediction that individuals who have expressed a desire for health services will subsequently engage with communications from providers of such services (e.g., return forms or attend appointments). Such behaviour is in line with their personal preferences (having already expressed a desire for a service) and the constraints facing them are low as many public health services are free at the point of use. The widespread approach to correspondence with patients in relation to funded health treatments is based on this belief. Correspondence from health service providers to patients emphasises the factual aspects of health services (such as the time and date of an appointment, or the deadline to return a form by). Such correspondence typically does not incorporate insights from psychology or from behavioural economics as to what motivates engagement with requests.

A notable exception to current practice is the use of findings from behavioural economics to inform the content of Short Message Service (SMS) reminders for hospital appointments (Hallsworth et al. 2015) in order to reduce “did not attend” (DNA) rates. A DNA occurs when a patient unexpectedly fails to attend an appointment. DNAs for hospital appointments lead to worse care for patients, the inefficient use of staff, and increased waiting times (Karter et al. 2004; Murray 2000). The use of SMS reminders helps to reduce DNAs and the application of behavioural economics to SMS content can improve effectiveness further. Nevertheless, there are some limitations to the use of SMS reminders. First, not all patients with a scheduled appointment receive a SMS reminder. It is not unusual for hospital administrators to possess accurate records of mobile telephone numbers for only

20% of individuals with appointments (Hallsworth et al. 2015). Second, and perhaps more fundamentally, a substantial proportion of people on waiting lists would not attend an appointment regardless of how well crafted a SMS reminder is. This is because waiting lists are not fully accurate and many people on waiting lists no longer require a treatment. Experience shows that a proportion of patients remain on waiting lists even though they may have had the treatment in another hospital, feel better, or no longer want a procedure.

Another way to reduce DNAs is to improve the accuracy of waiting lists by undertaking administrative validation. This is a process where administrators in a hospital write to patients on waiting lists to check if patients still require a procedure or wish to be removed from the waiting list. An advantage of administrative validation is that it can include all relevant patients on waiting lists, as hospital administrators possess accurate postal records. By using administrative validation to improve the accuracy of hospital waiting lists, DNA rates can be reduced as appointments are less likely to be offered to people who no longer require them. In addition, patient waiting times can be improved as procedures can be offered sooner to patients who require treatment (National Treatment Purchase Fund (NTPF) n.d.). A challenge for hospital administrators when undertaking administrative validation is that a substantial percentage of patients typically do not return (DNR) a “validation letter” to confirm whether the procedure is still required. DNR rates of 25% are common (A. Green, personal communication, June 15, 2017).

A potential way to reduce DNR rates to administrative validation is to deploy nudges in validation letters. A nudge is “any aspect of the choice architecture that alters behaviour in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler, R. H. and Sunstein, C. R. 2009:6). The use of nudges to improve the outcomes of public services has been endorsed by international organisations such as the Organisation for Economic Cooperation and Development (Lunn 2014), the World Health Organization (WHO Director-General 2022), and the World Bank (World Bank 2015).

However, we are not aware of any published literature that has tested the effect of incorporating nudges into validation letters. Chapter 4 of this thesis reports on an application of nudges to the administrative validation of waiting lists which altered the content of validation letters to include design features (such as specific phrases and presentation devices) with the aim to decrease the DNR rate. The study in Chapter 4 followed a collaborative approach involving a government department, state agencies responsible for hospital waiting lists, and a relatively small expert behavioural science advisory group drawn from universities and not-for-profit bodies. The redesigned letter included multiple nudges, many of which have not previously been used in a health context.

2 Valuing the Q in QALYs: Does Providing Patients' Ratings Affect Population Values?

Previously published as: Murphy, R. P., C. J. Boyce, P. Dolan, and Alex M. Wood. 2020. "Valuing the Q in QALYs: Does Providing Patients' Ratings Affect Population Values?" *Health Psychology* 39(1):37–45. doi: 10.1037/hea0000806.

Acknowledgements: The authors would like to thank Leonhard Lades and David Comerford for comments on an early draft of this paper. [Comments were provided as members of the internal review panel for the year 1 upgrade report of Robert P. Murphy]

2.1 Abstract

Objective: QALYs are used to measure the health benefits associated with treatments.

QALYs are derived from objective mortality data weighted by assessments made by the general population of the impact on health-related quality of life associated with particular health states. In this study a simple change is introduced to improve the validity of QALYs by giving raters information about how people living in the health states rate the health states.

Methods: Participants from the general population ($N = 155$) judged three health states using a standard valuation technique after being randomly allocated to one of two groups. The intervention group was given patients' mean ratings of their own health states from worst to best imaginable health (0-100 scale) before providing their valuations, whilst the control group was given this information only after providing their valuations. The participants in both groups also indicated whether patients' mean ratings were higher, broadly similar, or lower than they previously expected.

Results: When the mean ratings given by patients were higher (lower) than expected, participants in the intervention group provided significantly higher (lower) valuations than participants in the control group. These findings show that participants adjust their valuations of a health state in the direction of the appraisals of those experiencing that state.

Conclusions: Insofar as policy-makers are committed to valuing health states using valuations given by people from the general population, it is desirable to elicit more informed values by providing people with information on how patients rate those states.

Keywords: Quality-Adjusted Life Years, judgment, decision making, resource allocation

2.2 Chapter Introduction

Quality-Adjusted Life Years (QALYs) combine two attributes of value – quality of life and length of life – into a single number. Each health state is assigned a value between zero (for death) and one (for full health) and that value is then multiplied by how long the state lasts. For example, one year of life in full health is equivalent to one QALY. When combined with the costs of intervention, QALYs allow policy-makers to determine an efficient allocation of resources (i.e., the one that generates the most QALYs given the resources available to healthcare). The QALYs method is accepted internationally as the gold standard in comparing the impact of medical treatments (NICE 2013) and it is the preferred measure of benefit for health technology assessments in England, Wales, Scotland and the Netherlands (Karimi et al. 2017b). Concerns have been raised, however, about how best to capture the Q in the QALY; that is, how best to value quality of life. It is widely accepted that the preferences of the general public should be taken into account, but there are doubts about whether they take sufficient account of patients' experiences of specific health states (Dolan and Kahneman 2008; Drummond et al. 2009). This paper is the first to test the impact of a simple, quick, and low-cost way of briefing participants in health valuation surveys about the experiences of patients by providing patients' mean ratings of their conditions. It is also the first study to test whether the effect of giving patients' mean ratings of their conditions to participants from the general population interacts with participants' prior beliefs about how such patients appraise their conditions.

When computing QALYs for the purposes of health technology assessment, expected length of life (the LY in the QALY) is established by epidemiological research, whereas quality of life (the Q in the QALY) is estimated by asking people what value they ascribe to specific health states. Most countries elicit valuations from members of the general population, who are asked their preferences over different states of health that vary according to their severity (e.g., in terms of physical functioning and mental health). The use of

population values is consistent with the idea that those who pay for publicly funded healthcare, and who are also the prospective beneficiaries of it, should have some say in how the benefits afforded by that expenditure are valued.

2.2.1 Valuations of the Imagined Compared to the Experienced

A potential problem is that many members of the general population have no direct experience and only limited knowledge of the health states being judged (Kind and Dolan 1995). Furthermore, this lack of knowledge is not addressed in health valuation studies, where the descriptions of health states are typically limited to information on salient dimensions of health (Stiggelbout and De Vogel-Voogt 2008). Information on patients' experience of living in the health states is never provided. As a result, it has been argued that the valuations produced using the current approach are largely uninformed by what it is like to be in the health states so described (Brazier et al. 2005a; Karimi et al. 2017a).

Those with knowledge of what it is like to live in a particular health state may value that state very differently to people without knowledge of it. A review of 38 studies by De Wit, Busschbach, and De Charro (2000) found that patients provided higher values than the public in 22 studies (58%); the public provided higher values than patients in two (5%); there was no difference in 11 (29%), and there were contradictory findings in three (8%). A recent meta-analysis by Peeters and Stiggelbout (2010) also found that, across all studies examined, patients on average gave higher values. Differences in nonpatients' exposure to health states (through family, friends, work or volunteering) is associated with differences in perceptions of and valuations of health states (Schkade and Kahneman 1998; Ubel et al. 2001).

Stiggelbout and De Vogel-Voogt (2008) argued that interpretation and judgment are two key steps in the process of health state appraisal. According to the stress-appraisal model of (Lazarus and Folkman 1984), a person's primary appraisal upon encountering a health-related stimulus (in this instance a written description of a health state) will be whether the health state represents harm, a threat, or a challenge. The primary appraisals of people with

and without experience of what it is like to live in the health state are likely to differ. In the secondary appraisal, according to Lazarus and Folkman, people consider the resources (internal and external) available to them to cope with the health state, and again there are likely to be different perceptions. For example, nonpatients may be unaware of coping strategies that patients use and of the possibilities of adapting to a health state. As a result, for many health states, people with knowledge of what it is like to live in a given health state are likely to have less negative primary and secondary appraisals of that state than people without such knowledge, and hence to value that health state less severely.

2.2.2 Obtaining Experience-Informed General Population Valuations

One way to overcome a lack of knowledge among the general population is to provide respondents with information on patients' appraisals of their experience of living with a health state before respondents judge that health state. Indeed, studies have called for patients' judgments of what it is like to live in particular health states to be provided to the general population when they judge those health states (e.g., Fryback 2003; Menzel et al. 2002). This position is referred to as a "third position" that implies the use of "informed general population values" (Brazier et al. 2005). The approach is consistent with recommendations of the influential Washington Panel on Cost-effectiveness in Health and Medicine (Gold et al. 1996) which are, firstly, that fully informed members of the community would provide the best articulation of society's preferences for particular health states (Brazier et al. 2005), and secondly, that there is a need to create a better understanding in the general population of the experiences of patients living in different health states (Stiggelbout and De Vogel-Voogt 2008).

Indeed, developing ways of briefing people from the general population on the experiences of those in particular health states is identified as one of thirteen research priorities for the future of the QALY by the Consensus Development Group (Drummond et al. 2009). Despite this, little effort has gone into making people from the general population

more informed when undertaking health valuation studies (Brazier et al. 2005). In one notable exception, McTaggart-Cowan, Tsuchiya, O'Cathain, and Brazier (2011) examined the effect of providing the general population with audio recordings of three patients discussing adapting to life with rheumatoid arthritis. They found that provision of this information positively influences the valuation of this health condition. Audio recordings of patients discussing their experiences provide one way of briefing members of the general population on the experiences of those in particular health states, but this method is time and resource intensive (Dolan et al. 2013).

2.2.3 The Current Study

This study explored a novel way of briefing people from the general population on the experiences of people with a health condition, through providing patients' mean ratings of their condition, on a 0-100 scale, from worst to best health imaginable. For each of the three conditions that meet the study selection criteria – arthritis, colostomy and tinnitus – participants were provided with patients' mean rating of the health condition and asked whether the patients' mean rating of a given condition was higher, about the same, or lower than expected. The control group was briefed after rating the conditions to establish whether their prior belief of patients' mean rating was accurate, whereas the intervention group was briefed before rating the conditions to establish whether their prior belief of patients' mean rating was accurate *and* to inform their own appraisal of the conditions. For participants who believed patients' mean rating of a condition would be worse than it is, we expected the mean valuation of the condition to be higher in the intervention group than in the control group because the intervention group's misconceptions would be corrected in an upward direction. For participants who believed patients' mean rating of a condition would be better than it is, we expected the opposite pattern, and we did not expect an impact for participants whose prior beliefs of patients' mean rating were accurate. Thus, we expected to find a cross-over interaction, depending on the expectations about the patient ratings.

2.3 Methods

2.3.1 Participants

This study was conducted online in the United States using Amazon Mechanical Turk (MTurk). MTurk has been judged to be a source of high quality data (Crump, McDonnell, and Gureckis 2013) and online health valuation surveys have been deemed reliable (Bansback et al. 2012). Eligible participants were adult (aged 18 and over) members of MTurk, had not undertaken pilot surveys for this study, and had completed an informed consent form. To support the quality of data collection two additional criteria, often used by MTurk requesters, were applied: (a) participants needed to have previously completed at least 100 surveys and (b) they were required to have a 95% approval rating. Participants who completed the survey were compensated for their time, in line with MTurk norms. Ethics approval was provided by the University of Stirling Management School's Ethics Committee.

Consenting participants ($n = 227$) accessed the survey between 18th August to 4th September 2016. Of these, six declined to participate after reading the introduction, three in the intervention group discontinued the survey before receiving the intervention, 22 did not complete the survey (split evenly between the control and intervention groups), and 196 completed the survey. Two a priori criteria for exclusion from analysis were set: (a) not giving sufficient consideration to the task, defined as completion in less than 50% of the median time; and (b) responding with logically incoherent judgments, defined as rating one health condition at the anchor 100 (*best health you can imagine*). Of the 196 people who completed the study 155 met the inclusion criteria (67 control group; 88 intervention group). The Consolidated Standards of Reporting Trials (CONSORT) participant flow chart is shown in Figure 2.1.

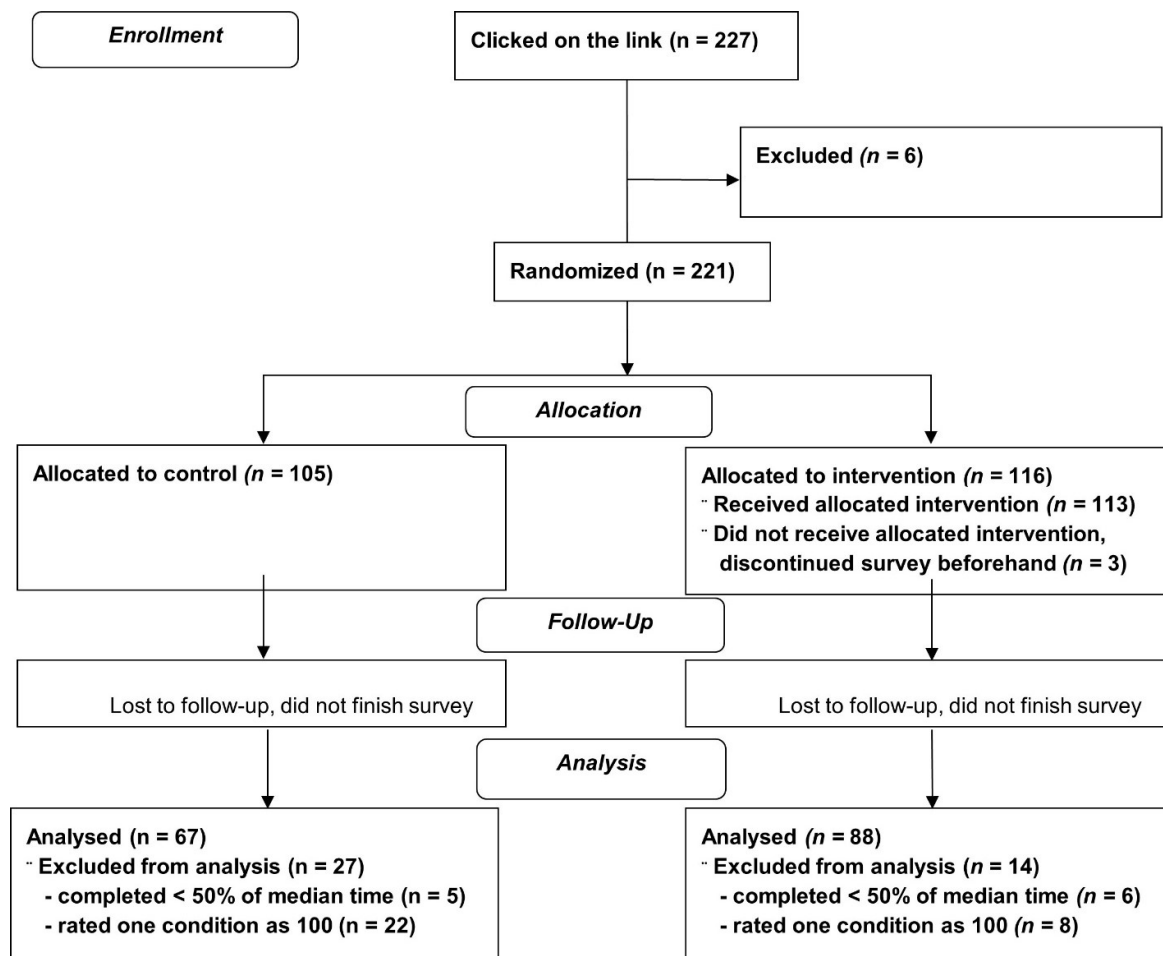


Figure 2.1 CONSORT Participant Flow Diagram.

To check that the study results were not dependent on differences in exclusions on criterion (b) between the control and intervention groups (23.40% versus 7.78%), the tests of the study hypothesis were re-run treating responses from these participants as missing data and imputing their valuations (differences did not seem substantial for noncompletion rates, 10.48% versus 12.07%, nor completion in less than 50% of median time, 5.32% versus 5.89%). For these participants the valuation was estimated through a multiple imputation technique (Rubin 1987) which imputes a series of missing values based on estimates from other observed variables and which more appropriately accounts for the statistical uncertainty in the imputations than many other commonly used imputation techniques, such as sample wide or item averages. Specifically, multiple imputation using chained equations (MICE: White, Royston, and Wood 2011) was used. Details on the specific MICE applied are

presented in the Supplementary File. As discussed in the Results, this did not have implications for the study conclusions.

2.3.2 Measures and Materials

Valuation technique. This study used the EQ VAS valuation technique, a standard method used in health-related quality of life research (Williams 1990 p. 200). Participants “warmed-up” by describing their current health using the two parts of the EQ-5D-3L, namely (a) its descriptive system (a generic health state descriptive system) and (b) its visual analogue scale (EQ VAS) – this is a vertical rating scale (thermometer) where the anchors are 0 (*worst health you can imagine*) and 100 (*best health you can imagine*). After this, participants used the EQ VAS scale to judge the hypothetical health conditions. A staff member from the EuroQol Research Foundation checked this study’s presentation of the instrument in a draft of the questionnaire and approved it subject to changing the format of one word (to ensure consistency with the instrument).

Health conditions. This study used the health conditions rheumatoid arthritis, colostomy, and tinnitus because they were the only health conditions which met the five study inclusion criteria. As international inventories of patients’ appraisals of their actual health conditions do not exist, two meta-analyses (Dolders et al. 2006; Peeters and Stiggelbout 2010) and a review (De Wit, Busschbach, and De Charro 2000) of patients’ and nonpatients’ judgments of ill health were searched to identify conditions which met the five selection criteria. The criteria were:

1. The health scenario was of a specific condition (rather than a generic health state) that was not predominantly specific to one gender.
2. The condition was appraised by patients of the condition rather than patients of a different condition or proxies such as patients’ family members or health professionals.
3. The condition was rated using the EQ VAS.

4. Studies showed that patients rate the condition higher than people from the general population do.

5. The condition was rated by a minimum of 50 patients. The intention was for participants to take the information provided into account, and not to discount it as unreliable. A sample size of at least 50 was a priori selected as a “rule of thumb” minimum sample which participants may require to regard the information as reliable. It also turned out to be a natural break point, as after criteria 1 to 4 were applied only four conditions remained; laryngeal-cancer which was rated by only 20 patients and the three conditions used in this study which were rated by more than 50 patients.

Questionnaire. Participants completed an online survey with seven sections as summarized below. The order in which participants completed the sections differed between the control and intervention groups.

1. Welcome. This explained why the study was being undertaken, stressed that participation was voluntary and that participants could withdraw at any stage, explained what partaking in the survey involved, confirmed that all information provided would be kept confidential, stated who was funding the survey, and provided an informed consent form.

2. Your health. As is standard practice in health valuation studies, participants described their own health using the EQ-5D-3L’s generic health state descriptive system and assessed their general health using its EQ VAS rating scale.

3. Description of health conditions and how to value them. Participants were provided with a description of each health condition along with an explanation of how to complete the valuation question. In line with established practice, participants were instructed to use the EQ VAS scale to rate the health state, from 0 for worst health imaginable to 100 for best health imaginable. This was followed by an instruction on where to enter their answer.

The description of each health condition consisted of (a) a summary description and (b) a detailed functional description (provided in box format, as is standard practice in health

valuation studies). The summary descriptions are presented below, and the detailed definitions are provided in the Supplementary File.

Arthritis is an illness that leads to inflammation in the joints; a joint is where two bones meet, such as your elbow or knee. This inflammation can lead to stiffness, swelling, redness, tenderness, and pain.

A colostomy is an operation involving the surgical redirection of your bowels through a hole created in your gut, called a stoma.

Tinnitus is the perception of noise or ringing in the ears.

The detailed description of each health condition consisted of five sentences: Each sentence described a salient feature of the condition. The descriptions of arthritis and colostomy were taken from published health valuation studies in which people from the general population were provided with the descriptions of the health conditions and asked to value the conditions (McTaggart-Cowan 2011; Smith et al. 2006). The description of tinnitus used in the one study found in which people from the general population rated tinnitus (Happich and von Lengerke 2005) was not used: The survey in that study was undertaken in German and an English translation of the description of tinnitus was not available. Therefore, a description of tinnitus from the website of the nonprofit Mayo Clinic, which is regarded as a leading hospital in the US, was used (Mayo Foundation for Medical Education and Research).

4. *Valuation of health conditions.* Participants rated the three conditions in random order. Each condition was rated on a 0-100 scale from worst to best health imaginable using the EQ VAS.

5. *Ease of completion.* Participants were asked how easy or difficult they found the specific task of rating the health conditions, and how easy or difficult they experienced the survey to be overall.

6. Presentation of patients' mean rating of a condition. In this section participants (a) viewed information on patients' mean ratings on the 0-100 scale; and (b) were asked whether the mean rating of people with each health condition was higher, about the same, or lower than expected. A respondent could in principle have provided different answers for different health conditions (e.g., indicated that they expected patients would have rated one condition higher, but indicated that they expected patients would have rated another condition lower). Patients' mean assessments of the three health conditions were provided in text and in a graphic. The text read: "When asked to choose a number between 0 and 100 that best represents their current health the average answer was: 63 for people with a colostomy, 60 for people with arthritis, 54 for people with tinnitus" (the last three items were presented as bullet points; the rating and the health condition text was bolded). The graphic showed an image of the EQ VAS 0-100 scale with patients' mean rating of each condition superimposed to the right of the scale. Information on patients' appraisal of a condition was also available to the intervention group when valuing a condition. The patient sample size in each survey was 184 for arthritis (McTaggart-Cowan et al. 2011), 95 for colostomy (Smith et al. 2006) and 208 for tinnitus (Happich and von Lengerke 2005).

7. Personal characteristics. Participants answered questions on their socio-economic status, namely: gender, age, highest educational attainment, and principal economic status.

2.3.3 Experimental Manipulation

Both groups saw the same questionnaire. The control group saw sections 1 to 7 of the questionnaire in order. That is, the process for the control group followed standard practice for health valuation studies; the only information provided at the point of judging each health condition was the description of the condition, and all information regarding how patients rated their own conditions was only presented after the participants had completed the valuation task so that it could not have affected their responses. The intervention group judged the health conditions after being informed of patients' ratings. That is, for the

intervention group Section 6: Presentation of patients' ratings came before Section 4: Valuation of health conditions and Section 5: Ease of completion. Participants were randomly assigned to the control group or the intervention group by a computerized random number generator.

2.3.4 Statistical Analysis

Analysis was conducted with a 2 (between: Group [control, intervention]) $\times$ 3 (within: Health Condition [arthritis, colostomy, and tinnitus]) $\times$ 3 (between: Prior Belief of Patients' Mean Rating ([imagined to be worse than is experienced, imagined to be similar to as is experienced, imagined to be better than is experienced])) mixed analysis of variance (ANOVA). The study prediction was that participants with inaccurate prior beliefs of patients' mean rating of a given condition would adjust their own valuation of a given condition in the direction of the mean rating of those living with a given condition. Therefore, the test of the study hypothesis was whether there was a significant interaction between prior belief of patients' mean rating and group. No other result of the ANOVA has importance for the test of the hypothesis.

2.4 Results

2.4.1 Descriptive Analyses

Participant characteristics. See Table 2.1 for participant characteristics.

Ease of completion. Respondents found the survey "Very easy" (29.03%), "Easy" (48.39%), or "Neither easy nor difficult" (16.77%); only 5.81% found it "Difficult" and none indicated that it was "Very difficult". Participants were also asked whether they found the specific judgment task easy (rating a health condition on a continuum from worst to best health imaginable). Only 2.58% "Disagreed" and 1.29% "Completely disagreed" that the rating task was easy, while 45.81% "Completely agreed", 40.00% "Agreed" and 10.32%

“Neither agreed nor disagreed”. These results suggest that the difficulty of the task had not posed a problem for the study.

Prior belief of patients’ mean rating. Participants who indicated that patients’ mean rating was a “higher rating than expected” for a given health condition were classified as having “imagined the condition to be worse than is experienced”. Participants who indicated that patients’ mean rating was “about the same rating as expected” were classified as having “imagined the condition to be similar to as is experienced”. Finally, if the respondent indicated that patients’ mean rating was a “lower rating than expected” they were classified as having “imagined the condition to be better than is experienced”. Response bias could have affected answers to this question because the intervention group answered the question about prior beliefs after being informed but prior to rating the health conditions, whereas the control group answered the question about prior beliefs after rating the conditions. For instance, a form of social desirability bias could have led participants in the control group to be more likely to indicate that patients’ mean rating of a condition (which was provided to them after they rated the condition) was broadly similar to as they thought it would be. Nevertheless, the answers to this question by the intervention group was not different to the answers by the control group, $\chi^2(2, N = 465) = 0.09, p = 0.958$. There is no evidence of a response bias.

Table 2.1 Participant Characteristics

Variable	Intervention	Control
Age		
18-19	3%	3%
20-29	33%	37%
30-39	32%	24%
40-49	8%	18%
50-59	14%	16%
60-69	8%	1%
Over 70	2%	0%
Highest educational attainment		
Not a high school graduate	2%	1%
High school graduate	13%	9%
Some college, but no degree	33%	27%
Associate degree	11%	19%
Bachelor's degree	30%	36%
Advanced degree or more	11%	7%
Principal economic status		
Employed or self-employed	64%	63%
Retired	5%	1%
Taking care of home	9%	13%
Student	13%	7%
Unemployed	3%	9%
Unable to work due to long-term sickness/disability	7%	6%
Total	88	67

There was, as expected, variation in participants' prior beliefs about how patients value health conditions (see Table 2.2). Participants indicated that their prior beliefs were "broadly similar" to patients' mean rating in less than half of the cases (around four in ten). In a substantial proportion of cases, health conditions were imagined to be worse than is reported by patients (around four in ten cases) or were imagined to be better than is reported by patients (around two in ten cases). This finding is consistent with the observation that using the current approach to obtain valuations from the general population yields values that are largely uninformed by what it is actually like to be in the health states (Brazier et al. 2005).

Table 2.2 Prior Belief by Group of Patients' Mean Rating of Health Conditions

Group	Prior belief of patients' mean rating, Number (%)		
	Imagined to be worse than is experienced	Imagined to be similar to as is experienced	Imagined to be better than is experienced
Intervention	96 (36.36)	108 (40.91)	60 (22.73)
Control	72 (35.82)	81 (40.3)	48 (23.88)

Valuations. The mean valuations are illustrated in Figure 2.2 (the 95% confidence intervals for a group that do not bound the mean of the other group indicate statistically significant differences). For both the intervention and control group, there appeared to be a positive relationship between prior beliefs of patients' mean rating of their health conditions and valuations of the conditions by the participants. This simply shows that the valuations participants gave of a given condition were strongly related to their prior impressions of how people with a given condition would rate it. The pattern in Figure 2.2 is consistent with the study prediction of a cross-over interaction between prior beliefs and provision of patients' information. Intervention-group participants who underestimated patients' mean ratings gave valuations that were higher than the valuations provided by control-group participants who underestimated patients' mean ratings (effect sizes averaged $d = 0.76$). For participants with broadly accurate prior beliefs there appeared to be no difference in the valuations given by the intervention group and the control group. While, intervention-group participants who overestimated patients' mean ratings gave valuations that were lower than the valuations provided by control-group participants who overestimated patients' mean ratings (effect sizes averaged $d = 0.38$). The next section formally tests for an interaction between prior belief of patients' mean rating and group.

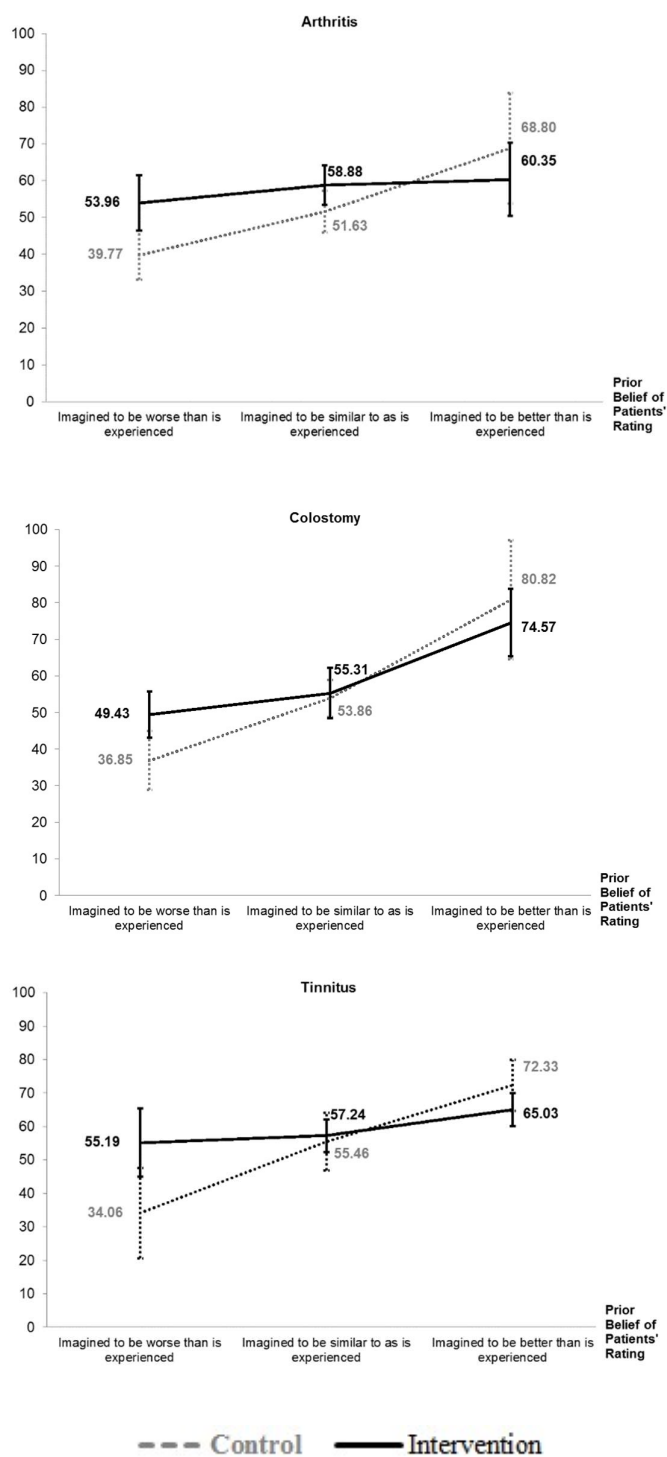


Figure 2.2 Mean valuation (EQ VAS) of conditions by prior belief of patients' mean rating and group.

Note. Error bars represent the 95% confidence intervals.

2.4.2 Test of Study Prediction

Analysis of the study prediction of a cross-over interaction between participants' valuation of a health condition and their expectations about patients' mean ratings of a condition was undertaken with a 2 (between: Group) $\times$ 3 (within: Health Condition) $\times$ 3 (between: Prior Belief of Patients' Mean Rating) mixed ANOVA (see Table 2.3). There was a main effect of prior belief about patients' mean rating, $F(2, 447) = 46.95, p < .001$. This means that participants own valuation of a given health condition was influenced by whether participants believed patients typically rate a given condition as worse, similar or better than the mean rating of that condition by patients. So, both groups' judgments were influenced by prior beliefs of patients' mean rating of the health conditions. There was a main effect for group, $F(1, 447) = 4.11, p = .043$, with higher mean valuations by the intervention group, 57.00 ($SD = 19.01$), than the control group, 52.37 ($SD = 23.81$), $d = 0.22$. However, this finding should be interpreted with caution given the study prediction (confirmed below) that prior beliefs will interact with group. The positive intervention effect on valuations pooled across the three health conditions occurred because the subgroup of participants who underestimated patients' mean ratings was larger than the subgroup of participants who overestimated patients' mean ratings, and there was a larger intervention effect size for the former subgroup.

Indeed, there was, as predicted, an interaction between prior belief about patients' mean rating and group, $F(2, 447) = 10.41, p < .001, \omega^2 = .16$. This means that the effect of prior expectations about health conditions on valuations of conditions made by participants from the general population was reduced when the descriptions of conditions were supplemented with patients' experience-based ratings of health conditions. This interaction between prior belief of patients' mean rating and group, $F(2, 555) = 9.67, p < .001$ remained when the ANOVA was re-run treating responses from participants with logically incoherent judgments (i.e., who rated one health condition as 100 or best health imaginable) as missing

data and imputing the VAS values using MICE (rather than excluding responses from such participants). Details are in the Supplementary File. The 3-way Group $\times$ Health Condition $\times$ Prior Belief of Patients' Mean Rating interaction was not significant, $F(8, 447) = 1.00$, $p = .437$. This finding was not consistent with the interaction between group and prior belief of patients' mean rating differing across the three health conditions (i.e., it was not consistent with the interaction being driven by one health condition only). The results lent support to the study hypothesis. Overall, the results showed that for both groups prior beliefs about patients' ratings of the health conditions influenced valuations of the conditions on a worst to best health imaginable continuum, but the effect of prior beliefs was reduced for the intervention group who had seen patients' mean rating of the conditions.

Table 2.3 Results of Mixed ANOVA

Predictor	Partial Sum	Mean			
	of Squares	<i>df</i>	Square	<i>F</i>	<i>p</i>
Prior belief of patients' mean rating	33777.71	2	16888.86	46.95	0.000
Group	1477.82	1	1477.82	4.11	0.043
Health condition	479.17	2	239.58	0.67	0.514
Group $\times$ health condition	101.33	2	50.67	0.14	0.869
Group $\times$ prior belief of patients' mean rating	7492.56	2	3746.28	10.41	0.000
Group $\times$ health condition $\times$ prior belief of patients' mean	2871.57	8	358.95	1.00	0.437
Residual	160805.90	447	359.74		
Total	210827	464	454.37		

2.5 Discussion

2.5.1 Strengths and Limitations

This study evaluates the effect of making a simple change to the current process through which the information used to estimate the Q in the QALY is obtained. This

approach is likely to be broadly acceptable among stakeholders on the basis that more faith can be placed in values obtained from the general population when those values are better informed. The approach used in this paper can be incorporated into existing representative surveys of people from the general population as it is simple for participants to understand, does not add substantial labor or processing costs, and can be applied to a wide range of health conditions and generic health states. This study shows that the provision of patients' mean rating of their condition influences the valuations of people from the general population with inaccurate perceptions of how patients typically rate a condition. The judgments of people with inaccurate perceptions of patients' mean rating of a given health condition change to align more with patients' reported experiences.

This study was designed to provide a proof of concept that it is possible to elicit more informed values of health states by providing people with information on how patients rate those states and it includes several limitations. There is the possibility that the intervention created a focusing effect (Schkade and Kahneman 1998) or an anchoring effect (Tversky and Kahneman 1992). In the case of a focusing effect the intervention group may have placed too much importance on patients' reported experience. With an anchoring effect, on the other hand, the intervention group's valuations may have stayed close to patients' mean rating even if that information was not relevant to the judgment task. Notwithstanding these concerns, providing patients' mean ratings seems warranted; doing so is arguably a requirement if people from the general population are to provide informed preferences. The informed preference theory of utility (Harsanyi 1997) contends that "in welfare economics and in ethics a person's utility function should be defined in terms of his hypothetical informed preferences" (p. 134) and that preferences are informed only when they are "freed from the distorting effects of factual errors" (p. 140). Other research shows that when valuing hypothetical health states people combine their interpretation of the health state description

with their own experiences of ill health and their imagination (Karimi et al. 2017b), but what they imagine as the consequences of ill health is not accurate (Karimi et al. 2017a).

2.5.2 Directions for Future Research

It will be important to examine the degree to which effects found in this study apply to a wider range of health conditions, for generic health states, for choice-based judgment methods and for a random sample of the general population. It would be useful to examine the direct mechanisms that lead valuations of health states obtained from the general population to be influenced by providing patients appraisals of their health states. Future studies adopting this study's approach should be powered to test not only for interactions between prior beliefs and information provision, but also to detect the effect of information provision on the mean valuations of individual health conditions. This study focused on three conditions that patients value higher than people from the general population do, but this study's approach should also be applied to conditions, such as mental health conditions, that patients do not value higher than nonpatients (Revicki, Shakespeare, and Kind 1996; Rosser and Kind 1978). It would also be useful to investigate whether there is an asymmetry in the size of the effect of providing patients' experience to people from the general population. Providing information on the experiences of patients may be more likely to reduce cognitive biases (such as loss-framing, failure to anticipate adaptation and focusing illusion) among respondents who misconceive an ill health state as worse (as opposed to better) than is typically experienced by patients.

More discussion between stakeholders on the possible approaches of obtaining QALY weights (i.e., using patient valuations, general population valuations or experience-informed general population valuations) would be helpful. Whether it is appropriate to use patients' judgments that are possibly influenced by adaptation to their conditions to inform resource allocation decisions is a topic that is likely to arise. The answer to this vexing question, in part, depends on whether the major elements of adaptation are perceived as admirable

achievements or are perceived as deficiencies. Menzel, Dolan, Richardson, and Olsen (2002) have argued that only when adaptation is comprised primarily of cognitive denial and suppressed recognition of full health should patients' actual valuations not be provided to people from the general population when judging health states. A better understanding of patients' experience of living with different forms of ill health (such as which major elements of adaptation or other factors have the most influence), and a fuller appreciation of the cognitive mechanism that influence judgment of ill health states by patients and by nonpatients is needed. While a number of helpful frameworks have been proposed to support the study of these issues (Brazier et al. 2005; Menzel et al. 2002; Stiggelbout and De Vogel-Voogt 2008; Ubel et al. 2003) it will be some time before a fuller picture emerges. In the meantime, it would be informative to obtain the views of the general population on the appropriateness of using experience-informed valuations to value the Q in the QALY.

2.5.3 Application

This study illustrates that it is possible to brief people from the general population of patients' experience of ill health in a way that can be incorporated into the current approach of obtaining QALY weights for health conditions. If this study's findings hold for additional health conditions, generic health states, and using choice-based judgments, then the provision of patients' experiences could be incorporated into the current approach to produce experience-informed general population QALY weights to capture the impact of medical treatments.

3 Do Misconceptions about Health-Related Quality of Life Affect General Population Valuations of Health States?

Previously published as: Murphy, RP, CJ Boyce, P. Dolan, GDA Brown, and AM Wood. 2023. “Do Misconceptions About Health-Related Quality of Life Affect General Population Valuations of Health States?” *Value in Health* 26(5):750–59. doi: 10.1016/j.jval.2022.10.009.

3.1 Abstract

Objective: Healthcare resource allocation decisions are often informed by the expected gains in patients' quality-adjusted life years (QALYs). Misconceptions about ill-health's consequences for quality of life may however affect evaluations of health states by the general population and hence affect resource allocation decisions informed by QALYs. We examine whether people selectively misestimate the quality of life consequences of moderate anxiety / depression as compared to other dimensions of health, and we test whether informing people of actual changes in quality of life associated with health states changes appraisals of their relative undesirability.

Methods: UK general population participants ($N = 1259$; in 2017) expressed preferences over moderate problems: of anxiety / depression, with self-care, of pain / discomfort. A randomized control trial design was used whereby a control group was given a functional description of each health state while two intervention groups were additionally given information on the actual differences in either life satisfaction (LS) or day affect (DA) associated with experiencing each health state.

Results: The LS (DA) group reported a higher preference for avoiding living with moderate anxiety / depression, being 13.4% (13.9%) more likely to choose it as most undesirable.

Conclusions: Informing people of the change in LS or DA associated with health states before they appraise them is a feasible way to obtain informed preferences.

Keywords: Judgment, decision making, resource allocation, Quality-Adjusted Life Years.

3.2 Highlights

- Previous research conjectures that misconceptions about ill-health's consequences for quality of life (QOL) may affect evaluations of health states by the general population.
- We found that people who were given information on the QOL consequences of experiencing each health state evaluated, reported a higher relative preference for avoiding living with moderate anxiety / depression.
- Our results show that providing information about QOL consequences of health states affects preferences, so policymakers may wish to consider supplementing existing QALY outcome measures with QALYs computed using general population preferences that are informed about the QOL consequences of health states.

3.3 Introduction

A fundamental objective of most public health systems is to determine which health treatments to fund to maximize the health of society given a limited budget. This objective can be achieved through cost-effectiveness analysis (CEA) followed by the use of appropriate decision rules such as a league table (Briggs and Gray 2000) or a threshold approach (Birch and Gafni 2004). A key decision in CEA is how best to define and capture different health outcomes using a single measure. The quality-adjusted life year (QALY) is a measure of health outcomes that captures both the increase in longevity arising from a treatment being evaluated and a patient's quality of life (QOL) during this time (NICE 2013). QALYs for a specific health state are obtained by multiplying expected average life-years gained from treatment by a utility value between -1 and 1 (Weinstein, Torrance, and McGuire 2009) which measures the strength of people's preferences for a specific health state relative to the alternative health states of death (equivalent to a value of 0) or perfect health (equivalent to a value of 1). The cost per QALY gained therefore measures the cost of treatment that generates the equivalent of one extra year of life in perfect health. The use of QALYs as the outcome measure in CEA is regarded as the gold standard outcome measure for CEA of health treatments, (Gold et al. 1996) and QALYs are used by Government authorities to inform decisions on coverage and pricing of treatments (Neumann and Greenberg 2009; Zhang and Garau 2020). However, concerns have been raised about the way in which the "quality" component of the QALY is calculated.

3.3.1 Difficulties with Forming QALY Weights Using General Population Preferences

The life-years gained from a treatment are obtained from clinical trials with patients, but the quality of life component or utility value associated with living in a health state is computed using preferences obtained by surveying members of the general population. The

current method assumes that people can accurately envisage what it would be like to live in the health states; participants in health state appraisal surveys are only provided with functional descriptions of the health states they are asked to appraise. While many members of the general population have experience of ill-health, the health states that people are asked to appraise are typically more severe than any that they may be experiencing at the time. As most people do not have knowledge of what life in the specific health states is like, they must use their imagination or other information to form a view on what living in the health states would be like (Karimi, Brazier, and Paisley 2017a). However, when imagining their future self in hypothetical health states, people may not have accurate expectations of what the negative consequences of ill-health on important domains of their life would be and on how their overall well-being would be impacted. This could affect the strength of people's preferences for hypothetical health states.

Studies find that the quality component of the QALY or the weights assigned to different dimensions of health differ depending on whether they are based on preferences obtained using current methods or based on the impact that health dimensions have on the actual well-being of people as measured by life satisfaction (Dolan and Metcalfe 2012), day affect (Dolan and Metcalfe 2012) or happiness (Mukuria and Brazier 2013). A common finding across these studies (Dolan and Metcalfe 2012; Mukuria and Brazier 2013) is that overall the impact of anxiety / depression is significantly worse than the impact of other health domains on people's well-being, but anxiety / depression is not ranked as the most undesirable health domain according to people's preferences. These empirical findings suggest that people may underestimate the impact of anxiety / depression on quality of life and that this in turn may affect their relative preferences for hypothetical health states.

3.3.2 Obtaining “Informed” General Population Preferences Over Health States

These problems have motivated calls to use “informed” preferences of members of the general population over hypothetical health states (Drummond et al. 2009). The suggestion is that respondents to health questionnaires should be provided with information on what living with specific health states is like before appraising those health states for CEA (Brazier et al. 2005). We are aware of only two studies that examine the effect on utility values by comparing preferences obtained where participants are only provided with a functional description of health states (i.e., the current method) to preferences obtained where participants are also provided with information on patients’ experiences of living in health states (i.e., an informed preferences method). McTaggart-Cowan, Tsuchiya, O’Cathain, and Brazier (McTaggart-Cowan et al. 2011) provided appraisers with audio-recordings of patients discussing adapting to life with arthritis, while Murphy, Boyce, Dolan and Wood (Murphy et al. 2020) provided participants with a summary measure of patients’ ratings of three health conditions (arthritis, tinnitus and a colostomy). A limitation of these two studies is that, while the health conditions examined cover physical function and pain, anxiety / depression was not included. The use of a summary measure of health status overcomes limitations of individual audio-recordings (e.g., that not all patients adapt to the same degree, and the time and attention required to listen to recordings, Dolan, Kavetsos, and Tsuchiya 2013). However, it remains limited in that it does not provide information on the underlying construct that the quality component of the QALY is intended to capture, namely health-related quality of life (HRQOL) which is defined as “the impact of the health aspects of an individual’s life on that person’s quality of life, or overall well-being ” (Gold et al. 1996:285).

3.3.3 The Current Study

Quality of life (QOL) or well-being can be defined either (a) clinically as the absence of disorder with observed symptoms, (b) subjectively by the person’s own opinion, or (c)

normatively by the satisfaction of externally defined criteria about what constitutes a person's well-being (Wood and Davidson 2020). As HRQOL is concerned with the impact of disorders on a person's QOL, information to convey the HRQOL of a person could be based on a person's own opinion or on externally defined criteria. In this study, we ensured that stated preferences over hypothetical health states were informed by knowledge of the well-being consequences of living in those health states ("moderate anxiety / depression", "moderate pain / discomfort", and "some problems with self-care"). We do this by providing a summary measure of the QOL consequences of living in the health states to members of the general population before they appraise the health states. Two widely used measures of QOL, drawn from each of the two approaches described above, are used in this study: life satisfaction (LS) and day affect (DA). LS has been defined as "a global assessment of a person's quality of life according to his chosen criteria" (Shin, D.C. and Johnson 1978). DA captures the extent of both positive and negative affect a person reports to typically experience for a set of externally defined emotions. Each summary measure briefly conveys QOL consequences of the hypothetical health states, as the information presented on the differences in LS/DA associated with each health states is from a study with a representative survey of the general population (Dolan and Metcalfe 2012).

In this study participants in the control group provide their preferences over health states after reviewing a functional description of each health state (i.e., in line with current methods) while participants in the intervention groups provide their preferences over health states after reviewing both a functional description of each health state and additional information on the consequences of each health state for either LS or DA. The study's pre-registered hypotheses (As Predicted: #6459) are: (1) more people from the general population will report that actual LS and DA consequences are bigger than expected for living with "moderate anxiety / depression" as compared with other dimensions of health, and (2) the

groups informed with LS or DA information will exhibit a lower preference for “moderate anxiety / depression” relative to other health states than will the control group.

3.4 Method

3.4.1 Sampling and Allocation Procedure

This study was conducted online in three of the four constituent countries of the United Kingdom (England, Scotland, and Wales). Online surveys have been deemed a valid method for health state valuation studies using simple appraisal techniques such as binary choice experiments (Mulhern et al. 2013). The study was designed to reflect the adult age profile and regional distribution of people living in the three countries, so the survey was stratified by six age groups and by region of residence (Scotland, Wales and nine regions in England). Participants in this study were adult members, aged 18 and over, of a panel of Internet users (“Survey Spot Internet panel”) operated by Survey Sample International (SSI). The panel maintains information on over 1 million unique member households recruited through “permission-based” techniques.

Participants self-enrolled, completed an informed consent form, and were compensated for their time if they completed the survey (it is customary to compensate participants in health valuation surveys). Ethics approval was provided by the University of Stirling Management School’s Ethics Committee. Participants were allocated to one of three groups by a computerized number generator using block randomization with a predetermined ratio of 1:1:1 (allocation concealment and double blinding were maintained throughout data collection). An a priori target of at least 500 respondents per group was set and was met for each group. The data were collected between 12th and 19th October 2017, and after applying exclusion criteria there were 1,259 participants. The Consolidated Standards of Reporting Trials (CONSORT) participant flowchart is shown in Figure 3.1.

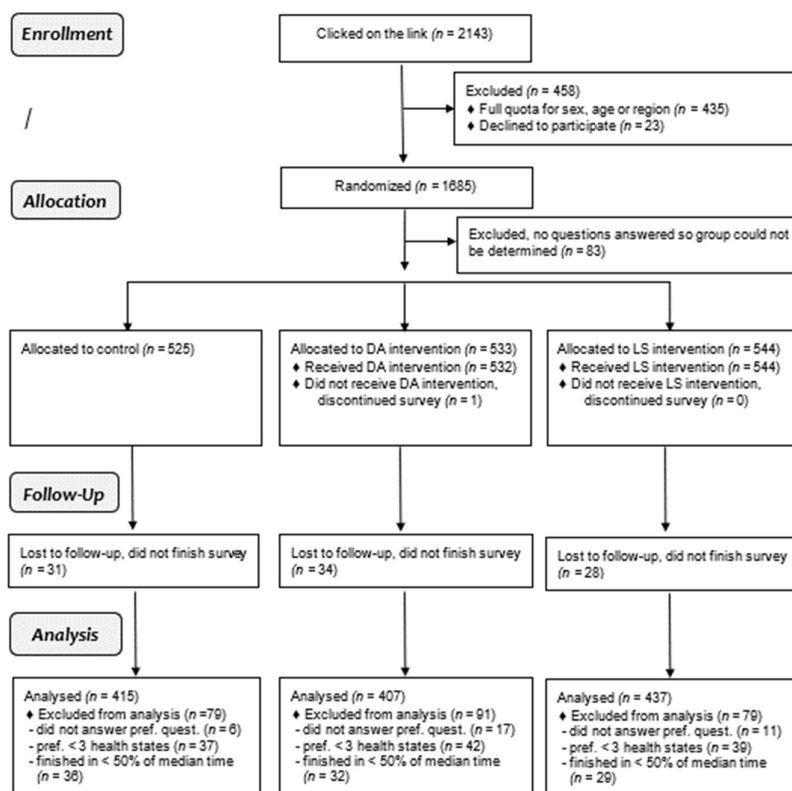


Figure 3.1 CONSORT Participant Flow Diagram

Note. “pref. < 3 health states” indicates where a respondent did not provide a 1st, 2nd and 3rd preference when choosing between the three health states.

3.4.2 Questionnaire

The questionnaire included seven sections, each section is described in the supplementary file, and details on Sections 3, 4 and 6 are below.

Description of health states. The health state descriptions used were based on descriptions from the EQ-5D-3L, which is one of the most widely used instruments in health valuation studies (EuroQol Research Foundation 2015). The descriptions were:

(1) Moderate anxiety / depression: No problems with walking about, no problems with self-care, no problems with usual activities, no pain or discomfort, moderately anxious or depressed.

(2) Moderate pain / discomfort: No problems with walking about, no problems with self-care, no problems with usual activities, moderate pain / discomfort, not anxious or depressed.

(3) Some problems with self-care: No problems walking about, some problems with self-care, no problems with usual activities, no pain or discomfort, not anxious or depressed.

The descriptions of the ill-health elements of (1) and (2) correspond exactly to the wording in the EQ-5D-3L descriptive system (i.e., “moderate pain / discomfort” and “moderately anxious or depressed”). Whereas the wording for (3) does not, as for consistency across health states the description used was “some problems with self-care” whereas the EQ-5D wording is “some problems washing or dressing” (i.e., the wording for each health state maintained a similar level of specificity for each dimension of health, “pain / discomfort”, “anxious or depressed” and “problems with self-care”). The dimensions of moderate pain or discomfort and some problems with self-care are used as these are the dimensions that people in the UK (Kind et al. 1998) and the US (Shaw, Johnson, and Coons 2005) are most and least likely to have direct experience of (prevalence of 30-40 and 4 per 100 respectively) and therefore provide a range of possible prior knowledge of the health states.

Preferences over health states. Participants were told: “We would like you to make a choice between health states, from most to least preferred to live in.” and instructed to: “Choose between these three health states, from 1 (most preferred) to 3 (least preferred). Make your choice by ticking 1 for your most preferred, 2 for second preference and 3 for your third preference.” The order in which the health states were listed was randomized. This study’s pre-registration form describes this current study and refers to a second possible experiment, in which participants would judge the health states using the time trade-off (TTO) method (Stolk et al. 2019). We did not run this second experiment and focused on the simple preference ranking task due to subsequent concerns that detection of a possible intervention

effect could be reduced by use of the TTO method as it involves a “complex iterative procedure” (Stolk et al. 2019) traditionally undertaken in face-to-face mode that might not lend itself as well to unsupervised online data collection. Studies suggest that compared to a face-to-face data collection mode, TTO tasks conducted online have more inconsistencies, decreased engagement, and larger variation,(Jiang et al. 2021; Norman et al. 2010) whereas both modes of data collection may be equally valid for simpler appraisal methods such as binary choice experiments (Mulhern et al. 2013). Participants were asked to “imagine that you will live in the health states described for up to 10 years and then die”, this was not technically required as the TTO method was not used but it was decided to adopt this approach as traditionally in health valuation studies respondents are asked to imagine being in an impaired health state for a fixed period of time (Versteegh and Brouwer 2016).

Presentation of association of health states with well-being. Participants (1) viewed information on the association of each health state with differences in LS or DA and (2) were asked if the association of each health state with differences in LS or DA was bigger, about the same, or smaller than expected. Participants in the intervention groups saw this information either for LS (first intervention condition) or for DA (second intervention condition). Global self-reported well-being measures are shown to have criterion-related validity as observed scores correlate with theoretically relevant variables (Hudson, Lucas, and Donnellan 2022). The information was provided in text (see the supplementary file) and in a graphic that presented the association of health states with differences in LS or DA alongside a vertical scale with the anchors 0% and -50%. Data on the association of health states with patients’ LS and DA were taken from Table 2 in Dolan and Metcalfe (Dolan and Metcalfe 2012).

3.4.3 Experimental Manipulation

For the control group sections 1 to 7 of the questionnaire appeared in order of section number. The functional description of each health state was therefore the only information about the health states that had been provided to the control group at the time when they stated their preferences over health states (i.e., the process of information provision for the control group followed standard practice). For the intervention groups, Section 6 (Presentation of association of health states with LS or DA) came before Sections 4 (Preference for health states and 5 (Clarity of tasks). Thus, the intervention groups expressed their preferences over the health states only after being informed of the association of those health states with differences in LS or DA.

3.4.4 Statistical Methods

Testing hypothesis 1: Testing hypothesis 1: First, the proportion of participants who believed that the actual LS or DA consequences of living in a health state were greater than they expected was calculated. Analysis of whether this proportion was highest for the health state of “moderate anxiety / depression” was then undertaken using Z proportion hypothesis tests (after applying the Bonferroni procedure to the significance level).

Testing hypothesis 2: The study’s preference measure is an ordered outcome variable (first, second, or third preference). We therefore used an ordered logistic model (Williams 2006) to compare the relative preferences for “moderate anxiety / depression” (over other health states) for each intervention group to the relative preferences of the control group by including information provision (INFO) as an explanatory variable (0 for control group, 1 for intervention group) in Equation 1 along with control variables for demographic characteristics, employment status, and life satisfaction (details are in Table 3.1). Tests were run separately for provision of LS information and DA information. Before the data were examined it was decided that this method of analysis was more appropriate than the use of

Spearman's rank order correlations envisaged in our pre-registration form as the use of an ordered logistic model facilitated the inclusion of control variables in our analysis.

$$P(C_j) = \beta_0 + \beta_1 INFO_i + \beta_2 DEMO_i + \beta_3 EMPLOY_i + \beta_4 LS_i + \varepsilon_i, j = 1, 2, 3, \quad i = 1, \dots, n \quad (1)$$

The model tested the relative likelihood of preferences for anxiety/depression falling into less preferred categories (because a preference of “1” is equivalent to “most preferred”). A significant positive co-efficient on the information provisions explanatory variable INFO would support the second hypothesis: A positive co-efficient indicates an increased likelihood of the preference being in higher-numbered (i.e., less preferred) categories, which indicates a lower preference for “moderate anxiety / depression” relative to the other health states. Effect size was estimated using marginal effects. Tests of the robustness of our findings did not change the study conclusions (details are in the supplementary file).

3.5 Results

3.5.1 Descriptive Analyses

Participant characteristics. The profile of the sample was: 52.50% female, mean age of 47.65 years (SD= 16.42, range: 15-85), 39.62% with a degree, 53.43% employed, and 45.57% married. The distribution of the sample characteristics was similar to that of the UK population (see Table 3.1).

Table 3.1 Participant and Population Characteristics

	Sample			Population
	Control	Inter. LS	Inter. DA	
Sex				
Female	52.3%	52.4%	52.8%	50.7%
Male	47.7%	47.6%	47.2%	49.3%
Age				
18-24 ^a	11.8%	12.1%	9.1%	10.9%
25-34	15.4%	17.8%	11.8%	17.2%
35-44	17.6%	16.5%	18.2%	16.0%
45-54	18.8%	16.0%	21.4%	17.8%
55-64	17.3%	15.6%	18.2%	15.0%
65+	19.0%	22.0%	21.4%	23.1%
Highest educational attainment ^b				
No qualifications	5.4%	4.4%	5.4%	22.4%
Level 1 qualifications	17.2%	22.7%	22.7%	13.4%
Level 2 qualifications	13.5%	12.5%	13.3%	15.4%
Level 3 qualifications	18.0%	16.0%	18.0%	12.1%
Level 4 qual. and above	41.9%	40.1%	36.8%	27.5%
Apprenticeships, entry level or other qualifications	3.9%	4.2%	3.7%	9.2%
Principal economic status ^c				
Employed	53.1%	53.4%	53.8%	60.9%
Unemployed	6.7%	5.3%	5.9%	2.7%
Economically inactive	40.0%	40.4%	40.0%	36.4%
None of the above	0.2%	0.9%	0.2%	n.a.
Relationship status ^d				
Married	46.1%	42.8%	48.0%	49.8%
Domestic partner	14.3%	17.1%	15.8%	n.a.
Unmarried	27.6%	25.9%	22.3%	36.2%
Divorced	7.9%	7.4%	8.9%	7.6%
Widowed	3.2%	3.5%	2.7%	6.4%
Prefer not to state/other	1.0%	3.2%	2.2%	n.a.

Note. Population figures are from the Office of National Statistics (ONS) www.ons.gov.uk. Unless stated otherwise population figures relate to mid-2017 and are for England, Wales and Scotland. Inter. = Intervention, LS = Life Satisfaction, DA = Day Affect. ^a Population figure for "18-19" age category is estimated based on two-fifths of the ONS figure for the 15-19 age category. ^b Population education figures are for 2011 for England and Wales. Detailed definitions are available on the ONS website, for illustration Level 1 includes 1-4 O Levels / CSE/ GCSEs, Level 2 includes 5+ O Levels, Level 3 includes 2+ A Levels/VCEs, Level 4 and above includes degree or higher degree. ^c Population figure for main activity is based on responses from questions in the ONS Labour Force Survey whereas the figures for this study are based on responses to one question "best describes your main activity", the former provides a narrower definition of unemployment. ^d Population figures for England and Wales, and figures for "Married" includes ONS figures for "Married" and "Civil partnered", and "Unmarried" is ONS figure for "Single".

Ease of completion. Respondents found the survey “Very Easy” (22.71%), “Easy” (41.06%), or “Neither easy nor difficult” (26.09%); only 8.53% found it “Difficult” and 1.61% found it “Very difficult”. Participants were also asked whether they found the specific preference ranking task easy. Only 6.68% “disagreed” and 1.27% “completely disagreed” with the statement that the preference ranking task was easy, while 42.64% “agreed”, 33.02% “completely agreed” and 16.39% “neither agreed nor disagreed”. These results suggest that the task difficulty did not pose a problem for participants. Re-running the analysis excluding participants who disagreed or completely disagreed that the preference ranking task was easy did not alter the study conclusions.

3.5.2 Test of Study Predictions

Hypothesis 1: Actual compared to imagined associations with LS or DA. Figure 3.2 shows whether the actual effects of each of the three health states on well-being were larger, about the same, or smaller than participants’ prior beliefs. As predicted, the proportion of participants reporting that the association of a health state with differences in DA or LS was bigger than expected was significantly higher for “moderate anxiety / depression” than for the other two health states (the 95% confidence intervals for a group that do not bound the mean of another group indicate statistically significant differences).

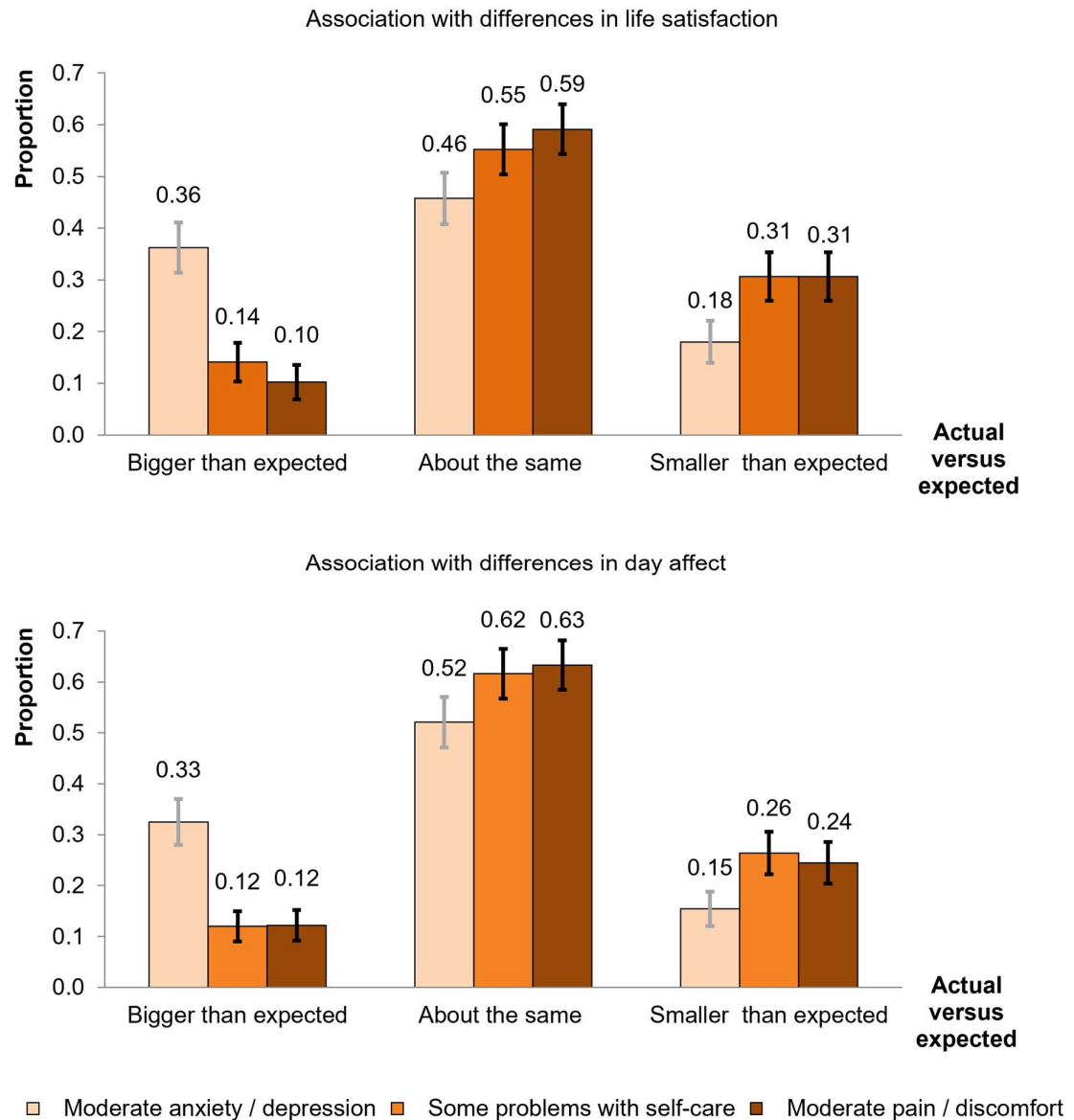


Figure 3.2 Actual Difference in Well-being Associated with Each Health State Compared to the Control Group's Prior Belief of Difference

Note. Error bars represent the 95% confidence intervals.

Hypothesis 2: Informed preferences compared to preferences not informed by LS or DA information. A visual illustration of the effect on stated preferences of providing LS or DA information is presented in Figure 3.3, which shows the proportion of participants in

the control and intervention groups who selected a health state as first, second or third preference. The most notable effect is that the proportion of participants who selected “moderate anxiety / depression” as the least preferred health state was higher in the intervention groups than in the control group (0.48 versus 0.35) with a Cohen’s $h = 0.255$ for the LS intervention and a Cohen’s $h = 0.265$ for the DA intervention.

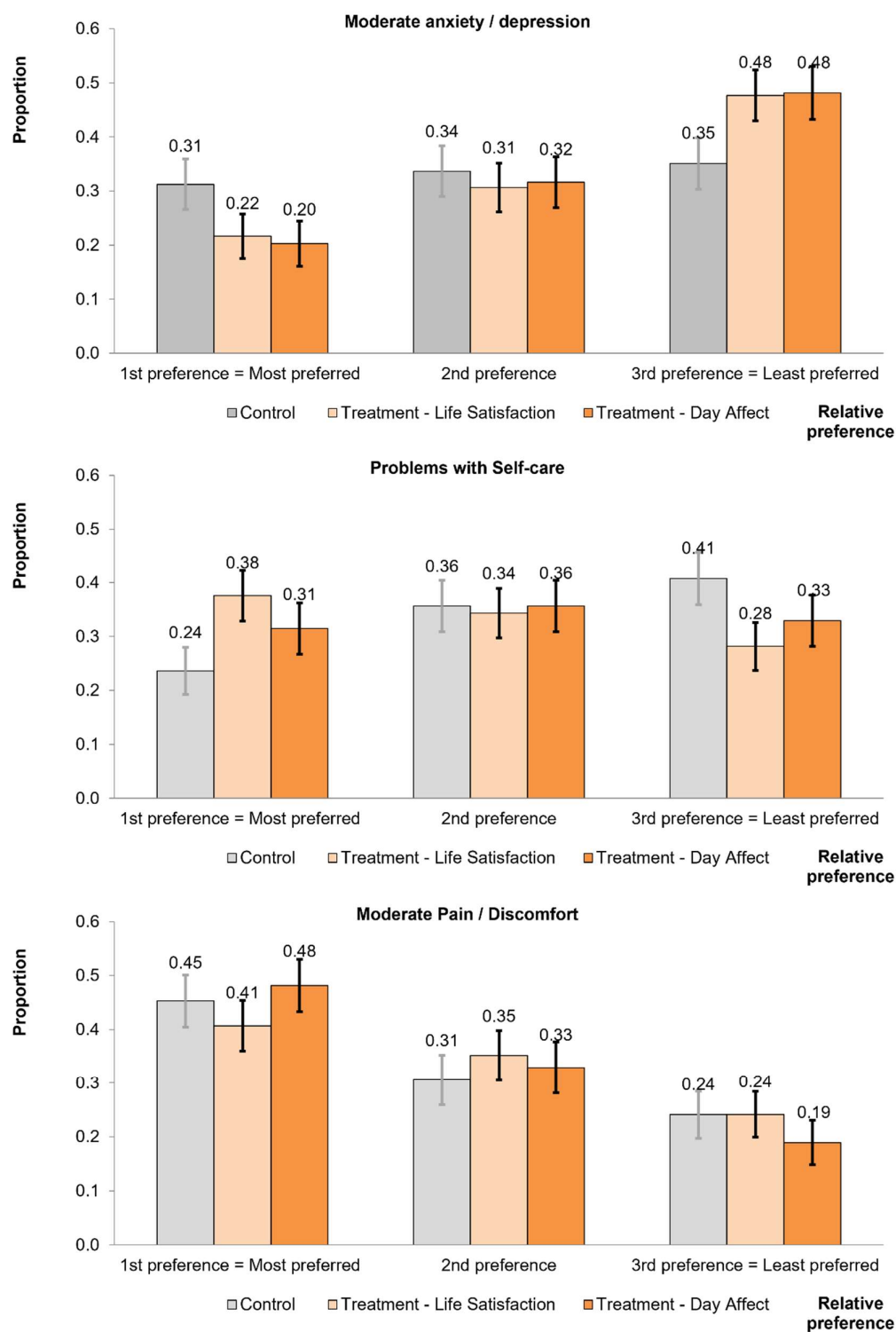


Figure 3.3 Proportion of Respondents Selecting Health State as 1st, 2nd or 3rd Preference by Group

Note. A lower preference (e.g., 2nd or 3rd) indicates a view that a health state is relatively less desirable to live in. Error bars represent the 95% confidence intervals.

To test the effect of information provision, while controlling for additional factors, we estimated ordered logistic models of the form in Equation 1 for the health state “moderate anxiety / depression”. These revealed a significant positive co-efficient on information provision (see Table 3.2). Thus, providing participants with LS or DA information prior to eliciting their preferences over health states reduced participants’ relative preferences for living with “moderate anxiety / depression”. That is, informed participants were more likely to select “moderate anxiety / depression” as the least preferred health state, or to select it as the least or second least preferred rather than most preferred. This applies for provision of both LS information, $z(814) = 4.19, p < .001$, and for DA information, $z(786) = 4.30, p < .001$.

Table 3.2 Ordered Logistic Model of Preferences for “Moderate Anxiety / Depression”

	LS information:		DA information:	
	Coefficient	Standard Error	Coefficient	Standard Error
Info. provision	0.558***	(0.133)	0.580***	(0.135)
Demographics				
Sex	-0.030	(0.141)	-0.102	(0.140)
Age	-0.036	(0.027)	-0.031	(0.028)
Age squared	0.000*	(0.000)	0.000	(0.000)
Relationship status	-0.237	(0.146)	0.033	(0.147)
Employed	-0.193	(0.160)	-0.110	(0.158)
Education				
2nd level	-0.219	(0.292)	0.244	(0.267)
Apprenticeship or entry level	-0.298	(0.475)	0.029	(0.531)
Degree or higher	-0.308	(0.295)	0.289	(0.272)
Life satisfaction rating	0.114***	(0.032)	0.056*	(0.032)
Constant				
Constant cut1	-0.995	(0.652)	-0.638	(0.677)
Constant cut2	0.469	(0.651)	0.818	(0.677)
Observations	814		786	
<i>Note.</i> Standard errors in parentheses. Base categories are no information on life satisfaction or day affect (information provision), female (sex), not married or with domestic partner (married or with domestic partner), not employed (employed) and no formal qualification or “other” (education). Info. = information; LS = life satisfaction; DA = day affect. Life satisfaction rating of participants was self-reported satisfaction with life on a 0–10 scale, with 0 denoting “not at all satisfied” and 10 denoting “completely satisfied” *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$				

This result suggests that, as predicted, provision of LS or DA information results in living with “moderate anxiety / depression” being judged as less preferred. Participants informed of the association of the three health states with differences in LS (DA), were 13.4%

(13.9%) more likely to choose “moderate anxiety / depression” as their least preferred health state and less likely to choose it as their second least preferred by 2.8% (3.0%), or as their most preferred by 10.5% (10.9%), see Table 3.3.

Table 3.3 Marginal Effects of Information Provision on Preferences for “Moderate Anxiety / Depression”

	1st Preference		2nd Preference		3rd Preference	
	LS Info.	DA Info.	LS Info.	DA Info.	LS Info.	DA Info.
Info. Provision	-0.105*** (0.025)	-0.109*** (0.025)	-0.028*** (0.008)	-0.030*** (0.009)	0.134*** (0.031)	0.139*** (0.032)
Demographics						
Sex	0.006 (0.027)	0.019 (0.026)	0.002 (0.007)	0.005 (0.007)	-0.007 (0.034)	-0.025 (0.034)
Age	0.007 (0.005)	0.006 (0.005)	0.002 (0.001)	0.002 (0.001)	-0.009 (0.006)	-0.007 (0.007)
Age squared	0.000* (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Married or with domestic partner	0.044 (0.027)	-0.006 (0.028)	0.013 (0.009)	-0.002 (0.008)	-0.057 (0.035)	0.008 (0.035)
Employed	0.036 (0.030)	0.021 (0.030)	0.010 (0.009)	0.006 (0.009)	-0.047 (0.039)	-0.027 (0.038)
Education						
2nd level	0.039 (0.049)	-0.049 (0.055)	0.015 (0.024)	-0.009 (0.007)	-0.054 (0.073)	0.057 (0.061)
Apprenticeship or entry level	0.054 (0.088)	-0.006 (0.111)	0.019 (0.029)	-0.001 (0.011)	-0.073 (0.115)	0.007 (0.122)
Degree or higher	0.055 (0.050)	-0.057 (0.056)	0.020 (0.024)	-0.011 (.008)	-0.075 (0.073)	0.069 (0.063)
Life satisfaction rating	-0.022*** (0.006)	-0.0106* (0.006)	-0.006*** (0.002)	-0.003* (0.002)	0.026 *** (0.008)	0.014* (0.008)

Note. Base categories are: no information on life satisfaction or day affect (information provision), female (sex), not married or with domestic partner (married or with domestic partner), not employed (employed) and no formal qualification or “other” (education). Info. = information; LS = life satisfaction; DA = day affect. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

3.6 Discussion

3.6.1 Strengths and Limitations

This is the first study to simultaneously obtain preferences informed by the experiences of others by providing information on the consequences of health states for LS or DA to participants from the general population, to obtain informed preferences over generic health states rather than specific health conditions (i.e., it uses the type of descriptions of ill-health used in HTAs), and to test how informing preferences affects participants' stated relative preferences over alternative dimensions of health. Key strengths of this study are the use of two different summary measures of QOL (an evaluative judgment of overall satisfaction with life, and a lived experience or emotions over the course of a day), and its use of an RCT with 1,259 members of the general population in England, Scotland, and Wales.

This study was designed as a proof of concept and has several limitations. There may be additional aspects to QOL that are not captured by the two measures we have used. The intervention may have created a focusing effect (Schkade and Kahneman 1998) whereby the intervention groups may have placed too much importance on the LS/DA consequences for patients. While acknowledging the potential for focusing effects, a previous study that obtained informed preferences (Murphy et al. 2020) argues that when obtaining preferences of the general population providing patients' information is warranted as what people imagine as the consequences of ill health is not always accurate (Karimi et al. 2017a). Indeed, preferences can only be said to be informed when they are "freed from the distorting effects of factual errors" (Harsanyi 1997). The extent of possible focusing effects associated with the provision of HRQOL information could be gauged by incorporating reflection and deliberation techniques (Shiell, Hawe, and Fletcher 2003; Stein et al. 2006) into future studies. Other limitations of this study include the use of an online internet panel (although

there is some evidence that online health valuation surveys are valid when simple appraisal techniques are used (Mulhern et al. 2013) and use of a simple preference ranking task. Moreover, the estimated LS/DA consequences of health states are taken from a 2012 study of US residents which might not apply to UK residents at the time of this study, and the preference ranking task did not allow for possible ties (this did not influence the study conclusions as the proportions were small, 7.05-7.89%, and evenly spread across the groups).

3.6.2 Directions for Future Research

This study shows that being more informed about the association between health states and differences in LS or DA results in living with “moderate anxiety / depression” being judged as less preferred relative to moderate problems with “self-care” or “pain / discomfort”, and so of increased priority for being addressed by health treatments. Several areas should be examined in future research. First, the extent to which this study’s findings generalize to preferences obtained using the time trade-off method and to a wider range of health dimensions should be examined. Second, more research is needed on how best to capture, and to convey, the well-being effects of ill-health to members of the general population as part of health valuation studies. This study used subjective well-being (SWB) measures of QOL but other measures of QOL should also be used. For instance, the construct of psychological well-being (PWB) as represented by Ryff’s (Ryff, C.D. 1989) six factors of positive relations with others, self-acceptance, purpose in life, autonomy, environmental mastery, and personal growth could be adopted as SWB and PWB, while highly correlated, are distinct constructs (Biaobin, Xue, and Lin 2004; Keyes, Shmotkin, and Ryff 2002; Linley et al. 2009). Given the aim of providing information on patients’ experiences is to enrich understanding of the impact of the health aspects on QOL it seems more appropriate to provide such information separately to, rather than within, health state descriptions. Third, more research is needed to better understand patients’ experience of living with ill-health and what influences their

appraisals of health. Whether it is normatively appropriate (in the context of resource allocation decisions) to provide people with information derived from the self-reported QOL of patients, given that self-reported QOL is likely to partly reflect adaptation to ill-health, depends in part on whether the adaptation that occurs can be viewed as admirable achievements or as deficiencies (Menzel et al. 2002). Fourth, there is a need for more evidence on the factors considered by members of the general population when appraising hypothetical health states (Baker and Robinson 2004; Brouwer et al. 2009; Karimi et al. 2017b; Krol et al. 2006; Myers et al. 2007; van Osch and Stiggelbout 2005; van der Pol and Shiell 2007) and for a better understanding of the cognitive mechanisms that influence such appraisals (Menzel et al. 2002; Stiggelbout and De Vogel-Voogt 2008; Ubel et al. 2003; Versteegh and Brouwer 2016).

4 Nudging to improve patient engagement with management of hospital waiting lists: a Randomised Controlled Trial

Acknowledgements: The authors would like to express our appreciation to: Mr Paul Bolger, Ms Martina Kinsella and Ms Alessandra Fantini, Department of Health; Ms Alison Green and Ms Helen Lenehan, National Treatment Purchase Fund (NTPF); Mr Gary Keenan and Ms Verona Walsh, RCSI Hospital Group; Mr John Doyle and Ms Aideen O'Callaghan, UL Hospital Group; members of the National IDPP Project Steering Group; Ms Brigid Doherty, CEO, Patient Focus; Mr Karl Purcell, Department of Public Expenditure and Reform; and Dr Christopher Boyce, University of Stirling.

4.1 Abstract

Objective: A commonly adopted intervention to help to reduce wait times for hospital treatment is administrative validation, where administrators write to patients to check if a procedure is still required. The did not return (DNR) rate to validation letters is substantial. We tested whether the return rate was increased by introducing nudges to validation letters.

Methods: Participants from eight public hospitals ($N = 2855$; in 2017) in Ireland were randomised to receive an existing (control group) or a redesigned validation letter including nudges (intervention group).

Results: Participants in the intervention group were more likely to return it than those in the control group. Control and intervention group DNR rates were 23.97% and 19.24%, $OR = .756$, $SE = .069$, $p = .002$. This is equivalent to 1 in 5 non-responders changing their behaviour because of the redesigned letter.

Conclusions: The redesigned letter increased patient compliance with the validation process. The redesign has subsequently been adopted by public hospitals in Ireland.

Keywords: Nudges, patient engagement, decision making, waiting lists

4.2 Introduction

Long waiting times for elective (non-emergency) specialist treatment in hospital can negatively affect patients' outcomes (Hurst & Siciliani, 2003; OECD, 2013). Negative outcomes can include deterioration in health status, difficulties in daily life due to pain or disability, and loss of income (Hurst and Siciliani 2003). Reducing waiting times for elective hospital treatments is a medium or high priority issue in most OECD countries (OECD 2020b). A commonly adopted intervention to reduce waiting times is to improve management of waiting lists. The idea is that “by eliminating inefficiency in the management of the list, the number of treatments for a given level of personnel and capital endowment can be increased” (Hurst and Siciliani 2003:29). Management of waiting lists can be improved by administrative validation, a process where administrators in a hospital write to patients on waiting lists to check if patients still require a procedure or wish to be removed from the waiting list. Experience shows that a proportion of patients on waiting lists no longer actually require treatment. They may, for instance, have had the treatment in another hospital, feel better, or no longer want a procedure. The removal of these patients from waiting lists reduces the size of waiting lists and enables procedures to be offered sooner to patients who require treatment. However, a challenge for hospital administrators is that a substantial percentage of patients typically do not return (DNR) a “validation letter” to confirm whether the procedure is still required. DNR rates of 25% are common (A. Green, personal communication, June 15, 2017).

A potential way to reduce DNR rates is to deploy nudges. A nudge is “any aspect of the choice architecture that alters behaviour in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler, R. H. and Sunstein, C. R. 2009, p. 6). An application of nudges to the administrative validation of waiting lists could involve altering the content of validation letters by including design features (such as specific phrases or presentation devices) with the aim to increase the proportion of patients who reply.

The use of nudges to improve the outcomes of public services has been endorsed by international organisations such as the Organisation for Economic Cooperation and Development (Lunn 2014), the World Health Organization (WHO Director-General 2022), and the World Bank (World Bank 2015). A recent literature review of the effectiveness of nudging (DellaVigna and Linos 2022) identified four studies in which the effect of the inclusion of nudges in physical letters was tested. These four studies included a range of nudges across different policy areas (and behaviours). They included the policy areas of health (Milkman et al. 2011), the environment (Rommel et al. 2015) and revenue collection (Fellner, Sausgruber, and Traxler 2013; Goswami and Urminsky 2016). However, that literature review did not identify, and we are not aware of, a primary study that has tested the effect of incorporating nudges into validation letters.

How-to guides have been developed to help policymakers and service providers to apply nudges, such as checklists of non-coercive influences on human behaviour (e.g., MINDSPACE by Dolan et al. 2012) or descriptions of ways to apply behavioural insights (e.g., EAST by the Behavioural Insights Team 2014). Policymakers and service providers can use such guides to increase their awareness of and, perhaps, their capability to apply nudges. However, there is little tailored guidance available on how best to design correspondence to promote certain behaviours in specific contexts. Furthermore, there is no guidance on how best to combine multiple nudges in a single communication.

Few studies have looked at how nudges combined within one communication might crowd each other out, that is if a combination of nudges that target different psychological mechanisms incorporated into a single treatment perform worse than each individual nudge implemented separately (Howley and Ocean 2022). Some results support the lack of crowding out effects, namely evidence in the domain of electricity consumption (Brandon et al. 2019) and in technology uptake in farming (Howley and Ocean 2022). However, some point to

possible crowding out as there is some evidence that inclusion of multiple nudges into letters does not necessarily increase uptake of influenza vaccination (Yokum et al. 2018).

This paper describes a study that redesigned and tested the impact of a waiting list validation letter on patient did not respond (DNR) rates. The study was part of a quality improvement project in Ireland, referred to as the Better Letter Initiative (BLI), that followed a collaborative approach involving a government department, state agencies responsible for hospital waiting lists, and a relatively small expert behavioural science advisory group drawn from universities and not-for-profit bodies (Murphy, R et al. 2020). The redesigned letter included multiple nudges, many of which have not previously been used in a health context. It has been argued that two of the ten most important nudges for policy are simplification and increases in ease and convenience (Sunstein 2014). A recent literature review of nudging identifies the use of increases in ease and convenience as one of the most used types of nudges in health (42% or 35/84 of interventions) but does not identify any studies that use simplification (Hummel and Maedche 2019). We used both types of nudges. We sought to achieve simplification by using plain English and by making important information salient by using bolding, and to increase ease and convenience by including a clear call for action and chunking of steps. We used four additional nudges that do not fall into the ten categories of nudges emphasised by Sunstein (2014). These included two nudges for which there is previous field evidence of positive effects, namely personalisation (Edwards et al. 2009) and messenger effects (Dolan et al. 2012). We also included two additional nudges: an apology and an articulation of altruistic intentions as well as the value of the service and resource involved in the validation process. There is no health context evidence on the effect of an apology but evidence from other sectors is mixed (Cheng and Tsai 2014; Munichor and Rafaeli 2007). We were not aware of evidence from field trials on the effect of articulation of altruistic intentions, the value of the service, and resource involved in the process but

laboratory evidence suggests this may positively influence client response (Bridger and Wood 2017).

The aim of the current study was to test the redesigned letter in a randomised controlled trial (RCT) in two hospital groups, one in the north-east and one in the mid-west of Ireland. Participants in the control group received a pre-existing validation letter used in that hospital group, while those in the intervention group received a redesigned validation letter. We hypothesised, straightforwardly, that the redesigned letter would lower DNR rates from patients who received the intervention letter compared to patients who received the control letter. From an applied perspective, if the intervention works, it would be important to know for which patients it occurs (individuals) and across contexts, so we also undertook pre-planned exploratory analysis (see *As Predicted: #98098*) to see if the intervention held across patient characteristics of sex and age, and healthcare characteristics of hospital group and medical specialty.

4.3 Method

4.3.1 Sampling and Allocation Procedure

The quality improvement project was conducted in two healthcare administrative areas in Ireland, referred to as “hospital groups”. In May 2013 the Minister of Health in Ireland announced a reorganisation of public hospitals into six hospital groups. Hospitals within each hospital group aim to operate as a single cohesive entity managed as one, to provide acute care for patients in their area, integrating with community and primary care. Each group includes between six and eleven hospitals and includes at least one major teaching hospital. The two hospital groups in this current study were the Royal College of Surgeons in Ireland (RCSI) Hospital Group and the University of Limerick (UL) Hospital Group. This current study included three hospitals in the RCSI Hospital Group (Our Lady of Lourdes Hospital, Drogheda; Cavan and Monaghan Hospital; Louth County Hospital) located in the north-east

of Ireland and five hospitals in the UL Hospital Group (University Hospital Limerick; St. John's Hospital Limerick; Nenagh Hospital; Ennis Hospital; Croom Hospital) located in the mid-west of Ireland. All adult patients on a waiting list for three months or more were included for 14 hospital specialties (listed in Section 1, Supplementary File).

Participants were randomly allocated to the control or intervention group by a computerised number generator, using block randomisation with a predetermined ratio of 1:1. Allocation concealment and double blinding were maintained throughout data collection. Randomisation was administered as follows: (a) a validation administrator working in the health service provided a researcher in the Department of Health with the total number of patients to be validated (e.g. 800), (b) a researcher in the Department of Health carried out randomisation with the random.org list randomiser and returned to a validation administrator a list on which each number (e.g. 1 to 800) was randomly allocated to the letter type (existing or redesigned), and (c) a validation administrator printed and posted validation letters according to this allocation.

In the absence of previous similar studies, power calculations were based on the effect sizes from three studies in which nudges were used to improve attendance at healthcare appointments (Behavioural Insights Unit (BIU) n.d.; Chadborn, T and Berry, D n.d.; Hallsworth et al. 2015). The simple arithmetic mean effect across these studies was 18.67%. Power calculations suggested that assuming a possible reduction in the DNR rate of 18%, a sample of 2,718 was required to test with 80% power at the 5% significance level. The required sample size was achieved with 2,870 participants (798 from RCSI Hospital Group and 2,072 from UL Hospital Group). The validation letters were posted from the 27th of September to the 24th of October 2017. The issuing of validation letters and the collection of data were undertaken by staff in the Health Service Executive (HSE) and the National Treatment Purchase Fund (NTPF) as part of service management in line with the statutory

functions of these state agencies. The lead author obtained access to an irreversibly anonymised data file containing responses to the postal validation, held by the Department of Health in compliance with data protection legislation. Ethical approval for statistical analysis of this data was provided by the University of Stirling Ethics Committee. After applying exclusion criteria there were 2,855 participants. The Consolidated Standards of Reporting Trials (CONSORT) participant flowchart is shown in Figure 4.1.

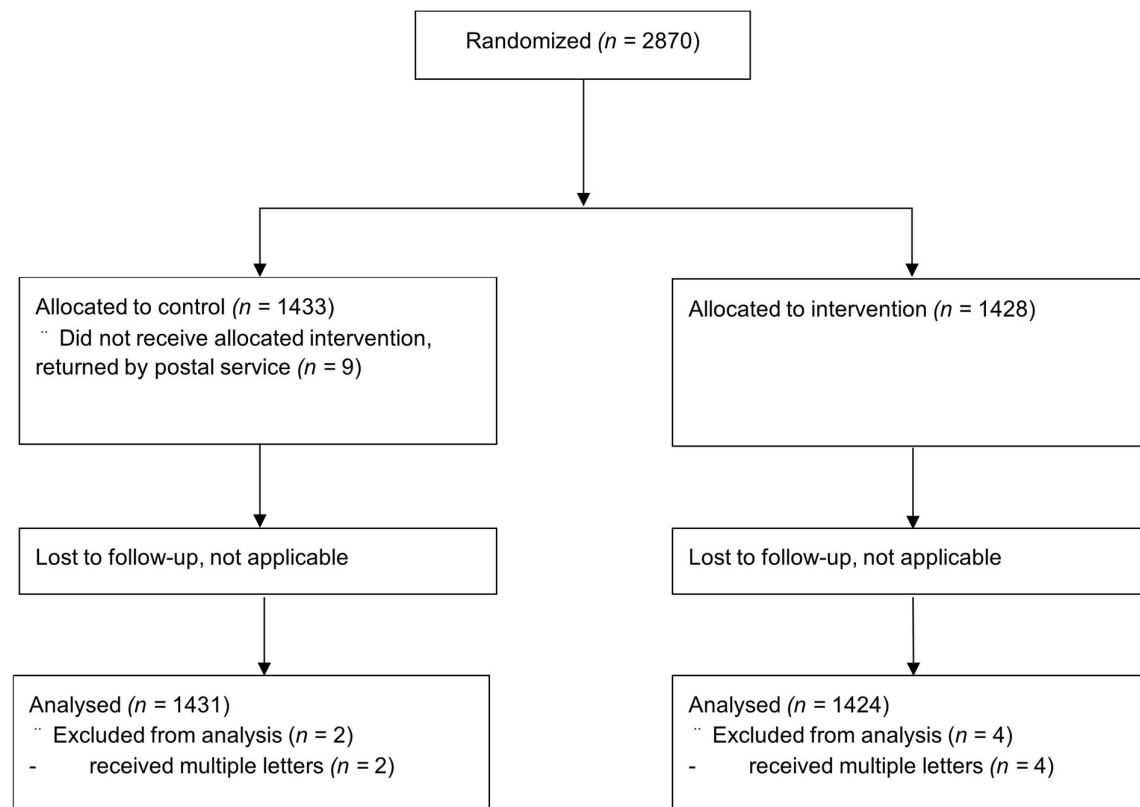


Figure 4.1 CONSORT Participant Flow Diagram.

4.3.2 Validation Letters

Each participant received either the control or intervention validation letter requesting confirmation of whether the relevant procedure was still required or not. In both conditions participants also received a prepaid addressed return envelope.

Control letter. At the time of the study, different hospital groups in Ireland used different validation letters. In some cases, different hospitals located in the same hospital

group used different validation letters. This resulted in the study using two control letters that matched formats of validation letters used in each hospital group before the project. The five hospitals in the UL Hospital Group already used an identical validation letter, so this was used as the control letter for participants from the UL Hospital Group (see Section 2, Supplementary File). The three hospitals in the RCSI Hospital Group used similar but not identical validation letters before the project. The manager and administrator responsible for validation for these hospitals created a common validation letter based a review of the pre-existing validation letters used in these hospitals, which was then used as the control letter for participants from RCSI Hospital Group (see Section 3, Supplementary File).

Intervention letter. The intervention letter (see Section 4, Supplementary File) included the following design elements: simplified language, a call for action, personalisation by addressing the patient by their first name, an apology for the wait, emphasis on the altruistic intentions of checking the waiting list and the value of the service and resource involved in the validation process, important information for the patient made salient by bolding, desired actions “chunked” in bullet points and an image, consequences of non-return, including observation made salient by bolding, and the name of the person sending the letter included to exploit any potential messenger effect. The redesign of the validation letter for the current study was informed by a review of existing practice in eight hospitals and a national validation protocol, literature on patient / customer engagement and increasing responses to surveys, and guides on how to apply behavioural insights and plain English. Research evidence used by the quality improvement team to inform the key design elements is presented in Table 4.1. Feedback on the redesign was provided from the national agency responsible for waiting lists, hospital service providers and managers (via a national group and staff in the participating sites), and a patient advocacy group. An iterative design process was followed with an expert behavioural science advisory group. Color was not part of the

design as color printing was not available in all HSE hospital sites and this would have limited scalability.

4.3.3 Outcome Measure

The outcome measure was the DNR rate of patients to validation letters, as recorded by each hospital group. For each letter sent, an administrative dataset includes whether the letter was returned by recipient or not. The recording process is undertaken by experienced staff and follows strict protocols and audit procedures (National Treatment Purchase Fund (NTPF) 2017). Measurement error in the data is therefore likely to be very small.

4.3.4 Statistical Methods

Testing the study hypothesis. We applied a binary logistic regression model with DNR as the outcome variable (0 for returned, 1 for DNR) and included letter type as an explanatory variable (0 for control group, 1 for intervention group). The model tested whether the intervention lowered the percentage of recipients not returning a validation letter. A significant odds ratio (OR) of less than one would support the hypothesis. In addition, a separate multilevel model, which does not assume the units of analysis are independent, was also undertaken and running this analysis does not alter the study conclusions (results available upon request).

Pre-planned exploratory analysis. We pre-registered two forms of pre-planned exploratory analysis: (i) whether intervention effects hold across patient characteristics of sex and age group, (ii) whether intervention effects hold across hospital characteristics of hospital group and specialty. These were examined by applying binary logistic regression models with DNR as the outcome variable (0 for returned, 1 for DNR) and letter type as an explanatory variable (0 for control group, 1 for intervention group) and testing if the intervention interacts with each patient or hospital characteristic.

Table 4.1 Nudge Design Elements Used in the Intervention Letter

Nudge Category ¹ / Design Element	Application in the Letter	Motivation
Simplification:		
Use of plain English.	A plain English practitioner simplified the language.	The easier it is for people to process information, the more likely they are to enact a behaviour (Dolan et al. 2012). Highlighting key features can draw people's attention to important information quickly (Dolan et al. 2012).
Salience.	Important information was made salient by bolding.	
Increases in ease and convenience:		
Call for action.	An action orientated heading: "Please reply to this letter".	Reducing barriers, including the time that it takes to understand what to do, makes a behaviour more likely Sunstein (2014).
Chunking.	The steps involved were broken into discrete tasks presented (a) in bullet points, and (b) an image.	Breaking tasks into chunks has been effective in supporting behaviours (Dolan et al. 2012).
Other nudges:		
Personalisation.	Recipients were addressed by their first name.	Field studies show people are more likely to respond to letters that use their first name (Edwards et al. 2009).
Messenger Effects.	Closed by a named staff member. Direct phone number of the staff member is provided.	We are heavily influenced by who communicates information, and also affected by the feelings we have towards the messenger (Dolan et al. 2012).
Apology.	"I apologise you are still waiting."	Mixed evidence on the effect of an apology on wait tolerance (Cheng and Tsai 2014; Munichor and Rafaeli 2007). Some pre-existing validation letters included an apology.
Altruism, Value, Resource.	The letter stated: "We want to provide our valuable services to our patients as soon as we can. That is why we are checking our waiting list"	Lab research shows that clearly articulating intention (altruism), cost and value of the service positively influences client response (Bridger and Wood 2017).

¹ Sunstein (2014) argues that two of the ten most important categories of nudges are simplification (noting "Many programs fail, or succeed less than they might, because of undue complexity.") and increases in ease and convenience (defined as "reducing various barriers (including the time that it takes to understand what to do) is often helpful.")

4.4 Results

4.4.1 Participant Characteristics

The profile of the sample was 51.10% female with a mean age of 53.78 years (SD= 21.07, range: 2-99). The distribution of the intervention and control groups was similar by sex, age group, and hospital group (see Table 4.2).

Table 4.2 Participant Characteristics

Variable		Control	Intervention
Sex	Female	741 (51.8%)	718 (50.4%)
	Male	690 (48.2%)	706 (49.6%)
Age group	0-17	69 (4.8%)	75 (5.3%)
	18-39	313 (21.9%)	293 (20.6%)
	40-59	425 (29.7%)	387 (27.2%)
	60-79	498 (34.8%)	499 (35.0%)
	80+	126 (8.8%)	170 (11.9%)
Hospital group	RCSI Hospital Group	397 (27.7%)	395 (27.7%)
	UL Hospital Group	1,034 (72.3%)	1,029 (72.3%)
Total		1,431	1,424

4.4.2 Study Prediction

The DNR rate was lower in the group that received the intervention letter than the control letter. Binary logistic regression revealed a significant odds ratio for letter type of less than one, $OR = .756$, $SE = .069$, $p = .002$, 95% CI [.632, .904].

The DNR rate of patients was 23.97% (343/1,431) for the control group compared to 19.24% (274/1,424) for the intervention group. The intervention letter resulted in a reduction in the DNR rate of 4.73 percentage points, or a relative reduction of 19.73%.

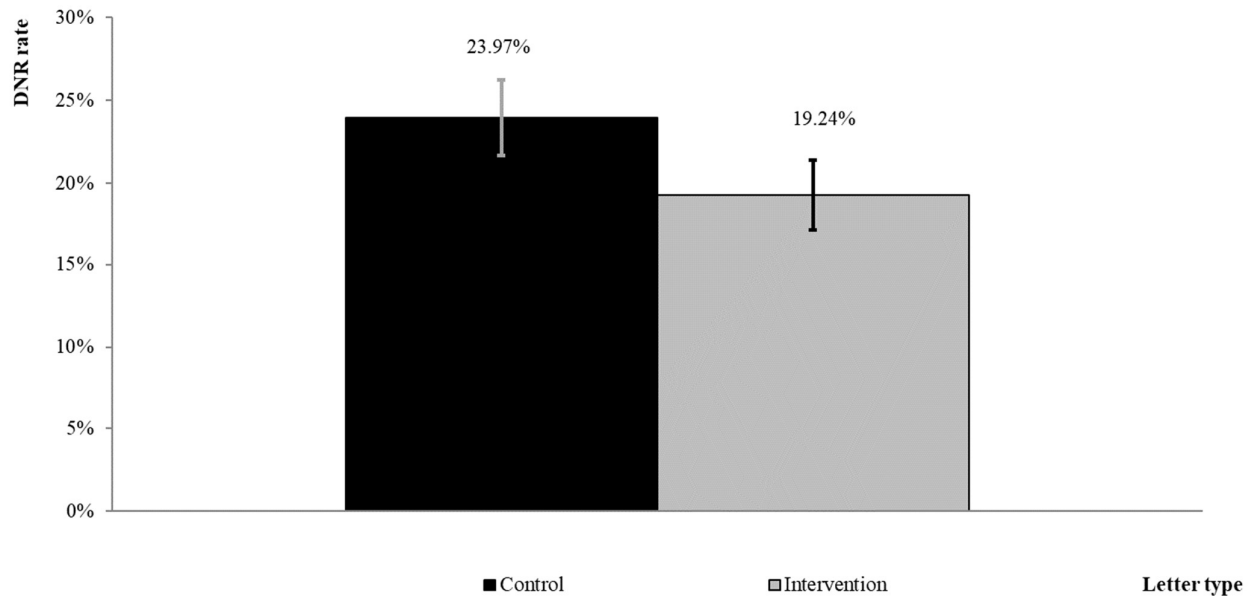


Figure 4.2 Did not return (DNR) rate by letter type.

Error bars represent the 95% confidence intervals.

4.4.3 Pre-planned Exploratory Analysis

Intervention effect by patient characteristics. A binary logistic regression model including a letter type and sex interaction term did not reveal a significant interaction, $OR = .849$, $SE = .155$, $p = .370$, 95% CI [.593, 1.215]. Directly comparing DNR rates between the control and intervention group for females and males suggests that the intervention held for both females and males (see Figure 4.3), although the difference for females was significant only at the 10% level.

To preserve power, we categorised age via a median split (<57 / ≥ 57 years). A binary logistic regression model including an interaction term for letter type and age revealed a significant interaction between letter type and age category, $OR = .681$, $SE = .130$, $p = .043$, 95% CI [.469, .989]. Separate DNR rates by age for the control and intervention group are shown in Figure 4.3, displaying the stronger effect of the intervention among older participants.

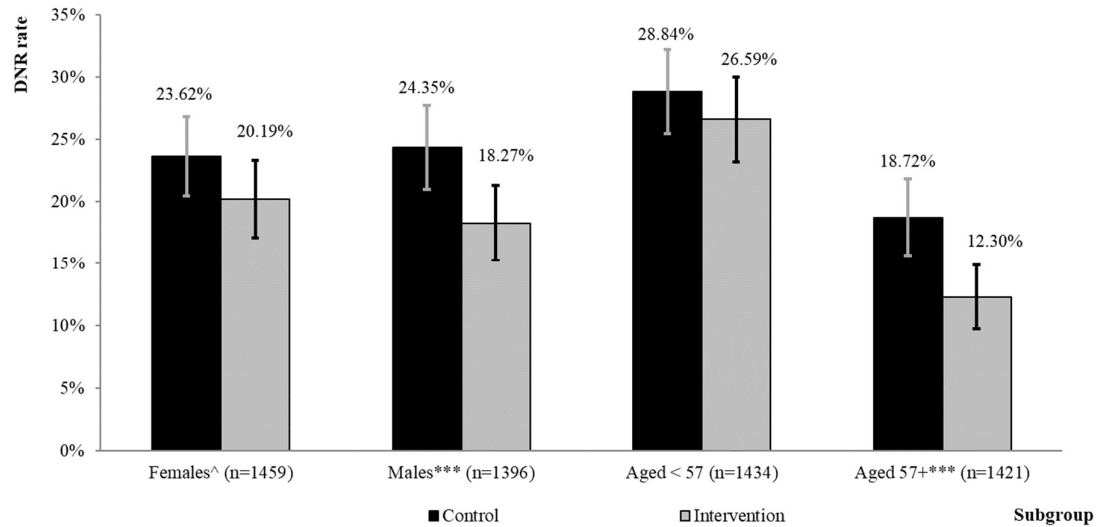


Figure 4.3 Did not return (DNR) rate by participant characteristic.

Error bars represent the 95% confidence intervals. ^ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. One-tailed tests.

Intervention effect by healthcare characteristics. A binary logistic regression model including a letter type and hospital group interaction term did not reveal a significant interaction between letter type and hospital group, $OR = 1.318$, $SE = .1.37$, $p = .171$, 95% CI [.888, 1.956]. It was not possible to run an equivalent regression analysis to test for interactions with specialty (which consisted of 14 levels) due to the small number of cells for many specialties. However, a direct comparison of DNR rates between the control and intervention groups found that the intervention effect was the expected sign for all six subgroups examined (relative reductions range of 14.61-31.60%) and that for four of the six the intervention effect achieved significance, see Figure 4.4.

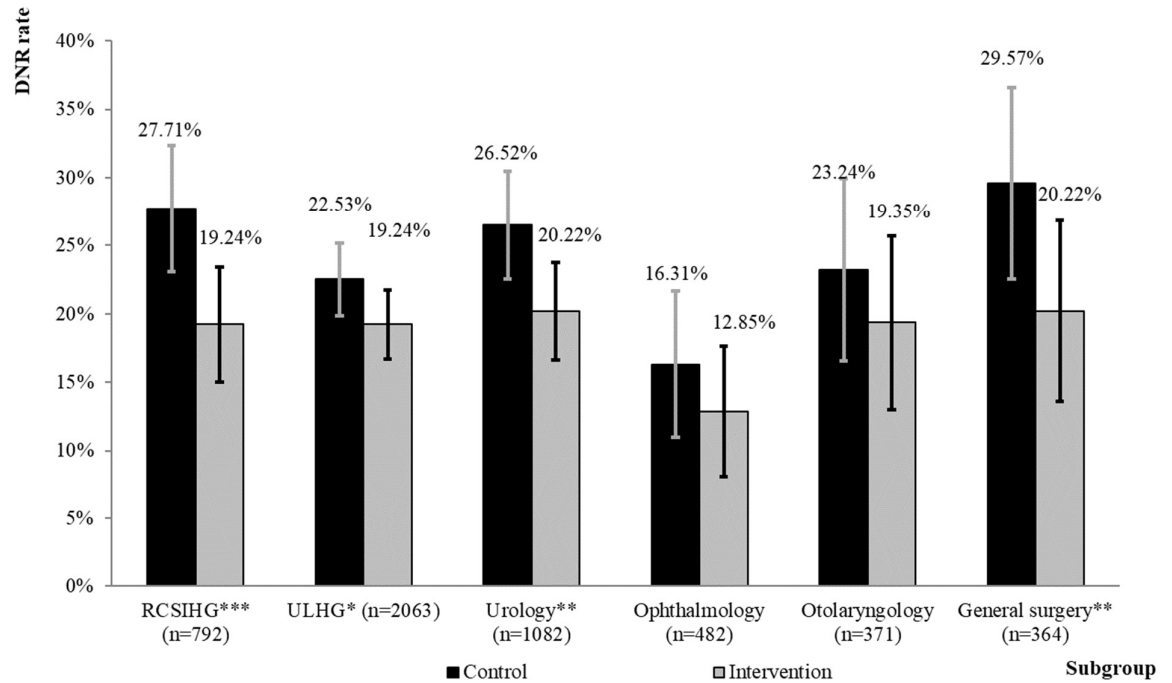


Figure 4.4 Did not return (DNR) rate by healthcare characteristic.

Error bars represent the 95% confidence intervals. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

One-tailed tests.

4.5 Discussion

4.5.1 Contribution to Knowledge

This is the first study to test whether the did not return (DNR) rate of patients can be decreased by incorporating nudges into letters used in the administrative postal validation of hospital waiting lists. A redesigned letter containing multiple nudges resulted in lower DNR rates by patients, a DNR rate of 23.97% in the control group compared to a DNR rate of 19.24% in the intervention group. The treatment effect of 4.73 percentage points is equivalent to 1 in 5 non-responders changing their behaviour because of the redesigned letter. HERE

A recent literature review of the effectiveness of nudging (DellaVigna and Linos 2022) identified four studies in which the effect of the inclusion of nudges in physical letters was tested. These four studies included a range of nudges across different policy areas and

behaviours (as discussed in the Introduction to this Chapter). Across the studies there were 12 intervention arms, eight interventions included a single nudge and four combined two or more nudges. The average treatment effect across the four studies was 1.67 percentage points. This suggests that the effect size from this current study is relatively large.

A recent literature review argues that the effectiveness of nudging may depend not only on the nudges used but also on how they are perceived by individuals and on the application context (Hummel and Maedche 2019). From an applied perspective, we found that the intervention appears to work cross the individual characteristics for which we have data of age and sex, and across the healthcare contexts for which we have data of hospital group and medical specialty. The finding of a stronger intervention effect for older adults (6.42 percentage points, a control group of a DNR rate of 18.72% compared to an intervention group DNR rate of 12.3%) was particularly welcome, as older people are a subgroup within the population with higher proportions of people with limited health literacy (Sorensen et al. 2015).

4.5.2 Strengths and Weaknesses

A strength of this study is that it was a randomised control trial with actual behaviour as the primary outcome, as opposed to behavioural intentions. The sample was large, at more than 2,800 patients. The intervention tested is also easily scalable. A weakness of this study is that the redesigned letter involved a combination of many nudges, so it is not possible to determine which specific nudges motivated the increase in engagement, nor to identify the psychological mechanisms behind the change in behaviour. This limitation often arises when behavioural scientists are engaged by policymakers to work towards a specific policy outcome; from an immediate policy perspective, *that* the intervention works matters more than *how* it works. Field trials can test only a limited number of interventions so there is an argument for combining multiple nudges into one letter on the assumption that one may work

and that none will interact negatively. This does not allow us to test mechanism but does increase the chance of having a positive impact. Either choice presents a risk, and this is a dilemma at the heart of scaling behavioural science interventions. Nevertheless, the study demonstrates the feasibility of hospital administrators engaging with behavioural scientists to apply nudges to a day-to-day implementation problem and to test against existing practice. Furthermore, the sample was not selected to be representative of inpatient and day case waiting lists nationally so the findings are not generalisable nationally. Nevertheless, the inclusion of waiting lists for eight hospitals from two hospital groups (one in the north-east and one in the mid-west of Ireland) and from 14 specialties along with the lack of within sample variation in effect size suggests that the findings may hold in other hospitals.

4.5.3 Research Impact

The Better Letter Initiative (BLI) quality improvement project had a substantial research impact, impacting on both nationwide practice and government policy. When describing the research impact, we use a framework (Belcher and Halliwell 2021) in which research outputs (products and services produced by research such as methods, discoveries, policy briefs, or networking) may lead to outcomes (a change in knowledge, attitudes, skills, or relationships which ideally manifest as a change in behaviour) which can generate realised benefits (a change in economic, social or environmental condition).

Outcome – National Practice. An important output from the BLI administrative validation project was a brief report which documented the impact on DNR rates of the redesigned validation letter. The report estimated that the use of the redesigned letter in bi-annual validation of 2017 waiting lists of three months plus for inpatient and day cases would result in at least 5,000 more patients responding upon receipt of a validation letter. This brief report was shared with the Head of the Unit in the NTPF with responsibility for waiting lists and with the Director in the Department of Health with responsibility for waiting list policy.

The research findings resulted in a change in the knowledge and attitudes of stakeholders, as the view emerged that using the redesigned validation letter would result in better use of resources and enable hospitals to meet the requirement to undertake postal validation cycles within given deadlines. In early 2018, after considering the headline results, the NTPF recommended a change in organisational practice in all public hospitals in Ireland, stating that hospitals should use the redesigned validation letter for postal administrative validation of inpatient and day case waiting lists. This recommendation, along with the redesigned letter template, was sent by the NTPF to all hospital group Chief Operations Officers; to all hospital Business Managers, Performance Managers, Schedule Care Leads, and Waiting List Leads; and provided to all attendees of the NTPF's Training and Development program for the Waiting List Management Protocol.

Outcome – National Policy. Another important output from the BLI administrative validation project was networking between the lead of the research team for the project with the Director in the Department of Health with responsibility for waiting list policy, and one of the topics discussed was the research process that was used as part of the BLI administrative validation project. To generate a sufficiently large sample, all non-maternity hospitals in the ULHG took part in the validation process. This was the first time that all relevant waiting lists in all hospitals of a hospital group were part of the same validation cycle. The use of this process in the BLI administrative validation project highlighted the benefit of a larger scale standardised process for the validation of waiting lists. Important benefits included the ability to validate a larger number of people on the waiting lists by taking advantage of economies of scale and the reduction in waiting lists that was achieved, as an estimated 20% of people on ULHG lists who replied indicated that they no longer required a procedure (subsequent research, Murphy, R and Taaffe, C 2020, showed that this was typically because they already had the procedure done). This change in knowledge led to a new national policy regarding the

organisation of the validation of hospital waiting lists. In June 2018 the Minister for Health wrote to the NTPF asking it (a) to develop a centralised process for validation and (b) to expand its validation remit to include outpatient waiting lists. The Minister's letter also requested that the NTPF use the learning from the BLI quality improvement project in this work. This change in policy was quickly implemented. By September 2018 a National Centralised Validation Unit (NCVU) was established within the NTPF which now works with public hospitals across Ireland to validate waiting lists using a streamlined and standardised process including a centralised postal service managed by the NCVU.

Outcome – Scaling Up of Behaviourally Informed Interventions. A common criticism of the application of nudges to public services is that, while hundreds of nudges have been tested over the past decades around the world, not many of these have been “scaled up” or successfully rolled out nationally. An outcome of the BLI administrative validation project was that two behaviourally informed letter templates (one for inpatient and day case procedure waiting lists and another for outpatient consultation waiting lists) were scaled up or rolled out nationally as they were used as the letter templates in the centralised postal service managed by the NCVU in the NTPF. For the validation of inpatient and day case waiting lists the NCVU adopted the redesigned validation letter that was tested during the BLI administrative validation project. Following expansion of the NTPF's validation remit to include outpatient waiting lists, the NCVU worked with the researchers and behavioural science advisory group from the BLI administrative validation project to design a letter for validation of outpatient waiting lists and used this new design (involving slight amendments to the redesigned inpatient/day case letter) for validation of outpatient waiting lists. Applying the terminology of the European Commission's classification of behavioural policy initiatives (Sousa Lourenco J et al. 2016), the BLI administrative validation project resulted in a behaviourally-tested application (i.e., the redesigned letter for validation of inpatient and day

procedure lists that was scaled up after the initial experiment) and a behaviourally-informed application (i.e., redesigned letter for validation of outpatient waiting consultation lists that was not tested in an experiment but was scaled up based on a review of previously existing behavioural evidence). During 2019 the redesigned letters were used in the validation of 266,493 waiting list places.

Realised Benefits. Patients are the main beneficiary of the BLI administrative validation project. Patients benefit in two ways. First, the use of the redesigned letters means that the validation process is clearer. We suspect that the higher response to the redesigned validation letter is because the content of the redesigned letter made it easier for a recipient to understand the importance of the validation process and what they were asked to do. Second, the more widespread checking of the accuracy of waiting lists on an ongoing basis improves patient waiting times by reducing the number of patients who do not attend for their appointment and by enabling hospitals to better manage waiting lists by improving data accuracy (National Treatment Purchase Fund (NTPF) n.d.). Between 2019 and 2021 the redesigned letters were used in the validation of 848,343 waiting list places.

5 Discussion

A fundamental objective of public health systems is to maximize the health of society given a limited budget. To achieve this objective, health policy decision makers need to decide which health treatments to fund (an ‘upstream’ issue), and how to support patients’ engagement with funded health treatments (a ‘downstream’ issue). This thesis applies a behavioural economics perspective to inform approaches to these two key issues. This chapter summarises the findings, discusses strengths and limitations of the thesis, highlights implications for future research, and considers possible applications of the approaches taken in the thesis.

5.1 Summary of Findings

5.1.1 Experience Informed Preferences for Ill-health

The use of the quality-adjusted life year (QALY) as a measure of benefit in cost-effectiveness analysis, as recommended by health economists, is widely used to inform decisions in public health systems on which health treatments to fund in order to maximize, within a limited budget, the health of society. The quality component of the QALY, or utility value, is based on people’s preferences for hypothetical health states obtained in surveys in which members of the general population are provided with functional definitions of the health states. However, it has been argued that the valuations of the general population produced using this approach are largely uninformed by what it is like to experience the health states described. This is because many members of the general population have no direct experience and only limited knowledge of the health states being judged. Furthermore, empirical findings suggest two important discords. First, members of the general population and patients’ preferences or mean valuations differ for the same health conditions. Second, the weights assigned to different dimensions of health (e.g., mobility, self-care, usual

activities, pain/discomfort, and anxiety/depression) as based on preferences of members of the general population obtained using current methods differ to those based on the impact that health dimensions have on the actual well-being of people. These issues have led to calls for the use of “experience informed general population values” for the quality component in QALYs. The suggestion is that respondents in health valuations studies should be provided with information on what living with specific health states is like before appraising those health states. This thesis includes two studies that explore the feasibility of obtaining informed general population appraisals of ill-health by providing a summary measure of patients’ experience of living with ill-health. These two studies also examine the effect of obtaining experience informed general population preferences and consider differences in prior beliefs across participants.

With regard to appraisals of hypothetical health conditions by members of the general population, Chapter 2 found: (a) a substantial share of members of the general population report having inaccurate views of patients’ mean ratings of health conditions; (b) a feasible way to obtain informed general population preferences for health conditions is to provide people with patients’ mean ratings of their health conditions before people appraise them; (c) there is an effect on the preferences of members of the general population of being informed of patients’ mean ratings of health conditions; (d) the direction of this effect depends on whether people have overly negative or overly positive prior beliefs of patients’ mean rating of the conditions –(i.e., participants adjust their valuations in the direction of the appraisals of those experiencing that conditions).

Regarding appraisals of hypothetical generic health states (i.e., which cover domains of health that are the type of descriptions of ill-health recommended for use by national health technology assessment guidelines developed in many countries) by members of the general population, Chapter 3 found: (a) members of the general population selectively underestimate

the effects of “moderate anxiety / depression” on well-being as compared to the effect of “moderate problems with self-care” and “moderate pain / discomfort” on well-being; (b) a feasible way to obtain informed general population preferences for health states is to provide people with a summary measure of the consequences of each health state, that outlines differences in either life satisfaction or day affect associated with living in each health state, before people appraise them; (c) informing people of actual changes in well-being associated with health states changes their appraisals of their relative undesirability; (d) being more informed about the association between health states and differences in well-being results in living with “anxiety / depression” being judged as less preferred relative to moderate problems with “self-care” or moderate “pain / discomfort”, and so of increased priority for being addressed by health treatments.

5.1.2 Increasing Engagement with the Management of Hospital Waiting Lists

This thesis argues that, as well as influencing the current approach to deciding which health treatments to fund, applying findings from behavioural economics can also influence the way in which people engage with funded health treatments. The widespread approach to correspondence with patients on funded health treatments is based on a belief that individuals who have expressed a desire for health services will subsequently engage with communications from providers of such services. For instance, they should attend appointments for health treatment, and if requested to, they should return forms in relation to future health treatment. Reflecting this belief, correspondence from health service providers to patients emphasises the factual aspects of health services (such as the time and date of an appointment, or the deadline for return of a form) but typically does not incorporate insights from psychology or from behavioural economics as to what motivates engagement with these requests. Despite the importance of hospital treatment for people’s health as well as low (or absence of) cost for treatment in public hospitals, “did not attend” (DNAs), when a patient

unexpectedly fails to attend an appointment, are a problem across public health systems. One way to reduce DNAs is to improve the accuracy of waiting lists by undertaking administrative validation, a process whereby administrators in a hospital write to patients on waiting lists to check if they still require a procedure or wish to be removed from the waiting list. Experience shows that a proportion of patients on waiting lists no longer require treatment, for instance, they may have had the treatment in another hospital, feel better, or no longer want a procedure. If waiting lists are accurate, then DNA rates can be reduced, as appointments are less likely to be offered to people who no longer require them. However, a challenge for hospital administrators when undertaking administrative validation is that a substantial percentage of patients typically do not return (DNR) a “validation letter” to confirm whether the procedure is still required (DNR rates of 25% are common).

Chapter 4 of this thesis therefore tested whether the DNR rate was reduced by introducing nudges to validation letters. The redesign approach used an iterative design process that was informed by research evidence, feedback on the redesign from the national agency responsible for waiting lists, hospital service providers and managers (via a national group and staff in the participating sites), a patient advocacy group, and an expert behavioural science advisory group. This led to a redesigned letter which included multiple nudges, many of which had not previously been used in a health context. The redesigned letter had the effect of reducing DNR rates, control and intervention group DNR rates were 23.97% and 19.24%. This is a relatively large effect. A treatment effect of 4.73 percentage points compared to an average treatment effect of 1.67 percentage points across four studies in which the effect of the inclusion of nudges in physical letters was tested across different policy areas and behaviours (DellaVigna and Linos 2022). The redesigned letter in the Chapter 4 was judged to increase patient compliance with the validation process, and it was

subsequently adopted (“scaled up”) as the national template for use in validation of waiting lists in Ireland.

5.2 Strengths and Limitations

A strength of this thesis is the methods used. The primary hypothesis of each study is tested by using the gold standard method of a randomized controlled trial (RCT). Furthermore, when implementing each RCT the good practice guidance within the CONSORT 2010 Statement is followed, including the 25-item checklist and the flow diagram. In addition, the study hypothesis of the two more recent studies is pre-registered (when undertaking the first study, Chapter 2, the author was not aware that pre-registration was good practice). The most appropriate statistical method is used to test each study’s primary hypothesis. Specifically, a 2 (between: Group) $\times$ 3 (within: Health Condition) $\times$ 3 (between: Prior Belief of Patients’ Mean Rating) mixed analysis of variance (ANOVA) is used for the rating outcome variable of Chapter 2. For the ordered outcome variable of Chapter 3 an ordered logistic model is applied, and for the binary outcome variable of returned or did not return of Chapter 4 a binary logistic regression model is applied. As a result, it is possible to assign causation to the intervention effects presented in this thesis and to form firm conclusions on each of the study hypotheses based on the findings presented. Another strength of this thesis is that each of the study samples are adequately powered to detect an effect.

Limitations of the study samples are that each study sample is not based on an entirely representative sample of the population and the health valuations survey technique of Chapters 2 and 3 is online rather than face to face. While the later limitation could potentially affect the validity of the findings in Chapters 2 and 3, the studies use simple appraisal techniques and there is evidence that online health valuation surveys are valid when simple appraisal techniques are used (Mulhern et al., 2013). National governments also view online

health valuation surveys as reliable, for example the valuations of EQ-5D health states used in Canada is based on a health valuation survey administered online (Bansback, Tsuchiya et al. 2012). Furthermore, both online samples were drawn from reliable sources. The source of the sample for Chapter 2, MTurk, is judged to be a source of high quality data (Crump, McDonnell et al. 2013). The source of the sample for Chapter 3, the Survey Spot Internet panel operated by Survey Sample International (SSI), includes information on over 1 million unique member households recruited through “permission-based” techniques.

The limitation of the samples not being fully representative of the target populations may influence the generalizability of the results. While the sample in Chapter 4 is not representative of inpatient and day case waiting lists nationally, the sample of 2,800 includes eight hospitals (across the north-east and the mid-west of Ireland) and 14 specialties which suggests that the findings may hold in other hospitals. In addition, although the sample of 1,259 for Chapter 3 is from an online panel, rather than a random sample of the population, the sample is stratified by six age groups and by region of residence (Scotland, Wales and nine regions in England) and reflects the adult age profile and regional distribution of people living in the three countries.

An additional strength of the applications tested in this thesis is that each application is designed so that it can be relatively easily scaled up or incorporated into existing practice. The information provided to obtain informed general population health valuations, the mean rating of the health states by people living in the health states (Chapter 2) or the association of the health states with differences in well-being (Chapter 3), can be incorporated into existing representative health valuation surveys of people from the general population. This information is simple for participants to understand, does not add substantial labour or processing costs, and can be applied to a wide range of health medical conditions and generic

health states. The redesigned waiting list validation letter of Chapter 4 has been subsequently adopted as national practice.

Therein lies another strength of this thesis, which is the research impact of the study in Chapter 4. It has substantial research impact, impacting on both nationwide practice and government policy, and realizing benefits for patients. Firstly, national organisational practice now requires all public hospitals in Ireland to use the redesigned letter template for validation of inpatient/day case waiting lists. Secondly, the findings of the study in Chapter 4 have led to a change in national policy. Based on the findings of this study, the Minister for Health wrote to the NTPF to develop a centralized process for validation, to expand its validation remit to include outpatient waiting lists, and to use the learning from the BLI quality improvement project in this work (this resulted in a new validation letter for outpatient waiting lists, co-designed with the research team, involving slight amendments to the redesigned and tested inpatient/day case letter). Between 2019 and 2021 the redesigned letters were used in the validation of 848,343 waiting list places. This results in realized benefits for patients as (a) the validation process is clearer for patients and (b) more accurate waiting lists improve patient waiting times by reducing the number of patients who do not attend their appointment.

A limitation of this thesis is that the studies do not identify the psychological mechanisms behind the change in behaviour or decision making. This limitation often arises in field trials which can test only a limited number of interventions, and a choice must be made between testing specific mechanisms or combining multiple nudges in the hope of increasing a positive impact. Nevertheless, Chapter 4 suggests that, contrary to previous evidence for uptake of influenza vaccination (Yokum et al. 2018), inclusion of multiple nudges in letters can increase positive behaviours. In the health valuation studies (Chapters 2 and 3) it is possible that the intervention may have created a focusing effect (Schkade and

Kahneman, 1998) whereby the intervention groups may have placed too much importance on the patient experience information provided. This thesis acknowledges the potential for focusing effects, but argues that when obtaining preferences of the general population, providing patients' information is warranted. This is because what people imagine as the consequences of ill health is not always accurate (Karimi, Brazier and Paisley, 2017a) and preferences can only be said to be informed when they are "freed from the distorting effects of factual errors" (Harsanyi, 1997). Suggestions on how to test for possible focusing effects are discussed in the next section.

5.3 Directions for Future Research

5.3.1 Further Testing of the Overall Effects of Experience Informed Preferences

This thesis shows that, for the health conditions / states examined, experience-informed general population preferences obtained differed to those obtained using the current approach. If these findings hold more generally then the use of experience-informed general population preferences would have implications for QALY values and resources allocation decisions informed by QALYs. Therefore, it will be important to examine the degree to which effects found in the health valuation studies of Chapters 2 and 3 apply to a wider range of health conditions / states, for traditionally used choice-based judgement methods such as the time trade-off method (Stolk et al. 2019) and for a random sample of the general population.

5.3.2 Testing of the Interaction between Prior Beliefs and Informing Preferences

This thesis found that the difference between experience-informed preferences and preferences obtained using the current approach depended on how information provided compared to participants expectations. Chapter 2 found that, when the mean ratings given by patients were higher (lower) than expected, participants in the intervention group provided

significantly higher (lower) valuations than participants in the control group, but when the mean ratings given by patients were broadly in line with participants' expectations there was no effect. As more research is undertaken to obtain experience informed general population preferences, it will be important to examine the extent to which the effect of the provision of information on what it is like to live in health states interacts with participants' prior beliefs of this. One immediate study that could be undertaken would be to re-run a version of the study of Chapter 3 but (a) using the choice-based judgement method of the time trade-off method to obtain preferences (which results in a quantitative value or utility for each health state assessed) and (b) to power the study to be able to test for each health state for a cross-over interaction, depending on the expectations about the consequences of health states for life satisfaction/day affect.

5.3.3 Communicating HRQOL to Obtain Experience Informed Preferences

Developing ways of briefing people from the general population on the experiences of those in particular health states is identified as a research priority for the future of the QALY by the Consensus Development Group Drummond, M., Brixner, D., Gold, M., Kind, P., McGuire, A., Nord, E., & Consensus Dev Grp. (2009). Toward a consensus on the QALY. *Value in Health*, 12, S31-S35. doi:10.1111/j.1524-4733.2009.00522.x. Relatively little effort has gone into examining the effect on health valuations of making people from the general population more informed Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., . . . Tsuchiya, A. (2005). Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy*, 4(4), 201-8. doi:10.2165/00148365-200504040-00002. The one notable exception prior to this thesis provided survey participants with audio recordings of three patients discussing how they adapted to life with rheumatoid arthritis (McTaggart-Cowan, Tsuchiya, O'Cathain, and Brazier (2011). This thesis introduced the use of summary measures to convey to members of the general population what health

states are like, patients' mean rating of a health condition (Chapter 2) and the well-being consequences of living in a health state (Chapter 3). The use of summary measures has the advantages that a relatively short amount of time and attention is required of survey participants to receive the information, and the information provides a representative account of how people experience a health condition/health state. This thesis also noted that the underlying construct that the quality component of the QALY is intended to capture is health-related quality of life (HRQOL) or "the impact of the health aspects of an individual's life on that person's quality of life, or overall well-being" (Gold et al. 1996:285). Chapter 3 of this thesis argues that it therefore follows that the impact of health states on well-being is the most appropriate information to provide to obtain experience informed preferences for hypothetical health states for use in resource allocation.

It was further noted that as HRQOL is concerned with the impact of disorders on a person's well-being, information to convey the HRQOL of a person could be based on (a) a person's own opinion on their well-being or (b) normatively by the satisfaction of externally defined criteria about what constitutes a person's well-being (Wood and Davidson 2020). Chapter 3 showed that it is feasible to brief participants in health valuation studies of the HRQOL consequences of health states by using two widely used measures of well-being drawn from each of the two approaches, life satisfaction (LS) which is a "a global assessment of a person's quality of life according to his chosen criteria" (Shin, D.C. and Johnson 1978) and day affect (DA) which captures the extent of both positive and negative affect a person reports to typically experience for a set of externally defined emotions. While DA is a widely used measure, it stems from a subjective well-being (SWB) conceptualization of well-being. There are other normative approaches to capture externally defined criteria about what constitutes a person's well-being, such as the construct of psychological well-being (PWB). While PWB and SWB are highly correlated, they are distinct constructs (Keyes, Shmotkin

and Ryff, 2002; Biaobin, Xue and Lin, 2004; Linley et al., 2009). Therefore, other normatively based measures of well-being, such as the PWB approach of Ryff's (Ryff, C.D. 1989) six factors of positive relations with others, self-acceptance, purpose in life, autonomy, environmental mastery, and personal growth, could be useful. Future research should therefore examine the influence on health assessments of obtaining informed general population preferences by providing information on the consequences of health states for PWB measures of well-being.

5.3.4 Informing Debate on How to Value the Q in the QALY

A number of possible approaches of obtaining QALY weights have been suggested in the literature. Most notably, using general population valuations, using patient valuations, or using experience-informed general population valuations. It would be useful for future research to stimulate and to inform discussion between stakeholders on these approaches. Areas where future research will help to inform this debate are outlined below.

Understanding influences on appraisals of ill-health. Our level of understanding of the factors considered and the cognitive mechanisms at play during appraisals of ill-health is relatively limited. There is a need, as researchers have argued (Menzel et al., 2002; Ubel, Loewenstein and Jepson, 2003; Brazier et al., 2005; Stiggelbout and De Vogel-Voogt, 2008; Versteegh and Brouwer, 2016), for a better understanding of the cognitive mechanisms that influence such appraisals by members of the general population.

In the context of resource allocation decisions, whether it is normatively appropriate to provide people with information on patients' mean rating of a health state or information derived from the self-reported well-being of patients (both of which are likely to partly reflect adaptation to ill-health) depends in part on whether adaptation that occurs can be viewed as admirable achievements or as deficiencies (Menzel et al., 2002). Therefore, further research is needed to better understand patients' experiences of living with different forms of ill health

(such as which major elements of adaptation or other factors have the most influence), and to obtain a greater appreciation of the cognitive mechanisms that influence judgement of ill health states by patients.

A few frameworks propose reasons for differences between patient and nonpatient valuations of health states (Menzel, Dolan et al. 2002, Ubel, Loewenstein et al. 2003, Brazier, Akehurst et al. 2005, Stiggelbout, De Vogel-Voogt 2008). These frameworks may be helpful to inform studies examining the influences on general population appraisals of hypothetical health states and on patient appraisals of ill-health.

Examining the mechanisms leading to changes in preferences as a result of being informed. It would be useful for future research to examine the direct mechanisms that lead to the changes in health state assessments as a result of being informed of the experience of people living in health states (i.e., what is it that leads to differences between experienced informed preferences and preferences obtained using the current approach).

Several cognitive biases (such as loss-framing, failure to anticipate adaptation and focusing illusion) may loom larger in the assessment of health conditions / states for which the general population on average value as worse than patients or among respondents who misconceive ill health states as worse than is typically experienced by patients. This could be examined by expanding the approaches taken in this thesis. For instance, the approach of providing mean patient ratings (as per Chapter 2) could be applied to conditions, such as mental health conditions, for which patients' mean rating of a health condition are not higher than nonpatients' rating (Revicki, Shakespeare et al. 1996, Rosser, Kind 1978). In addition, the approach of Chapter 3 should be run but with sufficient power to investigate whether there is asymmetry in the size of the effect of providing information on the well-being consequences of health states to people from the general population (i.e., whether the effect size differs between those with overly high or low estimates of well-being consequences).

Calls for the use of experience-informed general population preferences assume that providing information on the experience of people in ill-health will improve the validity of preferences obtained for hypothetical health states. However, there is the possibility that the information provided leads to unintended consequences or introduces biases in decision making. Regarding the possibility that the provision of patient information may introduce focusing effects to health assessments, this should be gauged by incorporating reflection and deliberation techniques (Shiell, Hawe and Fletcher, 2003; Stein et al., 2006) into future studies that otherwise replicate the approaches of Chapters 2 and 3.

Understanding stakeholders' views on approaches to valuing the Q in the QALY. It would be informative to obtain the views of the general population on the appropriateness of the alternative approaches to obtaining QALY weights (i.e., by using general population valuations, using patient valuations, or using experience-informed general population valuations). It would also be useful to obtain the views of policymakers and health service managers on the advantages and the disadvantages of using each of the above approaches to inform resource allocation. The views of health resource allocation decision makers should also be obtained on whether it would be most useful to conceive each approach as substitutes (i.e., in which decision makers are presented with QALYs where the Q is based on one approach only) or instead to take a complementarity perspective (i.e., where decision makers are presented with a set of QALYs using each approach).

5.3.4 Understanding the Mechanisms Leading to Reduced DNR Rates

It would be useful to examine the direct mechanisms that led to the change in behaviour as reported in Chapter 4. Of the multiple nudges included in the redesigned letter of Chapter 4, it would be particularly useful to examine separately the effect of two particular nudges on engagement with administrative validation, but also on engagement with health service appointment offers more generally. Firstly, an apology, for which we are not aware of

health context evidence and the evidence is mixed from other sectors (Cheng and Tsai 2014; Munichor and Rafaeli 2007). Secondly, the combined articulation of altruistic intentions, the value of the service, and resource involved in the validation process, for which laboratory evidence suggests this may positively influence client response (Bridger and Wood 2017). However, we are not aware of evidence from field trials.

5.4 Applications

5.4.1 Applications to Health Valuation Surveys

The studies in Chapter 2 and 3 demonstrate that it is possible to brief members of the general population of patients' experience of ill health in a way that can be incorporated into the current approach of obtaining QALY weights for health conditions (i.e., specific medical conditions) and generic health states (e.g., the dimensions of health on the EQ-5D descriptive system). The approach of providing a summary measure, in this thesis a measure of patients' mean rating of a health condition, and a summary measure of the well-being consequences of living in a health state, has the advantages that the information provides a representative account of how people experience a health condition/health state and the amount of time and attention required of survey participants to receive the information is relatively short.

The experience-informed preferences obtained in Chapters 2 and 3 differed to those obtained using the current approach. If the findings of Chapters 2 and 3 hold more generally (i.e., for additional health conditions/states, for techniques using choice-based judgements such as the TTO method, and for random sample of the general population), then the provision of patients' experiences could be incorporated into the current approach to health valuation surveys to produce experience-informed general population weights or utilities for QALYs. In the immediate term, and until there is a better understanding of influences on

general population and patient appraisals of ill-health, it would seem most appropriate for this approach to supplement rather than to substitute the existing approach. That is, resource allocation decision makers in health should be provided with both QALYs computed using the existing approach of general population weights and with QALYs computed using experience informed general population weights.

5.4.2 Applications to Engagement with Health Services

The broad approach used in Chapter 4 – a collaborative approach involving researchers in a Ministry, state agencies responsible for managing and delivering services, and a relatively small expert behavioural science advisory group drawn from universities and not-for-profit bodies – is an approach to determining how best to use findings from behavioural economics to inform engagement with health services that could be applied to a range of different public service areas. The study demonstrates the feasibility of hospital administrators engaging with behavioural scientists to apply nudges to a day-to-day implementation problem and to test against existing practice. Indeed, the project reported in Chapter 4 led to the creation of the Better Letter Initiative (BLI) which seeks to support better patient engagement with health services provided by hospitals in Ireland by incorporating findings from behavioural science into written communications used by health system administrators. The approach has been applied to five additional projects under the BLI. Using the classification of behavioural policy initiatives (Sousa Lourenco J et al. 2016), the BLI involves both behaviourally-informed and behaviourally-tested applications.

Two of the projects (now complete) involve a behaviourally-informed application, that is re-design that was not tested in a RCT experiment but was scaled up / adopted nationally based on a review of previously existing behavioural evidence, suggestions from an expert advisory group and feedback from stakeholders. Specifically, these project involved the application of findings from behavioural science to create the first NTPF/HSE

commissioning offer letter (where patients on a waiting list are offered earlier treatment in a private hospital paid for by the State), and the application of findings from behavioural science to the re-design of waiting list validation letters to include an online response option and to the design of an online portal for patients to respond (subsequently named the Patient Online Automated Response System, POLAR).

Three of the projects (one complete and two ongoing at the time of submission of this thesis) involve a behaviourally-tested application, that is the redesign is subject to an initial RCT experiment in anticipation of it being impactful and subsequently scaled up / adopted nationally. These involved a project to redesign an appointment offer letter (“To Come In”, TCI letter) used for inpatient/day case hospital appointments with the aim to increase patients’ confirmation upon receipt of a letter of their intention to attend and to reduce did not attend (DNA) rates on appointment day, a project to redesign SMS reminders for outpatient appointments with the objective to reduce DNA rates, and a project to redesign a TCI letter for outpatient hospital appointments to seek to reduce DNA rates.

The BLI is ongoing and the possibility to expand to additional applications remains under review. To increase the possibility of international or non-health applications of the approach taken in the BLI, a brief case study of the project reported in Chapter 4 is disseminated in international good practice hubs, specifically the OECD’s Observatory of Public Sector Innovation (OECD n.d.) and of the WHO’s Behavioural and Cultural Insights (BCI) Hub (WHO n.d.).

5.5 Conclusion

To maximise the health of society within a limited budget, decision makers in public health systems need to decide which health treatments to fund and how best to support engagement by patients with treatments. Current practice in these issues is heavily influenced

by the neoclassical model of rational choice theory. In this thesis we apply an alternative behavioural economics perspective to inform decisions on which treatments to fund and how to support engagement with services.

In many public health systems, decisions on which treatments to fund are informed by the expected gains in patients' quality-adjusted life years (QALYs). QALYs for a specific health state are obtained by multiplying expected average life-years gained from treatment by a utility value which measures the strength of people's preferences for a specific health state. Rational choice theory predicts that when members of the general population are asked to appraise hypothetical health states, the values they place on health states will be consistent with the utility they will obtain from each health state. In contrast, behavioural economics suggests that when imagining their future self in hypothetical health states, people may not have accurate expectations of what the negative consequences of ill-health on important domains of their life would be and on how their overall well-being would be impacted. This could affect the strength of people's preferences for hypothetical health states, and there may be a divergence between decision utility and experienced utility. Concerns have been raised about the way in which the quality component of QALYs is calculated, leading to calls for ways to obtain experience informed general population preferences.

With regard to appraisals of hypothetical health conditions / states by members of the general population, this thesis found: (a) a substantial share of members of the general population report having inaccurate views of patients' mean ratings of health conditions and of the consequences of health states for well-being, (b) a feasible way to obtain experience informed general population preferences for health conditions is to provide people with patients' mean ratings of their health conditions and to obtain experience informed preferences for health states is to provide people with a summary measure of the differences in either life satisfaction or day affect associated with living in each health state; (c) there is

an effect on the preferences of members of the general population of being informed of patients' mean ratings of health conditions and or a summary measure of the differences in well-being associated with living in each health state; (d) the direction of this effect depends on whether people have overly negative or overly positive prior beliefs of health conditions / states. Crucially, experience informed preferences differ to those obtained using the current method.

Applying rational choice theory to patient engagement with health services leads to the prediction that individuals who have expressed a desire for health services will subsequently engage with communications from such services providers (e.g., attend appointments and return forms). Such behaviours would be in line with people's personal preferences (having already expressed a need for a service), hospital treatment can improve people's health and treatment is provided at no or low cost to the patient. Existing health service management practice is consistent with this prediction. Correspondence from health service providers to patients emphasise the factual aspects of health services (such as the time and date of an appointment, or the deadline for return of a form) but typically does not incorporate insights from psychology or from behavioural economics as to what motivates engagement with these requests. The phenomenon of patients not attending hospital appointments without providing prior notice to hospital administrators is a problem across public health systems.

One way to reduce non-attendance is to improve the accuracy of waiting lists by undertaking administrative validation, a process whereby administrators in a hospital write to patients on waiting lists to check if they still require a procedure or wish to be removed from the waiting list. A challenge for hospital administrators is that a substantial percentage of patients typically do not return (DNR) a "validation letter", DNR rates of 25% are common. This thesis shows that DNR rates can be reduced by incorporating nudges into validation

letters. The redesigned and tested validation letter of Chapter 4 was subsequently adopted for use in validation of waiting lists by all public hospitals in Ireland. The broad approach followed in Chapter 4 of this thesis is being applied in Ireland to improve engagement with additional aspects of the management of health services.

This thesis shows that, as well as influencing the current approach to deciding which health treatments to fund, applying findings from behavioural economics can also influence the way in which people engage with funded health treatments. Taken together these studies show the importance of applying a behavioural economics perspective to inform decisions on how to maximise the health of society.

References

- Baddeley, M., M. Castells, and A. Guiora. 2019. *Towards a New Enlightenment?: A Transcendent Decade*. BBVA.
- Baker, R., and A. Robinson. 2004. “Responses to Standard Gambles: Are Preferences ‘Well Constructed’?” *Health Economics* 13(1):37–48. doi: 10.1002/hec.795.
- Bansback, Nick, Aki Tsuchiya, John Brazier, and Aslam Anis. 2012. “Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies.” *Plos One* 7(2):e31115–e31115. doi: 10.1371/journal.pone.0031115.
- Baron, J., DA Asch, A. Fagerlin, C. Jepson, G. Loewenstein, J. Riis, MG Stineman, and PA Ubel. 2003. “Effect of Assessment Method on the Discrepancy between Judgments of Health Disorders People Have and Do Not Have: A Web Study.” *Medical Decision Making* 23(5):422–34. doi: 10.1177/0272989X03257277.
- Behavioural Insights Team. 2014. *EAST: Four Simple Ways to Apply Behavioural Insights*.
- Behavioural Insights Unit (BIU). n.d. *Spotlight on Health Results: Behavioural Insights Short Report*. New South Wales, Australia: NSW Department of Premier and Cabinet.
- Belcher, B., and J. Halliwell. 2021. “Conceptualizing the Elements of Research Impact: Towards Semantic Standards.” *Humanities & Social Sciences Communications* 8(1). doi: 10.1057/s41599-021-00854-2.
- Biaobin, Y., Z. Xue, and Q. Lin. 2004. “SWB & PWB: The Distinction and Integration of Two Well-Being Research Orientation.” *Psychological Science-Shanghai* 27(4; ISSU 150):836–38.
- Birch, S., and A. Gafni. 2004. “The ‘NICE’ Approach to Technology Assessment: An Economics Perspective.” *Health Care Management Science* 7(1):35–41. doi: 10.1023/B:HCMS.0000005396.69890.48.

- Brandon, A., JA List, RD Metcalfe, MK Price, and F. Rundhammer. 2019. "Testing for Crowd out in Social Nudges: Evidence from a Natural Field Experiment in the Market for Electricity." *Proceedings of the National Academy of Sciences of the United States of America* 116(12):5293–98. doi: 10.1073/pnas.1802874115.
- Brazier, J., R. Akehurst, A. Brennan, P. Dolan, K. Claxton, C. McCabe, M. Sculpher, and A. Tsuchiya. 2005a. "Should Patients Have a Greater Role in Valuing Health States?" *Applied Health Economics and Health Policy* 4(4):201–8. doi: 10.2165/00148365-200504040-00002.
- Brazier, J., R. Akehurst, A. Brennan, P. Dolan, K. Claxton, C. McCabe, M. Sculpher, and A. Tsuchiya. 2005b. "Should Patients Have a Greater Role in Valuing Health States?" *Applied Health Economics and Health Policy* 4(4):201–8. doi: 10.2165/00148365-200504040-00002.
- Bremner, Karen E., Christopher A. K. Y. Chong, George Tomlinson, Shabbir M. H. Alihahi, and Murray D. Krahn. 2007. "A Review and Meta-Analysis of Prostate Cancer Utilities." *Medical Decision Making* 27(3):288–98. doi: 10.1177/0272989X07300604.
- Bridger, EK, and A. Wood. 2017. "Gratitude Mediates Consumer Responses to Marketing Communications." *European Journal of Marketing* 51(1):44–64. doi: 10.1108/EJM-11-2015-0810.
- Briggs, A., and A. Gray. 2000. "Economics Notes - Using Cost Effectiveness Information." *British Medical Journal* 320(7229):246–246. doi: 10.1136/bmj.320.7229.246.
- Brouwer, Werner B. F., Saskia Grootenboer, and Pedram Sendi. 2009. "The Incorporation of Income and Leisure in Health State Valuations When the Measure Is Silent: An Empirical Inquiry into the Sound of Silence." *Medical Decision Making* 29(4):503–12. doi: 10.1177/0272989X09336161.

- Chadborn, T and Berry, D. n.d. *Increasing Uptake and Improving Outcomes through Behavioural Insights: Initial Results and Methodologies, Health Check Results*.
London, England: Public Health England Behavioural Insights Team.
- Chater, Nick, and George Loewenstein. 2022. “The I-Frame and the s-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray.” *Behavioral and Brain Sciences* 1–60. doi: 10.1017/S0140525X22002023.
- Cheng, YH, and YC Tsai. 2014. “Train Delay and Perceived-Wait Time: Passengers’ Perspective.” *Transport Reviews* 34(6):710–29. doi: 10.1080/01441647.2014.975169.
- Crump, Matthew J. C., John V. McDonnell, and Todd M. Gureckis. 2013. “Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research.” *Plos One* 8(3):e57410–e57410. doi: 10.1371/journal.pone.0057410.
- Damschroder, L. J., B. J. Zikmund-Fisher, and P. A. Ubel. 2005. “The Impact of Considering Adaptation in Health State Valuation.” *Social Science and Medicine* 61(2):267–77.
- Damschroder, L. J., B. J. Zikmund-Fisher, and P. A. Ubel. 2008. “Considering Adaptation in Preference Elicitations.” *Health Psychology* 27(3):394–99.
- De Wit, G. A., J. J. V. Busschbach, and F. T. De Charro. 2000. “Sensitivity and Perspective in the Valuation of Health Status: Whose Values Count?” *Health Economics* 9(2):109–26. doi: 10.1002/(SICI)1099-1050(200003)9:2<109::AID-HEC503>3.0.CO;2-L.
- DellaVigna, S., and E. Linos. 2022. “RCTs to Scale: Comprehensive Evidence From Two Nudge Units.” *Econometrica* 90(1):81–116. doi: 10.3982/ECTA18709.
- Dolan, P., M. Hallsworth, D. Halpern, D. King, R. Metcalfe, and I. Vlaev. 2012. “Influencing Behaviour: The Mindspace Way.” *JOURNAL OF ECONOMIC PSYCHOLOGY* 33(1):264–77. doi: 10.1016/j.joep.2011.10.009.

- Dolan, P., and D. Kahneman. 2008. "Interpretations of Utility and Their Implications for the Valuation of Health." *Economic Journal* 118(525):215–34. doi: 10.1111/j.1468-0297.2007.02110.x.
- Dolan, P., G. Kavetsos, and A. Tsuchiya. 2013. "Sick but Satisfied: The Impact of Life and Health Satisfaction on Choice between Health Scenarios." *Journal of Health Economics* 32(4):708–14. doi: 10.1016/j.jhealeco.2013.04.002.
- Dolan, P., and R. Metcalfe. 2012. "Valuing Health: A Brief Report on Subjective Well-Being versus Preferences." *Medical Decision Making* 32(4):578–82. doi: 10.1177/0272989X11435173.
- Dolders, M. G. T., M. P. A. Zeegers, W. Groot, and A. Ament. 2006. "A Meta-Analysis Demonstrates No Significant Differences between Patient and Population Preferences." *Journal of Clinical Epidemiology* 59(7):653–64.
- Drummond, M., D. Brixner, M. Gold, P. Kind, A. McGuire, E. Nord, and Consensus Dev Grp. 2009. "Toward a Consensus on the QALY." *Value in Health* 12(Journal Article):S31–35. doi: 10.1111/j.1524-4733.2009.00522.x.
- Edwards, PJ, I. Roberts, MJ Clarke, C. DiGuseppi, R. Wentz, I. Kwan, R. Cooper, LM Felix, and S. Pratap. 2009. "Methods to Increase Response to Postal and Electronic Questionnaires." *Cochrane Database of Systematic Reviews* (3). doi: 10.1002/14651858.MR000008.pub4.
- EuroQol Research Foundation. 2015. *EQ-5D-3L User Guide*. Rotterdam: EuroQol Group.
- Fellner, G., R. Sausgruber, and C. Traxler. 2013. "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information." *Journal of the European Economic Association* 11(3):634–60. doi: 10.1111/jeea.12013.
- Fryback, D. G. 2003. "Whose Quality of Life? Or Whose Decision?" *Quality of Life Research* 12(6):609–10.

- Gilbert, D. T., E. C. Pinel, T. D. Wilson, S. J. Blumberg, and T. P. Wheatley. 1998. "Immune Neglect: A Source of Durability Bias in Affective Forecasting." *Journal of Personality and Social Psychology* 75(3):617–38. doi: 10.1037//0022-3514.75.3.617.
- Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein. 1996. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.
- Goswami, I., and O. Urminsky. 2016. "When Should the Ask Be a Nudge? The Effect of Default Amounts on Charitable Donations." *Journal of Marketing Research* 53(5):829–46. doi: 10.1509/jmr.15.0001.
- Grether, DM, and CR Plott. 1979. "Economic Theory of Choice and the Preference Reversal Phenomenon." *American Economic Review* 69(4):623–38.
- Hallsworth, M., D. Berry, M. Sanders, A. Sallis, D. King, I. Vlaev, and A. Darzi. 2015. "Stating Appointment Costs in SMS Reminders Reduces Missed Hospital Appointments: Findings from Two Randomised Controlled Trials." *PLOS ONE* 10(9). doi: 10.1371/journal.pone.0137306.
- Hallsworth, Michael. 2023. "A Manifesto for Applying Behavioural Science." *Nature Human Behaviour* 7(3):310–22. doi: 10.1038/s41562-023-01555-3.
- Hallsworth, Michael, and Elspeth Kirkman. 2020. *Behavioral Insights*.
- Happich, M., and T. von Lengerke. 2005. "Valuing the Health State 'Tinnitus': Differences between Patients and the General Public." *Hearing Research* 207(1–2):50–58. doi: 10.1016/j.heares.2005.04.002.
- Harsanyi, JC. 1997. "Utilities, Preferences, and Substantive Goods." *Social Choice and Welfare* 14(1):129–45.
- Hirschman, Daniel, and Elizabeth Popp Berman. 2014. "Do Economists Make Policies? On the Political Effects of Economics1." *Socio-Economic Review* 12(4):779–811. doi: 10.1093/ser/mwu017.

- Howley, Peter, and Neel Ocean. 2022. “Can Nudging Only Get You so Far? Testing for Nudge Combination Effects.” *European Review of Agriculture Economics* 49(5):1086–1112. doi: 10.1093/erae/jbab041.
- Hudson, Nathan W., Richard E. Lucas, and M. Brent Donnellan. 2022. “A Direct Comparison of the Temporal Stability and Criterion Validities of Experiential and Retrospective Global Measures of Subjective Well-Being.” *Journal of Research in Personality* 98(Journal Article):104230–104230. doi: 10.1016/j.jrp.2022.104230.
- Hummel, D., and A. Maedche. 2019. “How Effective Is Nudging? A Quantitative Review on the Effect Sizes and Limits of Empirical Nudging Studies.” *Journal of Behavioral and Experimental Economics* 80:47–58. doi: 10.1016/j.socec.2019.03.005.
- Hurst, Jeremy, and Luigi Siciliani. 2003. “Tackling Excessive Waiting Times for Elective Surgery.”
- Jiang, RX, J. Shaw, A. Muhlbacher, TA Lee, S. Walton, T. Kohlmann, R. Norman, and AS Pickard. 2021. “Comparison of Online and Face-to-Face Valuation of the EQ-5D-5L Using Composite Time Trade-Off.” *Quality of Life Research* 30(5):1433–44. doi: 10.1007/s11136-020-02712-1.
- Kahneman, D., and A. Tversky. 1979. “Prospect Theory - Analysis of Decision Under Risk.” *Econometrica* 47(2):263–91. doi: 10.2307/1914185.
- Kahneman, Daniel, 1934- author. 2011. *Thinking, Fast and Slow*. 1st ed. New York : Farrar, Straus and Giroux, [2011] ©2011.
- Karimi, M., J. Brazier, and S. Paisley. 2017a. “Are Preferences over Health States Informed?” *Health and Quality of Life Outcomes* 15(Journal Article):105–105. doi: 10.1186/s12955-017-0678-9.

- Karimi, M., J. Brazier, and S. Paisley. 2017b. "How Do Individuals Value Health States? A Qualitative Investigation." *Social Science & Medicine* 172(Journal Article):80–88. doi: 10.1016/j.socscimed.2016.11.027.
- Karter, AJ, MM Parker, HH Moffet, AT Ahmed, A. Ferrara, JY Liu, and JV Selby. 2004. "Missed Appointments and Poor Glycemic Control - An Opportunity to Identify High-Risk Diabetic Patients." *MEDICAL CARE* 42(2):110–15. doi: 10.1097/01.mlr.0000109023.64650.73.
- Keyes, CLM, D. Shmotkin, and CD Ryff. 2002. "Optimizing Well-Being: The Empirical Encounter of Two Traditions." *Journal of Personality and Social Psychology* 82(6):1007–22. doi: 10.1037//0022-3514.82.6.1007.
- Kind, P., and P. Dolan. 1995. "The Effect of Past and Present Illness Experience on the Valuations of Health States." *Medical Care* 33(4):AS255–63.
- Kind, P., P. Dolan, C. Gudex, and A. Williams. 1998. "Variations in Population Health Status: Results from a United Kingdom National Questionnaire Survey." *British Medical Journal* 316(7133):736–41.
- Krol, M., W. Brouwer, and P. Sendi. 2006. "Productivity Costs in Health-State Valuations - Does Explicit Instruction Matter?" *Pharmacoeconomics* 24(4):401–14. doi: 10.2165/00019053-200624040-00009.
- Lacey, Heather P., Angela Fagerlin, George Loewenstein, Dylan M. Smith, Jason Riis, and Peter A. Ubel. 2008. "Are They Really That Happy? Exploring Scale Recalibration in Estimates of Well-Being." *Health Psychology* 27(6):669–75. doi: 10.1037/0278-6133.27.6.669.
- Lazarus, R. S., and S. Folkman. 1984. *Stress, Appraisal and Coping*.

- Lichtenstein, S., and P. Slovic. 1971. "Reversals of Preference Between Bids and Choices in Gambling Decisions." *Journal of Experimental Psychology* 89(1):46-+. doi: 10.1037/h0031207.
- Linley, P. Alex, John Maltby, Alex M. Wood, Gabrielle Osborne, and Robert Hurling. 2009. "Measuring Happiness: The Higher Order Factor Structure of Subjective and Psychological Well-Being Measures." *Personality and Individual Differences* 47(8):878–84. doi: 10.1016/j.paid.2009.07.010.
- Lunn, Pete. 2014. "Regulatory Policy and Behavioural Economics."
- Marteau, Theresa M., David Ogilvie, Martin Roland, Marc Suhreke, and Michael P. Kelly. 2011. "Judging Nudging: Can Nudging Improve Population Health?" *BMJ* 342:d228. doi: 10.1136/bmj.d228.
- Mayo Foundation for Medical Education and Research. "Tinnitus."
- Mažar, Nina, and Dilip Soman, eds. 2022. *Behavioral Science in the Wild*. University of Toronto Press.
- McTaggart-Cowan, H., M. 2011. "Elicitation of Informed General Population Health State Utility Values: A Review of the Literature." *Value in Health* 14(8):1153–57. doi: 10.1016/j.jval.2011.05.046.
- McTaggart-Cowan, H., M. ,., A. Tsuchiya, A. O’Cathain, and J. Brazier. 2011. "Understanding the Effect of Disease Adaptation Information on General Population Values for Hypothetical Health States." *Social Science & Medicine* 72(11):1904–12. doi: 10.1016/j.socscimed.2011.03.036.
- Menzel, Paul, Paul Dolan, Jeff Richardson, and Jan Abel Olsen. 2002. "The Role of Adaptation to Disability and Disease in Health State Valuation: A Preliminary Normative Analysis." *Social Science & Medicine* 55(12):2149–58.

- Milkman, KL, J. Beshears, JJ Choi, D. Laibson, and BC Madrian. 2011. "Using Implementation Intentions Prompts to Enhance Influenza Vaccination Rates." *Proceedings of the National Academy of Sciences of the United States of America* 108(26):10415–20. doi: 10.1073/pnas.1103170108.
- Mukuria, C., and J. Brazier. 2013. "Valuing the EQ-5D and the SF-6D Health States Using Subjective Well-Being: A Secondary Analysis of Patient Data." *Social Science and Medicine* 77(1):97–105.
- Mulhern, B., L. Longworth, J. Brazier, D. Rowen, N. Bansback, N. Devlin, and A. Tsuchiya. 2013. "Binary Choice Health State Valuation and Mode of Administration: Head-to-Head Comparison of Online and CAPI." *VALUE IN HEALTH* 16(1):104–13. doi: 10.1016/j.jval.2012.09.001.
- Munichor, N., and A. Rafaeli. 2007. "Numbers or Apologies? Customer Reactions to Telephone Waiting Time Fillers." *Journal of Applied Psychology* 92(2):511–18. doi: 10.1037/0021-9010.92.2.511.
- Murphy, R. P., C. J. Boyce, P. Dolan, and Alex M. Wood. 2020. "Valuing the Q in QALYs: Does Providing Patients' Ratings Affect Population Values?" *Health Psychology* 39(1):37–45. doi: 10.1037/hea0000806.
- Murphy, R and Taaffe, C. 2020. *Patients' Reasons for Requesting Removal from Inpatient Waiting Lists During Administrative Validation. Research Services and Policy Unit Research Paper*. Dublin, Ireland: Department of Health.
- Murphy, R, Taaffe, C, Delaney, L, Lunn, P, Robertson, D, Ryan, H, and Wood, A. 2020. *The Better Letter Initiative: An Impact Evaluation of a Redesigned Waiting List Validation Letter. Research Services and Policy Unit Research Paper*. Dublin, Ireland: Department of Health.

- Murray, M. 2000. "Patient Care: Access." *BMJ (Clinical Research Ed.)* 320(7249):1594–96. doi: 10.1136/bmj.320.7249.1594.
- Myers, John, Steven McCabe, and Stephan Gohmann. 2007. "Quality-of-Life Assessment When There Is a Loss of Income." *Medical Decision Making* 27(1):27–33. doi: 10.1177/0272989X06297100.
- National Treatment Purchase Fund (NTPF). 2017. *National IDPP Waiting List Management Protocol*. Dublin, Ireland.
- National Treatment Purchase Fund (NTPF). n.d. "National Centralised Validation Unit." *Waiting List Validation*. Retrieved September 26, 2022 (https://www.ntpf.ie/home/Waiting_List_Validation.htm).
- Neumann, P. J., and D. Greenberg. 2009. "Is The United States Ready For QALYs?" *Health Affairs* 28(5):1366–71. doi: 10.1377/hlthaff.28.5.1366.
- NICE. 2013. *Guide to the Methods of Technology Appraisal 2013*. London, England: NICE.
- Norman, R., MT King, D. Clarke, R. Viney, P. Cronin, and D. Street. 2010. "Does Mode of Administration Matter? Comparison of Online and Face-to-Face Administration of a Time Trade-off Task." *Quality of Life Research* 19(4):499–508. doi: 10.1007/s11136-010-9609-5.
- OECD. 2017. "Behavioural Insights and Public Policy."
- OECD. 2020a. "Regulatory Policy and COVID-19: Behavioural Insights for Fast-Paced Decision Making."
- OECD. 2020b. "Waiting Times for Health Services."
- OECD. n.d. "Case Study Library." *Observatory of Public Sector Innovation (OPSI)*. Retrieved (https://oecd-opsi.org/case_type/opsi/).
- Oliver, A. 2019. "Towards a New Political Economy of Behavioral Public Policy." *Public Administration Review* 79(6):917–24. doi: 10.1111/puar.13093.

- Oliver, Adam, ed. 2017. "The Origins of Behavioural Public Policy." Pp. i–ii in *The Origins of Behavioural Public Policy*. Cambridge: Cambridge University Press.
- van Osch, SMC, and AM Stiggelbout. 2005. "Understanding VAS Valuations: Qualitative Data on the Cognitive Process." *Quality of Life Research* 14(10):2171–75. doi: 10.1007/s11136-005-6808-6.
- Peeters, Yvette, and Anne M. Stiggelbout. 2010. "Health State Valuations of Patients and the General Public Analytically Compared: A Meta-Analytical Comparison of Patient and Population Health State Utilities." *Value in Health* 13(2):306–9. doi: 10.1111/j.1524-4733.2009.00610.x.
- van der Pol, Marjon, and Alan Shiell. 2007. "Extrinsic Goals and Time Tradeoff." *Medical Decision Making* 27(4):406–13. doi: 10.1177/0272989X07302127.
- Revicki, D. A., A. Shakespeare, and P. Kind. 1996. "Preferences for Schizophrenia-Related Health States: A Comparison of Patients, Caregivers and Psychiatrists." *International Clinical Psychopharmacology* 11(2):101–8. doi: 10.1097/00004850-199611020-00004.
- Riis, J., G. Loewenstein, J. Baron, and C. Jepson. 2005. "Ignorance of Hedonic Adaptation to Hemodialysis: A Study Using Ecological Momentary Assessment." *Journal of Experimental Psychology - General* 134(1):3–9. doi: 10.1037/0096-3445.134.1.3.
- Rommel, J., V. Buttmann, G. Liebig, S. Schonwetter, and V. Svart-Groger. 2015. "Motivation Crowding Theory and Pro-Environmental Behavior: Experimental Evidence." *Economics Letters* 129:42–44. doi: 10.1016/j.econlet.2015.01.025.
- Rosser, R., and P. Kind. 1978. "Scale of Valuations of States of Illness - Is There a Social Consensus?" *International Journal of Epidemiology* 7(4):347–58. doi: 10.1093/ije/7.4.347.

- Rubin, D. B. 1987. "Frontmatter." Pp. i–xxix in *Multiple imputation for nonresponse in surveys*. John Wiley.
- Ryff, C.D. 1989. "Happiness Is Everthing, Or Is It - Explorations On the Meaning Of Psychological Well-Being." *Journal of Personality and Social Psychology* 57(6):1069–81. doi: 10.1037/0022-3514.57.6.1069.
- Schimmelpfennig, Robin, and Michael Muthukrishna. 2023. "Cultural Evolutionary Behavioural Science in Public Policy." *Behavioural Public Policy* 1–31. doi: 10.1017/bpp.2022.40.
- Schkade, DA, and D. Kahneman. 1998. "Does Living in California Make People Happy? A Focusing Illusion in Judgments of Life Satisfaction." *Psychological Science* 9(5):340–46. doi: 10.1111/1467-9280.00066.
- Schmidt, Ruth, and Katelyn Stenger. 2021. "Behavioral Brittleness: The Case for Strategic Behavioral Public Policy." *Behavioural Public Policy* 1–26. doi: 10.1017/bpp.2021.16.
- Shaw, J. W., J. A. Johnson, and S. J. Coons. 2005. "US Valuation of the EQ-5D Health States - Development and Testing of the D1 Valuation Model." *Medical Care* 43(3):203–20. doi: 10.1097/00005650-200503000-00003.
- Shiell, A., P. Hawe, and M. Fletcher. 2003. "Reliability of Health Utility Measures and a Test of Values Clarification." *Social Science & Medicine* 56(7):1531–41. doi: 10.1016/S0277-9536(02)00154-5.
- Shin, D.C., and D. M. Johnson. 1978. "Avowed Happiness as an Overall Assessment of Quality of Life." *Social Indicators Research* 5(4):475–92. doi: 10.1007/BF00352944.
- Simon, HA. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69(1):99–118. doi: 10.2307/1884852.

- Smith, D. M., R. L. Sherriff, L. Damschroder, G. Loewenstein, and P. A. Ubel. 2006. "Misremembering Colostomies? Former Patients Give Lower Utility Ratings than Do Current Patients." *Health Psychology* 25(6):688–95.
- Sorensen, K., JM Pelikan, F. Rothlin, K. Ganahl, Z. Slonska, G. Doyle, J. Fullam, B. Kondilis, D. Agraftiotis, E. Ueters, M. Falcon, M. Mensing, K. Tchamov, S. van Den Broucke, H. Brand, and HLS-EU Consortium. 2015. "Health Literacy in Europe: Comparative Results of the European Health Literacy Survey (HLS-EU)." *European Journal of Public Health* 25(6):1053–58. doi: 10.1093/eurpub/ckv043.
- Sousa Lourenco J, Ciriolo E, Rafael Rodrigues Vieira De Almeida S, and Troussard X. 2016. *Behavioural Insights Applied to Policy - European Report 2016*. JRC100146. Luxembourg (Luxembourg): Publications Office of the European Union.
- Stein, Ken, Julie Ratcliffe, Ali Round, Ruairidh Milne, and John E. Brazier. 2006. "Impact of Discussion on Preferences Elicited in a Group Setting." *Health and Quality of Life Outcomes* 4(Journal Article):22–22. doi: 10.1186/1477-7525-4-22.
- Stiggelbout, A. M., and E. De Vogel-Voogt. 2008. "Health State Utilities: A Framework for Studying the Gap between the Imagined and the Real." *Value in Health* 11(1):76–87.
- Stolk, Elly, Kristina Ludwig, Kim Rand, Ben van Hout, and Juan Manuel Ramos-Goni. 2019. "Overview, Update, and Lessons Learned From the International EQ-5D-5L Valuation Work: Version 2 of the EQ-5D-5L Valuation Protocol." *Value in Health* 22(1):23–30. doi: 10.1016/j.jval.2018.05.010.
- Straßheim, Holger. 2020. "The Rise and Spread of Behavioral Public Policy: An Opportunity for Critical Research and Self-Reflection." *International Review of Public Policy* 2(2: 1):115–28.
- Sunstein, CR. 2014. "Nudging: A Very Short Guide." *Journal of Consumer Policy* 37(4):583–88. doi: 10.1007/s10603-014-9273-1.

- Thaler, R. H. and Sunstein, C. R. 2009. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin Publishing.
- Thaler, Richard. 1985. "Mental Accounting and Consumer Choice." *Marketing Science* 4(3):199–214.
- Tversky, A., and D. Kahneman. 1974. "Judgment Under Uncertainty - Heuristics and Biases." *Science* 185(4157):1124–31. doi: 10.1126/science.185.4157.1124.
- Tversky, A., and D. Kahneman. 1992. "Advances in Prospect-Theory - Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5(4):297–323. doi: 10.1007/BF00122574.
- Ubel, P. A., G. Loewenstein, J. Hershey, J. Baron, T. Mohr, D. A. Asch, and C. Jepson. 2001. "Do Nonpatients Underestimate the Quality of Life Associated with Chronic Health Conditions Because of a Focusing Illusion?" *Medical Decision Making* 21(3):190–99.
- Ubel, P. A., G. Loewenstein, and C. Jepson. 2003. "Whose Quality of Life? A Commentary Exploring Discrepancies between Health State Evaluations of Patients and the General Public." *Quality of Life Research* 12(6):599–607. doi: 10.1023/A:1025119931010.
- Ubel, P. A., G. Loewenstein, and C. Jepson. 2005. "Disability and Sunshine: Can Hedonic Predictions Be Improved by Drawing Attention to Focusing Illusions or Emotional Adaptation?" *Journal of Experimental Psychology: Applied* 11(2):111–23.
- Versteegh, MM, and WBF Brouwer. 2016. "Patient and General Public Preferences for Health States: A Call to Reconsider Current Guidelines." *Social Science & Medicine* 165:66–74. doi: 10.1016/j.socscimed.2016.07.043.
- Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton University Press.

- Wang, Y., and FA Sloan. 2018. "Present Bias and Health." *JOURNAL OF RISK AND UNCERTAINTY* 57(2):177–98. doi: 10.1007/s11166-018-9289-z.
- Weinstein, Milton C., George Torrance, and Alistair McGuire. 2009. "QALYs: The Basics." *Value in Health* 12(Journal Article):S5–9. doi: 10.1111/j.1524-4733.2009.00515.x.
- White, Ian R., Patrick Royston, and Angela M. Wood. 2011. "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice." *Statistics in Medicine* 30(4):377–99. doi: 10.1002/sim.4067.
- WHO Director-General. 2022. *Behavioural Sciences for Better Health Initiative: Report by the Director General*. World Health Organization.
- Williams, A. 1990. "Euroqol - a New Facility for the Measurement of Health-Related Quality-Of-Life." *Health Policy* 16(3):199–208.
- Williams, Richard. 2006. "Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables." *Stata Journal* 6(1):58–82. doi: 10.1177/1536867X0600600104.
- Wilson, TD, and DT Gilbert. 2003. "Affective Forecasting." Pp. 345–411 in *Advances in Experimental Social Psychology*. Vol. 35, edited by M. Zanna.
- Wilson, TD, T. Wheatley, JM Meyers, DT Gilbert, and D. Axsom. 2000. "Focalism: A Source of Durability Bias in Affective Forecasting." *Journal of Personality and Social Psychology* 78(5):821–36. doi: 10.1037/0022-3514.78.5.821.
- Wood, Alex M., and Adam T. Davidson. 2020. "Why the Field of Moral Philosophy Must Guide Any Discussion on Well-Being." *World Psychiatry* 19(1):53–55. doi: 10.1002/wps.20717.
- World Bank. 2015. *World Development Report 2015: Mind, Society, and Behavior*. doi: 10.1596/978-1-4648-0342-0. Washington, DC: World Bank.

- Yokum, D., JC Lauffenburger, R. Ghazinouri, and NK Choudhry. 2018. “Letters Designed with Behavioural Science Increase Influenza Vaccination in Medicare Beneficiaries.” *Nature Human Behaviour* 2(10):743–49. doi: 10.1038/s41562-018-0432-2.
- Zhang, K., and M. Garau. 2020. *International Cost-Effectiveness Thresholds and Modifiers for HTA Decision Making*. Vol. OHE Consulting Report, London: Office of Health Economics. Available at: <https://www.ohe.org/publications/international-cost-effectiveness-thresholds-and-modifiers-hta-decision-making>. 1st ed. London: Office of Health Economics.

Supplementary Files

Supplementary Files for Chapter 2

Section 1: *Description of Multiple Imputation Using Chained Equations (MICE) Applied*

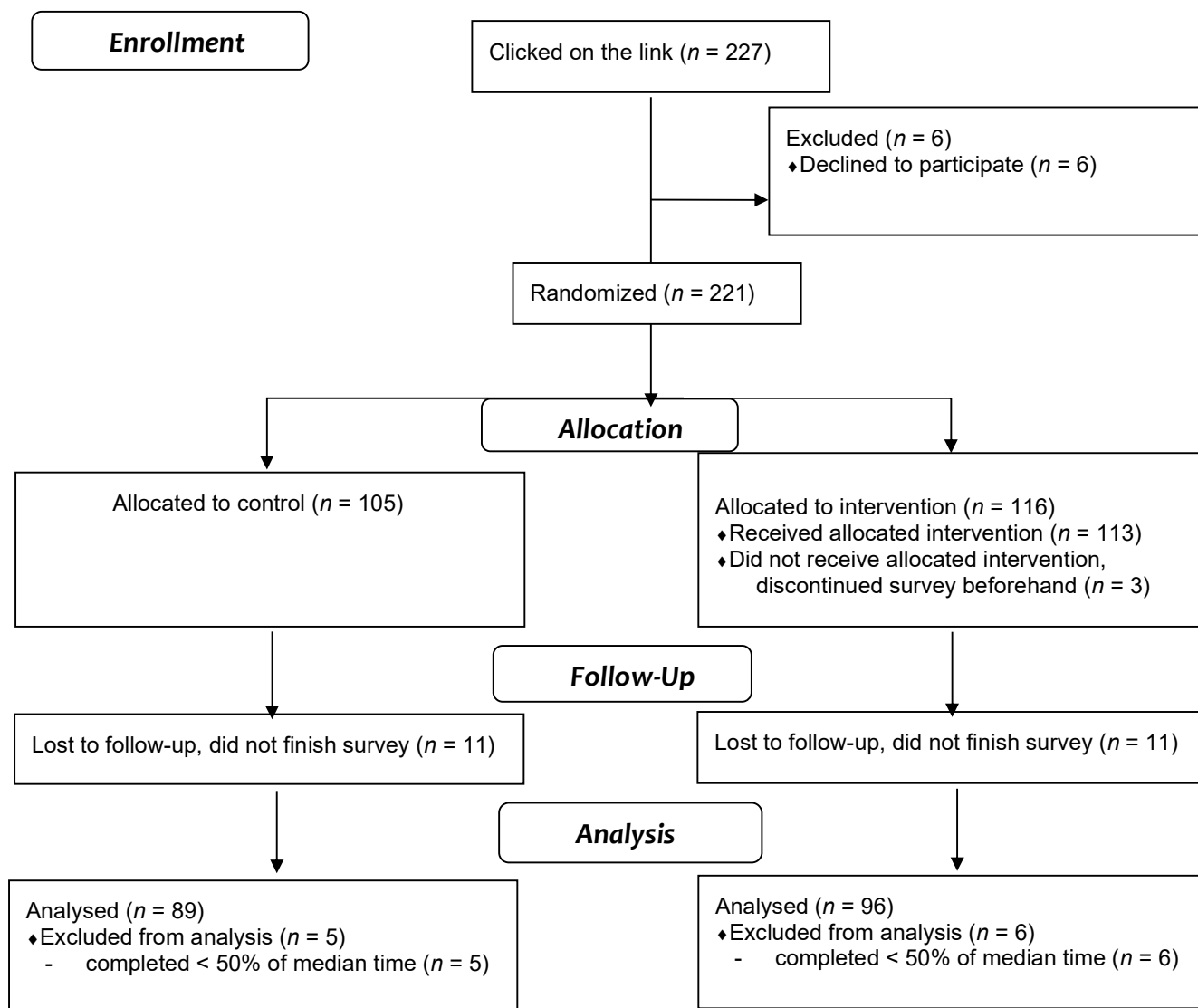
To check the study results were not dependent on the exclusion criterion (b), exclude participants with logically incoherent judgments determined as valuing one or more ill-health scenarios at the anchor 100 (*best health you can imagine*), rather than excluding these participants their VAS valuations were treated as missing data and their VAS valuations were imputed. For these participants the VAS valuation was estimated through a multiple imputation technique Rubin, D. B. (1987). Frontmatter. *Multiple imputation for nonresponse in surveys* (pp. i-xxix) John Wiley.

doi:<http://onlinelibrary.wiley.com/doi/10.1002/9780470316696>. which imputed a series of missing values based on estimates from other observed variables and which more appropriately accounts for the statistical uncertainty in the imputations than many other commonly used imputation techniques such as sample wide or item averages.

Specifically, multiple imputation using chained equations MICE White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377-399. doi:10.1002/sim.4067 was used. This is a technique whereby for each of the multiple imputations a series of sequential regressions are carried out in an iterative fashion. A predictive mean-matching approach was used, 20 imputations were obtained (based on 20 sequential iterations using MICE as recommended by StataCorp (2015), and these imputations were pooled to produce the final estimates. The variables used in the chained equations were health condition judged, prior belief of patients' ratings, group, age category, highest educational attainment, relationship status and principal economic status.

Section 2: CONSORT information including participants who provided a logically incoherent valuation (rated a condition at 100, “best health you can imagine”)

Figure S1: CONSORT 2010 Flow Diagram of Participants including participants who provided a logically incoherent valuation (rated a condition at 100, “best health you can imagine”)¹



Note. As a check on the results presented in the manuscript rather than excluding participants with logically incoherent judgments, determined as valuing one ill-health scenario at the anchor 100 (*best health you can imagine*), these participants’ VAS valuations were treated as missing data, their VAS valuations were imputed using MICE (details in the Supplementary Section 1), and the ANOVA analysis was re-run (results in Supplementary Section 3). We did not re-run the analysis including participants who completed the survey is less than 50% of median time (exclusion criterion (a)) due to a lack of confidence in the overall responses for these participants, plus exclusions for this criterion did not differ across the groups and so could not have caused the intervention effect.

Section 3: *Results of Mixed ANOVA including participants who provided a logically incoherent valuation (rated a condition at 100, “best health you can imagine”) and after applying MICE to impute these participants’ VAS valuations*

Table S1

Results of Mixed ANOVA including participants who provided a logically incoherent valuation (rated a condition at 100, “best health you can imagine”) and after applying MICE to impute these participants’ VAS valuations

Predictor	Partial Sum of		Mean		
	Squares	df	Square	F	p
Prior belief of patients’ rating	126947.68	2	63473.84	170.54	0.000
Group	3633.19	1	3633.19	9.76	0.002
Health condition	1756.60	2	878.30	2.36	0.095
Group x prior belief of patients’ rating	7197.48	2	3598.74	9.67	0.000
Group x health condition	1756.61	2	878.30	2.36	0.095
Group x health condition x prior belief of patients’ rating	4405.35	8	550.67	1.48	0.159
Residual	836298.45	2247	372.18		
Total	1032933.40	2264	456.24		

Note. To check the study results were not dependent on the exclusion criterion (b), exclude participants with logically incoherent judgments determined as valuing one or more ill-health scenarios at the anchor 100 (*best health you can imagine*), rather than excluding these participants their VAS valuations were treated as missing data and their VAS valuations were imputed using MICE (details in the Supplementary Section 1). We did not re-run the analysis including participants who completed the survey in less than 50% of median time (exclusion criterion (a)) due to a lack of confidence in the overall responses for these participants, plus exclusions for this criterion did not differ across the groups and so could not have caused the intervention effect.

Section 4: *Detailed definitions of health conditions***Arthritis**

You have some difficulty bending down to pick up clothes from the floor.

You have much difficulty climbing up 5 steps.

You have some difficulty lifting a full cup or glass to your mouth.

You have much difficulty standing up from a straight and armless chair.

You have moderate pain and discomfort.

Colostomy

Waste passes through your intestines and out the stoma into a bag, which you must empty several times a day.

If you wear relatively loose clothing, this bag will not be visible underneath your garments.

Occasionally, you will experience odours and noises caused by gas and waste passing through the stoma.

There is a chance that the colostomy bag may leak if it is allowed to fill past capacity.

Although you will be restricted from lifting very heavy weight, your daily activities will not otherwise be greatly affected.

Tinnitus

You experience a sensation of hearing sound when no external sound is present.

These types of phantom noises in your ears include ringing, buzzing, roaring, clicking, hissing.

The phantom noise may vary in pitch from a low roar to a high squeal, and you may hear it in one or both ears.

In some cases, the sound can be so loud it can interfere with your ability to concentrate or hear actual sound.

The sensation of hearing these sounds may be present all the time, or it may come and go.

Supplementary Files for Chapter 3

Section 1: *Description of Questionnaire*

Questionnaire

The questionnaire included seven sections as summarized below. Participants in control and intervention groups saw all sections of the questionnaire. For the control group the order of sections was as below, whereas for the intervention groups Section 6 appeared before Section 4.

1. Welcome. In this section the reasons for undertaking the study were explained. It was made clear that participation was voluntary and that participants could withdraw at any stage, a description of what participating in the survey would involve was provided, it was guaranteed that information provided would be kept confidential, survey funders were stated, and participants were requested to complete an informed consent form.

2. Your health: Consistent with standard practice in health valuation studies, in this section participants completed a warm-up exercise by describing their current health using the two parts of the EQ-5D-3L EuroQol Research Foundation. (2015). *EQ-5D-3L user guide*. (No. Version 5.1). Rotterdam: EuroQol Group. Retrieved from <https://euroqol.org/publications/user-guides/>. First, they described their general health using the EQ-5D-3L's generic health state descriptive system by selecting no problems, some problems, or severe problems for each of the five health dimensions of self-care, ability to undertake usual activities, pain and discomfort, and anxiety and depression. Second, they rated their own health using a visual analogue scale, the EQ VAS, which is a vertical rating scale with the anchors 0 (worst health you can imagine) and 100 (best health you can imagine).

3. Description of health states. A functional description of each health state was provided in this section. Consistent with standard practice in health valuation studies, the description of each health state was provided in a box with a line for each element of a health state. The health state descriptions used in this study were based on three health state descriptions from the EQ-5D-3L, which is one of the most widely used instruments in health valuation studies EuroQol Research Foundation. (2015). *EQ-5D-3L user guide*. (No. Version 5.1). Rotterdam: EuroQol Group. Retrieved from <https://euroqol.org/publications/user-guides/>. The health state descriptions were:

(1) Moderate anxiety / depression: No problems with walking about, no problems with self-care, no problems with usual activities, no pain or discomfort, moderately anxious or depressed.

(2) Moderate pain / discomfort: No problems with walking about, no problems with self-care, no problems with usual activities, moderate pain / discomfort, not anxious or depressed.

(3) Some problems with self-care: No problems walking about, some problems with self-care, no problems with usual activities, no pain or discomfort, not anxious or depressed.

The descriptions of “moderate pain / discomfort” and of “moderate anxiety / depression” correspond exactly to the wording in the EQ-5D-3L descriptive system (i.e., “moderate pain / discomfort” and “moderately anxious or depressed”). However, the wording for “some problems with self-care” does not, as for consistency across health states the description used was “some problems with self-care” whereas the EQ-5D wording is “some problems washing or dressing”. The same descriptions were provided to all groups.

4. Preferences over health states. Participants appraised and expressed their preferences over three health states (“moderate anxiety / depression”, “moderate pain / discomfort”, and “some problems with self-care”) in this section. Following standard practice in health valuation surveys Gudex, C. (Ed.). (1994). *Time trade-off user manual: The props and self-completion method*. [Adobe Digital Editions version]. Retrieved from <https://www.york.ac.uk/che/pdf/op20.pdf>, participants were instructed to: “imagine that you will live in the health states described for up to 10 years and then die” and then asked to provide their preferences over alternative health states.

Discrete choice experiments (DCEs) are typically used to obtain general population preferences over hypothetical health states Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., & Tsuchiya, A. (2013). Binary choice health state valuation and mode of administration: Head-to-head comparison of online and CAPI. *Value in Health*, 16(1), 104-113. and this study also elicits preferences using a DCE. Participants were told: “We would like you to make a choice between health states, from most to least preferred to live in.” and instructed to: “Choose between these three health states, from 1 (most preferred) to 3 (least preferred). Make your choice by ticking 1 for your most preferred, 2 for second preference and 3 for your third preference.” The order in which the health states were listed was randomized.

5. Clarity of tasks. Participants were asked in this section about task difficulty (how easy or difficult they experienced the choice task and the survey to be overall).

6. Presentation of association of health states with well-being. In this section participants (1) viewed information on the association of each health state with differences in life satisfaction or day affect and (2) were asked if the association of each health state with

differences in life satisfaction or day affect was bigger, about the same, or smaller than expected. Participants in the intervention groups saw this information either for life satisfaction (first intervention condition) or for day affect (second intervention condition). The information was provided in text (see Section 2) and in a graphic that presented the association of health states with differences in life satisfaction or day affect alongside a vertical scale with the anchors 0% and -50%. Data on the association of health states with patients' life satisfaction and day affect were taken from Dolan and Metcalfe (2012). The order in which sections 4 and 6 appeared were different for control and treatment groups. For the intervention groups Section 6 appeared directly before Section 4 and information on the association of health states with differences in life satisfaction or day affect remained available to the intervention groups when appraising a health state.

7. Personal characteristics. Detailed background characteristics were collected last as is typical in health valuation studies Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol group: Past, present and future. *Applied Health Economics and Health Policy*, 15(2), 127-137. doi:10.1007/s40258-017-0310-5. This section contained questions on socio-economic characteristics (gender, age, marital status, highest educational attainment, and main activity) and on personality. The personality information was collected for another study and was not analyzed for the current research.

Section 2: *Intervention Text*

Text Provided on Life Satisfaction Information

Before you make your choice we would like to let you know how the last three health states are associated with differences in people's life satisfaction. Please read the information below and then we will ask you a question.

Life satisfaction is a measure of how people evaluate their life as a whole or put another way it is a reflective assessment of people's satisfaction with their own life.

The average difference in **life satisfaction** associated with each health state was:

- **0%** (i.e. no change) for people with **some problems with self-care**.
- **2%** lower for people with **moderate pain/discomfort**.
- **13%** lower for people with **moderate anxiety/depression**.

The above are the differences in life satisfaction associated with being in each health state compared to not being in it. That is:

- some problems with self-care is not associated with lower life satisfaction compared to no problems with self-care.
- moderate pain/discomfort is associated with 2% lower life satisfaction compared to no pain or discomfort.
- moderate anxiety/depression is associated with 13% lower life satisfaction compared to no anxiety/depression.

These are the differences specific to each health state after taking other factors into account that could influence a person's life satisfaction such as age, gender, income, employment status and relationship status.

[The differences in life satisfaction associated with the three health states was presented in a graphic, a vertical scale with the anchors 0% and -50%, that appeared directly after the above text.]

Text Provided on Day Affect Information

Information on day affect was presented in the same format as the life satisfaction information above. Day affect was referred to as a positive feeling score. It was described as "A positive feelings score is the difference between the strength of the positive feelings (were friendly, happy, and calm) and negative feelings (were sad, lethargic, stressed, angry, tired, depressed, or worried) a person experiences in a day. A higher score reflects stronger experience of positive feelings and a lower score reflects stronger experience of negative feelings." The respective associations with day affect were 0% for some problems with self-care, 5% lower for moderate pain/discomfort and 20% lower for moderate anxiety/depression.

Section 3: *Model Assumptions and Robustness Checks*

The study's preference measure is an ordered outcome variable (first, second, or third preference). The model tested the relative likelihood of preferences for anxiety/depression falling into less preferred categories (because a preference of "1" is equivalent to "most preferred"). The coefficient estimates were obtained using a logit ordered model. When using a logit ordered model it is assumed that the coefficients describe the same relationship for all levels of the dependent variable — this is the proportional odds assumption Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, 6(1), 58-82. doi:10.1177/1536867X0600600104. The evidence did not suggest that this assumption of the ordered logistic model was violated: The results of a Brant test were non-significant for provision of LS information $\chi^2(10, N = 814) = 5.09, p = 0.885$ and for provision of DA information, $\chi^2(10, N = 786) = 14.34, p = 0.158$.

To examine the robustness of the test of hypothesis 2 we varied the parameters in the model: Equation 1 was estimated with health status (self-rated health on the EQ VAS scale) rather than the LS variable, without either of these variables, and with LS as a continuous rather than a categorical variable. To further test the robustness of our findings we ran a probit ordered model and we also ran an unconstrained model which does not make a proportional odds assumption Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, 6(1), 58-82. doi:10.1177/1536867X0600600104. None of the above alterations changed the study conclusions.

Supplementary Files for Chapter 4

Section 1: *Specialties in the sample by hospital group*

<u>RSCI Hospital Group</u>	<u>UL Hospital Group</u>
	Cardiology
	Dermatology
	Gastro-Enterology
General Surgery	General Surgery
Gynaecology	Gynaecology
	Maxillo-Facial
	Ophthalmology
	Orthopaedics
	Otolaryngology (ENT)
	Pain Relief
	Respiratory Medicine
	Rheumatology
Urology	Urology
	Vascular Surgery

Section 2: Control letter in UL Hospital Group (appeared on HSE headed paper)

Ospidéal na hOllscoile, Luimneach,
 Bóthar Naomh Neasáin, Tuar an Daill, Luimneach V94 F858
 Teil: 061 301111 Facs: 061 301165

University Hospital Limerick,
 St. Nessian's Road, Dooradoyle, Limerick V94 F858
 Tel: 061 301111 Fax: 061 301165

1st November 2017

<Title> <Patient First Name> <Patient Surname>

<Address 1>

<Address 2>

<Address 3>

<Address 4>

Patient Record Number: <>

Dear <Title> <Patient Surname>,

You are listed for a procedure on <Consultant Name> <Specialty> Waiting List. We want to ensure our waiting list is accurate and up to date. Accordingly, we would be grateful if you could complete this form and return it to us at your earliest convenience.

If we do not hear from you by **November 15th 2017**, we will assume that you no longer require the procedure and your name will be removed from the waiting list. Your Consultant and GP will be informed.

Yours sincerely,

Validation Officer, UHL**PLEASE UPDATE YOUR CONTACT DETAILS FOR OUR RECORDS**

Name:	
Address:	
DOB:	
Contact Telephone Number:	

Please tick the appropriate box:

Yes , I wish to remain on the waiting list		
No , I do not wish to remain on the waiting list		Please indicate the reason:

Signed: _____**Date:** _____

Section 3: Control letter in RCSI Hospital Group (appeared on HSE headed paper)

<Title> <Pt Forename> <Pt Surname>

<Pt. Address Line 1>

<Pt. Address Line 2>

<Pt. Address Line 3>

<Pt. Address Line 4>

<Current Date>

Ref: <MRN>**Re:<Consultant Last Name><Waiting List Name> Validation**

Dear <Title> <Pt Surname>

You have now been on the <Insert Specialty> waiting list for <Insert Wait Time Frame>. We regret that it has not been possible to offer you a date for your procedure but wish to assure you that your name is still on the list and hasn't been overlooked.

We find sometimes that patients obtain treatment elsewhere or no longer require their procedure. We periodically contact patients on our waiting lists to check whether this is the case.

We would be grateful if you could complete and return the attached validation slip to let us know if you still require a procedure.

If we do not hear from you within two weeks of the date of this letter, we will assume that you no longer wish to have this procedure, we will remove your name from the waiting list and inform your GP.

Yours sincerely,

<Insert Department/ Section Name>

<Insert Department/ Section Telephone>

Please tick:**Ref: <MRN>**☐ I still require this procedure☐ I no longer require this procedure. Please indicate the reason:

Signed: _____

Date: _____

Section 4: Intervention letter (appeared on HSE headed paper)

Mr Joe Bloggs

Strictly Private and Confidential

Hawkins House

Hawkins St

Dublin 2

2nd November 2017**Please reply to this letter**

Dear Joe

You are on our General Surgery waiting list for a procedure with John Smith. I apologise that you are still waiting. We want to provide our valuable services to our patients as soon as we can. That is why we are checking our waiting list.

	Read this letter
	Fill in the form
	Return this form

We need you to please:

1. Answer the question below and sign.
2. Return this page to us in the freepost envelope enclosed.

Please do this **even if** you have recently been in contact with the hospital.

If you **don't send** us back this page by **16th November 2017**, then we will take it that you do not require this procedure and **you will be removed** from our waiting list. Your GP (family doctor) will be informed.

Question: Do you still require this procedure? (tick one box only)

☐ Yes, I still require it ☐ No, I had it done elsewhere ☐ No, other reason

If "No, other reason" please give reason:

Please sign: _____ Medical Record No. 12345

If you have any questions about the above, please phone 01 635 3122.

Kind regards,

Jane Doe, Scheduled Care Department