# Exploration of the use of decision analytic modelling in low back pain and sciatica

BY JAMES ASHWELL HALL

A THESIS SUBMITTED TO

KEELE UNIVERSITY

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Research Institute for Primary Care and Health Sciences

David Weatherall Building,

Keele University,

Newcastle-under-Lyme

June 2020

**ABSTRACT**

Low back pain (LBP) is a global public health problem. Keele University developed the STarT Back tool to stratify LBP patients according to their risk of persistent disability, matching treatments to individual risk. A 12-month trial-based economic evaluation showed this stratified care model to be cost-effective. A recent trial, the SCOPiC trial, aimed to evaluate a modified stratified care model for sciatica patients consulting in primary care. However, the longer-term cost effectiveness of both care models is unknown.

To estimate the long-term cost-effectiveness of stratified care, two separate decision models were developed. The model conceptualisation process included expert consultations, and two systematic literature reviews assessing the use of decision analytic modelling in LBP and sciatica, and stratified care.

A de-novo state-transition cohort model was developed to estimate the cost-utility of stratified care for the management of LBP in primary care, from the NHS perspective, over a ten-year horizon. Model results provided support for the cost-effectiveness of the Keele stratified care model.

A de-novo individual-level simulation model was chosen to estimate the cost-utility of stratified care vs best usual care vs usual care for the management of those consulting with sciatica in primary care, from the NHS perspective, over a ten-year horizon. Model results suggest this model of stratified care is not cost-effectiveness relative to best usual care.

Both cost-effectiveness results were robust to structural assumptions, however, sensitivity analyses highlighted how assumptions regarding health states, long-term patient prognosis and EQ-5D values could affect cost-effectiveness results. Furthermore, the first Expected Value of

Perfect Parameter Information (EVPPI) analyses in decision modelling for LBP and sciatica highlight the value of further research exploring transitons between health states.

The thesis concludes with recommendations for modelling in low back pain and sciatica, including the need to strengthen modelling methodologies and fully explore structural and parameter uncertainty.

**ACKNOWLEDGMENTS**

This thesis would not have been possible without the guidance and support of my amazing supervisors, Sue Jowett, Kika Konstantinou, Martyn Lewis, and Raymond Oppong, and thanks also to Reuben Ogollah for his support during the first year of my work. Their psychological support and kindness will never be forgotten. I am especially grateful towards Sue, for her stewardship of the thesis and associated publications and presentations, and Kika for her clinical input, and for teaching me how to write and think in a manner acceptable towards clinicians, which must have required patience on her behalf! I have very much enjoyed the meetings and the time we've spent together, which has always been enjoyable.

There are many others who have contributed their professional expertise in the thesis, namely Pelham Barton who has advised on aspects of the modelling in the latter chapters and Kate Dunn for her feedback throughout the thesis. Thanks to the members of the Keele Stratified Care Research Group for their feedback on the initial drafts of models, as well as Katrina Humphreys, John Bedson, Elizabeth Cottrell, Adrian Chudyk, and Rohanna Burgess, for their clinical input into the modelling stages.

I am also grateful to my mother, sister, and best friend Natalija, and colleagues Ilias Goranitis and Alessandra Guariglia, all of whom have played their part in encouraging me forward in my academic career. Thanks also to the Health Economics Unit at University of Birmingham for making me feel welcome there, and for the technical skills I've developed since I joined the MSc programme in 2016. I'm grateful to have been part of such a good cohort of PhD students at Keele, and thank Hollie Birkinshaw especially for the chats, and administerial assistance.

## PUBLICATIONS RELATING TO THIS THESIS

The results presented in Chapter 4 have been previously given in the following peer reviewed paper, Hall J, Konstantinou K, Lewis M, Oppong R, Ogollah R, Jowett S (2019) Systematic review of decision analytic modelling in economic evaluations of low back pain and sciatica Applied Health Economics and Health Policy, doi: 10.1007/s40258-019-00471-w.

**Contents**

## LIST OF TABLES

**LIST OF FIGURES**

## LIST OF ABBREVIATIONS

| | |
|---|---|
| CBA | Cost-benefit analysis |
| CEA | Cost-effectiveness analysis |
| CUA | Cost-utility analysis |
| DAM | Decision analytic modelling |
| EQ-5D | EuroQoL five dimension |
| GP | General Practioner |
| GPC | Global perceived change |
| HCM | Human capital method |
| HRQoL | Health-related quality of life |
| ICER | Incremental cost-effectiveness ratio |
| IMPaCT Back | Implementation study to improve Patient care through Targeted treatment for back pain |
| ISM | Individual simulation model |
| LBP | Low Back Pain |
| NHS | National Health Service (UK) |
| NICE | National Institute for Health and Care Excellence |
| NRS | Numerical rating scale |
| NSAIDS | Non-steroidal anti-inflammatory drug |
| NSLBP | Non-specific low back pain |
| ONS | Office for National Statistics |
| PSA | Probabilistic sensitivity analysis |
| QALY | Quality-adjusted life year |
| RCT | Randomised controlled trial |
| RMDQ | Roland Morris Disability Questionnaire |

| | |
|---|---|
| SCOPiC | SCiatica Outcomes in Primary Care |
| STarT Back | Subgrouping for Targeted Treatment |
| VoI | Value of Information Analysis |
| WTP | Willingness to pay |

# Chapter 1: INTRODUCTION

## 1.1 Background

Low back pain (LBP) is a major global public health problem (Hoy et al., 2014). The often chronic nature of the condition places a significant burden upon individuals owing to disability and reduced wellbeing. The condition also carries significant societal costs resulting from high workplace absence and healthcare usage (Hoy et al., 2014).

The majority of LBP patients are managed in primary care, with treatment options agreed between patient and clinician. Keele University developed the STarT Back tool (Hill et al., 2011) to stratify LBP patients according to their risk of persistent disability, matching treatments to individual risk. A short-term clinical trial with 12 months follow-up demonstrated this stratified care model to be clinically effective compared with usual non-stratified care (Hill et al., 2011).

However, as available resources are limited, decision makers also must decide whether or not an intervention represents good value for money. An economic evaluation is the process by which the cost-effectiveness of an intervention is measured (Goodacre and McCabe, 2002). An economic evaluation parallel to the STarT Back trial, showed the Keele stratified care model also to be cost-effective over 12 months, leading to higher quality-adjusted life years (QALYs), lower NHS costs, and reduced time off work (Whitehurst et al., 2012; Whitehurst, et al., 2015). Whilst the STarT Back care model for LBP has been implemented by many commissioning bodies in England and endorsed in clinical guidelines (NICE, 2016), the longer-term cost-effectiveness of this approach remains unknown.

A stratified care model was also tested for sciatica patients in the SCOPiC (Sciatica Outcomes in Primary Care) trial (Foster et al., 2017). The results of the SCOPiC trial are

currently unpublished, nonetheless the trial revealed that stratified care was neither clinically nor cost-effective compared to usual non-stratified care (paper submitted for publication, under review) (K Konstantinou, 2019, personal communication).

Where a condition is considered chronic, or treatment impact long-term, decision makers generally prefer long-term cost-effectiveness evidence. LBP and sciatica are such conditions, with a natural history characterised by symptom recurrence. As a general rule, compared with a within-trial economic evaluation, decision analytic modelling (DAM) is considered a more appropriate tool for conducting an economic evaluation (Sculpher et al., 2006). The economist's preference for DAM stems predominantly from their ability to extrapolate outcomes over the long-term, synthesise available evidence and include all available comparator technologies and treatments in an analysis.

At present, there are no published decision analytic models which evaluate the cost-effectiveness of a stratified care approach for LBP and sciatica. This thesis is, therefore, concerned with exploring the use of decision analytic modelling in the clinical areas of low back pain and sciatica, as well as stratified care treatment approaches. The thesis then brings together these insights to perform two cost-effectiveness analyses of stratified care, one for the management of LBP, and one for the management of sciatica. As there was no superiority for stratified care in the SCOPiC trial, stratified care and non-stratified usual care in the trial were compared with usual care outcomes and costs found in a usual care cohort study.

## 1.2 Structure of this Chapter

In what follows, sub-section 1.3 provides a brief overview of the context as well as the rationale for the use of economic evaluations and cost-effectiveness analyses in healthcare decision making. Section 1.4 reflects upon the role and value of decision analytic modelling in economic evaluation. The purpose of both sub-sections is to briefly explore

the contextual justifications for performing a model-based cost-effectiveness analysis as subsequently used in the thesis.

Sub-sections 1.5 and 1.6 provide a brief review of the clinical areas of LBP and sciatica and discuss the challenges for treating both conditions in primary care. The Keele University stratified care approaches for treating LBP and sciatica in primary care are then discussed, including existing economic evidence for these approaches.

Sub-section 1.7 outlines the justifications for the thesis, 1.8 summarises aims and objectives, whilst 1.9 provides a summary to the chapter.

## 1.3 Economic Evaluation in Healthcare

In recent years, there has been a surge in demand placed upon health services, driven for the most part by an ageing population and increased prevalence of chronic diseases (Charlesworth and Johnson, 2018). Meanwhile, after the financial crisis of 2008, the political economy of most developed countries has been characterised by changes in attitudes towards government spending and the debt burden, the consequence of which has been a considerable slowing in real terms health service funding (Appleby, 2016). Facing increased demand and restricted resources, decision makers now more than ever face tough decisions about how to ration scarce resources.

Given the moral significance of allocating scarce resources, 'people, time, facilities equipment or knowledge' (Drummond et al., 2015), decision-makers often express a preference for explicit evidence-based processes for making resource allocation decisions (Donaldson and Mitton, 2009). Choices informed by a systematic consideration of the costs and consequences will nearly always produce superior outcomes to those based upon 'gut feelings', 'what we did last time', or 'educated guesses' (Drummond et al., 2015). Accordingly, agencies such as the National Institute for Health and Care Excellence (NICE), which issue guidance concerning whether various health technologies should be

available through the public healthcare system, explicitly demand systematic and evidence-based analysis of both benefits and costs of new treatments (NICE, 2013).

The demand from decision makers for such guidance has led to increased prominence for economic evaluation in healthcare (Gray et al., 2011; Kluge et al., 2007). In the healthcare context, an economic evaluation can be understood as a systematic comparison of the costs and benefits of an intervention(s), with the aim of providing explicit accountability in decision making (Al-Janabi et al., 2012). The fundamental idea lying at the heart of economic evaluation is that of opportunity cost, the cost, in terms of benefits forgone, of choosing one course of action over alternate uses of those resources (Drummond et al., 2015). The practical activity of performing an economic evaluation requires the analyst to identify, measure, value and compare the costs and consequences of alternatives being considered (Drummond et al., 2015). Provided with this information the decision maker should, in theory, be able to maximise the benefits available for a given amount of healthcare resource, the essence of the principle of allocative efficiency (Morris et al., 2012).

The undertaking of an economic evaluation can hold tremendous potential value for decision makers. Nonetheless, it must be acknowledged that what appears an objective endeavour is underpinned by considerable subjectivity. The analyst is required to make normative decisions which can profoundly influence the results of an evaluation (Morris et al., 2012). Normative economics is a strand of economics that deliberates upon value judgments regarding "economic fairness" or what the public policy ought to seek to achieve (Samuelson and Nordhaus, 2004). In health economics normativity is endemic because judgments will always be required regarding what is "fair" and what constitutes "costs" and "benefits" (Morris et al., 2012). Certainly, normative decisions will lie at the heart of this thesis.

The concepts briefly discussed above, as well as other important considerations will be explored in more detail in sub-sections of chapter three, which explores in full the rationale for and appropriate conduct of economic evaluations. This chapter now turns to consider the role of decision analytic modelling in economic evaluation.

## 1.4 Decision analytic modelling in Economic Evaluation

Randomised controlled trials (RCTs) are a common method of assessing the efficacy and effectiveness of healthcare interventions (Morris et al., 2012). Their clinical value derives from producing estimates of relative treatment effects within a population of interest, where randomisation minimises the risk of selection bias (Sculpher, 2015). However, an economic evaluation has a somewhat different purpose, aiming to compare not only outcomes but also costs of alternative treatments in order to inform decision-makers about resource allocation decisions (Petrou and Gray, 2011). Indeed, the application of a single RCT as a vehicle for economic analysis has been criticised as providing an inadequate basis for decision-making (Sculpher et al., 2006). A trial-based economic evaluation may not include all long-term costs and outcomes, use all available evidence, nor include the full range of comparator technologies (Morris et al., 2012; Sculpher et al., 2006).

To overcome these problems it is recommended that decision-analytic models should be used in most circumstances (NICE, 2013). Decision-analytic models, considered from the perspective of economic evaluation, can be defined as the application of mathematical relationships to compare expected costs and consequences of decision options over time, by synthesising information from multiple sources (Raiffa, 1968; Barton et al., 2004; Brennan et al., 2006).

A decision model uses mathematical relationships to define the likelihood of potential health consequences occurring for each comparator under evaluation. Each health consequence within the model has costs and benefits attached, meaning that the expected

cost-effectiveness of each option is a joint mathematical function of expected costs and benefits and the probability of each occurring (Briggs et al., 2006). Numerous techniques have been developed to enable decision models to extrapolate short-term trial data to longer-term horizons, include all relevant comparators in the analysis, and include the full range of available evidence (Ramsey et al., 2015; Sculpher et al., 2006; Sculpher and Drummond, 2005).

Decision modelling also facilitates the formal assessment of uncertainty (Briggs et al., 2006). Indeed decision analysis, from which decision modelling originates, has been defined as a systematic approach to decision making under uncertainty (Raiffa, 1968). Providing an estimate of the uncertainty of the results of an evaluation can be important, since the costs of incorrect decisions can be extremely high (Briggs et al., 2006).

Having established the justification for the use of decision-analytic models it must be acknowledged that concerns have been raised that inappropriate model selection is widespread in economic evaluation (Brennan et al., 2006). One systematic review of economic evaluations of screening for chlamydia trachomatis concluded that nearly all the models were methodologically flawed, with significant impact upon the validity of their results (Roberts et al., 2006). However, there are no published guidelines on modelling approaches in LBP or sciatica to assist with the development of a modelling methodology to underpin this thesis, nor are there many models from which to learn.

Even though both LBP and sciatica may often require long-term management, cost-effectiveness studies for both conditions tend to be conducted alongside short term clinical trials. For example, a systematic review of non-invasive and non-pharmacological interventions for LBP (Andronis et al., 2016), found 33 suitable studies met their inclusion criteria however only two (Kim et al., 2010; Norton et al., 2015) of these were decision analytic models. A review of the cost-effectiveness of management strategies for sciatica

(Lewis et al., 2011) found only one decision analytic model (Launois et al., 1994). This thesis will pay significant attention to consideration of what form an appropriate model for the condition and its treatment, would take.

## 1.5 Low back pain

Low back pain can be defined as pain, discomfort, or stiffness in the lower back region, which is commonly considered as the area below the costal margin and above the inferior gluteal folds (NICE, 2009; Weiner and Nordin, 2010). The condition can either be classified as specific or non-specific. Specific LBP is defined as that caused by a specific mechanism such as malignancy, infection, fracture, spondylitis or inflammatory disorder (NICE, 2009; Balague et al., 2012). Non-specific low back pain (NSLBP) meanwhile is defined as pain, tension, stiffness and/or soreness in the lower back region where no specific cause of the pain can be identified given current diagnostic tools (van Tulder and Waddell, 2005; NICE, 2009). Around 85% of patients presenting with LBP are thought to have (NSLBP) (Deyo and Weinstein, 2001). In approximately 60% of patients presenting with NSLBP, there is radiation of pain from the low back to the leg(s) (Hill et al., 2011).

LBP is a substantial international health concern, with a lifetime prevalence of 80–85% (WHO, 2003). In the United Kingdom (UK) around 14% of all primary care consultations are for LBP (Jordan et al., 2010). Whilst many episodes of LBP are short lived and many patients stop seeing their general practitioner (GP) in the first three months, the condition presents significant challenges for clinical management, with only 20-40% of patients reporting no pain or disability a year after first seeing a GP (Croft et al., 1998; Hestbaek et al., 2003). Maniadakis and Gray (2000) estimated the societal impact in terms of periods of work absence related to LBP to be between £7 and £12 billion, with NHS and community care costs in excess of £1 billion (1998 prices).

In making referral decisions regarding patients with NSLBP, clinical intuition is often used, despite the suggestion that this leads to inconsistent access to treatment (Hill et al., 2011; DoH, 2006). On the other hand, simply referring all patients for treatment is considered unnecessary and inefficient owing to the patient volume and cost of doing so (McGrail et al., 2001; Savigny et al., 2009; DoH, 2006). A one-size-fits-all primary care strategy is suboptimal because it ignores the heterogeneity in patient presentation of symptoms (van der Windt et al., 2008).

### 1.5.1 Stratified Care in LBP

Stratified care can be defined as the targeting of treatments according to biological or risk characteristics of subgroups of patients with similar characteristics (Hingorani et al. 2013). Subgrouping can be effective because it reduces variability in treatment, serving to improve treatment benefit, reduce harm and contribute to efficient health-care delivery (Sowden et al. 2018). Indeed research suggests subgrouping generates better outcomes than treatment based solely on clinical guidelines (Fritz et al., 2003).

In relation to LBP a longstanding research aim has been to develop effective means of subgrouping of LBP patients in an effort to improve patient outcomes (Cherkin et al., 2009). There is no exact consensus as to how LBP patients ought to be stratified and Keele University has pioneered one approach, the STarT Back approach (Hay et al., 2008; Hill et al., 2008). The STarT Back approach comprises subgrouping patients according to risk of persistent disability (low, medium, high risk) via a screening tool (the STarT Back tool); patients are then matched to appropriate treatments for their risk. The STarT Back tool was developed for, and validated with, primary care patients with LBP (with and without leg pain) and captures eight key modifiable physical and psychological prognostic indicators for persistent disabling symptoms using nine questions. In this sense, stratification here uses baseline information to make inferences about a patient's likely to response to

treatment to tailor their treatment decisions. Thus stratification can be differentiated from stepped care, where it is the actual response to previously offered treatment which guides future treatment (Hingorani et al. 2013).

The matched interventions delivered as part of the RCT investigating the effectiveness of this approach (Hill et al., 2011) are as follows. Those patients in the low-risk group receive advice, a 15-min educational video entitled 'Get Back Active' and given the Back Book. Medium-risk patients received further standard physiotherapy sessions to address symptoms and function (in addition to advice and booklet). High-risk patients further received psychologically informed physiotherapy interventions to address psychosocial obstacles to recovery (in addition to advice and booklet).

This stratified care approach was further tested in an implementation before-and-after study (Foster et al., 2014). The clinical and cost-effectiveness results of both the trial and implementation study demonstrated this approach was both clinically effective and led to greater QALYs, cost-savings to the NHS and reduction in time off work over 12 months follow up (Hill et al., 2011; Whitehurst et al., 2012; Foster et al., 2014; Whitehurst et al., 2015).

To-date, the approach has been recommended in UK treatment guidelines and has been adopted by a number of clinical commissioning groups in the UK as well as services overseas (NICE, 2016; Keele University, 2015). Despite the progressive implementation of the approach, the current economic evidence for the STarT Back approach for LBP (Whitehurst et al., 2012; Whitehurst et al., 2015) is based upon a single short-term trial finding, with longer-term cost-effectiveness unknown. Short trial-based economic evidence regarding treatments for LBP (and sciatica) studies is unlikely to capture the full extent of the associated costs and benefits (Andronis et al., 2016). As stated previously, where a

condition is predominately chronic, decision makers are likely to prefer evidence on the long-term cost-effectiveness of treatments.

### 1.6 Sciatica

Sciatica is a symptom of radiating pain from the low back to the leg, often extending to the foot (Valat et al., 2010; Navarro-Siguero et al., 2013). Patients may also have other leg symptoms such as pins and needles, numbness or leg muscle weakness (Fairbank, 2007; Valat et al., 2010; Qin et al., 2015). The most common reasons for sciatica are compression or irritation of a lumbar spinal nerve root (s) by a prolapsed or bulging disc or tightening of the spinal or lateral canal (spinal stenosis) (Kobayashi et al., 2005). Sciatica is known by a range of terms in the literature, such as lumbosacral radicular syndrome, radiculopathy, nerve root pain and nerve root entrapment or irritation (Konstantinou and Dunn, 2008). Sciatica is less prevalent compared with NSLBP, with a lifetime reported prevalence of between 12.2% and 43% (Konstantinou and Dunn, 2008). There is a link between LBP and sciatica, in that patients who previously had LBP were between 1.5 and 3 times more likely to develop a first incidence of sciatica (LeClerc et al., 2003; Kääriäl et al., 2011).

The true economic cost of sciatica is unclear. A cost of illness study from the Netherlands in 1991 estimated the impact of sciatica to be US $128m for hospital care, US $730m for absenteeism, and US $708m for disablement (van Tulder et al., 1995). Fitzsimmons et al. (2014) note the cost would be US $219m for hospital care and US $1.2bn for absenteeism (2013 prices).

Compared with patients with NSLBP, sciatica patients suffer more persistent and severe pain, higher and more prolonged levels of disability, higher absence from work, and require more healthcare resources (Konstantinou et al., 2013; Goode et al., 2011). One study suggested that 55% of patients still had some symptoms of sciatica after two years and 53% after four years (Tubach et al., 2004). This problem of persistence arises in part

from the fact that as sciatica becomes more chronic (>12 weeks) or recurrent, it becomes less responsive to treatment (Furlan et al., 2009).

The current clinical guidelines for the management of sciatica advocate a 'stepped' treatment approach, where patients initially receive educational materials, advice, and analgesic medications (NICE, 2016). For those not improving, referral to physiotherapy is recommended, for appropriate treatments such as exercise and manual therapy. Only patients with more persistent symptoms are referred to specialist services for further assessment and consideration of more invasive interventions such as spinal injections and surgery (NICE, 2016). Currently, there is no evidence to guide clinicians as to which patients may need more invasive treatments earlier in the presentation of sciatic pain, and there is variation in the treatment of sciatic patients in the UK, as noted by the U.K Spinal Taskforce (NHS, 2013). A model of stratified care for patients presenting in primary care with sciatic pain has been tested in an RCT, the SCOPiC trial (ISRCTN75449581) (Foster et al., 2017), HTA report in press.

### 1.6.1 Summary of the SCOPiC trial

The SCOPiC trial (Sciatica Outcomes in Primary Care) was a multi-center pragmatic assessor-blind, two-arm randomised controlled trial, set in primary care. The trial tested whether the stratified care model tested led to faster recovery and overall better outcomes for sciatica patients compared to usual non-stratified care and whether it was cost-effective.

In the SCOPiC trial the allocation of sciatica patients to one of three matched care pathways, was based on the combination of prognostic information, using the STarT Back tool, and information from the clinical examination. The details of the algorithm used to allocate sciatica patients in one of three groups are given in Konstantinou et al. (2019), and the matched treatments details are described in Foster et al. (2017). Briefly, patients were

allocated to group one, if their total score on the STarT Back tool was less than or equal to three out of a possible nine, and they were referred to primary care physiotherapy for management options of low treatment intensity. Using a combination of the STarT Back tool score (if ≥4) and a number of findings from the clinical examination, patients were allocated to group 2 or group 3. Those in group 2 were referred to physiotherapy for management options of higher treatment intensity, with those in group 3 being fast-tracked to a specialist spinal opinion and imaging tests.

As noted above, the results of the SCOPiC trial are at present unpublished, nonetheless the trial revealed that stratified care was neither clinically nor cost-effective compared to usual non-stratified care.

### 1.7 Rationale for the thesis

Where a condition is predominately chronic, decision makers are likely to prefer longer-term cost-effectiveness evidence which reflects the period over which the costs and /or effects of alternate options would be expected to differ (Drummond et al. 2015). The absence of evidence over the long-term cost-effectiveness of the STarT Back approach for LBP justifies the construction of the decision analytic model for stratified management of LBP in this thesis. In relation to sciatica, it was the view of the experts consulted (Appendix 8) that usual care in the SCOPiC trial was more reflective of best available care than usual care generally available to patients. As a consequence it was decided than an analysis should compare stratified care and 'best usual care' obtained in the SCOPiC trial, with that of 'usual care' in a cohort study.

Given the scarcity of decision analytic modelling around LBP and sciatica, no consensus over an appropriate methodology has emerged. Careful development of a modelling methodology for both conditions as well as the stratified nature of treatments will form a fundamental component of this thesis, and ensuring suitable model development will require

preliminary research. Therefore, it was necessary to carry out a systematic review of current modelling approaches for both conditions to help inform such a process. Given the stratification involved in the Keele approach, it was also necessary to ensure the modelling reflects this. A separate review of decision analytic modelling approaches to stratified care was therefore required.

Finally, having conducted research regarding modelling methodologies and built both decision analytic models, reflection upon this process can guide future modelling endeavours.

## 1.8 Aims, objectives and structure of the Thesis

### 1.8.1 Aims and Objectives of the Thesis

The thesis has two broad objectives, namely, to explore themes related to decision modelling in both conditions, as well as produce cost-effectiveness analyses of stratified care in both conditions.

Specifically, these broad objectives will be met by five specific aims;

1) To systematically review the current economic modelling literature in LBP and sciatica in order to explore decision modelling in both conditions.

2) To systematically review economic evaluations of stratified care/personalised medicine interventions in order to explore decision modelling in stratified care.

3) To conduct a decision model to estimate the long-term cost-effectiveness of a stratified care approach for LBP

4) To conduct decision model to estimate the long-term cost-effectiveness of a stratified care approach for sciatica

5) To draw upon the insights gained from the reviews and model building process to produce guidance on approaches to decision modelling in LBP/sciatica and for stratified care.

### 1.8.2 Structure of the Thesis

The rest of this thesis is structured as follows:

**Chapter Two** provides an overview of the clinical areas of both LBP and sciatica. The chapter includes definitions, a review of the epidemiological literature, detail on management of the conditions, and a review of their economic impact. The purpose of this chapter is threefold, providing context to the thesis, informing the appraisal of modelling studies included in the systematic reviews, and identifying key features of each condition that ought to be considered when constructing the decision analytic models.

**Chapter Three** sets out the broad rationale for the use of economic evaluation in healthcare funding decisions, as well as the justification for the specific use of decision analytic models as a method of economic evaluation. Other issues considered in this chapter are; how to select the appropriate model type, as well as best practices around model construction and execution. The content of chapter is primarily chosen to provide context, inform the systematic review, and guide the process of building and appraising the decision analytic modelling.

**Chapter Four** presents a systematic review of the modelling literature in both LBP and sciatica. The review provides a current statement regarding the use of decision analytic models in both conditions. This review also identifies critical considerations for the process of the building of the two de-novo models in this thesis.

**Chapter Five** is a literature review of the modelling literature on approaches to the economic evaluation of stratified care. The purposes of this review are identical to the clinical review, namely to inform the model building process.

**Chapter Six** presents the first decision analytic model, a Markov model to examine the cost-effectiveness of stratified care in a general back pain population compared to usual care. Analyses include a full examination of the uncertainty owing to structural, methodological, heterogeneity and parameter uncertainty, as well as value of information analysis.

**Chapter Seven** presents the second decision analytic model, an individual patient model, to examine the cost-effectiveness of stratified care in a sciatica population compared to best usual care and usual care. Analyses include a full examination of the uncertainty owing to structural, methodological, heterogeneity and parameter uncertainty, as well as value of information analysis.

**Chapter Eight** discusses the findings of the models presented in Chapters six and seven. It provides a discussion of the main findings, the policy implications, implications for future research, strengths and weaknesses of the research, and comparison with other papers. The chapter also answers the fifth objective of the PhD, to generate insights from the reviews and model building process to provide guidance on decision modelling in both low back pain and sciatica and stratified care.

## 1.9 Conclusion

In summary, this PhD thesis aims to produce an economic analysis of the long-term cost-effectiveness of the STarT Back approach for the management of LBP and the SCOPiC trial stratified approach for the management of sciatica. These objectives will be achieved by developing a modelling methodology, which will combine insights gleaned from systematic reviews and from clinicians and health economic modellers with the data

available in Keele cohort studies and trials. The thesis will subsequently reflect upon these processes and produce guidance for future modelling endeavours, both in LBP and sciatica in stratified care.

# Chapter 2: OVERVIEW OF THE CLINICAL AREA

## 2.1 Introduction

Low back pain (LBP) is one of the most significant contemporary global public health problems (Hoy et al., 2014). For most cases of LBP, it is not possible to find a specific cause for the pain and the term NSLBP (non-specific low back pain) is widely used (Hartvigsen et al., 2018). Globally, the prevalence of LBP is rising, driven by an ageing population, as well as trends in factors such as falling rates of physical activity and rising levels of obesity (Duthey, 2013). The condition presents major challenges for clinical management, is associated with high healthcare resource usage, absence from work, long-term incapacity, and is the leading cause of global disability, responsible for the most life years lost to disability (Buchbinder et al., 2013; NICE, 2016; Balague et al., 2012;, Hartvigsen et al., 2018, Buchbinder et al., 2018).

Sciatica (also called radicular pain or nerve root pain), a common variation of LBP which is due to spinal nerve root compression (Koes et al., 2007), is less prevalent when compared to NSLBP. However the impact upon patients is often worse than the impact of LBP alone (Konstantinou et al., 2013), and up to one-third of patients with severe sciatica continue to have significant pain at one year (Balague et al., 1999). Overall, reported rates of recovery at one-year for patients with sciatica range from 49% to 58% (Koes et al., 2007; Haugen et al., 2011; Konstantinou et al., 2018). Patients with sciatica report lower health-related quality of life than the general population and even lower than those suffering from cancer or heart failure (Laroche and Perrot, 2013). Whilst cost of illness studies are rare for sciatica, a Dutch study estimated that the cost of sciatica to society represents 13% of all LBP related costs (van Tulder et al., 1995), which translates to an annual impact to the UK economy of £268 million in direct medical costs and £1.9 billion in indirect costs (Foster et al., 2017).

This chapter aims to explore the epidemiology of both LBP and sciatica, their recommended management options reflecting updated NICE (2016) guidelines, alongside a brief review of the evidence which underpins these treatment options. The chapter concludes by reviewing the healthcare usage and broader economic impact associated with LBP and sciatica.

## 2.2 Epidemiology of low back pain

### 2.2.1 Prevalence

There is substantial literature on LBP prevalence rates, although there is no consensus over the exact rate, implying that mean estimates ought to be interpreted with caution (Hoy et al., 2012). Differences in the estimates often reflect different populations, alternate definitions, study designs, and data collection methods (Jones and MacFarlane, 2005; Balague et al., 2012). For example, where LBP was defined as requiring sick leave in the past six months, the prevalence was reported to be 8%, yet where defined as "pain lasting one-day" prevalence was estimated to be as high as 45% (Ozguler et al., 2000).

The WHO (2003) state that globally, the lifetime prevalence of NSLBP is estimated to be around 80–85%. Whilst other studies estimate that in industrialised nations, the lifetime prevalence is 60–70%, with a one-year prevalence of 15–45% (Duthey, 2013). A recent systematic review and meta-analysis found the mean lifetime prevalence of LBP to be 39% (Hoy et al., 2012).

There is no strong evidence that prevalence rates vary across age groups (Calvo-Munoz et al., 2013) with those around the age of 18 years having a similar prevalence to adults (Hoy et al., 2012) and those aged over 60 have similar prevalence to those in middle age (Fejer and Leboeuf-Yde, 2012). However, older age groups are more prone to experience severe pain. For example, the prevalence of severe back pain and loss of function is known to increase with age (Dionne et al., 2006; Dunn et al., 2013b).

### 2.2.2 Clinical course

Historically most studies had shown that most people with acute LBP recover in timely fashion (Balague et al., 2012). One review reported around half of patients who consult with acute NSLBP can expect to resume their normal activities within 4 to 6 weeks, by 12 weeks the recovery rate is approximately 90%, whilst only 6-10% of patients experience chronic pain and work incapacity (Von Korff et al., 2005).

However, literature emerged showing that that LBP does not necessarily fully improve for many, in fact a study by Croft et al., 1998 had already reported that of those experiencing pain and disability at baseline only 18% had fully recovered one year on. A systematic review reported that 40–50% of individuals with acute LBP still have symptoms at three months (May, 2010). Whilst other studies suggest that at least 40% of people seeking healthcare recover within a year of an episode (Hestbaek et al., 2003), and around one-third of patients had not recovered after one year, whilst amongst those still experiencing pain at 3 months 60% had not recovered at one year (Henschke et al., 2008). Dart et al. (2012) found that persistence of LBP up to 6 weeks reduces the probability of recovery. The seeming tensions in the findings arose not only from different definitions of improvement, but as Croft et al. (1998) found most patients with LBP were not returning to consult their doctor about their continuing pain, with only 8% continuing to consult for more than three months.

Studies with longer-term follow-up indicate that the number of patients reporting persistent symptoms represent a small minority. A Danish study with over 5-years follow-up, found around 10% reported more than 30 days of back pain at all points (Hestbaek et al., 2003). One Swiss study with annual follow-ups over five years indicated 14% of people suffering from back pain at all follow-up points, although only 35% back-pain free at all points (Kolb et al., 2011). These studies suggest that although most people will suffer from back

pain at some point during their life, it is less common to progress to chronic symptoms and long-term persistent back pain.

Overall evidence indicates that at a population level, the prognosis of acute LBP is generally good with highly likely substantial improvement in the first month (Hestbaek et al., 2003) whereas chronic pain may occur in 1 out of 5 patients seeking care for LBP (Weiner and Nordin, 2010).

#### 2.2.2.1 Recurrence

Recent evidence suggests that rather than experiencing isolated LBP episodes, individuals experience repeated episodes of pain throughout life (Dunn et al., 2013a). Large epidemiological studies show that recurrence is a common feature of LBP (da C Menezes Costa et al., 2012; Stanton et al., 2008). The recurrence rate or episode rate varies considerably because of the difficulties in estimation arising from the lack of a standardised definition (Weiner and Nordin, 2010). One systematic review estimated a one-year recurrence incidence to be between 24% and 80%, and also suggested about 60–70% of those who 'recover' from an episode have a recurrence within the following year (Hoy et al., 2010). At two years of follow-up the recurrence rate of NSLBP has been estimated between 5% and 80% (Carey et al., 1999; Campello et al., 1976; Vingard et al., 1976).

#### 2.2.2.2 Low back pain trajectories

Recent research suggests that most LBP patients tend to follow a particular trajectory of pain over time, with most patients displaying patterns reflecting little variation around their mean long-term pain (Dunn et al., 2013a; Dunn et al., 2006; Axen et al., 2011; Kongsted et al., 2015; Tamcan et al., 2010). The concept of pain trajectories is a refutation of the differentiation between acute and chronic LBP, instead seeing patients as having distinct pain trajectories over time (Dunn et al., 2013b).

Reviewing the literature in this area, a recent systematic review concluded that despite differential trajectories within different settings, there are general congruent patterns visible across cohorts and settings (Kongsted et al., 2016). The authors suggest that the majority of LBP patients in primary care do not experience either recovery or chronic severe pain (only one in five have persistent severe pain), but have patterns of pain of mild intensity or infrequent LBP episodes which are relatively stable over time.

Moreover, particular LBP trajectories are associated with different patient characteristics, suggesting that trajectory patterns might have the potential for supporting clinical decision-making. For example, treatments directed at a flare-up of LBP could be differentiated from interventions aimed at managing long-term LBP patterns (Kongsted et al., 2016). At present there are no established ways with which to differentiate group trajectories, and different authors use different subgroupings; Dunn et al. (2006) and Kongsted and Leboeuf-Yde (2010) group according to pain intensity; Tamcan et al. (2010) use pain characteristics, medication use, healthcare use and social and work limitations; and Axen et al. (2011) use clinical background variables. All studies described meaningful differences in trajectories according to these sub-groupings.

### 2.2.3 Risk factors for onset of LBP

NSLBP is multifactorial with medical, biomechanical, psychosocial, and socio-demographic risk factors at play (Ramond-Roquin et al., 2015b). There is reason to suggest that occupational factors contribute towards the development of LBP taking into account the higher prevalence of LBP within the working population (Ramond-Roquin et al., 2015a). For example, higher risk of LBP has been associated with heavy load lifting, exposure to vibrations, persistent standing, working more hours than planned, occupational driving, working in construction, bending and twisting and repetitive trunk movements

(Ramond-Roquin et al., 2015a; Duthey, 2013; Coeuret-Pellicer et al., 2010; Karacan et al., 2004).

Biological characteristics, such as height and obesity, have also been associated with an elevated risk of developing LBP (Andersson et al., 1999; Deyo and Weinstein, 2001; Hollingworth et al., 2002, Shiri et al. (2010a; 2010b)). Genetic components could also be important, with some suggestion that both LBP and intervertebral disc narrowing have a certain degree of heritability (Balague et al., 2012). Psychosocial factors may also potentially have a causal effect (Balague et al., 2012). Higher LBP incidence has been reported amongst those suffering from negative affectivity, low social support, low level of job control, high psychological demands, work dissatisfaction, stress, anxiety, and/or depression (Duthey, 2013).

### 2.2.4 Risk factors for poor outcomes and/or recurrence

There are many studies which have investigated factors associated with poor prognosis in LBP. These include pain severity, with higher baseline levels associated with poor outcome over time (Campbell et al., 2013), having pain in more than one site (Dunn et al., 2013b), being older (Henschke et al., 2008), having either limited or excessive levels of physical activity (Heneweer et al., 2009), coming from a lower socioeconomic background (Beneciuk et al., 2017) and having lower educational levels (Dionne et al., 2001)

Psychosocial factors are particularly implicated in poor LBP prognosis (Balague et al., 2012). One recent systematic review found that depression, psychological distress, self-rated psychosocial health, passive coping strategies, patients' expectations of recovery, and high levels of pain-related fear were each independently associated with poor outcome in LBP patients (Ramon-Roquin et al., 2011a;b). Maladaptive pain coping behaviours (including avoidance and catastrophising) and presence of psychiatric comorbid conditions

are also strong predictors of poor outcomes at one year (Chou et al., 2010; Chou et al., 2014).

## 2.3 Epidemiology of sciatica

### 2.3.1 Prevalence of sciatica

Of those patients who have LBP, between 20-35% suffer from sciatica (radicular pain) (Laroche and Perrot, 2013). However, there are fewer epidemiological studies on sciatica compared to LBP (Kääriäl et al., 2011) but a similar variability as regards prevalence estimates (Konstantinou and Dunn, 2008). This variation arises from a number of factors; differences between self-reported and clinically assessed symptoms; data collection methods; populations studied; time frames; and significant variability in the definitions of sciatica, particularly poor differentiation between "true sciatica" which is leg pain due to lumbar spinal nerve root involvement, and all other referred leg pain (Konstantinou and Dunn, 2008; Valat et al., 2010; Lewis et al., 2011). In a review of sciatica prevalence estimates, lifetime prevalence estimates were between 12.2% and 43%, annual period prevalence between 2.2% and 34%, and point prevalence estimates ranged from 1.6% to 13.4% (Konstantinou and Dunn 2008). In this review, there were three UK studies, all reported period prevalence reflective of experience of symptoms in the prior year. The estimates from both the Hillman et al. (1996) and Palmer et al. (2003) studies, 17.8% and 14.2%, fall around midway between the lowest and highest international period prevalence's found by the Konstantinou and Dunn (2008) review. Whilst in the study with the lowest UK prevalence (Lyons et al., 1994), 6.3%, sciatica was ascertained by clinical diagnosis. Generally, where clinical assessment or stricter case definitions are used, lower prevalence rates are reported (Heliovaara, 1987).

### 2.3.2 Recurrence

It was commonly thought that the clinical course of acute sciatica is favourable, and usually improves within 2–4 weeks regardless of treatment (van Tulder et al., 2010). However, literature indicates that compared to patients with NSLBP, sciatica patients suffer more persistent and severe forms of pain and loss of function, as well as less favourable outcomes, more prolonged disability and higher absence from work, consuming more health resources in the process (Konstantinou et al., 2013; Goode et al., 2011). One study, suggested that 55% of patients still had some symptoms of sciatica after two years and 53% after four years (which includes those who had recovered at year two but relapsed between years two and four) (Tubach et al., 2004). As sciatica becomes more chronic (>12 weeks) or recurrent, it becomes less responsive to treatment (Furlan et al., 2009) hence treatments that might prevent patients developing chronic forms of sciatica are imperative.

### 2.3.3 Risk factors for onset of Sciatica

Two studies on risk factors indicated that patients who previously had a case of LBP were between 1.5 and 3 times more likely to develop a first incidence of sciatica (LeClerc et al., 2003; Kääriäl et al., 2011). Another study reported that those who had a previous episode of severe LBP were 4.5 times more likely to report an incidence of sciatica (Riihimäki et al., 1994). Those in poor health are also three times more likely to suffer a first case of sciatica (LeClerc et al., 2003).

Various activities involving particular forms of movement are known to increase the risk of developing sciatica (radicular pain). Factors include constant heavy workload (Riihimäki et al., 1994; Miranda et al., 2002), frequent flexing and twisting the trunk, kneeling or squatting, raising arms above shoulder (Miranda et al., 2002; Laroche and Perrot, 2013). Other factors associated with a higher incidence of sciatica include obesity (Shiri et al., 2007, Rivinoja et al., 2011; Kääriäl et al., 2011), and smoking, both current and previous, (Kääriäl et al., 2011; Manninen et al., 1995; Miranda et al., 2002, Qiao et al.,

2000). Two studies reported that those who are physically active or moderately physically active, and those who walk moderately and actively, have higher risk of developing sciatica (Kääriäl et al., 2011; Miranda et al., 2002). Age is also known to be a factor, with evidence that the incidence rate for developing radicular pain peaks in the fifth decade and declines thereafter (Miranda et al., 2002; Laroche and Perrot, 2013). Height is a risk factor, with those over 180cm in height three times as likely to develop first incidence of sciatica (LeClerc et al., 2003).

There are few studies examining the impact of psychosocial factors on sciatica incidence, reporting conflicting results (Kääriäl et al., 2011). One Finnish study showed that low job control was associated with increased hospitalisations owing to intervertebral disc disorders (Leino-Arjas et al., 2004). Two other studies showed an association between sciatica and psychological distress and mental stress (Pietri-Taleb et al., 1995; Miranda et al., 2002). However, two other studies found no relationship between sciatica and psychosocial factors (Leclerc et al., 2003; Kääriäl et al., 2011). It may be that different definitions and means of measuring psychosocial factors contribute to conflicting findings.

### 2.3.4 Risk factors for poor outcome

There are a number of studies investigating prognostic factors for sciatica patients. Research on characteristics potentially associated with outcome in sciatica has identified a limited number of prognostic factors independently associated with outcome, mainly in studies of secondary care cohorts (Ashworth et al., 2011; Peul et al., 2008; Verwoerd et al., 2013; Konstantinou et al., 2018).

Only pain and condition-specific disability are consistently associated with having spinal surgery, which in this secondary care context is taken as a proxy of poor outcome for natural course and conservative management (Verwoerd et al., 2013). A UK primary care cohort of patients with suspected sciatica found that the impact of sciatic pain on patients,

and their expectation of non-improvement over time were independently associated with non-improvement (Konstantinou et al., 2018). Several other factors (e.g. age, gender, psychosocial factors) commonly thought to be associated with outcome in LBP, do not seem to be associated with outcome in sciatica presentations (Ashworth et al., 2011, Verwoerd et al., 2013, Konstantinou et al., 2018).

## 2.4 Management of low back pain and sciatica

In what follows, each of the potential treatment options for LBP and sciatica advocated in national clinical guidelines issued by NICE (2016) are discussed in turn, with a brief summary of the evidence base.

### 2.4.1 Use of stratification tools for guiding treatment

NICE (2016) recommends the use of stratification tools, such as the STarT Back tool (Hay et al., 2008; Hill et al., 2008) to assist clinical decisions about the management of patients with NSLBP or back pain with sciatica. The STarT Back tool estimates the risk of future poor outcome (low, medium or high). The overall approach combines the estimation of prognostic risk and matched treatments according to the level of risk (Hill et al., 2011). The STarT Back tool assesses a number of physical and psychological factors potentially associated with poor recovery from LBP, with treatment tailored to the individual patient. Patients at low risk of poor outcome are expected to improve and are likely to do well with minimal input comprising advice and reassurance. For patients at medium risk of future poor outcome, a course of physiotherapy treatment tailored to the individual patient's needs is recommended which may include advice, reassurance, exercise, or manual therapy techniques. For patients at high risk of a poor outcome, physiotherapy input is also recommended with more emphasis on addressing psychosocial obstacles to recovery, such as excessive worry about the condition and unhelpful pain-related fear of moving and physical activity.

### 2.4.2 Self-management

Where self-management strategies are appropriate, practitioners are advised to provide patients with advice and information specifically tailored to their needs at all steps of treatment pathways. This would include information on the nature of LBP and sciatica as well as encouragement to stay active and continue normal activities. All patients should receive information about treatment and prognosis (NICE, 2016).

Guidelines generally advocate remaining active, based upon evidence that continuing with daily activities is more effective than resting in reducing pain and improving functional status for patients with either acute or sub-acute LBP (Dahm et al., 2010). Where patients require bed rest to relieve severe symptoms, they ought to be encouraged to return to regular activity as soon as possible (Chou et al., 2007).

### 2.4.3 Exercise

Clinical guidelines (NICE 2016) recommend group exercise programmes, which could be biomechanical, aerobic, or mind–body or combination, can be considered on the NHS for people with a specific episode or flare-up of LBP with or without sciatica. Patient's specific needs, preferences and capabilities should be considered when choosing the form of exercise (NICE, 2016). This guidance reflects evidence suggesting that structured exercise programmes are clinically and cost-effective compared with usual care (NICE, 2016).

Research evidence indicates that individually designed exercise programmes offered in a supervised group setting are the most effective means of delivering exercise programmes (Hayden et al., 2005a). These measures are more effective in patients with chronic NSLBP but not in patients with acute back pain (Hayden et al., 2005).

Yoga has also been suggested as being useful in managing back pain. A meta-analysis of 10 RCTs found that yoga was associated with beneficial effects on short- and long-term pain and back-specific disability (Cramer et al., 2013).

In contrast, the evidence on physical activity and sciatica is conflicting. Reviewing RCTs on restricting physical activity, Lewis et al. (2011) report no difference in outcomes at various time points between bed rest and either advice to stay active or exercise. However, it makes intuitive sense that remaining active as able is likely to be overall more beneficial than resting for too long.

### 2.4.4 Manual therapies

Manual therapy, spinal manipulation, mobilisation or soft tissue techniques, can be offered but only as a component of a treatment package including exercise and/or psychological therapy (NICE, 2016). For acute LBP (duration <4 weeks), spinal manipulation administered by providers with appropriate training is associated with small to moderate short-term benefits (Assendelft et al., 2004). One systematic review found 'high-quality evidence' that spinal manipulation therapy was at least as effective as a variety of other interventions for reducing pain and improving function in patients with chronic LBP (Rubinstein et al., 2011). Studies on soft tissue techniques are exclusively taken from populations of LBP without sciatica, and reveal modest effects. In a review of these studies, only one of three studies found a clinically significant reduction in pain compared to sham (NICE, 2016).

### 2.4.5 Psychological interventions

NICE (2016) recommends cognitive behavioural approaches to be offered as part of a treatment package, including exercise, with or without manual therapy interventions. In a review of 22 studies across various psychological therapies including, cognitive behavioural therapy (CBT), self-regulatory therapy (SRT), behavioural therapy, and

multidisciplinary therapies, psychological therapy was found to be significantly superior to control treatments, including usual care and physiotherapy across all treatments across all time periods, all outcomes and all control treatments (Hoffman et al., 2007). Another review of 30 trials, found behavioural treatments more effective than usual care for short-term pain relief although no differences in the intermediate and long term (Henschke et al., 2010).

### 2.4.6 Combined physical and psychological programmes

A combined physical and psychological programme can be considered for people with persistent LBP or sciatica where there are either significant psychosocial obstacles to patient recovery or where previous treatment has been ineffective (NICE, 2016). Functional restoration with a cognitive-behavioural component and intensive interdisciplinary rehabilitation reduces work absenteeism due to LBP in occupational settings (Schonstein et al., 2003; Guzman et al., 2001). Patients with chronic LBP receiving multidisciplinary biopsychosocial rehabilitation experienced moderately less pain and disability, and more likely to be in employment, compared to those receiving usual care or physical treatment (Kamper et al., 2014). For patients with chronic disabling LBP, particularly those with psychosocial risk factors, intensive interdisciplinary or multidisciplinary therapy consisting of physical, vocational, and behavioural interventions provided by a multidisciplinary healthcare team seems more effective than standard care and is an important treatment option (Guzman et al., 2001; Karjalainen et al., 2001).

### 2.4.7 Return-to-work programmes

Returning to work and normal daily activities should be promoted and facilitated wherever possible (NICE, 2016). There is some evidence that physical conditioning as a component of a return to work strategy for workers with back pain has some effect compared to usual care. One systematic review of 25 RCTs, showed that for workers with chronic LBP,

physical conditioning reduces sick leave compared to usual care after 12 months (Schaafsma et al., 2013). For patients with sub-acute or acute LBP these interventions do not seem to be as effective when compared to usual care.

### 2.4.8 Pharmacological interventions

NICE (2016) acknowledges that the evidence base is weak but suggests the consideration of analgesic medications as appropriate in the management of LBP and sciatica, to facilitate recovery or maintenance of function. Analgesic options include non-steroidal anti-inflammatory drugs (NSAIDs) and weak opioids. Clinicians are advised to consider the lowest effective dose for the shortest possible period, informed by the risks and benefits of medication according to each patient's profile.

### 2.4.9 Invasive procedures

Referral for assessment for radiofrequency denervation can be considered for those with LBP where all three of the following conditions are satisfied; (i) where non-surgical treatment has not worked; (ii) where the origin of pain is considered to come from the medial branch nerve, (iii) and where localised back pain is considered to be moderate or severe at time of referral (NICE, 2016). Radiofrequency denervation is an option in people with chronic LBP after a positive response to a diagnostic medial branch block.

Consideration of epidural injections of local anaesthetic and steroid is recommended for those with acute and severe sciatica (NICE, 2016). A meta-analysis of 'good' RCTs showed epidural injections for the treatment of sciatica were 'significantly' better than inactive control at short-term follow up for reduction of pain and improving functional status, although no evidence of a difference in the longer and medium term (Lewis et al., 2011).

Spinal decompression can be considered for people with sciatica, where non-surgical treatment has failed to improve pain or function and where radiological findings are consistent with sciatic symptoms (NICE, 2016).

## 2.5 Healthcare usage and economic costs of low back pain and sciatica

Most of the costing literature on LBP comes from cost-of-illness studies, which predominantly take two forms, prevalence or incidence approaches. The prevalence approach takes a given year and derives the total annual cost of a certain disease. A more complex approach, the incidence approach, involves calculating lifetime costs of incident cases (Rice, 1994).

Published studies to-date use the prevalence approach to derive annual costs. Costs in these studies are commonly sub-divided into direct medical costs and indirect costs, the former relate to healthcare costs, whilst the latter relate to work absenteeism or productivity losses.

### 2.5.1 Direct costs of low back pain

In terms of healthcare usage, Maniadakis and Gray (2000) estimate 15% of the total costs relating to LBP fall within the healthcare sector, which amounts to £2.97bn in 2016 prices. However, the authors also found that 35% of these costs related to services provided privately, implying the NHS costs would be around £1.93bn in 2016 prices.

The NHS cost burden of LBP reflects in part the demands the condition places upon GPs. In the UK, a recent cohort study found that 5.9% of adults consult their GP about LBP each year and that 14% of all UK primary care consultations are for LBP complaints (Jordan et al., 2010). Similar results were also found in France (Plenet et al., 2010). The costs related to specific types of healthcare for LBP have not been subject to extensive investigation in the UK. A study conducted in the Netherlands, (Lambeek et al. 2011) found that 21% of expenses related to inpatient care, 1% to diagnostic evaluations, 25% to

outpatient care, 49% to physical therapy, chiropractor or massage, and 4% to prescription medication. There is trial-based evidence from the UK, for example in the usual care arm of a UK cost-utility study (Whitehurst et al., 2015) of the implementation of stratified care, the IMPaCT Back study, highest risk patients saw 15% of their total costs related to primary care consultations, 22% related to consultant consultation, 11% related to diagnostic tests and epidurals, 37% relate to consultations with physiotherapists, acupuncturists, osteopaths etc., 6% arose from prescription medication, and 8% arose from over the counter medications.

### 2.5.2 Indirect costs of low back pain

Studies investigating the costs associated with LBP consistently show that the indirect costs far outweigh the direct costs of treatment, with healthcare costs only estimated to be only 7–14% of the total cost of LBP (May, 2010). The extent of these indirect costs reflects the fact that LBP is the leading cause of activity limitation and work absence throughout much of the world. In the UK, LBP is the most common cause of disability in adults with an estimated 137 million work days lost per year (ONS, 2017). As the working population is particularly affected by LBP, it drives high work absenteeism and subsequent productivity losses, which are the predominant contributors to the considerable socioeconomic costs (Ramond-Roquin et al., 2015a). This can clearly be seen in the selected cost of illness studies shown in Table 2.1.

**Table 2.1 Costs associated with LBP (selected studies)**

| Lead author, country | Year | Total societal costs | Costs as a proportion of 2016 GDP | Indirect cost, % |
|---|---|---|---|---|
| Maniadakis, UK | 1998 | £19.77bn | 1.00% | 85 |
| Rizzo and Lou, USA | 1998 | £60.82bn | 0.38% | 53 |
| Van Zundert, Belgium | 1999 | £1.44bn | 0.38% | 84 |
| Ekman, Sweden | 2001 | £2.19bn | 0.52% | 84 |
| Walker, Australia | 2001 | £8.42bn | 0.85% | 89 |
| Weiser, Switzerland | 2005 | £6.68bn | 1.22% | 62 |
| Lambeek, Netherlands | 2007 | £3.65bn | 0.58% | 88 |

*Adapted from Harvigsen et al. (2017). Costs inflation-adjusted to 2016 prices using nation-specific inflator from World Bank Indicators, GDP deflator, (available at https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS, accessed 12th August 2019) converted to GBP using 2016 currency convertors from The World Bank Indicators, Official Exchange Rate (available at https://data.worldbank.org/indicator/PA.NUS.FCRF, accessed 12th August 2019).

Referring to Table 2.1, although comparisons between countries is not straightforward because of different costing perspectives, health systems, methodologies, and time periods, the results indicate the gravity of the economic burden of LBP, ranging from 0.38% of GDP in Belgium/USA to 1.22% in Switzerland. The table also highlights the significance of indirect costs in that burden, responsible for between 53% of the burden in the US to 89% of the burden in Australia, with indirect costs making up 85% of the burden in the UK.

Maniadakis and Gray (2000) quantified the burden of LBP in the UK, using estimates of prevalence, healthcare resource use, and unit cost data. Their study estimated that the cost of back pain in 1998 was £12.3bn (£19.86bn in 2016 prices) using the human capital methods (HCM), and £6.7bn (£10.77bn) using the friction cost approach. The HCM, which has been the traditional method of estimating productivity loss, estimates losses by multiplying some time loss from work absence by a wage rate. Whilst use of the HCM is common place in economic evaluation, it has been criticised for producing inflated estimates of productivity losses, and therefore overestimating societal costs (Koopmanschap et al., 1995). The friction cost method on the other hand assumes that productivity losses are limited only to the period of time taken to restore productivity levels, by training a replacement work (Koopmanschap et al., 1995).

Trial-based economic evidence would suggest that cost is associated with severity of presentation. For example, in the IMPaCT Back study, in the usual care arm, of those consulting with low back pain in primary care, high risk patients had £1459 of LBP-related work absence, even £1135 in medium risk patients, falling to £30 in low risk patients (Whitehurst et al., 2015).

### 2.5.3 Economic burden of sciatica

Although it is recognised that the impact of sciatica has on patients is more substantial than LBP alone (Konstantinou et al., 2013), the economic costs relating to sciatica are far less well understood. The only cost of illness study found in a literature search was from the Netherlands in 1991, and estimated that the cost of sciatica, to be 13% of the total costs of LBP, $128m of which were for hospital care, $730m for absenteeism, and US $708m for disablement (van Tulder et al., 1995). Converting these into UK prices, Foster et al. (2017) note that the UK costs would be £268m for hospital care and £1.9bn for absenteeism.

At the individual level, Kigozi et al. (2019) and estimate NHS costs associated with sciatica to be £312.85 per annum, with societal costs £1246.92, implying NHS costs are around 25% of the total costs. The two largest contributors to direct costs were NHS physiotherapy (£113.96) and GP consultations (£78.57), jointly responsible for 62% of total direct costs.

## 2.6 Summary

The purpose of this chapter was to provide an overview of LBP and sciatica. The chapter has explored aspects of both conditions which will be relevant to the modelling process, as well as what is known about the economic evidence in both conditions. The next chapter reviews the role of economic evaluation and decision modelling in healthcare.

# Chapter 3: OVERVIEW OF THE USE OF DECISION ANALYTICAL MODELLING IN ECONOMIC EVALUATION IN HEALTHCARE

## 3.1 Introduction

As stated in the introductory chapter, the role of economic evaluation is to aid decision-makers in making choices about which healthcare interventions to fund from scarce resources (Drummond et al., 2015). The purpose of this chapter is to explore the analytical techniques used by health economists, which are utilised in this thesis.

The chapter initially briefly restates the rationale for the use of economic evaluation in healthcare explored in the introduction, before moving onto consider both the philosophical origins of healthcare economic evaluation as well as the technical methodologies available to the analyst. The second half of the chapter investigates why a model-based evaluation can better meet the evidential needs of decision-makers than a trial-based evaluation, concluding, by considering the different types of decision model and particular circumstances under which each model may be regarded as appropriate.

## 3.2 Economic evaluation in healthcare

Economic evaluation in healthcare can be defined as a systematic framework of evidence synthesis, for quantifying the costs and benefits associated with choosing one intervention, service or policy over potential alternative courses of action (Drummond et al., 2015). Underpinning the use of economic evaluation in healthcare is the notion of resource scarcity - the idea that people, time, facilities, equipment, and knowledge are scarce (Drummond et al., 2015). This scarcity forces decision-makers to take difficult decisions about which often-costly treatments or technologies ought to receive funding (Cooper et al., 2006). The rationale for basing decision-making upon systematic analysis is clear, given that tradition, intuition, ideology, or incomplete evidence can lead to incorrect decisions and waste health system resources (Sculpher et al., 2006).

Agencies such as, the National Institute for Health and Care Excellence (NICE), which issue guidance concerning whether various health technologies should be made available through the public healthcare system, explicitly demand systematic evidence-based analysis of both costs and benefits of new treatments (Barton et al., 2004; Brennan et al., 2006). Accordingly, the purpose of economic evaluation in health is to provide decision-makers with the best available evidence upon which to base decisions about the efficient allocation of healthcare resources (Weinstein, MC, 2006; Petrou and Gray, 2011).

### 3.2.1 Philosophical Foundations of Economic Evaluation

Economics can be said to have two strands, normative and positive economics. Positivist economics uses empirical evidence to describe and explain relationships between economic phenomena (Morris et al., 2012). Normative economics is a strand of economics that deliberates upon value judgments regarding "economic fairness" or what the public policy *ought* to seek to achieve (Samuelson and Nordhaus, 2004).

Descriptive and predictive in nature, an economic evaluation is, in essence, a practical positivist technique, yet underpinned by strong normative theoretical foundations (Morris et al., 2012). In health economics, normativity is endemic because judgments will always be required regarding what is considered fair and what constitutes a cost and what constitutes a benefit, but also the perspective that the evaluation assumes (Morris et al., 2012). The different practical approaches discussed below (3.2.2) reflect profound philosophical differences.

#### 3.2.1.1 Welfarism

Traditionally economic evaluations in many policy domains, such as transport, the environment, infrastructure, are performed using Cost-benefit analysis (CBA). CBA itself is situated within the philosophical approach known as welfarism, which many health economists (e.g. Culyer, 1991; Culyer and Evans, 1996; Hurley, 1998) consider unsuitable

for economic evaluation. Welfarism can be defined as "the systematic analysis of the social desirability of any allocation of resources in terms of the utility obtained by individual people" (Morris, 2012, pg.205). To the welfarist, utility is considered the "wantability" of goods and services consumed (Dolan and Kahneman, 2008). Welfarism is therefore an individualistic approach, where individuals are the assumed best judges of their own welfare and strive to maximise utility accordingly (Anand, 2010). Accordingly, social welfare is understood to be a mathematical function of all individual utilities (Culyer, 1991). The implication being that governments ought to promote individual preference satisfaction; if individuals obtain more of what they desire, the economy provides higher utility (Simonetti et al., 2010). In health, the implication is that treatments should be supplied in accordance with individual preferences with paternalistic policy solutions deemed helpful (Brouwer et al., 2008). The role of the expert in the welfarist paradigm concerns the provision of information to assist individuals with their decisions (Seixas, 2017).

In traditional economics, a CBA attempts to value policy or an intervention, using prices revealed in markets. However, markets do not exist for health services in the UK, and therefore economists are required to establish a hypothetical 'willingness-to-pay' (WTP) valuation representing the price a patient would be 'willing-to-pay' for various treatments. Theoretically, in the health context using WTP ought to ensure the most cost-effective treatments are those which are more highly valued by the patient relative to their costs. Moreover, in an efficient market, the costs of providing those services would at most match the consumer's (patient's) willingness-to-pay. In traditional economics, where goods and services are supplied in accordance with consumer preferences, this is called allocative efficiency, which is the allocation that maximises welfare.

Whilst pharmaceutical firms may place monetary valuations on health treatments, such valuations are considered ethically unacceptable to health service professionals and the public (Birch and Donaldson, 2003). In fact, the primary objection in the literature to welfarism in health appears to be that basing welfare measurements on willingness-to-pay is indefensible because the ability to pay dictates that the allocation of healthcare resources could be 'skewed' towards the wealthy (Coast et al., 2008). Indeed, Coast et al. (2008) argue that had health economists not abandoned such an unpopular approach then economic evaluation may not have achieved such influence in healthcare decision-making.

#### 3.2.1.2 Extrawelfarism

Concerns regarding the use of the welfarist paradigm in health, have led to the development of a philosophical tradition known as 'extra-welfarism', from which cost-utility analysis (CUA) was developed. The extra-welfarist paradigm stems from Culyer's (1991) development of Sen's (1986) work on 'functionings' and 'capabilities'. Sen (1992) disputed the central tenant of welfarism, the concept of 'utility', arguing that utility is not an appropriate measure in any analysis since those in persistent deprivation will have preferences which reflect their own individual deprivation and ought not to be considered a true reflection of their true desires. For Sen, various human acts and states have value in themselves, not just in the extent to which they produce utility (Brouwer et al., 2008). Building on from this, Culyer (1991, pg15) argued that utility is too concerned with reactions to commodities, without considering what those commodities enable you to do. Extra-welfarism is an attempt to broaden that evaluative space, allowing that welfare can be derived not only from utility, but "extra" sources such as health, freedom, mobility, etc., the so-called "basic functionings".

Several consequences stem from these philosophical differences. Firstly, each paradigm attaches a different meaning to health. Whilst welfarism considers health only insofar as it

contributes to overall welfare, extra-welfarism on the other-hand offers the health economist an evaluative space pursuing health as an end in itself (Coast et al., 2008; Brouwer et al., 2008). For analytical purposes, this allows that an economic evaluation can consider information about health states directly in their analysis because ultimately health states do influence the preferred social state (Coast et al., 2008).

Secondly, the extra-welfarist can suggest society could be better off were it to embody some feature (health or wellbeing) not because it was preferred but because ethically it was right thing to do (Brouwer et al., 2008). Health can be weighted according to particularities, such as equity, wealth, need (defined by capacity to benefit), equality of access, potentially facilitating basic equality in healthcare provision (Brouwer et al., 2008).

Third, extra-welfarism also allows that outcomes could be weighted according to need. For example, Hurley (1998) contrasts effective market demand with 'effective need' which reflects the prospects for gaining health and unrelated to the consumer's ability to pay, thus shielding welfare judgments from the prevailing income distribution and the Pareto principle used within traditional welfarism. Theorists have argued that the concept of effective need could be particularly appropriate in health, which is commonly considered as not the type of good can be subject to initial inequities in distributions (Brouwer et al., 2008). The distribution of income may be of further importance given the correlation between income and health states and the use of healthcare services (Deaton, 2002).

Finally, perhaps most importantly for economic analysis, is that whilst the welfarist considers "interpersonal comparisons in the evaluative space of utility…impossible or meaningless" (Brouwer et al., 2008) extra-welfarism allows interpersonal comparability in terms of characteristics like health, or quality of life. In healthcare, NICE do mostly base their recommendations for allocation decisions upon cost-utility analyses, which express utility outcomes in a simple generic measure, the QALY, the quality-adjusted life year.

The QALY is a unit of life expectancy adjusted for the quality of life during those years and is conventionally calculated by multiplying the time spent in a given health state by the health-related quality of life relating to that health state (Bhattacharya et al., 2014).The use of a simple generic measure allows comparisons between patient groups and conditions, offering much-needed accountability to the decision-making process (Drummond et al., 2015) and avoids difficulties in policy formulation which may result from refusals to compare individual welfare and reliance on Pareto efficiency (Brouwer et al., 2008).

### 3.2.2 Techniques of economic evaluation

Underpinning the use of economics in health is the notion of opportunity cost, the idea that committing resources to the production of one good or service means forfeiting benefits from those resources not used in their next best alternative (Morris et al., 2012). In this context, the role of economic evaluation is to quantify the health gain achieved by a new treatment or policy with that forgone by the treatment displaced, which requires formal means of measuring the costs and benefits. Whilst the calculation of cost can only be expressed in monetary terms (although may cover costs accruing to a variety of different sectors), the measurement of the consequences of interventions comes in various different forms (Drummond et al., 2015).

There are five main analytic approaches for economic evaluation, each with their own means of evaluating the consequences of an intervention. Cost-minimization analysis (CMA) is used under very specific contexts where alternative treatments have equivalent clinical effectiveness facilitating the comparison of costs-per-course of treatment (Drummond et al., 2015). However given the rarity of circumstances in which it would be an appropriate means of analysis, CMA was pronounced as “(near) death” by Briggs & O’Brien (2001).

Cost-consequence analysis (CCA) is also less frequently used but can be a first step in an evaluation, insightful and easy to understand the CCA does not aggregate consequences into QALYs or cost-effectiveness ratios but presents a listing of associated costs and outcomes (Russell et al., 1996).

#### 3.2.2.1 Cost-benefit analysis

In many countries, health economists use cost-benefit analysis (CBA) which, as noted above, measures consequences in monetary terms. In the UK, CBA is a traditional analytic method used outside of health, in areas such as the environment or transport economics. Where used in health economics, CBA assigns a monetary value to the days lost to sickness, extra years of life, or medical complications (Drummond et al., 2015). Drummond et al. (2015) note that CBAs could be used in health economics within a broader social perspective. For example, a recent study suggested that one pound spent on mental health research gained a recurring 37p annual benefit for the economy in increased productivity and decreased healthcare costs (Economist, 2015). Similarly, a recent study estimated the implementation of a brief vocational advice service to improve work outcomes for MSK patients, would yield an overall societal benefit of £500 million for a cost of £10 million (Wynne Jones et al., 2018). However, owing to the criticisms of the welfarist paradigm the use of the CBA is unpopular in health economic evaluation in the UK, where the two most common frameworks for economic evaluation are cost-effectiveness analysis (CEA) and cost-utility analysis (CUA).

#### 3.2.2.2 Cost-effectiveness analysis

Cost-effectiveness analysis (CEA) is an analytical framework, which compares the health gain arising from a given treatment with the cost of that treatment (Bhattacharya et al., 2014). In a CEA, health gain is measured in natural units common to all comparators used in the analysis. Costs and effects of a given intervention are usually shown in the form of

an incremental cost-effectiveness ratio (ICER), which represents the difference in costs divided by the difference in effects, between options under comparison during a given period (Cooper et al., 2006).

CEAs are often conducted alongside clinical trials (Drummond et al., 2015). For example, an analysis might use a denominator such as 'reduction in the number of patients who fell' (Haines et al., 2013) or 'successfully treated patients' (Lewis et al., 2011). In such cases, the results can indicate whether an intervention can minimise the costs associated with achieving a certain level of health benefit, or maximise benefits holding costs constant (Drummond et al., 2015).

Standard cost-effectiveness decision rules consider that if an intervention is cheaper and more effective than the comparator(s), the intervention is dominant and unequivocally cost-effective (Briggs et al., 2008). However, if an option is more effective but also more costly than the comparator(s), the ICER will be calculated and compared with alternative uses of health services resources (Briggs et al., 2008). In a direct comparison, the preferred option will be that which delivers the lowest cost for a given unit of health outcome. Comparisons can also be made between the ICER and a 'cost-effectiveness threshold', a notional threshold value which decision-makers are prepared to pay for a unit of health (Briggs et al., 2008).

The most significant limitation of CEA is the inability to make direct comparisons across different areas of health, owing to the use of a measurement of effect which is disease-specific (Briggs et al., 2008). Consequently, it is difficult to compare the opportunity costs, for example, with funding a programme to prevent falls at the expense of not funding surgery for sciatica. This is especially problematic where programmes receive their funding from the same budget (Drummond et al., 2015). To make the comparison of the benefits gained from a new intervention with those lost from any displaced existing

programme; some generic measure of benefit relevant to all interventions is required (Drummond et al., 2015).

#### 3.2.2.3 Cost-utility analysis

Cost-utility analysis (CUA) is essentially a cost-effectiveness analysis which uses the QALY, quality-adjusted-life-year, as the effect measure (Briggs et al., 2008). Presentation of results in a CUA comes in the form of the ratio of incremental costs to incremental effectiveness measured by quality-adjusted life years gained (Cooper et al., 2006). For a CUA, this will be the cost-per-QALY gained (Drummond et al., 2015).

A major advantage of CUA is that it allows comparison of interventions across different areas of healthcare, for example, comparison between treatments for cancer and treatments for back pain, facilitating the assessment of the opportunity cost of implementing one over the others. Cost-utility analyses are now the most widely published form of economic evaluation (Drummond et al., 2015). In part, this owes to the publication of influential guidelines advocating such methods. For example, NICE recommend cost-utility analysis as the appropriate method for economic evaluation, *'For the reference case, cost–utility analysis is the preferred form of economic evaluation. This seeks to establish whether differences in expected costs between options can be justified in terms of changes in expected health effects. Health effects should be expressed in terms of QALYs' (NICE 2013, p.37).* Since the early 2000s, NICE has advocated a cost-effectiveness threshold of £20,000 to £30,000 per-QALY (McCabe et al., 2008).

### 3.2.3 Sources of evidence

When conducting an economic evaluation, it is common to see either a trial-based evaluation, or an evaluation via a decision analytic model. In what follows, the advantages of the use of a decision analytic model are explored in contrast to the trial-based analysis.

#### 3.2.3.1 Trial-based evaluations

The trial-based economic evaluation considers only the results of a single clinical trial, where the collection of economic data (e.g. resource utilisation, quality of life values) takes place alongside clinical trials (Petrou and Gray, 2011). Indeed, given their essential role in generating evidence for the evaluation of healthcare programmes, some researchers have argued RCTs ought to provide a vehicle for economic analysis (Drummond et al., 2015). There are several benefits of doing so. Firstly, trial-based economic evaluations provide robust estimates of relative treatment effect, where appropriate randomisation minimises the risk of selection bias (Sculpher et al., 2006). Second, trial-based economic evaluations not only broaden the evidence base on particular interventions but also can prove timely where decision-makers can use evidence regarding the value of a drug to make early adoption decisions (Glick et al., 2014).

Third, statistical and econometric techniques can utilise individual patient-specific data, for example, to analyse the relationships between specific events and health related quality of life (HRQOL), and/or perform subgroup analysis (Drummond et al., 2015; Petrou and Gray, 2011). There may also be good quality information on the costs of interventions and resource use, as well as good patient level utility data. Finally, given typically substantial fixed costs incurred in collecting clinical data, the marginal cost of collecting economic data maybe modest (Petrou and Gray, 2011).

Whilst there are advantages to using trial-based evaluations, they may well not always meet the requirements for which economic analytic framework must fulfil. For reasons explored below, almost without exception, trial-based economic evaluations often fail to satisfy these demands (Sculpher et al., 2006).

3.2.3.2 Criticism of the trial-based paradigm.

An influential paper by Sculpher et al. (2006) called into question the ability of trial-based evaluations to fulfil the evidentiary needs of healthcare decision makers. The main objections raised by Sculpher and colleagues are now discussed in turn.

*Synthesis*

Consistent with principles of evidence-based medicine (e.g. Sacket, 1996) economic evaluations which seek to inform the decision-making process must use and explicitly synthesise all relevant evidence on the decision problem (Sculpher et al., 2006; Briggs et al., 2008). One of the criticisms of trial-based economic evaluation is that a trial-based analysis may ignore other important information derived from trials, meta-analyses, and observational studies, including relevant insights into risks of complications and adverse events (Sculpher et al., 2006; Petrou and Gray, 2011). Decision analytic modelling can provide a full synthesis of the available evidence, even where the rates are in non-constant forms and/or there are no direct comparisons of treatments (e.g. Ades et al., 2006).

*Consideration of all relevant comparators*

Many diseases may be amenable to or require a number of interventions and any evaluation to inform decision-making ought to evaluate all the available treatments (Sculpher et al., 2006). However, typically, RCTs often consider only two comparators - the new technology plus the 'standard' intervention, often-usual care (Sculpher et al., 2006). Furthermore, economic evaluations conducted alongside trials undertaken for drug registration purposes often have a placebo or older therapy as comparator, neither of which represents relevant or current practice in the selected jurisdiction for the condition under investigation (Drummond et al., 2015). A related issue is that for many chronic diseases, therapies are sequential, with few RCTs comparing alternative treatment sequences (Sculpher et al., 2006). Where appropriate options are not included in the analysis, the

result will be a partial analysis, which could potentially result in inappropriate adoption decisions (Sculpher et al., 2006). It is likely, therefore, that data will require synthesis from several clinical studies. A decision-analytic model using appropriate statistical methods provides the perfect framework for such synthesis (e.g. Sutton and Abrams, 2001; Ades et al., 2006).

*Appropriate time horizon*

Any evaluation seeking to inform decision-making must adopt a time horizon sufficient to capture differences in economic outcomes between options (Petrou and Gray, 2011). For example, if the treatment is more effective in the longer-term, not only will the patient gain additional QALYs but may also incur additional costs (Sculpher et al., 2006). Some RCTs cover the lifetime horizon, especially in trials for the treatment of terminal conditions. However, follow-up periods in RCTs are generally shorter than necessary for economic evaluation (Ramsey et al., 2015). To provide reliable estimates of cost-effectiveness the analyst should construct a decision model to structure the extrapolation of costs and effects beyond the length of the trial (Sculpher et al., 2006; Briggs et al., 2008).

A related issue concerns the suitability of common endpoints used in trials for economic evaluations (Ramsey et al., 2015). In general, outcome measures in RCTs reflect some clinically meaningful measure of treatment efficacy. Nowadays, it is standard practice to collect economic data alongside the study however; these economic aspects are commonly 'piggybacked' on and viewed as an afterthought (Glick et al., 2014; Drummond et al., 2015). Owing to the clinical focus of many pharmaceutical trials the outcome measures are commonly intermediate biological markers, e.g. total blood cholesterol. For cost-effectiveness analysis, the link to final health outcomes will have to be quantified, which can be achieved through decision modelling informed by clinical evidence (Drummond et al., 2015).

*Uncertainty*

A fundamental component of economic evaluation for decision-makers is the need to indicate how uncertainty in available evidence translates into decision uncertainty (Briggs et al., 2008). Trial-based economic evaluations can characterise uncertainty, but only regarding the evidence contained in that trial (Sculpher et al., 2006). Failure to include evidence from other sources on relevant parameters and functions of parameters can lead to less precision in parameter estimates, which, all things being equal will result in over-estimates of the expected value of information (Sculpher et al., 2006).

If the decision maker requires an assessment of variability in the results rather than simple mean outputs, stochastic models may be necessary instead of deterministic models (Brennan et al., 2006). NICE now expects an analysis to contain probabilistic sensitivity analysis (PSA) to allow quantification of the uncertainty in mean outputs owing to parameter uncertainty (Brennan et al., 2006). Probabilistic decision-analytic models can present this uncertainty to decision makers at a minimum by facilitating the production of cost-effectiveness acceptability curves (CEACs), but the probabilistic output also facilitates value of information (VoI) analysis (Sculpher et al., 2006). VOI is essentially a quantitative method of assessing the marginal cost and marginal value of further studies, and further translate it into information about the optimal design of additional research (Briggs et al., 2008; Wilson, 2015).

In summary, a trial-based evaluation can provide valuable evidence relating to the disease and technology of concern, the first stage of the evaluative process (2006). However, in matters relating to syntheses, comparators, time horizons, and uncertainty, the trial-based economic evaluation will usually not provide an adequate basis to inform decision-makers charged with regulatory and reimbursement responsibilities (Briggs et al., 2008; Petrou and Gray, 2011). Evidence from RCTs should sit within a second evaluative phase, this broad

framework of evidence synthesis focussed upon decision-makers objectives (Sculpher et al., 2006). This second phase is the essence of the role of decision-analytic modelling; firstly to synthesise evidence for the production of robust estimates of the relative cost-effectiveness of specific healthcare options, and then to quantify the variability and uncertainty associated with decision options (Briggs et al., 2008).

#### 3.2.3.3 Decision-analytic modelling

Considered from the perspective of economic evaluation, a decision analytic model applies mathematical relationships to compare expected costs and consequences of decision options over time, by synthesising information from multiple sources (Raiffa, 1968; Barton et al., 2004; Brennan et al., 2006). Over the past two decades there has been an increased prominence for decision modelling in economic evaluation (Weinstein, 2006; Sculpher et al., 2006; Briggs et al., 2008; Petrou & Gray, 2011; Drummond et al. 2015). For example, in guidance over how to conduct methods of technology appraisal, NICE (2013) explicitly state that "most" technology appraisals require decision analytic models, for the very reasons discussed above.

## 3.3 Types of decision-analytic models

Given the critical role of decision-analytic modelling in economic evaluation, attention now turns to consider which model is appropriate in what contexts.

The responsibility of specifying model structure usually lies with the economic analyst rather than the decision-maker, this is a significant responsibility given the influence of the various assumptions of each model type upon a model's results (Brennan et al., 2006). However, to-date few published model-based economic evaluations attempt to justify their model choice, and guidance on the choice of model structure to assist the analyst in choosing an appropriate model is insufficient (Brennan et al., 2006; Peñaloza Ramos et al., 2015). The papers used in the following section, whilst containing important overarching

points, were mostly drawn from early 2000-2012 when computer processing limited the application of more complex modelling approaches to some degree.

Existing literature advises that model structure should be consistent with a coherent theory of the health condition under evaluation (Philips et al., 2006). Whilst data availability may limit or refine the model's structure, data availability should not be an overriding factor in the development of the structure of a model (Philips et al., 2006). Briggs et al. (2008) offers four critical considerations to make when structuring the model; (i) Have health-related events occurring over time been considered? (ii) Do event risks change over time? (iii) Does intervention effectiveness change over time? (iv) Is the probability of health-related events over time dependent upon patient history?

The consensus across existing guidance favours the simplest model given the study objectives and natural history of the disease and treatment pathways (Brennan et al., 2006; Philips et al., 2006; Sculpher et al., 2000; Karnon, 2003; Barton et al., 2004; Koopman et al., 2001). The discussion below develops some of the most important themes identified in these papers. The discussion is structured around Figure 3.1, a reworking of Barton et al's (2004) flowchart, which the authors developed to assist researchers in identifying appropriate modelling techniques for health economic evaluations.

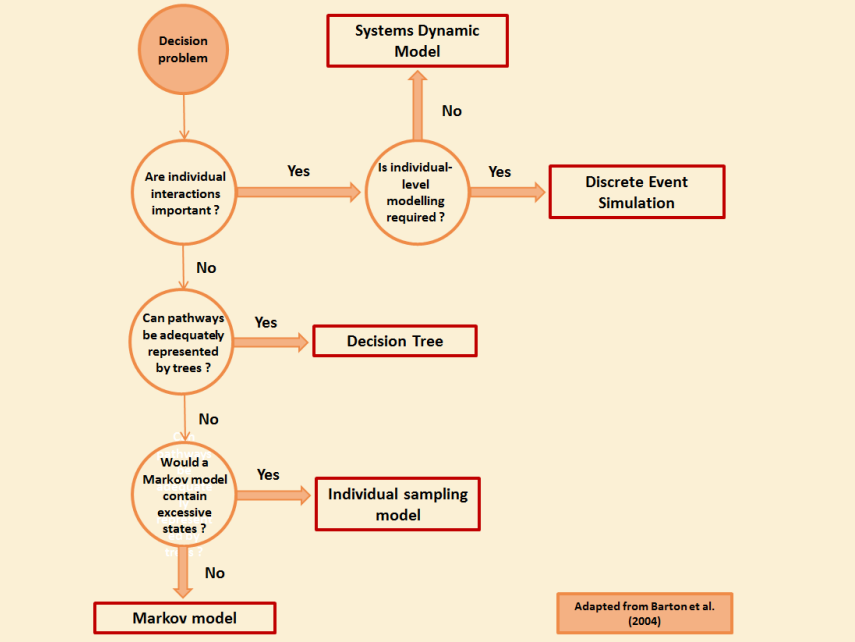



**Figure 3-1 Flowchart to help identify appropriate decision-analytic model**

According to Barton et al. (2004), the first consideration in choosing model structure should include whether individual interactions are important.

### 3.3.1 Models without interaction

Most modelling for health economic evaluations does not involve representing interaction between individuals (Brennan et al., 2006). Interaction, in the context of modelling, is conceptually distinct from interactions in statistics. A modelling interaction refers to the fact that in some health conditions when one has the health state, they are likely to pass it on to another - the concept of an "interaction" is therefore fundamental to infectious disease modelling. Traditionally, on the assumption of independence between individuals, health economists have used state transition models, decision trees and/or Markov models, to perform their modelling (Brennan et al., 2006). Indeed ISPOR guidance (Roberts et al., 2012; Siebert et al., 2012) states that a state transition model is the reasonable choice of model, so long as using states for the decision problem is logical, individual interactions are irrelevant, and the population is a closed cohort. A cohort model sees groups of

individuals with a particular health condition followed over time with no new individuals added to the group as time progresses.

#### 3.3.1.1 Cohort models without interaction

Where a cohort model is appropriate and interaction is not required, the analyst should first consider whether a decision tree could adequately represent patient pathways (Barton et al., 2004).

***Decision trees***

If the decision problem is short term, using a decision tree is generally sufficient (Barton et al., 2004). The value of the decision trees lies in their simplicity, transparency, and ability to clarify options of interest (Petrou and Gray, 2011). The simplicity of a decision tree is its key advantage, which is easier to develop and understand and thus easier to validate (Barton et al., 2004). Simple models that still accurately reflect disease progression and healthcare delivery to the extent needed by a given decision problem are to be ultimately desired (Barton et al., 2004; Brennan et al., 2006).

The basis of the decision tree is the representation of alternative treatment pathways displayed explicitly by various branches of the decision-tree, as demonstrated in figure 3.2 in relation to Sciatica. The decision problem here relates to which strategy is most cost effective to manage patients with sciatica.

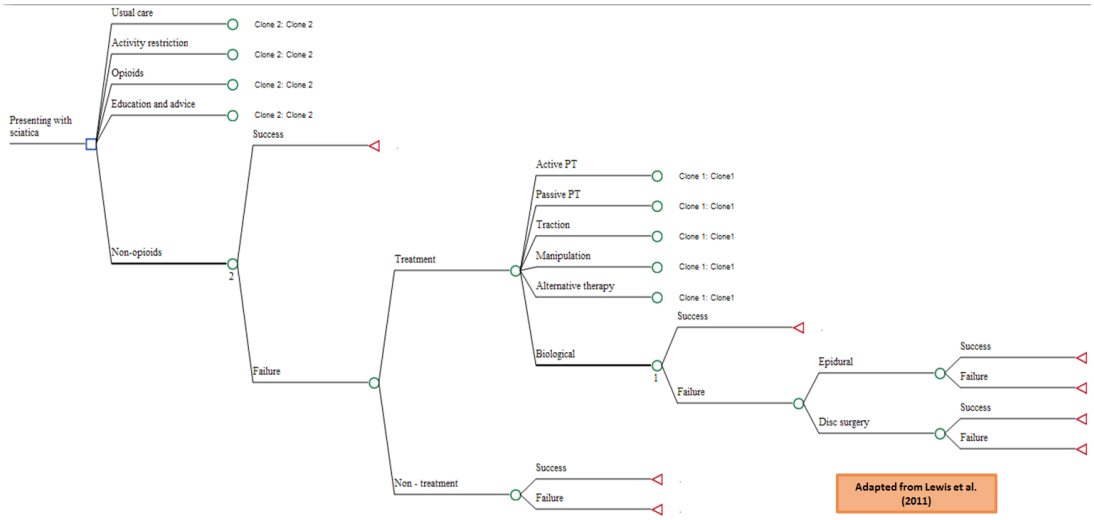



**Figure 3-2 Decision Tree schematic of economic evaluation to identify cost-effectiveness of treatment combinations for sciatica (Lewis et al., 2011)**

The tree begins with a decision node (denoted by a square), from which the five treatment options follow, usual care, activity restriction, opioids, education and advice, and non-opioids. The decision node is always structured according to the nature of the decision (Cooper et al., 2006). The pathways, which follow from each option, represent a series of logically ordered alternative events, denoted by branches emanating from further decision nodes (denoted by the black circular symbols). This decision tree shows only the branches for the non-opioids option, but the denotation of (+) following the other four treatment options is modelling shorthand to tell us the structure we see emanating from non-opioids will repeat for each of the other four treatment options.

The alternatives at each chance node must be mutually exclusive and their probabilities should sum exactly to one (Barton et al., 2004). The endpoints of each pathway are denoted by terminal nodes (triangular symbols) to which values or pay-offs, such as costs, life years, or quality-adjusted life years (QALYs), are assigned (Brennan et al., 2006). In this decision tree, we can see that all successes are terminal nodes; failure only has a terminal node attached if a patient reaches a third treatment, which can only be disc

surgery or epidural. Once the probabilities and pay-offs are entered, the decision tree is "averaged out" and rolled back, allowing the expected values of each option to be calculated (Briggs et al., 2008). In the example above, success and failure is assigned a quality of life score independent of treatment; therefore the most effective interventions will be those who maximise the success whilst minimising failures.

There are several limitations to using decision trees. A major limitation is the lack of explicit time variable, making it difficult to deal with time-dependent elements of an economic evaluation (Drummond et al., 2015). Second, decision trees are not suitable to model recursion or looping. Attempting to incorporate recurring events to represent chronic diseases, for example, can cause decision trees to become complex with numerous lengthy pathways (Petrou and Gray, 2011). Moreover, given that the time horizon of the model should extend far enough into the future to reflect all crucial differences between the strategies under evaluation (Philips et al., 2006), decision trees will often not allow the analyst to capture intervention effects into the future. Where these issues are present, analysts will commonly select a Markov model over a decision-tree (Brennan et al., 2006; Karnon and Brown, 1998).

***Markov models***

Markov models alone or in combination with decision trees are the most common models used in economic evaluations. There are three crucial advantages of using a Markov model in relation to a decision tree. Firstly, they permit recurrence (a Markov model is essentially a recursive decision-tree), secondly, they allow for patient progression through the model to be time-dependent, and thirdly they allow modelling of chronic diseases over the lifetime (Briggs et al., 2008; Petrou and Gray, 2011). Relative to sampling models, cohort models may be preferred if the decision problem can be adequately captured within the

cohort model, this is because of the transparency, efficiency, computational expense, ease of debugging, and ability to conduct value of information analysis (Roberts et al., 2012).

In a Markov model, a homogenous cohort of patients reside in one of a finite number of health states at any point in time, and patients transition between those health states over a series of discrete time cycles (Cooper et al., 2006; Drummond et al., 2015). The probability that a patient remains in a state or moves to another in each cycle depends upon a set of defined transition probabilities (Petrou and Gray, 2011). The transition probability depends only on the state in which the patient is at the start of the cycle - a major assumption within Markov models, known as the Markovian assumption.

Patients in any given state may only complete a single transition per cycle (Cooper et al., 2006). The nature and quantity of health states and the duration of these cycles usually relate to the decision problem and condition-specific health processes (Petrou and Gray, 2011). Guidelines suggest that states reflect clinical classifications of disease, remaining life expectancy, and allow transitions to occur that are consistent with the clinical problem and intervention effects (Philips et al., 2006; Briggs et al., 2008; Roberts et al., 2012).

For example, the Markov model shown in Figure 3.3 shows a state transition diagram used for assessing the cost-effectiveness of acupuncture for treating patients with chronic LBP; accordingly the health states are temporal classifications of low back pain.

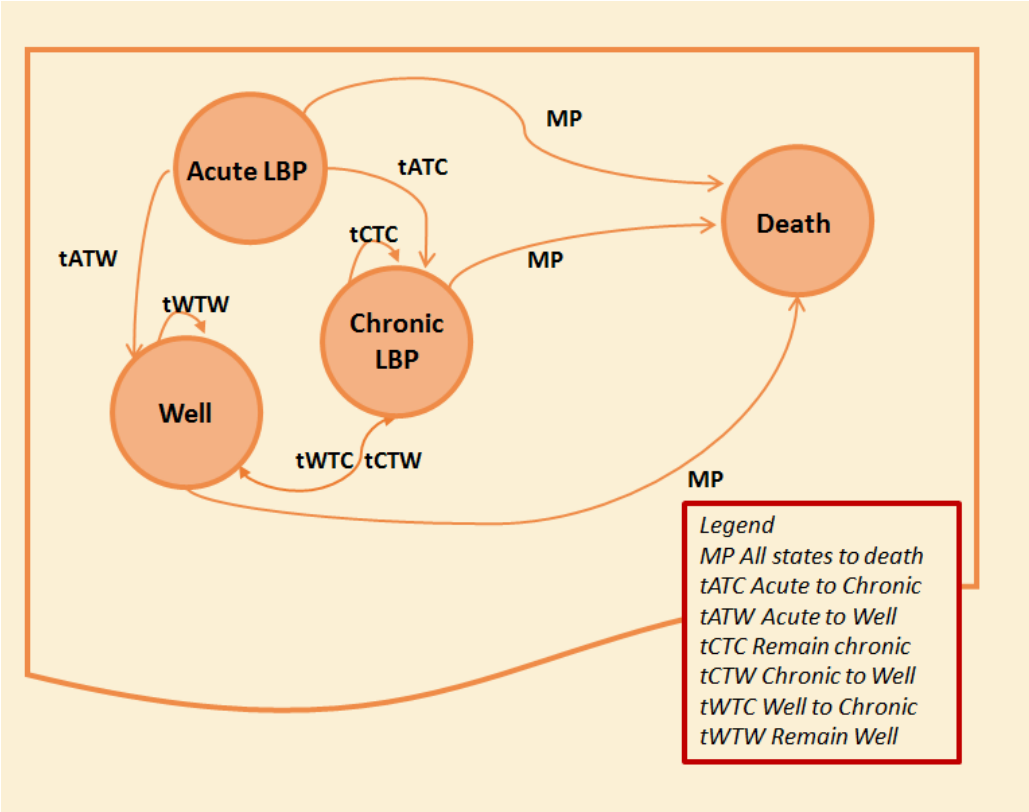



**Figure 3-3 A Markov model for low back pain by Kim et al. (2010)**

Here, we can see that patients will occupy one of four health states, Acute LBP, Chronic LBP, Well, and Death. For each cycle, patients may move between the states as designated in the transition matrix above, e.g. the movement from Chronic LBP to Well is represented by the probability tCTW. Note that movements from each state at the end of each cycle must sum to 1. In this case, movements from aLBP to other health states, denoted by tATC, tATW, and MP, must always sum to 1.

In order to end the Markov process, some condition must be set. The model could end either within a set number of cycles, a proportion of patients accumulating or passing through a specified state, or the entire population reaching an absorbing state that cannot be left (e.g. dead) (Petrou and Gray, 2011). The calculation of expected costs and outcomes involves summing each cost and outcome for each cycle across each health state, and

weighting by the proportion of the cohort expected to be in each state (Briggs et al., 2008). If the model time horizon exceeds one year, then discounting is required to estimate expected costs and outcomes in terms of their present value (Petrou and Gray, 2011). Discounting in this context relates to the need to discount future costs and benefits to presents costs, and is performed to account predominantly for differences in how future benefits and costs are valued in the present (Drummond et al. 20015).

In some situations, it may become appropriate to combine the decision-tree with a Markov model, which can be useful where some initial or short-term event affects the proportion of the patient cohort passing into some health state at the end of the treatment or event (Drummond et al., 2015). The Markov model would then estimate the expected quality-adjusted survival duration and lifetime cost conditional on the patient's status at the end of the decision-tree (Briggs et al., 2008). This could be an efficient and transparent way to model diagnostic test strategies and short-term treatments.

The considerations raised by the flowchart of Barton et al. (2004) show the major distinctions that an analyst must take when selecting a model. However, there are a number of other considerations which an analyst must make. For example, a modeller must decide whether parameter values in their models are to be mean values, or whether this could be restrictive and the aim may be to give representation to randomness and patient heterogeneity (Brennan et al., 2006). Markov models can allow for both deterministic models by using expected values or can take a stochastic form by using simulated random transitions (Brennan et al., 2006).

The primary limitation of Markov models however, is the Markovian assumption that transition probabilities depend only on the current health state, independent of and thus ignoring time spent in a given state or previous patient history (Barton et al., 2004; Petrou and Gray, 2011). The Markovian assumption can be severely limiting where these aspects

are strong determinants of future disease progression, such as where a disease has important comorbidities (Philips et al., 2006).

Solutions to overcoming this limitation could be to introduce a number of non-temporary health states so called “tunnel states”, or to introduce temporary states for which patients may only enter for one cycle, or have a series of temporary states visited in a fixed sequence (Gray et al., 2010). However, whilst either approach is feasible in a simple model, it is easy to see how modelling increasing number of ‘adverse events’ and ‘post-adverse event’ states will require an ever-increasing number of states, potentially creating models which lack transparency and are slow to run (Brennan et al., 2006).

#### 3.3.1.2 Individual models without interaction

Decision trees and Markov models track patient cohorts in progressing through the states simultaneously; however, the analyst may wish to distinguish one individual from another. If it is impossible to represent the decision problem in a manageable quantity of health states such that fundamental characteristics of the decision problem are captured, then the cohort state-transition model may need to be rejected.

***Individual Sampling Models***

As stated in the previous section, to modify a Markov model to overcome the Markovian assumption, an ever-increasing number of states are required. An alternate solution is to use an individual sampling model (ISM) for which the Markovian assumption either does not apply or can be relaxed. An ISM minimises the number of states by varying transition probabilities according to a patient’s heterogeneous attributes (Barton et al., 2004). As a result, individual patients progress through such models according to their multiple risk factors and patient histories, which may also evolve over time (Weinstein, 2006). Applying Monte Carlo simulation generates individual transitions partly dependent upon random number generation (Roberts et al., 2012). As patients progress through the model, an

accumulated history of transitions, costs, and health outcomes is produced. This feature allows the ISM to model dynamic intervention strategies, where future decisions depend on present and/or historical patient characteristics, whilst updating patient characteristics in the process (Brennan et al., 2006).

Unlike Markov models, simulation models can also simulate the patient's time-to-next-event, as well as generating a duration for which the patient spends in the given state (Brennan et al., 2006). In a discrete event simulation model patients can, occupy given states for a variable time-period and experience multiple events in parallel (Brennan et al., 2006). Individual simulation models can also model individual characteristics as continuous variables, whereas Markov models or decision trees categorise continuous variables. Indeed the flexibility of the ISM in relation to a Markov model has led to an increase in the use of simulation of individual patients in economic evaluation, especially as computing capabilities have improved (Brennan et al., 2006).

There are several limitations associated with the use of ISMs. For example, ISMs require additional evidence to populate models due to the increase in the potential number of parameters. The major problem with the individual sampling model is, however, that they require time-consuming replications to arrive at stable estimates of outcomes of interest, a time which will only increase with the size of the population modelled (Jaime-Caro et al., 2012; Brennan et al., 2006). Computation time may become even more problematic if probabilistic sensitivity analyses or value-of-information analyses are required (Siebert et al., 2012), although hardware and software innovations are improving computation times.

### 3.3.2 Models with interaction

Interactions are particularly significant where modelling infectious diseases because the risk to the individual of becoming infected depends upon how many others already have the disease (Brennan et al., 2006). Modelling interactions can also more accurately

represent situations where treatment choices made for one patient affect that which is available for another (Barton et al., 2004). Where interaction is important, discrete event simulation (individual level) and system dynamics models (cohort) should be used.

### 3.3.3 Assessment of uncertainty

Having considered why decision analytic models are employed, as well as the various model options, it is worth revisiting the idea of uncertainty. As noted, uncertainty analysis forms a vital component of any informative economic evaluation and will be fundamental to the analyses performed in this thesis. Uncertainty is endemic in economic evaluation because it is impossible to predict with total certainty what costs and effects of a particular treatment will be (Bojke et al., 2009). Modellers should distinguish between the four principal types of uncertainty - methodological, structural, heterogeneity, and parameter (Philips et al., 2006). Each must be differentiated, assessed and commented upon within an evaluation.

#### 3.3.3.1 Methodological

Methodological uncertainty relates to whether particular analytic steps taken in the analysis are most appropriate, for example, the discount rate that has been used. Methodological uncertainty can be addressed by running alternative versions of the model with different methodological assumptions.

#### 3.3.3.2 Structural

There is no standard definition of structural uncertainty, perhaps reflecting its recent identification. To differentiate from the methodological uncertainty, Ghabri et al. (2016) suggest structural uncertainty relates to uncertainty regarding model structure and the fundamental assumptions contained within. Meanwhile, Bojke et al. (2009) suggest structural uncertainty reflects the uncertainty arising from various simplifications and scientific judgments which form part of the process of constructing a decision model. On

these understandings, structural uncertainty may consider issues such as choice of health states, how those states are defined, assumptions about possible movements between those states, or perhaps even the assumptions one makes regarding how patients move over time.

Similar to the manner with which methodological uncertainty is treated, structural uncertainties should be addressed as sensitivity analyses using alternative structural assumptions. This would involve re-running the model with a series of alternative extrapolation techniques and presenting the results under each scenario. For example, sensitivity analysis could consider the effect of changing the values used to extrapolate long-term effects, or changing the choice of health states.

#### 3.3.3.3 Heterogeneity

It is essential to distinguish between uncertainty resulting from the process of sampling from a population, and variability due to the heterogeneity arising from systematic differences between population subgroups (Philips et al., 2006). In the case of heterogeneity, it is advisable to run the model separately for different subgroups (Philips et al., 2006).

#### 3.3.3.4 Parameter

Parameter uncertainty reflects the uncertainty over the true values of parameter estimates in any given model (Briggs et al., 2008). There is disagreement about how to appropriately explore parameter uncertainty, sometimes called second-order uncertainty (e.g. Sculpher et al., 2000; Soto et al., 2002). Philips et al. (2006) discuss these debates, and settle on the following recommendations;

Where the incorporation of data into models takes the form of point estimates - univariate or multivariate sensitivity analysis may be undertaken to explore parameter uncertainty, with ranges used in sensitivity analysis stated and justified. However, standard approaches, such as varying one or multiple parameters simultaneously to assess the impact upon

results are limiting because choosing the parameters to vary and by how much is essentially arbitrary. Philips et al. (2006) endorse the position of Briggs (2000) that probabilistic sensitivity analysis is far more appropriate in handling parameter uncertainty because it can evaluate the combined impact of uncertainty across all parameters. Moreover, it allows for value-of-information analysis to be run, which can explicitly assess the effects of decision uncertainty.

Parameter variability is differentiated from parameter uncertainty. Variability reflects situations where the parameter mean could systematically vary in relation to the patient characteristics or treatment location. This variability should be assessed using standard sensitivity or scenario analysis, where the value of the given parameter has its mean value and distribution re-specified. As with structural uncertainty, the analysis should be re-run with newly specified parameters (Philips et al., 2006).

### 3.4 Critical appraisal of modelling quality

Finally, it is worth noting that there is significant variation in both the conduct and reporting of health economic analyses (Neumann 2005). Several authors have written on the subject of the limitations and restrictions of using decision modelling in health economic analysis (Sheldon, 1996; Buxton, 1997; Kuntz and Weinstein, 2001). There are particular concerns regarding model structures, the data chosen to populate the model, and transparency. As a result, rather than ignoring modelling, methodological guidelines were created to guide the analyst in generating models which have transparent methods, assumptions and data, all of which reflect available evidence (e.g. McCabe and Dixon, 2000; Sculpher et al., 2000).

Philips et al. (2006) have reviewed and contrasted these recommendations to develop a framework for systematic assessment of the quality of decision-analytic models, commonly referred to as “the Philips checklist”. The checklist develops three key themes;

'Structural' which relates to the scope and mathematical structure of the model; 'Data' which concerns issues pertaining to data identification and appropriately addressing uncertainty; and finally 'Consistency' which refers to the overall quality of the model. This checklist has been included in Appendix 1.

## 3.5 Summary

The purpose of this chapter was to provide an overview of the use of economic evaluation and decision-analytic models specifically. The chapter has explored the philosophical origins of economic evaluation, and how that manifests in the practical analytic techniques available to the economic analyst. The necessity of using decision-analytical models in economic evaluation was explored with reference to their advantages over trial-based analyses. The specific types of decision model available to the analyst were outlined, as well as particular situations conducive to their use. The next chapter systematically reviews decision analytic models of LBP and sciatica.

# Chapter 4: SYSTEMATIC REVIEW OF DECISION ANALYTIC MODELLING IN ECONOMIC EVALUATIONS FOR LOW BACK PAIN AND SCIATICA

## 4.1 Introduction

The preceding two chapters provided an overview of both the clinical area and decision analytic modelling in economic evaluation. In what follows, the focus turns towards the practice of decision analytic modelling, specifically, within LBP and sciatica. The purpose of this review chapter is to inform the model development processes detailed in the empirical chapters.

The review has four specific objectives:

1. Document and classify existing model-based economic evaluations for treatment and management of LBP and sciatica.

2. Critically appraise current modelling techniques, analytical methods, data inputs, and model structure, using narrative synthesis.

3. Identify examples of good modelling practice.

4. Identify currently unresolved methodological problems and gaps in the literature.

Particular deliberation focusses upon model choice and representation of the condition within health states, origins of key parameter values, and methods used for extrapolation of treatment effect, resource use and utility values.

The results presented here have been previously given in the following peer reviewed paper, Hall J, Konstantinou K, Lewis M, Oppong R, Ogollah R, Jowett S (2019) Systematic review of decision analytic modelling in economic evaluations of low back

pain and sciatica Applied Health Economics and Health Policy, doi: 10.1007/s40258-019-00471-w.

## 4.2 Methods

### 4.2.1 Search strategy

Systematic literature searches were conducted, according to a pre-specified protocol, in order to identify economic evaluations of interventions for treatment or management of low back pain and sciatica. The protocol for this systematic review was developed using the PRISMA-P checklist (http://www.prisma-statement.org/Extensions/Protocols.aspx). Articles were identified using database searches with studies subsequently identified by reference searching also included.

Searching took place in the following health databases: OVID INTERFACE (MEDLINE, EMBASE, PsychINFO), EBSCO INTERFACE (Cumulative Index to Nursing and Allied Health Literature (CINAHL), Allied and Complimentary Medicine Database (AMED), EconLit), COCHRANE LIBRARY INTERFACE (Database of Abstracts of Reviews of Effects (DARE), Cochrane Database of Systematic Reviews (CDSR), Health Technology Assessment (HTA) Database, NHS Economic Evaluation Database (NHS EED)), and THOMSON REUTERS INTERFACE (Web Of Science).

In developing the search strategy, economic terms were based upon a strategy developed by the NHS Centre for Reviews and Dissemination at the University of York, itself based upon the SIGN (Scottish Intercollegiate Guidelines Network) approach. Clinical terms reflected strategies taken by other systematic reviews of economic evaluations in LBP and sciatica (Andronis et al., 2017; Lewis et al., 2011). The search strategy was subsequently refined with the assistance of the supervisory team (SJ, KK, RaO, ReO) with input from Dr. Nadia Corp, Research Associate: Systematic Reviews at Keele University. Search terms used for 3 databases (MEDLINE, PsychINFO, NHS EED) are included in Appendix

2. Initial database searching took place during January 2017 with the review updated in February 2019.

### 4.2.2 Inclusion and exclusion criteria

The protocol specified that studies would be included in this review if undertaking cost-effectiveness, cost-consequence, cost-benefit or cost-utility analyses. Studies could consider any treatment or management approach for patients with LBP and sciatica, providing the study stipulated the use of decision analytic modelling. Reviews would be eligible for inclusion providing the publication also contained a unique economic model.

Inclusion criteria:

1. Model-based economic evaluation studies using cost-effectiveness, cost-consequence, cost-benefit, cost-utility, or cost-minimisation analysis.
2. The study must either have:

   (i) Specified the use of a decision model, a decision tree, a Markov model, an individual sampling (or patient level) model, a Monte Carlo model or simulation, a discrete event simulation, dynamic transition model, or systems dynamic model.

   Or

   (ii) stated the use of a (possibly unspecified) economic model.
3. Studies must consider any treatment for any patient group in any setting for LBP patients, and any sciatica treatment which patients in the SCOPiC trial could viably receive.

Exclusion criteria:

1. Any economic evaluation which does not include decision analytic modelling, e.g. trial-based evaluations.
2. Studies which do not fully report methods, such as conference abstracts or editorials.

3. Studies not in English language.

### 4.2.3 Data selection and extraction

A two-stage exclusion process was employed. The first reviewer (JH) excluded studies in accordance with the protocol exclusion criteria. Due to the volume of studies, 10% of these excluded studies were independently checked by one other reviewer (SJ). Studies considered potentially relevant were then subdivided into "included" and "possible" by the first reviewer (JH). A second reviewer (SJ) checked the suitability of "included" studies, whilst all "possible"' studies were independently reviewed by all four other reviewers (SJ, KK, RaO and ReO) for relevance. The five authors (SJ, KK, RaO, ReO, and JH) then reached consensus regarding the inclusion of the final studies.

Data were extracted according to a pre-specified protocol. Namely, nine tables, each representing different aspects of the economic evaluation and model construction process; (i) summary of studies, (ii) analytical characteristics, (iii) study rationale and results, (iv) model design, (v) model scope, (vi) modelling assumptions, (vii) model inputs, (viii) sensitivity analysis, (ix) validity checks, and (x) quality appraisal. The process of quality appraisal was undertaken using the Philips checklist, with each study scored according to how adequately the study fulfilled each aspect of the checklist. A second reviewer (SJ) checked 25% of the quality appraisal to ensure no major disagreement.

## 4.3 Systematic Review Results

Figure 4.1 illustrates the process of selecting and identifying studies eligible for inclusion in the review.

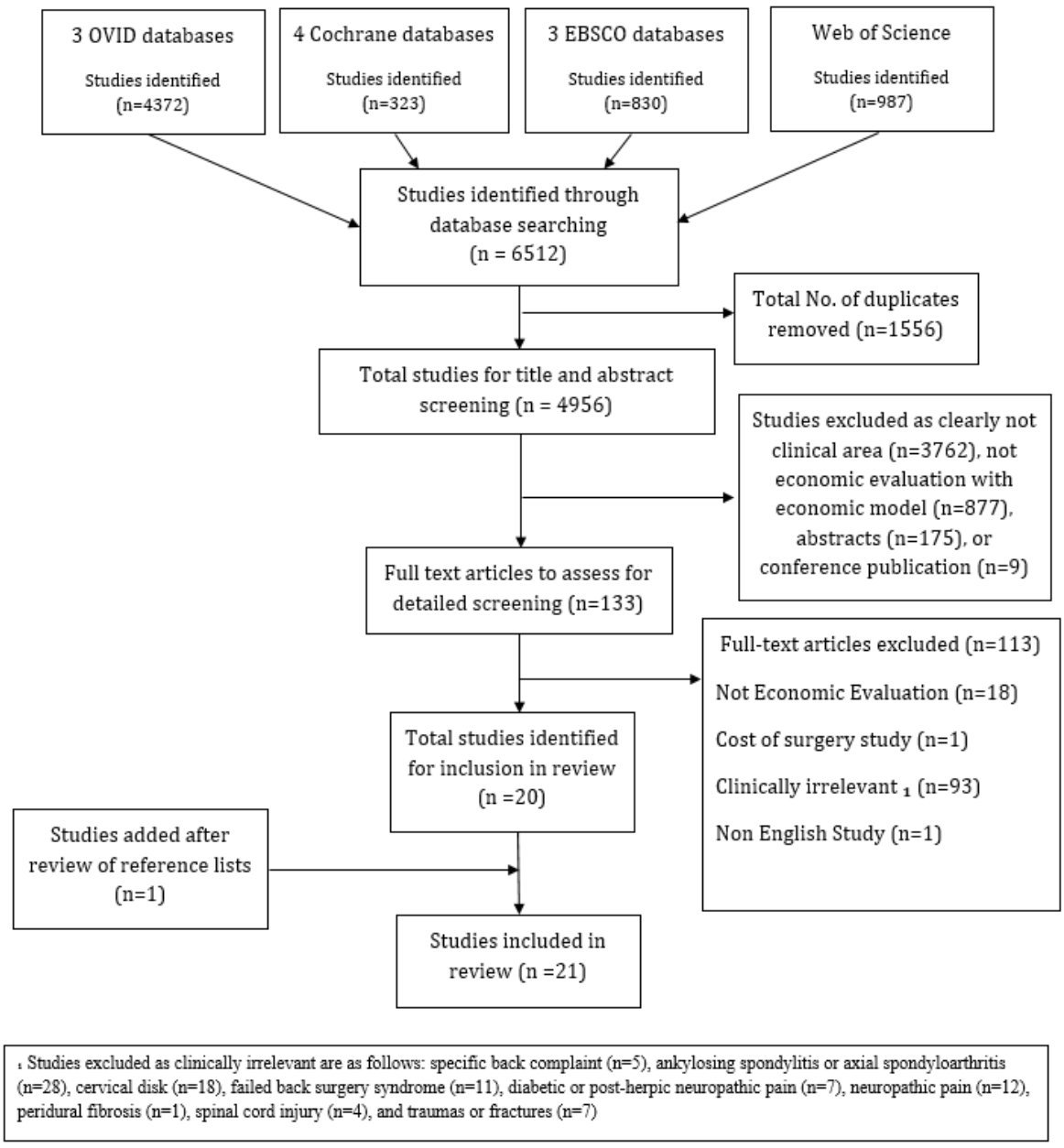



**Figure 4-1 PRISMA flow diagram showing study selection for inclusion in the systematic review**

6512 records were imported into Endnote, of which 1556 were duplicates. Of the 4956 unique studies, 4823 were excluded in accordance with the protocol exclusion criteria by the first reviewer (JH). 3762 studies were excluded because they did not reflect the clinical area, 877 were not economic evaluations with an economic model, and 175 were abstracts and nine conference publications. 10% (or 400) of these excluded studies were independently checked by one other reviewer (SJ). The 133 titles that were considered potentially relevant were then subdivided into "included" and "possible" by the first

reviewer (JH). A second reviewer (SJ) checked the suitability of "included" studies, whilst all "possible"' studies were independently reviewed by all four other reviewers (SJ, KK, RaO and ReO) for relevance. In Figure 4.1, publications deemed "clinically irrelevant" were studies that the first two authors (JH) and (SJ) were not able to determine were sciatica or non-specific LBP, these were excluded following further consultation with the clinician on the team (KK) regarding whether or not the various conditions described in the studies, constituted a diagnosis of sciatica or non-specific LBP. Having reviewed all of the possible studies independently, the five authors (SJ, KK, RaO, ReO, and JH) then reached consensus regarding the inclusion of the final 20 studies, and one additional study (which was referenced in one of the original 20) was also added.

### 4.3.1 Overview of studies

Table 4.1 provides an overview of the studies included in this review. Studies were classified according to whether patients had LBP or sciatica, as each condition requires a potentially different modelling approach. Sciatica studies were further subdivided according to whether they contained a non-surgical comparator, as modelling studies solely for surgical treatments could require a unique structure. Accordingly, tables show (i) studies with treatments for LBP, (ii) studies considering *at least* one non-surgical treatment for sciatica, and (iii) studies evaluating solely surgical treatments for sciatica.

There were only five models for treatment or management of LBP, three for chronic low back pain and two for acute. These studies considered a variety of interventions, including heat wraps, cognitive behavioural therapy (CBT), acupuncture and pharmaceutical treatments. Nine modelling studies (of eight unique models) included a non-surgical treatment option for sciatica; most studies include models where conservative care is used a comparator (n=8). These studies considered various pathologies pertaining to sciatic pain, commonly lumbar disc herniation and lumbar spinal stenosis (LSS). The English

HTA (Health Technology Assessment) model (Lewis et al., 2011; Fitzsimmons et al., 2014) considered the full range of treatment options available for UK patients with sciatica. Finally, seven models solely evaluating surgical treatments for populations with sciatica were included, again featuring various forms of sciatic pain. Studies evaluated either specific instruments/implants used in surgical procedures or different types of surgery.

Nearly all included studies are recent publications, eighteen published after 2010 with only three before 2010. Twelve of the studies were conducted in the US, four in the UK, with the remaining six from Canada, South Korea, France, Japan, Iran and Australia.

**Table 4.1 Overview of studies included in this review**

| Author | Country | Condition and Population | Intervention | Comparator(s) |
|---|---|---|---|---|
| **Low back pain decision modelling studies** | | | | |
| Lloyd et al. (2004) | U.K | Adult patients, with acute nonspecific LBP | Heat wrap therapy (ThermaCare® Procter and Gamble Ltd.). | Paracetamol, Ibuprofen |
| Kim et al. (2010) | South Korea | Cohort of 60-year old females with CLBP | Acupuncture plus routine care options | Routine care (NSAIDs, heat therapy, electrotherapy and lumbar traction) |
| Wielage et al. (2013a) | U.S.A | CLBP patients | Duloxetine | Celecoxib, Naproxen, Pregabalin, Oxycodone APAP, Oxycodone ER, Tapentadol, Tramadol |
| Wielage et al. (2013b) | Canada | CLBP patients | Duloxetine | Celecoxib, Naproxen, Pregabalin, Hydromorphone, Oxycodone ER, Amitriptyline |
| Norton et al. (2015) | U.S.A | Adult CLBP patients | CBT with educational materials | Educational materials on managing pain, activity and symptoms. |
| **Sciatica decision modelling studies** | | | | |
| Launois et al. (1994) | France | Patients with radicular pain caused by lumbar disc herniation | Chemonucleolysis | Surgical discectomy |

| Author | Country | Condition and Population | Intervention | Comparator(s) |
|---|---|---|---|---|
| Lewis et al. (2011) | U.K | Patients presenting with sciatica | Full range of Sciatica treatments used in the U.K | |
| Skidmore et al. (2011) | U.S.A | Patients at least 50-years old with moderately impaired LSS | Decompression using the X-STOP ® Interspinous Spacer | Conservative care (epidural, supplemented by NSAIDS, oral steroids, physical therapy, or spinal manipulation).<br>Laminectomy |
| Fitzsimmons et al., (2014) | U.K | See Lewis et al. (2011) | | |
| Koenig et al. (2014) | U.S.A | Patients of various age cohorts, with lumbar disc herniation | Lumbar discectomy | Non-surgical treatments |
| Udeh et al. (2015) | U.S.A | Patients with moderate to severe LSS who failed conservative therapy. | *mild*®, ESI or laminectomy surgery | Standard treatment for LSS patients after failure of conservative therapy |
| Igarashi et al. (2015) | Japan | Patients with moderate or severe LBP alongside neuropathic pain | Pregabalin | Usual care (standard analgesic) |

| Author | Country | Condition and Population | Intervention | Comparator(s) |
|---|---|---|---|---|
| Parker et al. (2015) | U.S.A | Patients with diagnosis of LSS who have completed six months of conservative treatment. | Minimally-invasive interspinous spacer. | CC comprised of (physical therapy, NSAIDs, mild opioids, and epidural injections).<br>DS |
| Tapp et al. (2018) | U. S. A | Patients with LSS with no previous surgery | Minimally-invasive interspinous spacer | CC<br>DS |
| **Sciatica decision modelling studies – surgical treatments** | | | | |
| Kuntz et al. (2000) | U.S.A | Patients with degenerative lumbar spondylolisthesis and LSS | Non-instrumented fusion and instrumented fusion | Laminectomy without fusion |
| Kim et al. (2012) | U.S.A | Patients with LSS who failed conservative treatment | Lumbar decompression without fusion | Lumbar decompression with fusion |
| Parkinson et al. (2012) | Australia | Patients with axial back pain and/or radicular pain who failed conservative treatment | Lumbar AIDR | Lumbar fusion.<br>Anterior lumbar interbody fusion.<br>PLF |

| Author | Country | Condition and Population | Intervention | Comparator(s) |
|---|---|---|---|---|
| Schmier et al. (2014) | U.S.A | Patients with moderate to severe sciatica and low-grade generative spondylolisthesis | Coflex® interlaminar stabilization inserted following decompressive surgical laminotomy | Instrumented posterolateral lumbar fusion |
| Bydon et al. (2015) | U.S.A | Patients with degenerative spondylolisthesis | Posterior lumbar interbody fusion or transforaminal lumbar interbody fusion | Non-interbody fusion and posterolateral fusion |
| Vertuani et al. (2015) | U.K and Italy | Patients with degenerative lumbar spinal conditions | MIS | Open Surgery |
| Yaghoubi et al. (2016) | Iran | Patients requiring surgery for treatment of LSS | Dynamic Interspinous Spacer (Coflex®) and Static Spacer (X-STOP ®) | Laminectomy |

Abbreviations: AIDR (Artificial intervertebral disc replacement); CBT (Cognitive behavioural therapy); CC (Conservative care); CLBP (Chronic low back pain); DS (Decompression surgery); ESI (Epidural steroid injections); HRQoL (Health Related Quality of Life); ICER (Incremental cost-effectiveness ratio); LBP (Low back pain); LSS (Lumbar spinal stenosis); mild® (Minimally invasive lumbar decompression); MIS (Minimally invasive surgery); NHS (National Health Service); NSAID (Nonsteroidal anti-inflammatory drug); PLF (Posterolateral fusion); QALY (Quality adjusted life year).

### 4.3.2 Purpose and Results of the Models

Of the 21 studies included, 18 were applied cost-effectiveness or cost-utility studies with the explicit purpose of comparing a variety of treatment options. Three studies took a methodological angle; Koenig et al. (2014) reviewed the impact of incorporating productivity costs into an economic evaluation for lumbar discectomy; Tapp et al. (2018) sought to understand factors affecting the long-term cost-effectiveness of interspinous spacer devices; and Parkinson et al. (2012) reviewed the extent to which the cost-effectiveness of lumbar artificial intervertebral disc replacement was driven by choice of comparator.

The cost-effectiveness results can be summarised as follows. For episodes of acute LBP, heat wrap dominates paracetamol and ibuprofen (Lloyd et al., 2004). In Korea, acupuncture appears to be a cost-effective means of managing chronic LBP, relative to routine care (Kim et al., 2010). Duloxetine is suggested to be a cost-effective means of managing chronic LBP pain in the U.S and Canada, and dominates most other pharmacological comparators, although it is unclear whether or not it is cost-effective relative to Naproxen (although there are further author correspondences regarding the modelling of tapentadol) (Wielage et al., 2013a; 2013b). CBT (Cognitive Behavioural Therapy) is cost-effective relative to educational materials in treating chronic LBP in Canada (Norton et al., 2015). Societal analyses considerably decreased the cost per QALY for the most effective treatments (i.e. made the intervention more cost-effective) (Kim et al., 2010; Wielage et al., 2013b).

In sciatica, a comparison of two pharmacological approaches for managing patients with severe LBP pain alongside a neuropathic component showed that pregabalin was cost-effective compared to usual care treatment with standard analgesia (Igarashi et al., 2015). All other sciatica studies considered some form of surgery as a comparator, with the length

of time that patients spend receiving conservative care, and ordering of treatments, seeming to determine the cost-effectiveness of surgery. For example, in a review of over 100 different potential treatment combinations for patients presenting with sciatica, Lewis et al. (2011) and Fitzsimmons et al. (2014) show that in the UK, stepped care approaches based on initial treatment with non-opioids are the most cost-effective, whilst referring patients who fail an initial treatment to surgery, is unlikely to be cost-effective. However, evidence from the U.S suggests that surgery after one course of failed treatment or extended duration of symptoms could be cost-effective. In one study of patients with moderate to severe lumbar spinal stenosis (LSS), who failed conservative therapy, minimally invasive lumbar decompression maybe cost-effective ($43,760 per QALY) relative to standard non-surgical treatment (Udeh et al., 2015). Compared with ongoing conservative care, Skidmore et al. (2011) show that for moderately impaired patients with LSS, decompression surgery using a spacer is cost-effective relative to non-surgical care and dominates laminectomy. Amongst LSS patients complete six months of conservative treatment without improvement, Parker et al. (2015) show that the minimally invasive interspinous spacer and decompression surgery are highly cost-effective relative to conservative care at $16,300 per QALY and $15,200 per QALY respectively. Tapp et al. (2018) show that over ten years, for patients with LSS and no previous surgery, minimally invasive procedures using a spacer and decompression are cost-effective relative to usual care at $25,000 per QALY and $30,874 per QALY respectively. Although the evidence on surgery for patients with lumbar disc herniation suffering functional limitations, suggests discectomy may not be cost-effective relative to non-surgical care ($52,416 per QALY), although the inclusion of societal costs decreases the incremental cost-effectiveness ratio (ICER) to $35,146 per QALY (Koenig et al., 2014).

Of the papers which evaluate solely surgical techniques, three U.S studies suggest that various spinal fusion techniques are not cost-effective relative to other surgical techniques

without fusion. Kuntz et al. (2000) found laminectomy without fusion more cost-effective, Kim et al. (2012) lumbar decompression without fusion (Kim et al., 2012), and Schmier at al. (2014) Coflex® interlaminar stabilization following decompressive laminotomy. Furthermore, evidence from Iran supports the cost-effectiveness of the Coflex® relative to the X-Stop and laminectomy (Yaghoubi et al., 2016). In studies evaluating only means of performing fusions, for U.S patients with degenerative spondylolisthesis, interbody fusions were found cost-effective relative to non-interbody fusion, at $9,883.97 per QALY (Bydon et al., 2015). Amongst Australian patients with radicular pain who failed conservative treatment, posterolateral fusion was deemed to be the most cost-effective surgical approach for lumbar fusion or artificial intervertebral disc replacement (AIDR) (Parkinson et al., 2012). Finally, in the UK and Italy, minimally invasive surgery dominated open surgery for one- or two-level lumbar spinal fusion in the treatment of degenerative lumbar spinal conditions (Vertuani et al., 2015).

**Table 4.2 Study purpose and findings**

| Low back pain decision modelling studies | | | |
|---|---|---|---|
| Author | Sponsor | Purpose of Model | ICER Value / Main findings |
| Lloyd et al. (2004) | Proctor and Gamble Health Sciences Institute | Establish cost-effectiveness of treating an episode of LBP with heat wrap | Heat wrap dominates comparator, at £48.72 per successfully treated patient. Paracetamol next best at £131.63 per successful patient |
| Kim et al. (2010) | None | Establish cost-effectiveness of usual care with acupuncture compared to usual care alone | Acupuncture cost-effective vs routine care (ICER 3,421,394 KRW/ QALY). Inclusion of indirect costs lowers ICER to 1,349,463 KRW/ QALY |
| Wielage et al. (2013a) | Eli Lilly and Company, Indianapolis, USA | Establish cost-effectiveness of Duloxetine in CLBP patients. | Duloxetine cost-effective treatment for LBP compared to all but generic NSAIDS. Duloxetine ICER of $59,473 vs Naproxen. |
| Wielage et al. (2013b) | Eli Lilly and Company, Indianapolis, USA | Establish cost-effectiveness of Duloxetine in CLBP patients | Naproxen least expensive. From societal perspective Celecoxib ICER of $19,881/QALY, and duloxetine ICER $43,437/QALY relative to Naproxen. Others dominated. |
| Norton et al. (2015) | None | Evaluate the cost-effectiveness of CBT for treatment of persistent CLBP patients | ICER of $5855/QALY for CBT vs advice alone at ten years. |

| Sciatica decision modelling studies – non-surgical treatments | | | |
|---|---|---|---|
| Author | Sponsor | Purpose of Model | ICER Value / Main findings |
| Launois et al. (1994) | None | Evaluate both the costs and effectiveness of discectomy and Chemonucleolysis for lumbar disc herniation | Chemonucleolysis dominates discectomy. More effective, and 9,126 Francs cheaper |
| Lewis et al. (2011) | None | Estimate the cost-effectiveness of treatment regimens for sciatica patients | Stepped approaches based on initial treatment with non-opioids most cost-effective regimens relative to direct referral to surgery. Referring patients who fail initial treatments to surgery unlikely to be cost-effective |
| Skidmore et al. (2011) | Medtronic, Inc Sunnyvale, CA. | Evaluate cost-effectiveness of decompression with the X-STOP Interspinous Spacer compared to conservative care, and laminectomy, to treat LSS | X-STOP cost-effective when compared with CC (ICER $17,894/QALY. X-STOP spacer dominant compared with Laminectomy |
| Fitzsimmons et al. (2014) | None | See Lewis et al. (2011) | |
| Koenig et al. (2014) | American Academy of Orthopaedic Surgeons | Does inclusion of productivity costs impact cost-effectiveness of lumbar discectomy? | Consideration societal costs reduces the ICER for discectomy from $52,416/QALY to $35,146/QALY over 4 years |

| Author | Sponsor | Purpose of Model | ICER Value / Main findings |
|---|---|---|---|
| Udeh et al. (2014) | Outcomes Research Department of Cleveland | Evaluate cost-effectiveness of three options to treat LSS | *mild*® most cost-effective ($43,760/QALY), ESI next best at additional $37,758/QALY. Laminectomy least cost-effective ($125,985/QALY) |
| Igarashi et al. (2015) | Pfizer Inc | Cost-effectiveness of pregabalin for chronic LBP with accompanying neuropathic pain | Pregabalin cost-effective relative to usual care, ICERs ¥2,025,000/QALY. Inclusion of societal costs decreases ICER to ¥1,435,000/QALY |
| Parker et al. (2015) | VertiFlex, Inc. | Compare cost-effectiveness of conservative care and decompressive surgery to a new minimally-invasive interspinous spacer. | CC had the lowest cost at $10,540, but also lowest QALY increase (0.06). ICER for Spacers compared to CC was $16,300/QALY and for DS was $15,200/QALY |
| Tapp et al. (2018) | Agency for Healthcare Research and Quality and from the National Institute for Arthritis, Musculoskeletal and Skin Diseases | Characterize the factors affecting the long-term cost-effectiveness of interspinous spacer devices relative to decompression surgery | DS cost effective relative to CC at $25,000/QALY. Spacer cost-effective relative to decompression at $89,500/QALY |

| Sciatica decision modelling studies – surgical populations | | | |
|---|---|---|---|
| Author | Sponsor | Purpose of Model | ICER Value / Main findings |
| Kuntz et al. (2000) | None | Assess cost-effectiveness of different types of laminectomy. | Laminectomy with non-instrumented fusion costs $56,500/QALY *versus* laminectomy without fusion. ICER for instrumented fusion compared with non-instrumented fusion was $3,112,800/QALY |
| Kim et al. (2012) | W.Garfield Weston Foundation | Determine cost-utility of decompression with and without instrumented fusion for lumbar spondylolisthesis patients. | Compared with decompression alone, decompression plus instrumented fusion cost $185,878/QALY. |
| Parkinson et al. (2012) | Australian Department of Health and Ageing | Establish cost-effectiveness of lumbar AIDR and how choice of comparator impacts result. | AIDR cost-saving compared with lumbar fusion ($1600/patient). However anterior lumbar interbody fusion and PLF were less costly by $2155 and $807. Not all comparators had cost/QALY. PLF dominates AIDR on cost/QALY. |
| Schmier et al. (2014) | Unclear – authors work consultancies. | Determine cost effectiveness of Coflex® interlaminar stabilization vs instrumented posterolateral lumbar fusion for treating LSS and spondylolisthesis | QALYs higher for Coflex patient's vs fusion. Costs lower for Coflex compared to fusion, at $15,182 compared to $26,863 for the fusion control |

| Author | Sponsor | Purpose of Model | ICER Value / Main findings |
|---|---|---|---|
| Bydon et al. (2015) | None | Compare cost-effectiveness of interbody fusion vs posterolateral fusion for patients with lumbar spondylolisthesis. | ICER for the interbody fusions vs non-interbody fusion, $9,883.97/QALY |
| Vertuani et al. (2015) | Medtronic | Evaluate cost-effectiveness of minimally invasive versus open surgery techniques for lumbar spinal fusion. | MIS dominant compared with open surgery, yielding cost savings and improved HRQoL. Total cost saving per procedure €973 for Italy and €1666 for the UK, with an improvement of 0.04 QALYs over 2 years |
| Yaghoubi et al. (2016) | None | Evaluate cost-effectiveness of Dynamic Interspinous Spacer (Coflex®) and Static Spacer (X-Stop) versus laminectomy. | ICER for X-stop and Coflex versus laminectomy was US$ 665.9/QALY and US$ 780.7/QALY. X-stop the most cost-effective treatment strategy. |

Abbreviations: AIDR (Artificial intervertebral disc replacement); CBT (Cognitive behavioural therapy); CC (Conservative Care); CLBP (Chronic low back pain); DS (Decompressive surgery); ESI (Epidural steroid injections); HRQoL (Health-related quality of life); ICER (Incremental Cost-Effectiveness Ratio); LBP (Low back pain); LSS (Lumbar spinal stenosis); MIS (Minimally invasive surgery); NSAID (Nonsteroidal anti-inflammatory drug); PLF (Posterolateral fusion); QALY (Quality-Adjusted Life Year);

### 4.3.3 Model Analytical Characteristics

#### 4.3.3.1 Analytical characteristics of models for LBP

The sole cost-effectiveness analysis (CEA) (Lloyd et al., 2004) considered 'successfully treated patients' as the benefit measure and did not discount as this was a short-term decision tree. The remaining four studies were identified as cost-utility analyses (CUA), having presented the benefits of their interventions in 'cost-per-QALY' terms. These four studies discounted both costs and benefits as the time horizon was beyond one year. The five studies took a variety of perspectives, one private payer, two societal, one commercial payer, and one NHS.

#### 4.3.3.2 Analytical characteristics of models for non-surgical sciatica treatment

Most studies identified themselves as either cost-effectiveness or cost-utility studies and discounted correctly given their time frame. However, Skidmore et al. (2011) did not discount, as the study did not clarify a time horizon, it is impossible to determine whether this was the correct approach.

Studies took numerous different perspectives, two used the UK NHS, three took the payer perspective, one took the societal perspective and three the US Medicare perspective. Two studies (Igarashi et al., 2015; Koenig et al., 2014) performed their analysis from both a payer and societal perspective.

#### 4.3.3.3 Analytical characteristics of models for surgical treatment for sciatica

Included studies were either cost-utility or cost-effectiveness analyses. Most reported cost-per-QALY, although Yaghoubi et al. (2016) used cost-per- Visual Analogue Score (VAS) score reduction and Parkinson et al. (2012) a number of cost-per measures including QALYs, but also cost-per-discontinuation, -overall treatment success and -success in reducing condition specific disability (using the Oswestry Disability Index – ODI).

Three studies discounted correctly, two (Yaghoubi et al., 2016; Vertuani et al., 2015) did not discount but had an unclear time horizon, whilst Parkinson et al. (2012) and Bydon et al. (2015) did not use discounting despite having a time horizon beyond one year.

As in previous studies, models took a variety of different perspectives; three took the healthcare perspective, one the societal, one the hospital, and one the third-party perspective. Bydon et al. (2015) did not state which perspective they took.

**Table 4.3 Analytical Characteristics**

| Low back pain decision modelling studies | | | | | |
|---|---|---|---|---|---|
| Author | Type of Analysis | Perspective | Currency / Price Year | Benefits | Discounting |
| Lloyd et al. (2004) | Cost-Effectiveness | UK NHS | GBP 2004 | Successfully treated patients | None |
| Kim et al. (2010) | Cost-Utility | South Korean Societal | Korean Won 2009 | QALYs | 5% costs and benefits |
| Wielage et al. (2013a) | Cost-Utility | Private Payer | US dollar 2011 | QALYs | 3% costs and benefits |
| Wielage et al. (2013b) | Cost-Utility | Quebec Societal | Canadian dollar 2011 | QALYs | 5% costs and benefits |
| Norton et al. (2015) | Cost-Utility | US Commercial payer | US dollar 2015 | QALYs | 3% costs and benefits |
| **Sciatica decision modelling studies – non-surgical treatments** | | | | | |
| Launois et al. (1994) | Cost-Utility | Unstated | French Franc 1990 | QALYs | 5% costs and benefits |
| Lewis et al. (2011) | Cost-Effectiveness | UK NHS | GBP 2009 | Successfully treated patients / QALYs | None |
| Skidmore et al. (2011) | Cost-Effectiveness | US Societal | US Dollar 2009 | QALYs | None |
| Fitzsimmons et al. (2014) | Cost-Effectiveness | UK NHS | GBP 2009 | Successfully treated patients / QALYs | None |

| Author | Type of Analysis | Perspective | Currency / Price Year | Benefits | Discounting |
|---|---|---|---|---|---|
| Koenig et al. (2014) | Cost-Effectiveness | US Private Payer / Societal | US Dollar 2009 | QALYs | 3% costs and benefits |
| Udeh et al. (2014) | Cost-Effectiveness Analysis | Medicare | US Dollar 2013 | QALYs | 3% costs and benefits |
| Igarashi et al. (2015) | Cost-Utility | Japanese payer and societal | Japanese Yen price year unclear | QALYs | None |
| Author | Type of Analysis | Perspective | Currency / Price Year | Benefits | Discounting |
| Parker et al. (2015) | Cost-Effectiveness | Payer | US Dollar 2014 | QALYs / cost-per-patient | 3% costs and benefits |
| Tapp et al. (2018) | Cost-Effectiveness | Medicare | US Dollar price year unclear | QALYs | 3% costs and benefits |
| **Sciatica decision modelling studies – surgical treatments** | | | | | |
| Kuntz et al. (2000) | Cost-Effectiveness | US Societal | US Dollar 1997 | QALYs | 3% costs and benefits |
| Kim et al. (2012) | Cost-Utility | U.S Hospital | Canadian Dollar 2010 | QALYs | 3% costs and benefits |
| Parkinson et al. (2012) | Cost-Effectiveness and Cost-Utility | Healthcare | Australian Dollar 2011 | QALYs /discontinuation/ overall success/ ODI success. | None |

| Author | Type of Analysis | Perspective | Currency / Price Year | Benefits | Discounting |
|---|---|---|---|---|---|
| Schmier et al. (2014) | Cost-Effectiveness | Third party | US Dollar 2013 | QALYs | 3% costs and benefits |
| Bydon et al. (2015) | Cost-Effectiveness | Unstated | US Dollar 2010 | QALYs | None |
| Vertuani et al. (2015) | Cost-Effectiveness | Healthcare Payer | Euro 2013 | QALYS | None |
| Yaghoubi et al. (2016) | Cost-Effectiveness | Iran Healthcare provider | US Dollar price year unclear | VAS scores | None |
| Abbreviations: NHS (National Health Service); ODI (Oswestry Disability Index); QALY (Quality-Adjusted Life Year); | | | | | |

### 4.3.4 Model Design and Structure

#### 4.3.4.1 Model Type

Table 4.4 summarises the characteristics of model design and structure. Model selection generally reflected study time horizon rather than condition-specific health processes. For example, across both conditions, five studies modelled decision trees and all had a time horizon between one to two years. Of the models with an identifiable modelling approach, the majority used Markov modelling; four out of five LBP models, six out of ten non-surgical sciatica models, and four out of seven surgical sciatica models. Two models, one for LBP (Norton et al., 2015) and one for sciatica (Launois et al., 1994) used a decision tree prior to their Markov model.

A number of studies had ambiguous modelling methodologies. Kuntz et al. (2000) used a model consistent with their claim to have used Markov modelling; however, without a diagram, the structure of their model is unclear. Three studies (Skidmore et al., 2011; Vertuani et al., 2015; Schmier et al., 2014) did not state their choice of model. In the case of Schmier et al. (2014) their diagram and inputs were suggestive of a Markov approach. Despite declaring the use of a model, it is uncertain that Vertuani et al. (2015) used one; it certainly appears QALYs have been derived from mean utility values of patients receiving either treatment, regardless of treatment outcome.

Despite reportedly using a Markov model, Parkinson et al. (2012) used a similar approach for the calculation of their QALY totals. They state overtly that QALY values reflected mean values of treated patients at the trial endpoint regardless of the outcome.

#### 4.3.4.2 Cycle Length and Time Horizon.

A variety of time horizons were employed in both conditions. Of the four Markov models for LBP, three stated their cycle length and all used three-month cycles (Wielage et al.,

2013a; Wielage et al., 2013b; Kim et al., 2010). Only the Wielage and colleagues papers used the lifetime horizon for their analysis.

Of the non-surgical treatments for sciatica, a variety of cycle lengths and time horizons were used. The two decision trees took either a one-year time horizon used by the HTA model or a two-year horizon (Udeh et al., 2014). Of the five Markov models, Parker et al. (2015) and Launois et al. (1994) used 3-month cycles, with the former running their model for two years and the latter for seven years. Igarashi et al. (2015) used a one-month cycle in their model lasting only one year. The other Markov model by Koenig et al. (2014) used one-year cycles lasting four and eight years. Tapp et al. (2018) did run their model for ten years but did not state the cycle length. The horizon and cycle length used by Skidmore et al. (2011) was unclear, as was the model type.

Of the surgical treatment for sciatica models, generally, the time horizons were longer. Four used a time horizon of five years or more (Kuntz et al., 2000; Bydon et al., 2015; Kim et al., 2012; Schmier et al., 2014). However, only Kim et al. (2012) and Parkinson et al. (2012) had easily identifiable cycle lengths, the former using a one year cycle for a ten year time horizon and the latter a one month cycle for a two-year horizon. Of the models which did not state cycle length, two did not stipulate a time horizon (Yaghoubi et al., 2016; Vertuani et al., 2015).

#### 4.3.4.3 Health States, Model Events and Risk Factors

There was considerable diversity in states used to represent health across models, with states usually chosen to reflect model purpose, such as “success” or “failure” of treatment for example, rather than the underlying health processes.

##### LBP models

Wielage et al. (2013a) used four simple states - ‘treatment’, ‘death’, ‘adverse events’ and ‘post-adverse events’, with the two adverse event states representing an amalgamation of a

comprehensive range of adverse events. The focus on adverse events reflected the impact upon health of adverse events associated with pharmaceutical interventions. The authors (Wielage et al., 2013a; 2013b) also used age-dependent utilities and mortality risk in their models.

The model by Norton et al. (2015) opted for a simpler structure, using three states, 'improved', 'not improved' and 'dead'. The model did allow for recurrence but included no dependent risk factors or adverse events. Similarly, Kim et al. (2010) allowed for recurrence and had no adverse events. In terms of states, they used four, 'acute LBP', 'chronic LBP', 'well' and 'dead', each possessing its own distinct utility level. However, their model classified a patient with two episodes of back pain within five years to have "chronic LBP". The model also allows for age and gender-associated mortality risk.

Lloyd et al. (2004) took a different approach using a decision tree, reflecting their acute decision problem. The tree begins with 'successful treatment' or 'not', with success representing a level of clinically relevant pain reduction. Treatment failure is followed by 're-consultation' or 'not', with those seeking re-consultation either 'treated with physiotherapy' or 'with NSAIDs'. Each state has a risk of adverse events, although no mortality or risk factors were included in the model.

### Non-surgical sciatica treatment models

Amongst these sets of models, six of the eight models with identifiable structures used "treatment success" to structure their models regardless of model type. In addition, all models allowed for representation of either recurrence of symptoms or reoperation (itself implying recurrence of symptoms).

For example, the decision tree in the HTA model used over one hundred different branches allowing for various treatment options which either succeed or fail. Initial treatment failures received a second-line treatment, which if resulting in another failure would allow

the patient to be ‘referred to disc surgery’ or ‘epidural (injection)’ each possessing its own probability of ‘success or ‘failure’. Their model focusses on three pathways for treatment. The first pathway considers management within primary care including, usual care education/advice, activity restriction, non-opioids or opioids. The second pathway represents the stepped care approach, which includes use of various intermediate treatments administered in primary or secondary care, including manipulation, traction, passive physical therapy modalities, active physical therapy modalities, alternative treatments (for example acupuncture) and biological agents. This second pathway also allows for more invasive treatments, epidural injections followed by disk surgery, if the intermediate treatments fail to resolve symptoms. The third pathway is immediate referral for surgery following initial treatment failure in primary care.

Similar, albeit simplified structures, with a reduced range of treatments, are present in the decision trees of Launois et al. (1994) and Udeh et al. (2014), as well as the Markov model by Parker et al. (2015). Koenig and colleagues perform further simplification, using only four states, ‘successful’ or ‘unsuccessful’ outcomes, ‘death’ or ‘revision (surgery)’. The Koenig model was also the only study to model mortality risk according to age and gender.

Two models did not structure using success. Tapp et al. (2018) used an alive/dead model with allowance for surgery in the surgical arm. Whilst Igarashi et al. (2012) opted to use four states reflecting pain severity, ‘mild or no pain’, ‘moderate pain’, ‘severe pain’ with movement between all states in the first two months (see Figure 4.2). Although, with movement between pain states not allowed between months 3 to 11, long-term symptom recurrence was not be modelled.

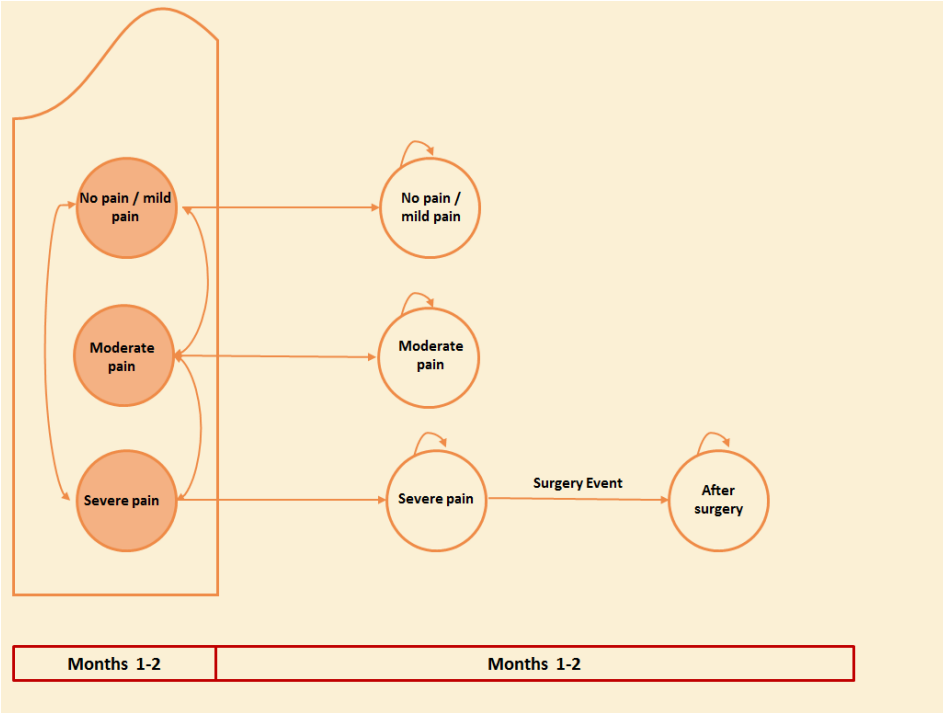



**Figure 4-2 Cohort simulation model used by Igarashi et al. (2015)**

Surgical sciatica treatment models

Some of these studies had unclear model structure and methodologies. Vertuani et al. (2015) do not describe states or events in their model. The names of the states used by Kuntz et al. (2000) in their Markov model are unclear; although it is clear they have modelled complications, recurrence, and re-operation rates somehow. Similarly, the exact quantity of states, and their names, in the Schmier et al., model was unclear. The approach of Yaghoubi et al. (2016) was identifiable, although presented a simple decision tree, with three treatment options and two branches attached to each for success or failure.

The four models identifiable as state transition models (Kim et al., 2012: Parkinson et al., 2012; Bydon et al., 2015; Tapp et al., 2018) had identifiable quantities and descriptions of their health states, as well as key events. Parkinson et al. (2012) and Bydon et al. (2015) structured their Markov model around treatment success, with the former using a

comprehensive range of states relating to surgical procedures and possible outcomes, and the latter a comprehensive range of adverse events. Tapp et al. (2018) using two different Markov models used an alive/dead model for their conservative care, with the addition of 'surgery' and 'complication' states for their surgical model.

Kim et al. (2012) had a simple structure with only four states, 'unwell', 'well', 'no improvement' and 'death' (,Figure 4.3). Their model presented all important events, including relapse, reoperation (including the possibility that a patient may not have a re-operation despite worsening of symptoms), clinical worsening, clinical improvement, general and perioperative death.

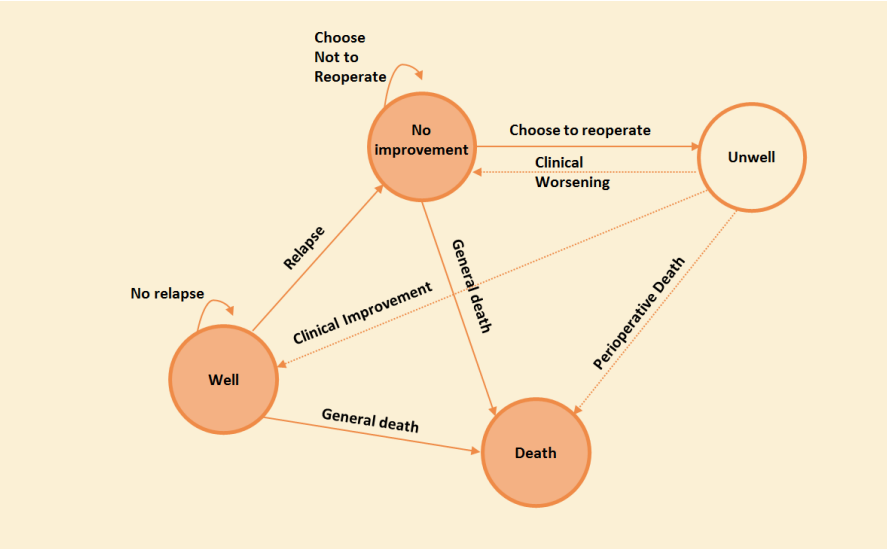



**Figure 4-3 Markov state diagram used by Kim et al. (2012)**

**Table 4.4 Model Design and Structure**

| Low back pain decision modelling studies | | | | | | | |
|---|---|---|---|---|---|---|---|
| Author | Model Type | Cycle Length | Time Horizon | No. of States | Names of key states | Key Events | Dependent risk factors |
| Lloyd et al. (2004) | Decision tree | Unclear | 1 episode | 6 | Successfully treated or not, Consultation or not, Refer to physio or treat with NSAID | AEs related to the treatments | None |
| Kim et al. (2010) | Markov model | 3 months | 5 years | 4 | Acute LBP, Chronic LBP, Well, Death | Recurrence of symptoms | Age and gender related mortality risk. |
| Wielage et al. (2013a) | Markov model | 3 months | Lifetime | 4 | Treatment, Adverse Event, Post-Adverse Event, Death. | Numerous persistent and transient AEs | Age dependent utilities, mortality, and risk of AE. |
| Wielage et al. (2013b) | Markov model | 3 months | Lifetime | 4 | Treatment, Adverse Event, Post-Adverse Event, Death | Numerous persistent and transient AEs | Age dependent utilities, mortality, and risk of AE. |
| Norton et al. (2015) | Decision tree and Markov model | Unclear | 1 year and 10 years | 3 | Improved, Not improved, Dead | Recurrence of symptoms | None |
| **Sciatica decision modelling studies – general treatments** | | | | | | | |
| Launois et al. (1994) | Decision tree and Markov model | 3 months | 7 years | 8 | Success or Failure, Maintained or Free Survival; Re-Op, Deterioration no reoperation or Definitive failure | Both treatments may fail. Symptoms recurrence. Reoperations are possible. | None |
| Lewis et al. (2011) | Decision tree | 1 episode | 12 months | Over 100 | First line treatments succeed or fail. Failures have 2nd treatment. 2nd failures have possible disc surgery or epidural, each with success or failure. | Stepped model, with over 100 different treatment possibilities | None |
| Skidmore et al. (2011) | "Economic model", type unclear. | Unclear | Unclear | Unclear | Unclear | Treatment success, re-operation, various AEs | None |
| Fitzsimmons et al. (2014) | See Lewis et al. (2011) | | | | | | |

| Author | Model Type | Cycle Length | Time Horizon | No. of States | Names of key states | Key Events | Dependent risk factors |
|---|---|---|---|---|---|---|---|
| Koenig et al. (2014) | Markov model | 1 year | 4 and 8 years | 4 | Satisfactory outcome, Unsatisfactory outcome, Death, and Revision. | Revision possible for surgery. Patients may not leave states after 3 months. | Age and gender related mortality risk |
| Udeh et al. (2014) | Decision-tree | Unclear | 2 years | 21 | Various reflecting type of treatment, complications during treatment, and any further treatments. | Model allows for complications, revision and treatment failure. | None |
| Igarashi et al. (2015) | Markov model | 1 month | 12 months | 4 | Mild or no pain, Moderate pain, Severe pain, After surgery. | Recurrence of symptoms | None |
| Parker et al. (2015) | Markov model | 3 months | 2 years | 7 | 7 states with failure or success attached. Conservative care, DS, Continue post-DS, Spacer implant, Continue post-spacer, (DS and continue post-DS after spacer). | All treatments allow for failure. | No death in the model. |
| Tapp et al. (2018) | 2 unique Markov models for CC and surgery | Unstated | 3 and 10 years | CC:2 Surgery: 4 | CC Markov: alive, dead<br>Surgery model: Surgery, post-surgery, post major complication | Model allows for reoperation and complications | Probability of surgery varies by model stage |
| **Sciatica decision modelling studies – Surgical populations** | | | | | | | |
| Kuntz et al. (2000) | Stated as Markov, but no diagram | Unclear | 10 years | Unclear | Unclear. | Surgical complications, recurrence of symptoms, and reoperation. | Age adjusted mortality risk |
| Kim et al. (2012) | Markov model | 1 year | 10 years | 4 | Unwell, Well, No improvement, and Death. | Re-operate, Clinical worsening or improvement, perioperative or general death, relapse. | None |

| Author | Model Type | Cycle Length | Time Horizon | No. of States | Names of key states | Key Events | Dependent risk factors |
|---|---|---|---|---|---|---|---|
| Parkinson et al. (2012) | Markov used for costs. Utility values are pooled. | 1 month | 2 years | 9 | Initial surgery, Successful surgery, Failed surgery; Replace, Remove without replace, Revise, Supplemental fixation, Other re-operation > successful surgery following re-operation | Allows for revised surgery and for reoperation. | Death not included in model |
| Schmier et al. (2014) | Undefined, resembles a Markov. | Unclear | 5 years | Unclear | Surgery, Short-term postoperative, Long-term postoperative. | Complications, revisions, and treatment failure | None |
| Bydon et al. (2015) | Markov model | Unclear | 9 years? | 4 (AE's merged) | Re-operation or No re-operation. PLF may have an additional reoperation or not. | Re-operations. Plus a plethora of complications. | None |
| Vertuani et al. (2015) | Unclear | Unclear | Unclear | Unclear | Unclear | Unclear | Unclear |
| Yaghoubi et al. (2016) | Basic decision Tree | Unclear | Unclear | 2 | Success or Fail. | None | None |
| Abbreviations: AE (Adverse events); CC (Conservative care); CLBP (Chronic low back pain); DS (Decompressive surgery); LBP (Low back pain); NSAID (Nonsteroidal anti-inflammatory drug): PLF(posterolateral fusion) | | | | | | | |

### 4.3.5 Key assumptions regarding modelling and extrapolation.

Given the absence of data to directly populate the models, a large number of assumptions about missing data and extrapolation, as well as model structure were utilised in all studies (see Table 4.5).

#### 4.3.5.1 Assumptions and extrapolation for LBP models.

For the LBP studies, Wielage et al. (2013a; 2013b) and Kim et al. (2010) have made assumptions regarding absent data. For example, given the absence of utility values, the latter has assumed that patients treated with acupuncture patients have a higher QoL than those treated in usual care. They have also made the structural assumption that a single recurrence of pain in the 'well' state moves someone into 'chronic' LBP, meaning a patient experiencing two flare-ups of pain in 5 years is considered to have CLBP.

All studies in this review extrapolated using a blend of trial data, literature and assumption. For example, Norton et al. (2015) used literature to derive the probability of recurrence over time and supplemented this with an assumption that the long term efficacy of CBT declined by a rate of 20% per year. Meanwhile, Kim et al. (2010) assumed a constant relative risk of moving from chronic to well between treatments over the long term and used literature to identify long term recurrence. Both models by Wielage and colleagues were able to produce a lifetime model by adjusting the adverse event (AE) profiles according to age-dependent relative risks. Age-dependent utility values and mortality risks were also used.

#### 4.3.5.2 Assumptions and extrapolation for models concerning non-surgical treatments for sciatica.

Some studies made a number of assumptions regarding parameter values, specifically on recurrence, surgical success, re-operation and adverse events. However, far more prevalent within this category of studies were structural assumptions. For example, the HTA model

made several assumptions regarding how the patient receives their treatment, such as assuming the healthcare for treatment failures occurs outside the NHS. Koenig et al. (2014) assumed those treated non-surgically never receive a surgical intervention. Launois et al. (1994) assume discectomy failure did not lead to re-operation. Parker et al. (2015) made assumptions about the settings within which the procedures would take place, as well as the length of treatment. Udeh et al. (2014) made numerous assumptions about treatment sequencing.

Three models attempted extrapolation. Koenig et al. (2014) attempted an extrapolation of four-year data to eight years but did not specify the assumptions underpinning this process. Igarashi et al. (2015) simply extrapolated an eight week trial to 52 weeks. Tapp et al. (2018) set all usual care costs to zero and analysed the costs and disutility associated solely with surgery and related complications. Therefore they required only the assumptions that long-term rates of surgery and re-operation from year 4-10 were equal to those in the third year, as with the utility values. The HTA model explicitly ruled out extrapolation on the grounds of inadequate evidence. The remaining studies used a time horizon which matched the time horizon of their source data.

4.3.5.3 Assumptions and extrapolation for models of surgical treatments for sciatica. Structural assumptions were the most common form of assumption in this category of models, especially regarding surgical pathways and patient behaviour. Most of the studies made assumptions about absent parameter values. For example, Kuntz et al. (2000) assumed that LSS patients have the same utility as patients with ‘severe LBP’ as well as assuming that patients recovering from treatment have a utility the same as an individual in perfect health. Bydon et al. (2015) reference Kuntz and colleagues as providing the utility values for patients with the specific condition of lumbar spondylolisthesis. However, they did not state that the Kuntz utility values are those for severe LBP.

In performing extrapolation, Kim et al. (2012) categorised their health states to represent either “well” or “unwell”, plus a state for “return of symptoms” (see Figure 4.3), with long term mean utility values associated with these states derived from their one-year observational cohort study. In order to justify the use of these one-year values across ten years, they reference a study by Weinstein et al. (2009) who suggest that utility values derived at one year are relatively stable across four years. Kim et al. (2012) then assume that utility values remain further stable from four to ten years albeit adjusted downwards by 3% per year to account for ‘ageing, clinical deterioration and comorbidities’ (although they suggest this as their justification for discounting their benefits). The authors, then set movements between those states over time, with the ten-year reoperation rate for transitioning from “no improvement” to “re-operation” based upon an administrative database which they also use for the ten-year probabilities of moving from “well” to no “improvement”. They derived their ten-year clinical improvement, for patients moving from “unwell” to “well” from the satisfaction rate from the senior surgical authors’ practice.

Schmier et al. (2014) extrapolated two-year trial data to treatment outcomes over five years based upon the assumption that probabilities do not change substantially over five years. They set utility values associated with success or failures of treatment, and then assumed that utility values achieved at two years could be extrapolated to five years.

Of the remaining studies, one (Kuntz et al., 2000) attempted extrapolation but used unjustified assumptions, two (Parkinson et al., 2012; Beydon et al., 2015) did not engage in extrapolation as their data matched their time horizon, and two (Vertuani et al., 2015; Yaghoubi et al., 2016) were problematic to analyse because of the manner with which the authors described their methods.

**Table 4.5 Assumptions and Extrapolation**

| Low back pain decision modelling studies | | |
|---|---|---|
| Author | Key Assumptions | Methods for Extrapolation |
| Lloyd et al. (2004) | (i) various AE's were unrelated to treatment, (ii) most heat wraps bought over counter, (iii) 24% exempt from prescription charge, (iv) AE's have GP consultation | 50% of patients re-consult for treatment, likelihood of physiotherapy referral was 18% |
| Kim et al. (2010) | (i) usual care and treatment has same cost and effect, (ii) recurrence of LBP in well state is considered chronic, (iii) CLBP patients treated by acupuncture have higher utility than usual care, (iv) same relative risk between treatments over time. | Treatment effectiveness measure for acupuncture vs usual care assumed to have same relative risk over time. Long term recurrence from literature. |
| Wielage et al. (2013a) | (i) equal efficacies for various comparators owing to shortages of data, (ii) AE rates for oral treatments similar across MSK conditions, (iii) patients receive medication every day. | AEs extrapolated using initial rates combined with age-dependent risks derived from literature. Age and AE dependent mortality used, calculated from literature. |
| Wielage et al. (2013b) | (i) equal efficacies for various comparators owing to shortages of data, (ii) AE rates for oral treatments similar across MSK conditions, (iii) patients receive medication every day, (iv) utility value can be derived from pain scores. | AEs extrapolated using initial rates combined with age-dependent risks derived from literature. Age and AE dependent mortality used, calculated from literature. |
| Norton et al. (2015) | (i) gradual loss of efficacy of CBT over time by 20%. | Extrapolating CBT effectiveness assumed probability of recurrence over 10 years was 0.60 - a rate 'reflected in literature'. Utilities assumed the same in respective states over 10 years as in a 1-year study. |

| Sciatica decision modelling studies – general treatments | | |
|---|---|---|
| Author | Key Assumptions | Methods for Extrapolation |
| Launois et al. (1994) | (i) discectomy failures not re-operated, (ii) Unreferenced and unjustified assumptions regarding the rate of recurrence for reoperation of 3% in year 1, 2% in year 2, and 1% in the following years. | Most extrapolation based upon studies included in literature review |
| Lewis et al. (2011) | (i)patients managed through one of 3 pathways, (ii) ultimate treatment failures resort to therapies outside health system, (iii) inactive control assumed no NHS cost, (iv) utility gains continue for 12 months, (v) each prescription required a GP | No extrapolation on grounds on inadequate evidence |
| | consultation, more for analgesics (vi) not seeking further treatment following | |
| Lewis et al. (2011) (cont.) | failure assumed no cost, (vii) no reduction in utility for 2nd failure, (viii) combined therapies as effective as stand-alone treatments. | No extrapolation on grounds on inadequate evidence |
| Skidmore et al. (2011) | (i) treatment-related AE's treated the same regardless of initial treatment, (ii) success and re-op rates the same for LAMI with fusion and without fusion, (iii) utility of LAMI does not change one-year after postop. | No extrapolation required, trial same as the model time horizon. |
| Fitzsimmons et al. (2014) | See Lewis et al. (2011) | See Lewis et al. (2011) |
| Koenig et al. (2014) | (i) patients treated non-surgically never receive surgical treatments. | Extrapolations all based upon literature and fully sourced. |

| Author | Key Assumptions | Methods for Extrapolation |
|---|---|---|
| Udeh et al. (2014) | (i)If no relief anytime within 2 years after mild procedure, procedure was considered failed and patients receive surgical option, (ii) same for surgical intervention patients failing who receive repeat surgical intervention, (iii) no further treatment option considered for serial epidural patients, (iv) serial epidural injections offer minimal relief | No extrapolation required, trial data and sources match model time horizon |
| Igarashi et al. (2015) | (i) in months 3-11 patients are "stuck" in the states they exit month 2 at, (ii) assuming pregabalin was no longer effective after discontinuation, (iii) postsurgical pain severity score assumed to be 2 as "confirmed by independent Japanese clinicians", (iv) surgery assumed to occur only after the 3rd month of treatment and only to those in severe pain (ii) | Pain beyond 8-week extrapolated to 1 year based on pain scores observed in clinical trials and studies. |
| Parker et al. (2015) | (i) spacer follow-up service utilization, physician visits, medications, diagnostics assumed same as DS patients, (ii) spacer procedures assumed to be performed in the hospital outpatient setting, (iii) various assumptions about procedure length | No extrapolation required, trial same as the model time horizon |
| Tapp et al. (2018) | (i) annual spacer failure rate for years 4–10 assumed to be the same as year 3, (ii) Utility for the spacer group assumed equal to that of decompression. | Assumed re-operation rates were the same years 4-10 as in year 3. Assumed utilities are the same years 4-10 as year 3. Usual care costs are set to zero, so only costs of surgeries and complications are included. |

| Sciatica decision modelling studies – surgical populations | | |
|---|---|---|
| Author | Key Assumptions | Methods for Extrapolation |
| Kuntz et al. (2000) | (i) postoperative mortality from fusion independent of instrumentation use, (ii) multiple assumptions regarding rates of AE's, (iii) assumptions regarding healing of fusion, (iv) 65% of laminectomy patients assumed to experience clinical improvement, (v) probability of clinical improvement independent of fusion healing, (vi) assumed annual rate of symptom recurrence twice the annual reoperation rate, (vii) stenosis patients have same utility as "severe back pain", (viii) post-surgery assumed the same utility as perfectly health | Patients who did not experience improvement at 6 months one half assumed to have another operation and assigned cost and disutility for laminectomy with non-instrumented fusion. 60% of patients who underwent a second operation assumed to experience relief of their symptoms. Among patients who improved at 6 months 2.3% had another operation owing to worsening symptoms. Annual rate of symptom recurrence assumed twice the annual re-op rate (4.65%) |
| Kim et al. (2012) | (i)patients have no prior spinal surgery at entry, (ii) no direct consideration of adverse events, except for perioperative mortality although cost data cover any in-hospital costs incurred for postsurgical AE's, (iii) patients experiencing surgical failure or recurrence have possible reoperation after a year in no improvement | Various long-term studies used to derive movements between health states. For their utilities they use trial data, referencing a source suggesting that outcomes achieved at 1-year are maintained for 4-years, then assume utility is constant over 10 years |
| Parkinson et al. (2012) | (i) all types of artificial discs / bone grafts assumed equally effective, (ii) removal without replacement is not an option for certain types of surgery, (iii) supplemental fixation is not an option for certain surgeries, (iv) for fusion, hardware replacement same approach as taken initially, (v) only one re-operation is considered, following which patients enter 'successful surgery post re-operation' | No extrapolation required because of the time horizon and nature of the sources used |

| Author | Key Assumptions | Methods for Extrapolation |
|---|---|---|
| Schmier et al. (2014) | (i) assumptions regarding the no. of months each complication affected utility scores, (ii) 68% of cases were one-level, and the remainder were two-level, (iii) costs set at Medicare value plus 20% | 24 month TE's and utilities assumed the same continuously through five years. Utilities weighted for experience of specific complications. Complication rates from trial data extrapolated using published sources. |
| Bydon et al. (2015) | No stated assumptions. However assume patient's recovering from surgery have same utility as perfectly well patients. Moreover, utility values for lumbar spondylolisthesis are not actually for this condition, but for severe low back pain. There is no acknowledgment of this assumption. | No extrapolation required, cohort study same as model time horizon. |
| **Sciatica decision modelling studies – surgical populations (cont'd)** | | |
| Bydon et al. (2015) | No stated assumptions. However assume patient's recovering from surgery have same utility as perfectly well patients. Moreover, utility values for lumbar spondylolisthesis are not actually for this condition, but for severe low back pain. There is no acknowledgment of this assumption. | No extrapolation required, cohort study same as model time horizon. |
| Vertuani et al. (2015) | (i) blood transfusion was necessary for OS but not for MIS, (ii) surgical wounds were drained postoperatively in 20% of MIS and 92% of OS patients following surgery, (iii) use of bone grafts or substitutes same for MIS and OS, (iv) antibiotic treatment assumed necessary after SSIs, (v) for OS re-op after post-surgical complication in MIS, assumed patient had same hospital stay as initial OS patient, (vi) additional expenses due to certain complications assumed similar in both procedures. | Temporal features of model unclear. |

| Author | Key Assumptions | Methods for Extrapolation |
|---|---|---|
| Yaghoubi et al. (2016) | Unstated | Totally unclear |

Abbreviations: AE (Adverse events); CBT (Cognitive behavioural therapy); CEA (Cost-Effectiveness Analysis); CLBP (Chronic low back pain); CUA (Cost-Utility Analysis); DS(Decompression surgery); ESI (Epidural steroid injections); GP (General Practitioner); LAMI (Laminectomy); LSS (Lumbar spinal stenosis); MIS (Minimal Invasive Surgery); MSK( Musculoskeletal); NHS (National Health Service); NSAID (Nonsteroidal anti-inflammatory drug); OS (Open Surgery); SSI (Surgical Site Infection); TE (Treatment Effect).

### 4.3.6 Parameter sources and methods of derivation

#### 4.3.6.1 Parameter sources for treatment efficacy.

Overall, derivation of treatment efficacy for the LBP models was performed in keeping with best, or at least good practice. The two studies by Wielage and colleagues used meta-analysis of trial data to derive their treatment effects, as did Kim et al. (2010) who also used two additional cohort studies to derive their treatment parameters. Two other studies (Norton et al., 2015; Lloyd et al., 2004) used the results of a single trial to derive their parameters, although both trials have adequate sample sizes.

Of the non-surgical treatment models for sciatica, only the HTA model used a systematic review and meta-analysis. Tapp et al. (2018) used the Medicare Provider Analysis and Review database for complication and re-operation rates, with expert opinion for extrapolation. The remaining studies used literature review (Launois et al., 1994) or trial data with adequate sample sizes (Koenig et al., 2014 Igarashi et al., 2015; Tapp et al., 2018). Two studies used a combination of both trial data and literature (Skidmore et al., 2011; Parker et al., 2015) although both trials had small sample sizes. For example Skidmore et al. (2011) used data from only 131 patients, with parameters for first-line laminectomy derived from 21 patients who had a second line LAMI.

In the surgical treatment models, two studies used meta-analysis from a systematic review (Parkinson et al., 2012; Yaghoubi et al., 2016). Three of the studies were driven almost exclusively by relatively small studies, a prospective cohort study of 150 (Kim et al., 2012) an RCT of 150 (Schmier et al., 2014), and data on 137 patients collected retrospectively (Bydon et al., 2015). Kuntz et al. (2000) used a mixture of six prospective and observational studies.

#### 4.3.6.2 Parameter sources for recurrence

Not all studies reported a recurrence rate, although generally recurrence and reoperation rates were derived from existing literature (Norton et al., 2015; Kim et al., 2010; Igarashi et al., 2015; Launois et al., 1994; Tapp et al., 2018; Koenig et al., 2014; Kim et al., 2012; Bydon et al., 2012). Three studies combined literature, with trial data (Skidmore et al., 2011) and expert opinion (Kuntz et al., 2000; Schmier et al., 2014). Parkinson et al. (2012) derived their revision and re-operation rate from their systematic review and meta-analysis of RCT's. Three studies had unclear sources for recurrence or reoperation (Udeh et al., 2014; Yaghoubi et al., 2016; Vertuani et al., 2015).

#### 4.3.6.3 Utility values

The actual utility values used as inputs in studies in this review are presented with their sources shown in Table 4.6 and discussed in the following subsection, 4.3.6.4.

There were some differences in the utility values used in the LBP studies, likely partly explained by their specific population. The meta-analyses of chronic LBP patients who were prescribed pharmacological treatments, showed that chronic LBP patients had utility values of between 0.7282 (Pregabalin) and 0.7688 (Naproxen) (Wielage et al., 2013a; 2013b). Meanwhile, the other studies suggested lower values; Kim et al. (2010) used 0.62 and 0.65 for chronic LBP patients on usual care and acupuncture, respectively, although acute LBP (0.85) and "well" states (0.96) were considerably higher; and Norton et al. (2015) used much lower scores, with an improving chronic LBP patient having utility of 0.640, and a non-improver having utility of 0.59.

There was some consistency across some of the sciatica decision models which incorporated conservative care and used utility values independent of treatment. Igarashi et al. (2015) report that a sciatica patient without pain would have a utility of 0.867, whilst "severe pain" would be 0.611, somewhat consistent with the scores used by Koenig et al.

(2014) for treatment of herniated intervertebral disc. Both studies are consistent also with the value used for "improvement" in sciatica patients by Lewis et al. (2011) and Fitzsimmons et al. (2014), although the "non-improvement" score in the latter is lower than other scores at 0.37.

Skidmore et al. (2011) use utility scores for each treatment which are weighted averages of both improvers and non-improvers on each treatment, and are weighted also for likelihood and disutility of adverse events. Their weighted average for conservative care (which has only a 4.8% success rate), is a utility value of between 0.61 and 0.65 over the duration of the model. Given that non-improvers are the predominant constituent of the conservative care group, this value is consistent with the utilities of the other non-improving patients in the prior four studies discussed. The values used by Tapp et al. (2018) for post –surgery, which represents a patient after surgery without complications, is slightly higher at 0.77, although the values also account for the disutility associated with complications and recurrence separately.

For the solely surgical studies, the two studies which used the Beaver Dam Study, applied a utility value of 0.79, for symptomatic spinal stenosis patients in one study (Koenig et al., 2014) and for lumbar spondylolisthesis patients with a negative outcome in another (Yaghoubi et al., 2016). Both studies also used a utility value of 0.97 for having a positive outcome. These values are far higher than the two other studies to use"improve" / "non-improve" to differentiate patients. For patients with lumbar spondylolisthesis; Kim et al. (2012) used 0.74 for an improvement with fusion and no-fusion, compared to non-improvement of 0.50 and 0.54 respectively, baseline patients had a utility value of 0.58, and Schmier et al. (2014) used a value for "clinical success" of 0.692, whilst "failure" had utility 0.552 and "worsening pain" 0.599.

Parkinson et al. (2013) used much lower baseline pre-operation utilities of 0.42 and 0.36, although possibly explained by their specific population, patients with sciatica who had failed conservative treatment. After two years their patients who had AIDR had average utility of 0.67 and patients who had fusion had utility of 0.69. These scores are similar to the patients undergoing spinal fusion to which Vertuani (2015) assigned utilities of 0.72 after two years following minimally invasive surgery and 0.68 following open surgery.

#### 4.3.6.4 Parameter sources for utility values

The calculation of utility values in both studies by Wielage et al. (2013a; 2013b) was consistent with best practice guidelines, deriving pain scores for various pharmaceutical treatments for CLBP and translating those into utility values weighted for age and sex. Norton et al. (2015) used a large RCT (n=701) to derive EQ-5D data to convert to utility scores for both CBT and usual care arms. Kim et al. (2010) used survey data and a large RCT to derive utility values for acupuncture and usual care.

For the non-surgical treatment models for sciatica, nearly all studies used questionnaire or trial data and converted these into utility values, although many used small sample sizes. Launois et al. (1994) used Rosser coefficients from a patient survey (n=146) with no information or reference provided for the survey. Skidmore et al. (2011) used SF-6D data from an RCT with a small sample (n=131), with the utility value for first line laminectomy derived from 21 patients who had a 2nd line laminectomy second to CC or X-STOP. Parker et al. (2015) used SF-6D values from two separate studies for each of their arms, although both studies had small sample sizes.

Similarly, derivation of utility values for the HTA model appears not entirely appropriate. Firstly, they used EQ-5D scores obtained from a very small RCT (van den Houdt et al., 2008). Second, the same utility values for “successful” or “failed” treatment was applied for all interventions in the model. Secondly, the utility value used for “successful

treatment" in the HTA model, 0.83, is the *highest mean* utility value achieved by a specific intervention, early surgery, during the one year duration of the RCT. Third, the "treatment failure" score in the HTA model represents the utility at baseline in the RCT, 0.37.

Igarashi et al. (2015) converted pain scores taken from a non-interventional trial into utility values, but these scores were from eight weeks follow-up which they proceeded to use as one-year utility values. Koenig et al. (2014) used the utility values from a previous economic evaluation (Malter et al., 1996) of treatments for herniated intervertebral disc. However Malter did not use utility values for patients treated for herniated intervertebral disc as Koenig et al. (2014) claim, but had actually applied utility results for 83 patients with "severe LBP" from the Beaver Dam study (Fryback et al., 1993). Finally, Udeh et al. (2014) used 'QALY gains' reported in various trials of patients with mild or moderate LSS, and adjusted their QALY gains downwards by 25% as patients in their model had more severe LSS although this approach was not justified in their paper. Furthermore, all of their treatment QALY gains originated from very small sample sizes, the ESI QALY gain derived from 39 patients. Moreover, they compound this problem by erroneously calculating the QALY gain for ESI, reporting a QALY gain of 0.21 per 2-months when the paper they took this information from clearly stated this QALY gain of 0.21 was for a 3 month period.

For the surgical treatment models, two more studies (Kuntz et al., 2000; Bydon et al., 2015) used utility values for "severe LBP" patients from the Beaver Dam study (Fryback et al., 1983) as utility values for their LSS and spondylolisthesis patients. Kuntz et al. (2000) did acknowledge that using LBP utility values could be inappropriate for LSS patients; Bydon et al. (2015) failed to acknowledge that the utility values were not for his target population. Both studies used a utility value of 0.97 for symptom resolution, the utility score from the Beaver Dam study for individuals with perfect health.

Two papers based their utility values upon small samples. Parkinson et al. (2012) used EQ-5D values derived from a small RCT (n=150) and converted them into utility values. Kim et al. (2012) supplemented utility values from their own small prospective cohort (n=115) study with published literature. However, as noted above, the authors acknowledge extrapolation of utility values should have accounted for 'ageing, deterioration, and comorbidities' which they confused with the need to discount benefits.

One study (Yaghoubi et al., 2016) derived utility values from VAS scores from a systematic review and meta-analysis of studies of surgery for spinal stenosis, where patients received either dynamic or static implants. Finally, Vertuani et al. (2015) converted EQ-5D scores from the Swedish National Registry (n=2437) into utility scores, although further evaluation is not possible as it is unclear how they have taken data from the original source.

#### 4.3.6.4 Parameter Sources for costs

Most of the actual costing of the resource use into monetary terms comes from national health cost databases, which represent the best practice standard for deriving costs (Philips et al., 2006). There was however a variety of methods employed to estimate resource use.

For the LBP models, Wielage et al. (2013a; 2013b) used published literature and expert opinion was used to derive resource usage. Two studies (Norton et al., 2015; Kim et al., 2010) used trial data supplemented by literature. Lloyd et al. (2004) used manufacturer costs, as well as published studies and the BMA formulary.

For the sciatica non-surgical treatments, most studies used expert opinion, with only two studies (Koenig et al., 2014; Tapp et al., 2018) managing to obtain their resource use from a database. Although, Tapp et al. (2018) only calculate the cost of surgery but no other healthcare expenditures relating to back pain.

The HTA model used ‘clinical opinion’ to determine resource use. Similarly, Igarashi et al. (2015) used a physician based internet survey. Skidmore et al. (2011) used estimates provided by a panel of experts. Two studies contacted patients, Launois et al. (1994) used their unreferenced patient survey, and Parker et al. (2015) used a telephone interview. Finally, Udeh et al. (2014) used three published studies to derive resource use.

The surgical population models also used a variety of methods. Two studies were able to use administrative databases to derive resource use, Kuntz et al. (2000) using a previous study deriving their values from one hospital database and Kim et al. (2012) used one of their author’s hospital financial department. Bydon et al. (2015) used their 137 case reviews supplemented by data from Kuntz et al. (2000). Parkinson et al. (2012) used expert opinion to derive resource usage. Schmier et al. (2014) use various published sources, analysis of the Medicare Limited Data, and expert opinion to estimate their resource usage.

Two studies appear to have derived costs directly, without describing resource use. Vertuain et al. (2015) use systematic review and meta-analysis for costing, whilst Yaghoubi et al. (2016) derived costs from the literature, patient bills, and manufacturer costs.

#### 4.3.6.5 Methods and Parameter sources for calculating Societal Costs

Six of the nineteen distinct models used the societal perspective for their model (Table 4.6a). Three models (Kim et al., 2010; Skidmore et al., 2011; Wielage et al., 2013b) performed their analysis solely from societal perspective with three (Kuntz et al., 2000; Igarashi et al., 2015; Keonig et al., 2014) including a societal analysis as a sensitivity or subsidiary analysis.

Igarashi et al. (2015) based their productivity costs in their analysis on the Work Productivity and Activity Impairment (WPAI) questionnaire adapted for LBP. The “work productivity” component of the WPAI includes absenteeism and presenteeism, providing

an estimate of the overall work impairment in line with friction cost approaches. To estimate LOP, productivity was defined as a percentage from 0% to 100% and mapped to pain scores, such that each point change in pain score represented a change in productivity. The authors estimate their costs using mean monthly income in Japan.

Two studies (Wielage et al.,2013b; Koenig et al., 2013) used variants of the human capital approach to estimate their indirect costs. Both studies estimated the amount of days lost owing to the condition and then transformed that into a monetary value using different measures of national wages. Kim et al. (2010) attempted a similar approach, although their losses only stemmed from the treatment sessions.

Two studies (Skidmore et al., 2011; Kuntz et al., 2000) claimed that health-related quality-of-life measures already capture the impact of disability upon lost income, and therefore including some estimate of lost wages would be double counting. Kuntz et al. (2000) do at least model the impact of relaxing this assumption and include the impact of lost wages on cost-effectiveness in a sensitivity analysis.

#### 4.3.6.6 Parameter Sources for Adverse Events

Only half of the papers included states for adverse events in their models, most of which were surgical treatment models for sciatica.

Of the LBP studies, Wielage et al. (2013a; 2013b) derived rates of AE's related to pharmaceutical treatments from their systematic review and meta-analysis as well as expert opinion. Lloyd et al. (2004) also used expert opinion to judge whether the cause of certain events could be attributed to the treatment. For non-surgical treatment models for sciatica, Skidmore et al. (2011) and Tapp et al. (2018) derived probability of AE's, and complication rates in the case of the latter, from Medicare claims databases. Parker et al. (2015) obtained rates directly from a small trial and prospective spinal registry. Udeh et al. (2014) used eight different sources, a combination of different studies and databases.

For the surgical populations, Kuntz et al. (2000) used numerous different sources of different types; Bydon et al. (2015) used their 137 patient institutional series; Schmier et al. (2014) used the results of a small RCT and extrapolated those results using expert opinion. Vertuani et al. (2015) used a systematic review and meta-analysis to derive rates of AE's and complications. The methods of Yaghoubi et al. (2016) were unclear.

**Table 4.6 Model Parameters**

| Author | Treatment Efficacy | Recurrence | QoL | Utility Values | Resource Use |
|---|---|---|---|---|---|
| **Low back pain decision modelling studies** | | | | | |
| Lloyd et al. (2004) | Successful or unsuccessful treatment **SOURCE:** pivotal trial of heat wrap (n=371) by Nadler et al. (2002) | N/A | Successfully treated patients (meaningful reduction in NRS pain scores and RMDQ) **SOURCE:** Nadler et al. (2002) | No utility values provided | **SOURCES:** Assumptions regarding number of painkillers taken, number of GP and physio appointments |
| Kim et al. (2010) | Movements between acute, chronic, well and death states. Treatment effect assumed to have same relative risk over time **SOURCES:** Cohort studies by Grotle et al. (2005) (n=123) and Cassidy et al. (2005) (n=1100). "Chronic" to "Well" in both treatments from meta-analysis of RCT's. | "Well" to "Chronic" **SOURCE:** Cassidy et al. (2005) | **SOURCES: "**Acute LBP" AND "Well" from Korean National Health and Nutrition Survey. CLBP from pragmatic RCT by Witt et al. (2006) (n=11630) | Well 0.96<br>Acute LBP 0.85<br>CLBP Usual Care 0. 62<br>CLBP Acupuncture 0.65 | **SOURCES:** Resource usage derived from 2 pragmatic trials (Witt et al., 2006; Thomas et al., 2006) |

| | | | | | |
|---|---|---|---|---|---|
| Wielage et al. (2013a) | 3-month discontinuation and post-discontinuation rates<br>**SOURCE:** Meta-analysis of CLBP and OA trials<br>AEs extrapolated using age-dependent risks derived from literature | N/A | Utilities derived from pain scores, age/sex weighted<br>**SOURCES:** Pain scores from meta-analysis of CLBP trials. Age/sex utility weights from US National Health Measurement Study | CLBP on Duloxetine<br>0.7541<br>CLBP on Celecoxib<br>0.7688<br>CLBP on Naproxen<br>0.7688<br>CLBP on Pregabalin<br>0.7282<br>CLBP on Oxycodone<br>0.7628 | **SOURCES:** Resource use provided by expert opinion. Costs associated with AE's from the Agency for Healthcare Research and Quality database and published literature |
| Wielage et al. (2013b) | 3-month discontinuation and post-discontinuation rates<br>**SOURCE:** Meta-analysis of CLBP and OA trials<br>AEs extrapolated using age-dependent risks derived from literature | N/A | Utilities derived from pain scores, age/sex weighted<br>**SOURCES:** Pain scores from meta-analysis of CLBP trials. Age/sex utility weights from Canadian community health survey | See Wielage et al. (2013a) | **SOURCES**: Physician and drug costs from IMS-Brogan Provincial Formulary Database.<br>Cost of AE's from published literature, IMS-Brogan Database, and Ontario Costing Analysis Tool |

| | | | | | |
|---|---|---|---|---|---|
| Norton et al. (2015) | Initial treatment success, long-term relapse and improvement<br>**SOURCES:** Back Skills Training Trial (n=701)<br>Assumed gradual loss of efficacy for CBT by 20% | **SOURCES:**<br>Mortimer et al. (2006); Hestbaek et al. (2003); Enthoven et al. (2004)<br>Ten-year recurrence was 0.60 as 'reflected in literature' | Utilities derived from EQ-5D scores<br>**SOURCE:** Back Skills Training trial<br>Utilities assumed the same in respective states over 10 years as in a 1-year study | LBP Improved 0.640<br>LBP Not-improved 0.592 | **SOURCES:** Intervention costs from Back Skills Training trial<br>Healthcare costs for "similar patients" identified using the Ingenix Impact Research Database. Symmetry Episode Treatment Grouper used to identify medications associated with LBP |
| **Sciatica decision modelling studies** | | | | | |
| Launois et al. (1994) | Success, Deterioration<br>**SOURCES**: Literature review of various types of studies<br>Extrapolation based upon studies in literature review | Recurrences and re-operations<br>**SOURCES:**<br>6 studies identified in literature review | Utilities come from conversion of Rosser coefficients<br>**SOURCE:** Rosser coefficients from a "survey of 146 patients" who underwent chemonucleolysis and surgery | No utility values stated in the paper | **SOURCE:** Resource usage obtained from "the survey"<br>Administrative costs, with laboratory and radiology examinations added, unsourced |
| Lewis et al. (2011) | Success or failure of treatments<br>**SOURCE:** Systematic review of treatment | N/A | Annual utilities derived from 6-12 week EQ-5D scores<br>**SOURCE:** RCT (n=283) by van den Hout et al. (2008) | Sciatica - Improved 0.83<br>Sciatica - Not Improved 0.37 | **SOURCES:** Resource use based upon "clinical opinion from members of the clinical team" |

| | | | | | |
|---|---|---|---|---|---|
| | effectiveness for sciatica treatments. Pair-wise Meta-analysis and mixed-treatment comparison | | | | |
| Skidmore et al. (2011) | Successful treatment **SOURCES:** CC and X-STOP success from an RCT (n=131) (Zucherman et al., 2005). Success for laminectomy comes from literature | Re-op rate **SOURCES**: CC and X-STOP from RCT (Zucherman et al., 2005). Laminectomy from 'published literature' | Utility values derived from SF-36 **SOURCES:** Zucherman et al. (2005). | CC 0.61 – 0.65* XStop 0.62 – 0.79* Laminectomy 0.53 - 0.67* | **SOURCES:** Resource use from 'expert panel' estimates. |
| Fitzsimmons et al., (2014) | | See Lewis et al. (2011) | | | |
| Koenig et al. (2014) | Satisfaction with treatment **SOURCES:** Randomised observational study, the SPORT trial (n=743) (Weinstein et al., 2006; Weinstein et al., 2008) | Revision **SOURCES:** Three observational studies (Osterman et al., 2006; | **SOURCE:** Utilities come straight from an economic evaluation for treating herniated intervertebral disc (Malter et al. 1996), originally from the Beaver Dam health | Satisfactory outcome 0.89 Unsatisfactory outcome 0.56 Revision surgery 0.69 | **SOURCE:** Surgery frequency from 2009 Medicare claims database. Medical resource use from SPORT trial (Toteson et al., 2008) |

| | | | | | |
|---|---|---|---|---|---|
| | Extrapolations all based upon literature and fully sourced | Weinstein et al., 2006; Atlas et al., 2005). | outcomes study (Fryback et al. 1993). | | |
| Udeh et al. (2014) | Relief of symptoms<br>**SOURCE:** Unclear | Revisions<br>**SOURCE:** unclear | **SOURCES:** ESI QALY gains from Whynes et al. (2014). DS gains from an RCT (n=91) by Glassman et al. (2012) and trial (n=601) by Tosteson et al. (2011). Values reduced by 25% as patients in this study had 'severe' LSS. For 'mild' ODI scores from 4 trials (n=301) were converted to utility scores | Authors only provide QALY Gain | **SOURCES:** Resource use appears to be from 3 published studies in the literature |
| Igarashi et al. (2015) | Movements between health states<br>**SOURCE:** 8 week study by Taguchi et al. (2015) (n=331). Surgery risk from Medical Data Vision Co database (n=69,325) | Recurrence of symptoms in months 1-2<br>**SOURCE:** Taguchi et al. (2015) | NRS Pain scores from trial converted to utility values.<br>**SOURCE:** Taguchi et al. (2015)<br>Extrapolated 8 week pain scores to 52 weeks, citing a previous study as justification | CLBP with neuropathic component<br>No / mild pain 0.867<br>Moderate pain 0.739<br>Severe Pain 0.611 | **SOURCES:** Resource use from physician internet based survey of clinicians |

| | | | | | |
|---|---|---|---|---|---|
| Parker et al. (2015) | Success or failed treatment **SOURCES:** DS estimates from prospective spinal Registry. CC estimates from prospective study by Parker et al. (2014) (n=100). Spacer data from Spacer trial (Patel et al., 2014) (n=129) | N/A | Utility values derived from SF-36. **SOURCE:** DS estimates from prospective spinal Registry (n=129). CC estimates from Parker et al. (2014) (n=100). Spacer data from Spacer trial (Patel et al., 2014) (n=129) | Authors only provide QALY Gain | **SOURCES:** Resource use for follow up care for CC and DS patients collected by telephone interviews. Spacer assumed the same |
| Tapp et al. (2018) | Re-operation or complication **SOURCES**: Medicare Provider Analysis and Review database for complication and re-operation within 3 years [uncited]. Reoperation 4-10 years, for spacer expert opinion, and for | *Re-operation was the major treatment efficacy (see column left)* | Utility values are EQ-5D **SOURCES:** Utilities for CC, decompression, and fusion taken from pooled SPORT trial (Tosteson et al. 2008) and observational study results (Yano et al. 2008) (n=634). Spacer utility assumed equal to | Conservative Care / Pre Surgery 0.71 Post-surgery 0.77 Post - Major surgical complication 0.55 Major complication - 0.08 Non-major complication -0.04 | Costs stated directly, no resource use as such. **COST SOURCES:** CC costs assumed as zero for incremental purposes. Spacer and decompression surgical costs, as well as costs of complications taken directly from Medicare Provider |

| | | | | | |
|---|---|---|---|---|---|
| | decompression 4 cohort studies | | decompression. Disutility associated with complications based upon expert opinion. | Expert opinion | Analysis and Review database [uncited]. |
| **Sciatica decision modelling studies – surgical treatments** | | | | | |
| Kuntz et al. (2000) | Clinical improvement and fusion healing rate **SOURCES:** Mix of prospective and observational studies Extrapolation used literature and assumptions | Recurrence **SOURCE:** Assumptions and Deyo et al. (1993) Extrapolation used literature plus assumptions | Utility scores from time-trade-off technique **SOURCE:** (Fryback et al. 1996). | Symptoms of spinal stenosis 0.79 CLBP 0.79 Symptom free 0.97 | **SOURCE:** Previous study by Katz et al. (1997) who used a hospital cost accounting system, in one Boston hospital |
| Kim et al. (2012) | Clinical improvement or worsening, death, relapse **SOURCE:** Death rates from Deyo et al. (2010). All other transition probabilities from their observational study (n=150) | Reoperation **SOURCE:** Their observational study | Combined utility values from their study with literature **SOURCES:** Their observational study and observational study by Toteston et al. (2008) (n=601) Referenced a source suggesting outcomes achieved at 1-year are maintained for 4-years, authors then assume | Patients with lumbar spondylolisthesis Baseline 0.58 Decompression Improve 0.74 Decompression Not improve 0.50 | **SOURCE:** Costs derived from the authors hospital financial department |

| | | | | | |
|---|---|---|---|---|---|
| | | | utility is further constant over 10 years | Decompression Fusion Improve 0.74<br>Decompression Fusion Not Improve 0.54 | |
| Parkinson et al. (2012) | Success or failure of surgery **SOURCE:** Systematic review and meta-analysis of RCT's | Revision, Re-operation, other surgical outcomes **SOURCES**: Systematic review and meta-analysis of RCT's | Utilities derived from EQ-5D **SOURCE:** RCT (n=150) by Berg et al. (2009) | AIDR Pre-OP 0.42<br>AIDR @ 1 year 0.71<br>AIDR @ 2 years 0.67<br>PLF / PLIF Pre-Op 0.36<br>PLF / PLIF @ 1 year 0.63<br>PLF / PLIF @ 2 years 0.69 | **SOURCES:** Resource use based upon expert opinion, Medicare Benefits Schedule database, their SR, Australian department of Health publications |
| Schmier et al. (2014) | Clinical success **SOURCES:** Initial rates come from an RCT (n=322) by Davis et al. (2013). Extrapolated via published sources, Medicare data, and | Revisions and complications **SOURCES:** Published sources, Medicare | Utility scores converted from ODI scores **SOURCES:** The RCT by Davis et al. (2013) extrapolated using expert opinion. | Lumbar spinal stenosis<br>Clinical success 0.692<br>Clinical failure 0.552<br>New or worsening pain 0.599 | **SOURCES:** Expected treatment patterns derived from published sources, analysis of the Medicare Limited Data, and expert opinion |

| | | | | | |
|---|---|---|---|---|---|
| | expert opinion. 24 month treatment effect assumed the same continuously through five years | data, and expert opinion. Extrapolated using published sources | 24 month utilities assumed the same continuously through five years | | |
| Bydon et al. (2015) | Resolution of symptoms **SOURCE:** Retrospective data on 137 patients from a single institutional series | Re-operation rates **SOURCE:** The 137 patient institutional series | Utility values taken directly from Kuntz et al. (2000) **SOURCE:** Kuntz et al. (2000) from Fryback et al (1996). | Lumbar spondylolisthesis Positive outcome 0.97 Chronic back pain / Neurologic deficit 0.79 | **SOURCE:** The 137 patient institutional series, supplemented by Kuntz et al. (2000) |
| Vertuani et al. (2015) | No treatment effects as such. Their model appears more of an amalgamation of costs and QALY's | N/A | EQ-5D **SOURCE:** Swedish National Registry for Lumbar Spine Surgery Report 2008 (n=2437) | MIS for Spinal Fusion after 2-years 0.72 Open Surgery for Spinal Fusion after 2-years 0.68 | **SOURCE:** Resource use based upon systematic literature review and meta-analysis |
| Yaghoubi et al. (2016) | Success or failure of surgery **SOURCE:** Meta-analysis and SR | N/A | Reported as VAS scores **SOURCE:** Meta-analysis and SR | No utility values provided | **SOURCE:** Costs are derived directly from literature, "the bill of 30 patients in Tehran" and manufacturer costs |

Abbreviations: AE (Adverse events); CBT (Cognitive behavioural therapy); CC (Conservative care); CLBP (Chronic low back pain); DS (Decompression surgery); EQ-5D (EuroQoL-5D); ESI (Epidural steroid injections); GP (General Practitioner); LBP (Low back pain); LSS (Lumbar spinal stenosis); MIS (Minimally invasive surgery); NRS (Numerical rating scale); OA (Osteoarthritis); QALY (Quality-adjusted life year); ODI (Oswestry Disability Index); RCT (Randomised controlled trial); RMDQ (Roland Morris Disability Questionnaire); SF-36 (Short Form (36) Health Survey); SR (Systematic review); VAS (Visual Analogue Scales)

**Table 4.6a – Methods for Calculating Indirect costs**

| Author | Condition | Costs included | Costing methods | Sources |
|---|---|---|---|---|
| Wielage et al. (2013b) | LBP | Indirect costs owing to LOP from both LBP and treatment related AE's | Human capital approach; valued using No. of hours worked per week, minimum wage and retirement age | CLBP LOP from German study (Becker et al. 2010). AE LOP from guidelines for return to work (Work loss Data Institute, 2010). |
| Author | Condition | Costs included | Costing methods | Sources |
| Kim et al. (2010) | LBP | Direct non-medical costs (patient costs). | Indirect non-medical= Daily Wage * %Economically active | Direct non-medical from KNHNS (2007) data Indirect non-medical from Ministry of Employment and Labour |

| | | | | |
|---|---|---|---|---|
| | | Indirect non-medical costs (lost wages) | * %employed * No. treatment sessions | |
| Skidmore et al. (2011) | General sciatica | "Indirect costs of lost productivity is not estimated directly but instead is implicitly included incorporated in utility values" | | |
| Koenig et al. (2013) | General sciatica | Indirect costs arising from missed workdays and loss to household earnings | Inferred effect upon earnings and workdays from change in patient functional status | Missed workdays and income come from NHiS. Change in functional status from observation study (Weinstein et al. 2006). |
| Igarishi et al. (2015) | General sciatica | Indirect costs. "Work productivity component" of WPAI provides estimate of both productivity | % productivity was mapped to pain scores, each % change in pain score resulted in | WPAI adapted for LBP (WPAI:CLBP-NeP). Costs estimated based on mean monthly income in Japan. |
| Author | Condition | Costs included | Costing methods | Sources |
| Igarishi et al. (2015) cont. | | losses from absenteeism and | Estimate of productivity loss. | |

| | | | | |
|---|---|---|---|---|
| | | presenteeism<br><br>(Lofland et al. 2004) | | |
| Kuntz et al. (2000) | Sciatica surgery | Annual cost of lost<br><br>wages for patients | Unclear | Unclear |
| Abbreviations: AE (Adverse event); CLBP (Chronic Low back pain); KNHNS (Korean National Health and Nutrition Survey); LBP (Low back pain); LOP (loss of productivity); NHIS (National Health Interview Survey); WPAI (Work Productivity and Activity Impairment) | | | | |

### 4.3.7 Sensitivity Analyses

The scope of sensitivity analyses was limited. All studies aside from Launois and colleagues performed some form of deterministic sensitivity analysis. Most deterministic analyses addressed parameter uncertainty by varying at least one parameter, and most studies undertook analysis to address different forms of methodological and structural uncertainty.

Five studies attempted sub-group analysis. Wielage et al. (2013a; 2013b) ran their models for different age and risk groups; Igarashi et al. (2015) for patients with different initial pain level; Koenig et al. (2014) used different age groups and inpatient / outpatient treatment mix; and Skidmore et al. (2011) ran their model for various different patient age-groups.

Five studies (Wielage et al., 2013a; 2013b; Kim et al., 2010; Igarashi et al., 2015; Parker et al., 2015) attempted Probabilistic Sensitivity Analysis (PSA), with each study varying a significant number of parameters in their PSA. One study (Lloyd et al., 2004) bootstrapped costs and effects.

Nine studies attempted best case / worse case analysis. Two LBP studies (Norton et al., 2015; Lloyd et al., 2004) considered situations where their parameter values would produce the highest costs for the intervention. Of the non-surgical sciatica treatments, the HTA model engineered various best case / worst case cost scenarios, whilst Launois et al. (1994) looked at the effect of high / low estimates of treatment efficacy. Skidmore et al. (2011) considered the impact of increasing the utility of the comparator as well as if re-operation success with the X-STOP was lower. Udeh et al. (2014) investigated the impact of high utility values for comparators. Whilst Tapp et al. (2018) looked at the impact of a high cost scenario with a lower utility value scenario for surgery. Only one of the surgical treatment models conducted best case/ worst case analysis, Kim et al. (2012) considered

the effect upon cost-effectiveness of using extreme values of numerous different parameters, including utility values, clinical effectiveness and revisions.

Seven studies performed threshold analysis, four of which were non-surgical sciatica models. The HTA model investigated the required cost of non-opioids to alter the cost-effectiveness decision, they also identified the surgery success rate and the relative ratio of utility values required to change the result. Parker et al. (2015) considered the required utility value of spinal decompression such that the cost-effectiveness decision would change. Finally, Udeh et al. (2014) considered how many ESI's would need to be administered per-treatment course, as well as what degree of QALY gains would be required, to alter the cost-effectiveness decision. Three models for surgical treatments considered threshold analysis. Schmier et al. (2014) claimed to have run various threshold scenarios although there was no reporting of the results. Parkinson et al. (2012) estimated the relative treatment success required to change the decision. Vertuani et al. (2015) investigated what level of cost of minimal invasive surgery would need to take to alter the cost-effective result.

Only one study attempted to perform Value of information Analysis (Kim et al., 2010). In their analysis they attempted to place a value on the benefits likely to accrue from further research into the cost-effectiveness of acupuncture. At the generally accepted generally accepted societal threshold for willingness to pay at 20,000,000 KRW per QALY, they estimate a population EVPI of 120,000,000,000 KRW, around £80m in 2020 prices. This appears to be low, given the size of their population and model horizon, but the exact parameters of the calculation are not provided.

In terms of the different kinds of analysis, studies generally assessed the uncertainty associated with three main areas. As discussed in the chapter on health economic modelling, 'Parameter uncertainty' is the name given to the uncertainty the analyst has

regarding the real value of any of the input parameters in the model, such as transition probabilities, costs or utility values (Mahon, 20124). As noted, many of the analyses in these studies address parameter uncertainty to some degree, although few performed probabilistic analysis.

Methodological uncertainty generally refers to normative choices regarding which evaluative approach optimises decision making, for example choice of perspective, of costs, or discount rate (Bilcke et al., 2011). In this review it was common to vary in a univariate manner the cost perspective (Kim et al., 2010; Fitzsimmons et al., 2014; Lewis et al., 2011; Kim et al., 2012; Schmier et al., 2014).

'Structural uncertainty' refers to the appropriateness of what is imposed by the model framework (Bojke et al., 2009). Common approaches in this regard were to explore different assumptions regarding the treatment pathway or setting in which treatment was delivered (Lloyd et al., 2004; Fitzsimmons et al., 2014; Lewis et al., 2011; Skidmore et al., 2011; Parker et al., 2015; Kim et al., 2012; Schmier et al., 2014; Vertuani et al. (2015).

**Table 4.7 Sensitivity Analysis**

| Author | Deterministic | Sub-group Analysis | PSA | Best case / Worst case | Threshold Analysis | Value of Information |
|---|---|---|---|---|---|---|
| **Low back pain decision modelling studies** | | | | | | |
| Lloyd et al. (2004) | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ |
| Kim et al. (2010) | ✓ | ✗ | ✓ | ✗ | ✗ | ✓ |
| Wielage et al. (2013a) | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ |
| Wielage et al. (2013b) | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ |
| Norton et al. (2015) | ✓ | ✗ | ✗ | ✓ | ✗ | ✗ |
| **Sciatica decision modelling studies – general treatments** | | | | | | |
| Launois et al. (1994) | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ |
| Lewis et al. (2011) | ✓ | ✗ | ✗ | ✓ | ✓ | ✗ |
| Skidmore et al. (2011) | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Fitzsimmons et al. (2014) | ✓ | ✗ | ✗ | ✓ | ✓ | ✗ |
| Koenig et al. (2014) | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ |
| Udeh et al. (2014) | ✓ | ✗ | ✗ | ✓ | ✓ | ✗ |

| Author | Deterministic | Sub-group Analysis | PSA | Best case / Worst case | Threshold Analysis | Value of Information |
|---|---|---|---|---|---|---|
| Igarashi et al. (2015) | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ |
| Parker et al. (2015) | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ |
| Tapp et al. (2018) | ✓ | ✗ | ✗ | ✓ | ✗ | ✗ |
| **Sciatica decision modelling studies – Surgical populations** | | | | | | |
| Kuntz et al. (2000) | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Kim et al. (2012) | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Parkinson et al. (2012) | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ |
| Schmier et al. (2014) | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ |
| Bydon et al. (2015) | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ |
| Vertuani et al. (2015) | ✓ | ✗ | ✗ | ✗ | ✓ | ✗ |
| Yaghoubi et al. (2016) | ✓. | ✗ | ✗ | ✗ | ✗ | ✗ |

### 4.3.8 Model Validity Checks

Table 4.8 highlights that models within this review did not commonly perform validity checks. Only two studies (Kim et al., 2010; Kim et al., 2012) reported internal validity, noting that they had tested their model output to check that it was sensible. Bydon et al. (2015) offered brief reflection upon whether aspects of their model results were realistic.

The quality appraisal dimension for validity (Philips dimension C1-C2 (iii) in Table 4.9), show that where validity checks were performed they were rarely done in accordance with best practice. Adequate discussion of the standard academic practice of comparing study results with previous studies occurred in only 13 out of 21 studies. Tapp et al. (2018) perhaps provided the best example of this, comparing their results with other economic evaluations in this area, as well as considerable exploration in the discussion regarding re-operation rates in other studies.

Only one study (Kim et al., 2012) provided discussion of external and internal validity, comparing their results with other models and checking that the output of their model produced realistic results.

**Table 4.8 Validity checks**

| Low back pain decision modelling studies | | |
|---|---|---|
| Author | External Validity | Internal Validity |
| Lloyd et al. (2004) | Unreported | Unreported |
| Kim et al. (2010) | Authors acknowledge the absence of other modelling studies, and therefore compare their results with those of other trial-based evaluations. | Output produced by the usual care option matched with real data on chronic low back pain |
| Wielage et al. (2013a) | Unreported, but acknowledge this is due to absence of comparative information. | Unreported |
| Wielage et al. (2013b) | Unreported, but acknowledge this is due to absence of comparative information. | Unreported |
| Norton et al. (2015) | Compared to previous trial-based evaluations. | Unreported |
| **Sciatica decision modelling studies – general treatments** | | |
| Launois et al. (1994) | Unreported | Unreported |
| Lewis et al. (2011) | Unreported | Unreported |

| Author | External Validity | Internal Validity |
|---|---|---|
| Skidmore et al. (2011) | Comparison with other modelling studies | Unreported |
| Fitzsimmons et al. (2014) | Unreported | Unreported |
| Koenig et al. (2014) | Comparison with other trial-based evaluations. | Unreported |
| Udeh et al. (2014) | Unreported | Unreported |
| Igarashi et al. (2015) | No prior models. Have compared to previous trial-based evaluations. | Unreported |
| Parker et al. (2015) | Results compared with one other model based evaluation, as well as to trial-based evaluations. | Unreported |
| Tapp et al. (2018) | Considerable exploration in the discussion regarding re-operation rates in other studies. | Unreported |
| **Sciatica decision modelling studies – surgical populations** | | |
| Kuntz et al. (2000) | Some discussion of counter-intuitive results. | Unreported |
| Kim et al. (2012) | Extensive comparison with both model and trial based evaluations. Compared their results with deviation from previous findings. | Compared recurrence with Canadian recurrence rates. Utilities and re-operation rates compared with real world data to ensure mathematical logic of model was correct. |

| Author | External validity | Internal validity |
|---|---|---|
| Parkinson et al. (2012) | Some discussion of previous trial-based evaluations. | Unreported |
| Schmier et al. (2014) | Unreported | Unreported |
| Bydon et al. (2015) | Unreported | Limited discussion of whether values in the model are realistic |
| Vertuani et al. (2015) | Unreported | Unreported |
| Yaghoubi et al. (2016) | Vague comparison with other studies and data sources | Unreported |

### 4.3.9 Quality appraisal

Table 4.10 demonstrates significant inconsistency in the overall quality of modelling papers in this review. The 'Total' column on the right-hand side of Table 4.10 represents a score out of twenty (one point for each of the twenty papers) for each dimension of the Philips checklist. In order to score 20 each Philips dimension must have been satisfied sufficiently in every paper in the review. In the table, full compliance with the Philips criteria is represented by the '✓'which scores one point, partial compliance receives a '~' and scores half of a point, and a '✗' scores zero points and represents a non-compliant or poorly compliant study.

Generally, the structural components of the modelling were of reasonable quality. Indeed for Philips dimensions concerned with structure (S1a-S9) there are only six dimensions which at least 50% of papers failed to meet adequately; S1c, stating the primary decision maker; S3b, sources used to developed the structure of the model; S5b, all feasible and practical options evaluated; S5c, justification of exclusion of all feasible options; S7a, sufficient time horizon; and S9, cycle length stated and justified in terms of the condition.

The dimensions relating to data are of a much lower quality. Data identification methods (D1a-D1e) scored badly, especially assessment of data quality (D1d) which scored only full compliance in four studies. The scoring for baseline data (D2a) was especially poor driven by the omission of half cycle correction (D2a(iii)) and lack of justification for doing so (D2a(iv)). Treatment effect (D2b) scores were low owing to poor documentation and justification of extrapolation methods (D2b(ii)), continuing treatment effects (D2b(iv)) and exploring of alternate assumptions in subsequent sensitivity analysis (D2b(iii) and D2b(v)). For reasons discussed in the previous section, the utility values used in the models (D2d(i)) are rarely considered fully appropriate, despite all four LBP models using appropriate values. Indeed, none of the sixteen sciatica models used entirely appropriate utility values.

Costing transparency, sourcing, and discounting (D2c) scored well generally, although as reported in Section 5.3.6 few studies used real world clinical data to obtain resource use.

Some of the lowest scores in the review came for the characterisation of uncertainty (D4). Only four studies (Wielage et al., 2013a; 2013b; Koenig et al., 2004; Skidmore et al., 2011) scored adequately in barely half of the dimensions for uncertainty. Dimensions concerned with internal and external validity (C1-C2C) scored even lower as noted in the previous sub-section.

## Table 4.9 Quality Appraisal

| | Low Back Pain Models | | | | | Sciatica non-surgical treatments | | | | | | | | Sciatica surgical treatments | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Philips Dimension | Wielege (2013a) | Wielege (2013a) | Norton | Lloyd | Kim et al. (2010) | Fitzsimmons | Lewis | Igarashi | Launois | Koenig | Skidmore | Parker | Udeh | Kuntz | Bydon | Kim et al. (2012) | Yaghoubi | Schmier | Parkinson | Vertuani | Total |
| S1a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | **20** |
| S1b | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | **20** |
| S1c | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ | ~ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | **3.5** |
| S2a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | ✗ | ✓ | ✗ | ✓ | **15** |
| S2b | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | **?** | ✓ | **17/19** |
| S2c | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ~ | ✓ | ✗ | ✓ | ✗ | ~ | **17** |
| S2d | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ~ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | **18.5** |
| S3a | ✓ | ✓ | ✓ | ~ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ~ | ✓ | ~ | ✓ | ~ | ✗ | ✓ | ~ | ✗ | **15.5** |
| S3b | ✓ | ✓ | ~ | ~ | ~ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ | ~ | ✗ | ~ | ✓ | ✓ | ✗ | ~ | ✗ | ✗ | **10** |
| S3c | ✓ | ✓ | ✗ | ✓ | ~ | ✓ | ✓ | ~ | ✗ | ✓ | ✓ | ~ | ✓ | ~ | ✓ | ✓ | ✗ | ✗ | ✓ | ~ | **13.5** |
| S4a | ✓ | ✓ | ✗ | ✓ | ~ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ~ | ~ | ~ | ✗ | ✗ | ✓ | ✗ | **13** |
| S4b | ✓ | ✓ | ✗ | ~ | ~ | ✓ | ✓ | ~ | ✗ | ~ | ✓ | ~ | ✓ | ~ | ~ | ~ | ✗ | ✓ | ~ | ✗ | **11.5** |
| S5a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | **18** |
| S5b | ✓ | ~ | ✗ | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ | ~ | ~ | ~ | ✓ | ✗ | ✗ | ✗ | ✗ | ~ | ~ | ✗ | **8** |
| S5c | **N/A** | ✓ | ✗ | **N/A** | ✗ | **N/A** | **N/A** | ✗ | ✓ | ✗ | ✗ | ✗ | ~ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | **5.5/16** |
| S6 | ✓ | ✓ | ✓ | ✓ | ✓ | ~ | ~ | ✓ | ~ | ✓ | ~ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ | **16** |
| S7a | ✓ | ✓ | ~ | ✗ | ~ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ~ | ~ | ~ | ✗ | ✗ | ✗ | ✗ | **4.5** |
| S7b | ✓ | ✓ | ✓ | ✗ | ✓ | ~ | ~ | ~ | ~ | ~ | ✓ | ✓ | ✓ | ~ | ✓ | ✓ | ✗ | ~ | ~ | ~ | **13.5** |
| S8 | ✓ | ✓ | ~ | ~ | ~ | ✓ | ✓ | ~ | ✓ | ✗ | ✓ | ✓ | ✓ | ~ | ✓ | ~ | ✗ | ~ | ✓ | ✗ | **13.5** |
| S9 | ✓ | ✓ | ✗ | **N/A** | ~ | **N/A** | **N/A** | ✓ | ✗ | ~ | ✗ | ~ | ✗ | ✗ | ✗ | ✓ | ✗ | ~ | ~ | ✗ | **6/17** |

| | Low Back Pain Models | | | | | Sciatica non-surgical treatments | | | | | | | | Sciatica surgical treatments | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Philips Dimension | Wielege (2013a) | Wielege (2013a) | Norton | Lloyd | Kim et al. (2010) | Fitzsimmons | Lewis | Igarashi | Launois | Koenig | Skidmore | Parker | Udeh | Kuntz | Bydon | Kim et al. (2012) | Yaghoubi | Schmier | Parkinson | Vertuani | Total |
| D1a | ✓ | ✓ | ✓ | ~ | ✓ | ✓ | ✓ | ✗ | ✓ | ~ | ~ | ✗ | ✗ | ✗ | ✗ | ~ | ✓ | ✓ | ~ | ~ | **12** |
| D1b | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ~ | ✗ | ✗ | ✗ | ✗ | ~ | ✓ | ✓ | ✗ | ✗ | **9** |
| D1c | ✓ | ✓ | ~ | ✗ | ✓ | ✗ | ✗ | ✗ | ~ | ~ | ✓ | ~ | ✗ | ✗ | ✗ | ✓ | ~ | ✓ | ~ | ✗ | **9** |
| D1d | ~ | ~ | ~ | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ~ | ✗ | ✗ | ~ | ~ | ✓ | ✗ | ~ | **7.5** |
| D1e | ✓ | ✓ | **N/A** | **N/A** | **N/A** | ✓ | ✓ | ✗ | **N/A** | **N/A** | ✓ | N/A | N/A | **N/A** | **N/A** | **N/A** | **N/A** | ~ | ~ | N/A | **6/8** |
| D2 | ✓ | ✓ | ✓ | ✓ | **?** | ✓ | ✓ | ~ | ~ | ✓ | ✗ | ~ | ✓ | **?** | ✓ | ~ | ✓ | **?** | ✗ | ✗ | **12.5/17** |
| D2a(i) | ✓ | ✓ | ~ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ~ | ✓ | ~ | ✗ | ~ | ~ | ~ | ✓ | ✓ | ~ | ~ | **13** |
| D2a(ii) | **?** | **?** | **?** | **?** | ✓ | ✓ | ✓ | **?** | ✓ | **?** | **?** | **?** | **?** | ✗ | **N/A** | ? | ✗ | **?** | **?** | ? | **4/6** |
| D2a(iii) | ✗ | ✗ | ✗ | N/A | ✗ | N/A | N/A | ✗ | N/A | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | **0/16** |
| D2a(iv) | ✗ | ✗ | ✗ | N/A | ✗ | N/A | N/A | ✗ | N/A | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | **0/16** |
| D2b(i) | ✓ | ✓ | N/A | N/A | ✓ | ✓ | ✓ | N/A | ✓ | N/A | N/A | ✓ | ✗ | ✗ | ✗ | N/A | ✓ | N/A | ✓ | N/A | **9/12** |
| D2b(ii) | ~ | ~ | ✓ | ✗ | ✗ | N/A | N/A | ~ | ~ | ✗ | N/A | N/A | N/A | ~ | ✗ | ~ | ✗ | ✓ | N/A | ✗ | **5/14** |
| D2b(iii) | ? | ? | ✓ | ✗ | ✗ | N/A | N/A | ✗ | ✗ | ✗ | N/A | N/A | N/A | ~ | ✗ | ~ | ~ | ✗ | N/A | ✗ | **2.5/12** |
| D2b(iv) | ~ | ~ | ✗ | ✗ | ✗ | N/A | N/A | ~ | ✓ | ✗ | N/A | N/A | N/A | ~ | ✗ | ✓ | ✗ | ✗ | N/A | ✗ | **4/14** |
| D2b(v) | ? | ? | ✓ | ✗ | ~ | N/A | N/A | ✗ | ✗ | ✗ | N/A | N/A | N/A | ~ | ✗ | ~ | ~ | ✗ | N/A | ✗ | **4/12** |
| D2c(i) | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ~ | ~ | ✓ | ~ | ✓ | ~ | ✓ | ✗ | ~ | ✗ | ✓ | ✓ | **14.5/20** |
| D2c(ii) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | **20** |
| D2c(iii) | ✓ | ✓ | ✓ | N/A | ✓ | N/A | N/A | N/A | ✓ | ✓ | ~ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✓ | ✗ | **13/16** |
| D2d(i) | ✓ | ✓ | ✓ | N/A | ✓ | ✗ | ✗ | ~ | ~ | ✗ | ~ | ~ | ~ | ✗ | ✗ | ~ | N/A | ~ | ✗ | ? | **7.5/17** |
| D2d(ii) | ✓ | ✓ | ✓ | N/A | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | N/A | ✓ | ✓ | ✓ | **17/18** |

| | Low Back Pain Models | | | | | Sciatica non-surgical treatments | | | | | | | | Sciatica surgical treatments | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Philips Dimension | Wielege (2013a) | Wielege (2013a) | Norton | Lloyd | Kim et al. (2010) | Fitzsimmons | Lewis | Igarashi | Launois | Koenig | Skidmore | Parker | Udeh | Kuntz | Bydon | Kim et al. (2012) | Yaghoubi | Schmier | Parkinson | Vertuani | Total |
| D2d(iii)j | ✓ | ✓ | ✓ | N/A | ✓ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ~ | ✗ | ✓ | N/A | ✓ | ✗ | ✗ | **10/18** |
| D3a | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ~ | ✓ | ~ | ~ | ✓ | ✓ | ✓ | **18.5/20** |
| D3b | ✓ | ✓ | N/A | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ~ | ✓ | ~ | ✗ | ✓ | ✓ | ~ | ✓ | ✓ | ? | **11.5/18** |
| D3c | ✗ | ✗ | ✓ | ~ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ~ | ✓ | ✓ | ✓ | ~ | ~ | ✗ | ✓ | ✓ | **14** |
| D3d | ✗ | ✗ | ✗ | N/A | ~ | ✓ | ✓ | ~ | N/A | N/A | **N/A** | ~ | ~ | **N/A** | N/A | N/A | **N/A** | N/A | N/A | N/A | **4/9** |
| D3e | ✗ | ✗ | ✗ | N/A | ✓ | N/A | N/A | ~ | N/A | N/A | N/A | ~ | ~ | **N/A** | N/A | N/A | N/A | N/A | N/A | N/A | **2.5/7** |
| D4(i) | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ~ | ✗ | ✓ | ✓ | ~ | ~ | ✗ | ✗ | ~ | ✗ | ✗ | ✗ | ✗ | **6** |
| D4(ii) | N/A | N/A | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | N/A | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | **1/17** |
| D4a | ✓ | ✓ | ✓ | ✗ | ~ | ✗ | ✗ | ✓ | ✗ | ✓ | ✓ | ✗ | ~ | ✗ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ | **9** |
| D4b | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | **7** |
| D4c | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ | **7** |
| D4d(i) | ~ | ~ | ✗ | ~ | ✓ | ~ | ~ | ~ | ✗ | ✗ | ~ | ✓ | ✓ | ~ | ✗ | ~ | ✗ | ✗ | ~ | ✗ | **8** |
| D4d(ii) | ~ | ~ | ✗ | ✓ | N/A | ✗ | ✗ | ✗ | ~ | ~ | ~ | N/A | ~ | ~ | ✗ | ✗ | ✗ | N/A | ✗ | ~ | **5/17** |
| C1 | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ~ | ✓ | ✗ | ✗ | ✗ | ✗ | **2.5** |
| C2a | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | ~ | N/A | N/A | N/A | N/A | ✓ | N/A | ✓ | ✗ | ✗ | N/A | N/A | **2.5/5** |
| C2b | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ~ | ✗ | N/A | N/A | N/A | N/A | ✗ | ✓ | ✓ | ✗ | ✗ | N/A | **5.5/15** |
| C2c | ~ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ | ✓ | N/A | ✓ | ✓ | ✓ | ✗ | N/A | ✗ | ✓ | ~ | ✗ | ✓ | ✗ | **9/18** |

## 4.4 Discussion

### 4.4.1 Statement of principal findings

### 4.4.2 LBP models

#### 4.4.2.1 Model structure

Choice of time horizon should be justified, and reflect important differences between comparators not be driven by data availability (Philips et al., 2006; Jaime-Caro et al., 2012; Karnon et al., 2003; Sibert et al., 2012). The two models by Wielage et al. (2013a; 2013b) were the only models to use the lifetime horizon, possibly because data were readily available for their specific decision problem. For the other three studies, the use of medium-term time horizons could be permissible, given that many treatments for LBP do not provide a long-term treatment effect.

In terms of model type, four LBP models used Markov state transitions, suggesting that these authors considered this approach appropriate for modelling LBP. However, in situations in which an early event or a patient characteristic determines future patient pathways, and the number of health states required might be unwieldy, individual-level microsimulation methods should be considered (Philips et al., 2006).

Health states should adequately reflect the condition-specific health processes (Philips et al., 2006). On this basis, the three-state approach used by Norton et al. (2015), i.e. ‘improved’, ‘not improved’ and ‘dead’ might be considered an oversimplification. While using ‘improved’ and ‘not-improved’ will to some degree always incorporate the pain associated with each treatment, unless utility values are always collected separately for each treatment, interventions that deliver higher rates of long-term improvement at a lower level of utility will be advantaged by using health states reflecting ‘improvement’.

For LBP, an approach starting from the structure used by Kim et al. (2010) could be the most appropriate, the authors chose health states reflecting a temporal classification of LBP, i.e. ‘acute LBP’, ‘chronic LBP’, ‘well’ and ‘dead’. However, the structural validity of their particular model is limited by the structural assumption that one recurrence of back pain in the ‘well’ state can move the patient into a ‘chronic LBP’ state. This leads to the possibility that a patient with only two episodes of LBP across 5 years would be considered to have ‘chronic LBP’. Nonetheless, the modelling of degrees of symptoms is preferred over a dichotomy as it is likely to produce results that better reflect the patient experience. It is also worth noting that it is likely there is heterogeneity in pain severity within temporal classifications.

Current research evidence suggests that a potentially more appropriate categorisation of patients with LBP could reflect pain severity as well as the rate of recurrence (Dunn et al., 2013). For example, Dunn et al. (2013) show that pain level after 1 year is predictive of pain level at 7 years, and patients are categorised into three groups, according to persistence and severity, i.e. ‘no or occasional pain’, ‘persistent mild pain’, and ‘persistent severe pain’, with a fourth category used for those who show no consistent pattern. Given the relative consistency of symptoms over time it is perhaps surprising more studies have not engaged in extrapolation over longer-periods.

Ultimately, model structure and health state selection should involve consultation with subject experts and stakeholders (Philips et al., 2006). While experts were clearly involved in the construction of these models, none of the studies clearly justified or discussed issues related to their model structure. The lack of discussion around choice of health states, or model choice and time horizon, is problematic for improving the representation of both conditions in model form as such subjective components of the modelling process should

be predicated on a clear understanding and subsequent critique of the principles upon which such decisions are made.

#### 4.4.2.2 Model data

High-quality information on resource use in LBP patients does not seem to be available; however utility values for the LBP models were obtained in accordance with best practice (Kim et al., 2010; Norton et al., 2015; Launois et al., 1994; Lloyd et al., 2004).

Four LBP models extrapolated over a longer period than their data allowed, although only Wielage et al. (2013a; 2013b) attempted lifetime extrapolation. Having evidence available to them, both studies modelled across the lifetime by adjusting the AE profiles according to age-dependent relative risks, and used age-dependent utility values directly obtained from the available literature.

However, often in decision problems relating to LBP, evidence of long-term treatment effect is likely to be unavailable because trials commonly only span across 1 year of patient observation. Nonetheless, Norton et al. (2015) and Kim et al. (2010) show how an incomplete evidence base can still be extrapolated in a decision analytic model by using assumptions and expert opinion in addition to the literature, and offer methods which can be used to guide the necessary assumptions to be made in the models in this thesis. Both studies defined their health states independently of treatment, and established utility values for those specific health states, therefore requiring only information on the long-term movement of patients between health states. Lacking data on long-term treatment efficacy, Norton et al. (2015) assumed a gradual loss in efficacy (resolution of symptoms) over time of 20% per annum. Similarly, Kim et al. (2010) extrapolated short-term treatment efficacy, assuming short-term relative risks between the treatment arms remained constant over time.

Parameter behaviour over time, such as the long-term treatment effect, often represents the largest source of uncertainty within a model (Mahon et al., 2014). Accordingly, studies performing extrapolation should be expected to undertake rigorous examination of temporal uncertainty (Philips et al., 2006). However, none of the five LBP models really addressed temporal uncertainty, despite the importance of extrapolation assumptions used by Norton et al. (2015) and Kim et al. (2010). Consequently, the reader is left without an understanding of how uncertainty over the long-term treatment effect could impact the cost effectiveness of these treatments for LBP.

### 4.4.3 Sciatica non-surgical treatment models

#### 4.4.3.1 Structure

Three studies used Markov modelling and three used decision trees. However, given that surgery was a comparator in all models, and would be expected to improve long-term outcomes for sciatica patients, short time horizons modelled within a decision tree may seem unsuitable in this condition. The use of individual sampling models (ISM) could be of real value in this condition, given that candidacy for surgery is likely to be event-dependent, e.g. having a failed previous treatment, and/or time spent in severe pain.

Treatment guidelines for sciatica in the UK follow a stepped pathway, in that patients can receive more invasive treatments dependent on prior treatment failures. Most models did allow between one and three stepped treatment failures before they allow surgery to take place (e.g. Lewis et al., 2011; Launois et al., 1994; Igarashi et al., 2015; Skidmore et al., 2011; Parker et al., 2015).Lewis et al. (2011) and Fitzsimmons et al. (2014) not only presented the stepped nature of the treatment pathway but also provided 100 different treatment combinations. Yet, the complexity of the representation of the treatment pathway appears in contradiction to the simplicity of the estimates used for the utility values, which in turn may limit the validity of the study results. While there is clearly a need to represent

the stepped pathway when modelling a presurgical period of the treatment pathway, it is doubtful that a typical model should consider all combinations of possible treatments unless that is the specific goal of the analysis, as was the case in the Lewis et al. (2011) and Fitzsimmons et al. (2014) publications.

If the aim of the analysis is to compare an intervention with usual care, then it might be more efficient to have a comparator that represents common/usual practice. Indeed, most models only compared two or three treatment options, one of which included conservative care. Two studies (Skidmore et al., 2011; Parker et al., 2015) used a comparator that reflected a specific combination of usual care treatments based on an observed combination of treatments. For example, Skidmore et al. (2011) used 'conservative care' as their comparator, defined as at least one ESI, supplemented by non-steroidal anti-inflammatory drug [NSAIDs], oral steroids, analgesics, physical therapy, or spinal manipulation therapy. Their data on treatment were derived from a trial where patients received usual care, as considered appropriate for the individual patient. Finally, Parker al. (2015) refer to conservative care as physical therapy, pain medications (NSAIDs, mild opioids), and ESIs, as guided by clinical judgement of the treating physician. Their data were derived from analysis of institutional registry data, and trial data.

The approach used by the Institute for Clinical and Economic Review (2011) in its model of treatments for lumbar disc herniation, even though grey literature are not included in the review, is nonetheless worthy of consideration. It uses a Markov model that allows patients to continue on some specific combinations of usual care treatments (identified by systematic review), and also allows a specific proportion of the cohort to move into receive a discectomy. Upon receiving a surgical procedure, patients can then improve, receive a second reoperation, or suffer a complication.

Assuming that the model requires the entire pathway to be modelled, we would advocate taking the best of these approaches, by representing the stepped nature of the treatment pathway by initially defining usual care as the treatment combination which evidence shows that sciatica patients initially receive. If the model took the form of an ISM or a discrete event simulation, it would be possible to use tracker variables to track the length of time a patient remained symptomatic, and/or received the described initial usual care. After some defined period of non-improvement, a patient could either become eligible for a more intense usual care treatment option (a second step), and/or upon failure of that second treatment become eligible for surgery. As well as time spent in the receipt of usual care, a patient's candidacy for surgery could also be a function of the unique characteristics of that patient, although the degree of complexity will be limited by available data.

With regard to the choice of health states, similar to the LBP models, all but one model (2015) used health states relating to treatment success. The use of health states such as 'improved' and 'not improved' are perhaps more appropriate in models for sciatica, in which perhaps interventions such as spinal injections and/or surgery might be expected to provide a more pronounced and sustained treatment response, on average, when compared with treatment effects for non-specific LBP. However, as noted above, with respect to modelling, if such health states are used, this will necessitate that utility values are collected for each intervention as some interventions could deliver their improvement at very different utility levels. None of the models in this analysis undertook this endeavour. A related issue is that all of the studies use different definitions of success or improvement, making comparison between studies difficult.

Guidance suggests that where health states reflect the treatment pathway effect, this ought to be justified, and alternative methods of doing so explored in a sensitivity analysis (Mahon et al., 41); however, none of the studies justified their approach or explored

alternate methods. The only model in this review to provide health states based on pain severity (Igarashi et al., 2015) provides a starting place for how pain severity could also be used to conceptualise health states for sciatica. No pain/mild pain, moderate pain and severe pain are used as health states. Nonetheless, the inability to move between health states during a period of 3–11 months and the model's short 12-month timeframe, somewhat diminishes its potential as an example of best practice if it is accepted that the analytical framework ought to be over a longer period. While 'improved' or 'success' models are likely to be conceptually linked with pain severity to some degree, the potential implications for cost-effectiveness estimates of using an 'improved'- or 'success'-based model over one based on pain states could be one area of future research.

#### 4.4.3.2 Data

Only three of nine models attempted extrapolation, although those that did can offer insights into potential methods. For example, Tapp et al. (2018) extrapolated from 4 to 10 years by determining the long-term rate of reoperation for decompression from the literature, and lacking data on long-term reoperation with the spacer, assuming the rate was identical to decompression. They also used the same utility values over time, and assumed that complication rates were the same regardless of first surgery or reoperation. However, their study cannot be held as an example of best practice given their cost estimates are unlikely to be representative, as they do not include costs of other treatments aside from surgery, and conservative care and post-surgical care carry zero costs.

The fact that only three of nine non-surgical sciatica models attempted extrapolation was typically justified by the claims made by Fitzsimmons et al. (2014) of a "lack of evidence regarding relapse and recurrence rates" making "it difficult to extend the analysis beyond this time period". It is certainly the case that, where estimates extend well beyond known available data, the accuracy of the estimates may be questionable. However, as stated

above, best practice guidelines in decision analytic modelling state that data availability should not define the time horizon of the model (Philips et al., 2006). The apprehension regarding extrapolation is perhaps somewhat unfounded given that two of the LBP studies above (Kim et al., 2010; Norton et al., 2015), and one sciatica model (Tapp et al., 2018), made simple assumptions regarding the long-term treatment effect. Moreover, there are many techniques to infer values for unobserved model parameters (Mahon et al., 2014). For example, Mahon and colleagues advocate fitting parametric functions through statistical methods, and/or using expert opinion to derive some probabilistic assessment of the likelihood a parameter function takes a certain shape. Nothing approaching this level of sophistication was attempted in any of these papers in this review. And moreover, as was the case with the LBP models, none of the models undertaking extrapolation attempted even a basic assessment of temporal uncertainty.

Given the productivity costs associated with both LBP and particularly sciatica, it might be expected to see more analyses performed from a societal perspective, at least as a secondary evaluation. However, as shown in Table 4.3, and detailed in Table 4.6a, in this review only six models performed some form of societal analysis, which were of varying rigour. The most detailed method (Igarashi et al., 2015) used the established methodology of the WPAI scale (Lofland and Pizzi, 2004), a validated method to assess lost productivity that assesses the losses due to both absenteeism and presenteeism, although, for a full societal analysis, other non-medical costs should also be included, as per the paper by Kim et al. (2010).

The utility values used in this class of models reflect to some degree the lack of availability of utility data for sciatica patients. The methods used by Lewis et al. (2011) and Fitzsimmons et al. (2014) to calculate utility values show why it is not necessarily advisable to have values based on the success or failure of a treatment because an

appropriate analysis would require many unique utility values. As this information is clearly not available, the authors used the same utility values for 'success' or 'failure' of all treatments in their study, which would bias against those treatments that delivered higher utility gains where successful, and/or treatments that minimise utility losses where they fail. Moreover, while their value of 0.83 for treatment improvement seems consistent with other studies (Igarishi et al., 2015; Koenig et al., 2014), their 'non-improvement' score of 0.37 is significantly lower than values used for non-improvement in most other studies in this review. This highlights the need for consistent health state selection and definition in the models. Admittedly, the authors of both studies (Lewis et al., 2011; Fitzsimmons et al., 2014) acknowledged the problems with utility values, which they attempted to address using alternate scenarios in sensitivity analyses, although the results of these analyses were not presented, and differential EQ-5D scenarios between treatments were not tested.

There were also concerns in relation to the derivation of resource use estimates, which ought to originate from real clinical practice instead of clinical trials (Soto et al., 2002). However, for the sciatica non-surgical treatments, only Tapp et al. (2018) obtained some of their cost estimates from a database reflecting actual patient healthcare use, and half had to rely on expert opinion, suggesting a problem with the availability of high-quality resource use information. This is understandable given how challenging it can be to identify and attribute visits when conditions may be mentioned only incidentally.

### 4.4.4 Sciatica surgical treatment models

#### 4.4.4.1 Structure

While it is commendable that the time horizons considered by the surgical intervention models were longer than the non-surgical models, this group of models was replete with methodological problems. Of the models with identifiable structure, five of seven were Markov models, which are recommended for decision problems, such as modelling

surgical procedures, where model horizon is longer, and/or the model contains time-varying transition probabilities (Roberts et al., 2012). In the case of models that begin from a surgical process, a Markov model is sufficient, provided it allows the possibility of future reoperation following the recurrence of symptoms. The need for an ISM is less pressing than in the case of a sciatica model including a preoperative period, although allowing a risk of recurrence to be dependent on individual characteristics would offer the model more flexibility.

Given that all comparators in this group were surgical procedures and did not feature conservative care, health states generally are reflective of the 'success' or 'failure' of surgery rather than health-specific processes. As noted above, there is still a need among these models to use consistent health states and definitions for comparability. The structure used by Kim et al. (2012) could provide a basic template for developing a more sophisticated model structure. Their model health states were potentially more appropriate because of the use of health states such as 'well', 'unwell' or 'return of symptoms', which partially reflect the condition-specific health processes, and captured all important *events* for sciatica patients, e.g. relapse, reoperation (including the flexibility of choosing whether or not to reoperate), clinical worsening and improvement, and general and perioperative death.

#### 4.4.4.2 Data

The Kim et al. (2012) structure can also provide a framework within which to engage in extrapolation processes. Using the health states of 'well', 'unwell', or 'return of symptoms', they then derived long-term utility values associated with these states, independent of treatment approach, from their 1-year observational study. To justify the use of these 1-year values across 10 years, Kim et al., referenced a study by Weinstein et

al. (2009), who suggested that utility values derived at 1 year are relatively stable across 4 years, and then assumed further stability to 10 years, although the authors then adjusted the values downwards by 3% per year to account for clinical deterioration rather than present net value. Where possible, the impact of age on the utility of the patient population should be modelled using available data or a data-driven assumption. Regardless, having established their utility values, they then required only annual transition probabilities between states for 10 years, which they derived using an administrative database study and data from the senior surgical author's practice.

As with the sciatica non-surgical models, the quality of the utility values was generally poor, as would be expected given that they are for similar populations. However, while it is to be lauded that attempts were made to derive patient- and treatment-specific utilities, small sample sizes (Schmier et al., 2014; Parkinson et al., 2013) and the use of LBP utility values from the Beaver Dam study (Kuntz et al., 2000; Bydon et al., 2015) were a common limitation of this group of publications. The use of the Beaver Dam values of 0.97 for being symptom-free (2000) or having a positive outcome (Bydon et al., 2015), and 0.79 for having CLBP, are quite different to the utility values used in other studies.

The use of administrative databases in five studies (Kuntz et al., 2000; Kim et al., 2012; Schmier et al., 2014; Bydon et al., 2015; Parkinson et al., 2013) facilitates precise calculation of resources associated with each treatment, and therefore offers more confidence in the accuracy of the costs associated with these surgical procedures.

### 4.4.5 Strengths and weaknesses of the study

This is the first systematic review to identify, document and classify model-based economic evaluations of treatments for LBP and sciatica. It is possible that the search

criteria may prove restrictive in that model-based economic evaluations faced exclusion if they did not contain both economic and modelling terms. The search strategy may have been improved by including the term 'utility', however the use of a very broad search strategy was employed, including 'economic' or 'model' as standalone terms, with the specific aim of increasing the potential number of studies. The breadth of these search criteria, as well the variety of databases used, is a key strength of this review, and the addition of a single paper identified from reference lists demonstrates that the search was exhaustive.

### 4.4.6 Implications for researchers, clinicians and policymakers

This chapter has identified flaws, and suggests opportunities, in models evaluating interventions for LBP and sciatica. Concerns relating to studies in this review include not only modelling across inadequate time horizons but also the inappropriate use of utility data, calculation errors, a lack of transparency regarding methodologies, and the failure to consider the extent to which uncertainty and assumptions limit the applicability and generalizability of the results.

Overall, the current cost-effectiveness evidence is indicative of the uncertainty around the clinical-effectiveness evidence on these treatments to date. Most of the studies included in this review report on the limitations of available effectiveness data in order to populate models. Policymakers' attention is directed to the sensitivity analyses in these studies, which in some cases help with accounting for the uncertainty of model parameters. Longer follow-up in trials, and the collection of health-related quality-of-life scores, would help in reducing the uncertainty around the long-term cost effectiveness of treatment.

In considering the cost-effectiveness results from studies that included a non-surgical comparator, it is evident that surgery after the failure of conservative care could be cost effective; however, there does not appear to be a consensus regarding at what stage

surgical procedures might become cost effective. This could be a potentially valuable research priority, alongside factors that influence the cost effectiveness of surgery following these repeated failures of conservative care. It is also noted that research needs to explore the implications of using different health states in both conditions.

Health economists and modellers developing models in both conditions also need to be more willing to explore the implications of extrapolation of treatment effect over an appropriate time horizon. The thesis by Mahon (et al., 2014) provides a comprehensive review of the methods that can be used to infer parameters where they are unobserved. Additionally, guidance is available on how to capture the associated uncertainty relating to extrapolation of unobserved treatment parameters in sensitivity analyses. National Institute for Health and Care Excellence (NICE) methods guidance advocates scenario analyses with (1) nil treatment effect over the unobserved period; (2) treatment effect during the unobserved period is set equal to the observed period; and (3) treatment effect diminishes over time (NICE, 2013).

Future models should pay particular attention to the methodological challenges raised here to ultimately help enable more useful comparisons between treatments. Until modellers produce more high-quality modelling studies, consistent with modelling guidelines (e.g. Philips et al., 2006; Siebert et al., 2012; Soto et al., 2002; Roberts et al., 2012; Husereau et al., 2013), the standard of discourse necessary to stimulate methodological improvements in these areas will be severely restricted.

## 4.5 Conclusion

This review identified a number of insights to inform the modelling process, including some examples of good practice. The principle finding of interest to inform the modelling chapters is that whilst only limited extrapolation has been done, with no exploration of

temporal uncertainty, studies have engaged with extrapolation using trial-based evidence, and combining with literature and assumptions.

Nonetheless the standard of modelling is currently inadequate to provide much needed economic evidence in both conditions. In order to improve the standard of modelling, future model-based analyse ought to adopt strong methodologies which follow good practice recommendations and guidelines, as well as being transparent about the model, structure and parameters used with uncertainties being adequately addressed.

# Chapter 5: SYSTEMATIC REVIEW OF DECISION ANALYTIC MODELLING IN STRATIFIED CARE: Using Osteoporosis as a case study.

## 5.1 Introduction

Chapter 2 explored the clinical and economic value of stratified care treatments for low back pain, whilst Chapter 4 systematically reviewed decision analytic models undertaken to-date in treatments for low back pain and sciatica. In the latter review, it can be noted that there was an absence of modelling studies containing stratified treatment pathways for either condition. Therefore, there is a need to understand how health economists have attempted to model stratified treatments and understand how this could differ from modelling non-stratified care.

Various guidelines have been produced to guide modellers in creating high quality and consistent economic evaluations in the area of stratified treatments (e.g. Degeling et al., 2017). However, these guidelines focus upon stratification by means of genetic testing which has its own specific methodological debates, such as, what is the utility value associated with a patient knowing that they carry a genetic mutation? Or, can the model be constructed such that the uncertainty of the sensitivity and specificity of the test are represented? Whilst these are interesting debates, these reviews and guidelines do not provide information on specific methodological issues which may arise from having to model stratified care pathways within decision analytic models.

There are a number of methodological questions which this review aims to explore.

1. Which type of decision analytic model is commonly employed to represent the stratification in the treatment pathway?
2. Which health states are used to represent the process of determining the stratified risk group in the model?

3. Upon stratification does each risk group have a unique model structure?
4. What methods had health economists used to identify parameters for each risk (stratification) groups?

In order to answer these questions a review of decision analytic models in stratified care approaches in musculoskeletal disease was undertaken. The review had three specific objectives:

To identify, document and classify existing model-based economic evaluations of stratified care pathways in musculoskeletal diseases in terms of i) their modelling techniques ii) data inputs, and iii) structure of the models. Narrative synthesis will be used to summarise and explain findings.

## 5.2 Methods

### 5.2.1 Search strategy

The protocol designed for this systematic review was developed by considering the PRISMA-P checklist (http://www.prisma-statement.org/Extensions/Protocols.aspx). The protocol specified that studies that were considered relevant for inclusion in this review if they were an economic evaluation undertaking decision analytic modelling of any stratified treatment for any musculoskeletal condition. Stratified treatment was understood in broad terms, to include personalised medicine, targeted treatments, and other associated terms.

After a scoping search, which identified a large number of potentially relevant studies, the review was restricted to include only osteoporosis studies. Face-to-face discussions were held between all authors regarding which conditions ought to be included in the review. Osteoporosis was selected on the basis that, of all the conditions identified by the review, the stratification and subsequent matching of patients to a matched treatment occurred in a manner most similar to that of the stratified care approach considered in this thesis.

Articles were identified using database searches, with subsequent studies identified in the literature and backward citations were included. The following health and health economic databases were used to conduct searching: OVID INTERFACE (MEDLINE, EMBASE, PsychINFO), EBSCO INTERFACE (Cumulative Index to Nursing and Allied Health Literature (CINAHL), Allied and Complimentary Medicine Database (AMED), EconLit), COCHRANE LIBRARY INTERFACE (Database of Abstracts of Reviews of Effects (DARE), Cochrane Database of Systematic Reviews (CDSR), Health Technology Assessment (HTA) Database, NHS Economic Evaluation Database (NHS EED)), THOMSON REUTERS INTERFACE (Web Of Science).

In developing the search strategy, economic study terms were created using the SIGN strategy developed by the NHS Centre for Reviews and Dissemination at the University of York. Terms relating to stratified medicine were developed after a discussion with academic experts in the stratified care research group at Keele, as well as input from supervisors, and from Dr. Nadia Corp, Research Associate: Systematic Reviews at Keele University, who also helped further define the strategy. Search terms used for 2 of the databases, (Embase and NHS EED) are included in Appendix 3. Database searches were conducted during August 2017; searches covered the lifetime of the databases and were limited to English papers.

### 5.2.2 Inclusion and exclusion criteria

Studies were included if they were economic evaluations that contained any form of decision analytic model. In order to be considered a modelling study, articles had to state that they had used any of the modelling types discussed in Chapter 4. Studies could include any treatment for osteoporosis providing they were evaluating a stratified treatment approach, either in comparison to another method of stratification or a non-stratified approach.

Inclusion criteria:

1. Model-based economic evaluation studies using cost-effectiveness, cost-consequence, cost-benefit, cost-utility or cost-minimisation analysis.
2. To be considered a modelling study the paper must either (i) specified to have used either a decision tree, Markov model, decision analytic model, individual sampling or patient (simulation) model, dynamic transition model, or (ii) declared to have used an economic model, even if specific model type was unspecified.
3. Any stratified treatments for patients with osteoporosis in any care setting.

Exclusion criteria:

1. Any economic evaluation which does not include decision analytic modelling, e.g. trial-based evaluations.
2. Conference abstracts, editorials.
3. Studies not in English language.
4. Studies using stratification tools to perform sub-group analysis or consider particular patient cohorts, without directly comparing a stratified care pathway with a non-stratified treatment.

### 5.2.3 Data selection and extraction

Records obtained from the databases were imported into ENDNOTE. Duplicates were removed and stored. A two stage exclusion process was used. Firstly, titles were scanned on the basis of title and abstract, with clearly irrelevant studies excluded. Studies were then placed into three folders named “included”, “possible”, and “excluded”. SJ checked that studies placed into the “included” folders were indeed suitable for inclusion. At the second stage, the list of ‘possible’ studies was reviewed independently by SJ and KK to scan for relevance using the criteria reported above.

## 5.3 Systematic Review Results

Figure 5.1 illustrates the process of selecting and identifying studies eligible for inclusion in the review.

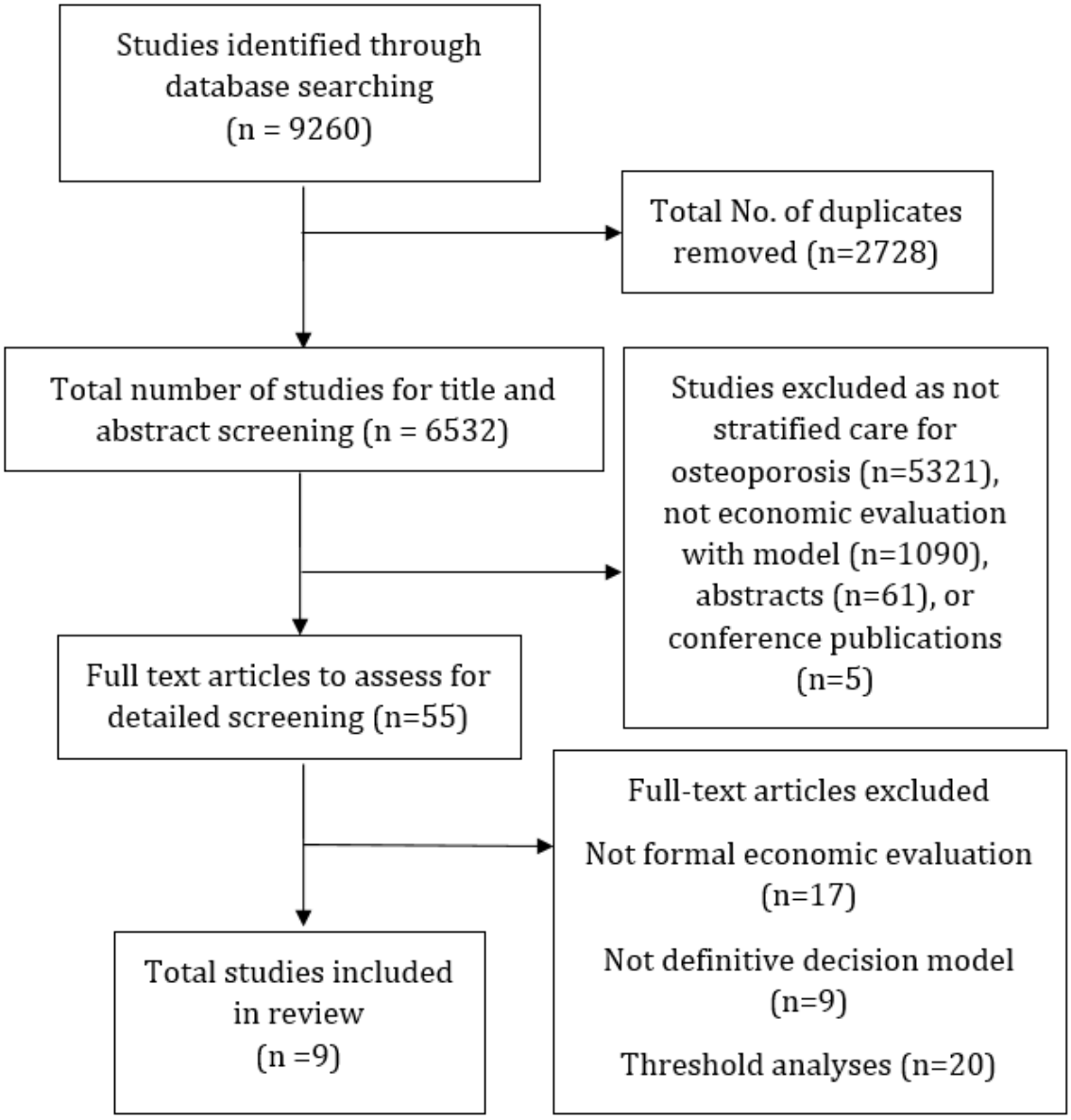



**Figure 5-1 PRISMA flow diagram showing study selection for inclusion in the systematic review**

9260 studies were imported into Endnote, of which 2728 were duplicates. Of the 6532 unique studies, 6477 were excluded in accordance with the protocol exclusion criteria, by

the first reviewer. 10% of these excluded studies, or 600, were checked independently by the second reviewer (SJ). The 55 possible studies were all independently reviewed by all four other reviewers (SJ, KK, RaO, and ML). 17 studies were excluded on the basis they did not perform an economic evaluation, 9 studies were economic analyses but did not include a clear decision modelling methodology, and 20 studies were excluded on the basis they were threshold or sub-group analyses not direct comparisons between a stratified care pathway and another method of stratification, and/or a non-stratified treatment.

### 5.3.1 Overview of studies

Table 5.1 provides an overview of the modelling studies included in this review. The table reports results for nine studies concerned with stratified treatment approaches for osteoporosis.

All included studies were stated to be cost-effectiveness analyses. Note that, Stevenson et al. (2007) is a publication of the model built in the HTA Monograph. All studies had the aim of evaluating screening methods to identify the most efficient means of stratifying the prescribing of treatments for osteoporosis. The treatments considered were either strontium ranelate (Borgstrom et al., 2010; Stevenson et al., 2007), oral bisphosphonate treatment commonly alendronate (Mueller and Gandjour et al., 2008; Nayak et al., 2011; 2012; Schott et al., 2007; Ito et al., 2009; 2014), or hormone replacement therapy (Nagata-Kobayashi et al., 2002).

Five of the papers (Schott et al., 2007; Nayak et al., 2011; Ito et al., 2009; Ito et al., 2014; Mueller and Gandjour et al., 2006) had two stage stratification of patients, which involved patients having some initial test to determine their suitability for some further (more expensive) evaluation. The remaining four papers were concerned with identifying the cost-effectiveness of using a particular sole screening mechanism to determine whether patients should receive the treatment for osteoporosis.

Common comparators were no screening (Schott et al., 2007; Nayak et al., 2012), no screening and no treatment (Stevenson et al., 2007), watchful waiting (Ito et al., 2014), and universal scans (Schott et al., 2007; Ito et al., 2009).

Table 5.1 Overview of studies included in this review

| Author | Country | Osteoporosis Treatment | Population | Test Type | Comparators | Purpose | Type of Economic Evaluation |
|---|---|---|---|---|---|---|---|
| Borgstrom et al. (2010) | U.K. | Strontium ranelate | Postmenopausal women with clinical risk factors for fracture | FRAXA (R) algorithm and T-scores for BMD | No treatment | To identify cost-effective thresholds | Cost-effectiveness Analysis |
| Ito et al. (2009) | U.S | Alendronate | White males aged 70 and over with no previous osteoporotic fracture | Treatment based upon selective bone densitometry using the OST. **DOUBLE STRATIFICATION** | No bone densitometry and universal bone densitometry | To evaluate the cost-effectiveness of performing bone densitometry according to risk stratification | Cost-Utility Analysis |
| Ito et al. (2014) | U.S | Alendronate | Rural women aged 65 years and above with no previous fracture | Treatment initiated based upon screening using the BMD. Patients with Osteopenia were subsequently assessed using FRAX **DOUBLE STRATIFICATION.** | Watchful waiting | To assess the cost-effectiveness of various strategies for rural women with limited access to duel-energy X-ray absorptiometry | Cost-effectiveness Analysis |
| Mueller and Gandjour (2008) | Germany | Alendronate | Women of the general population aged 50–90 years | Screening with QUS as a pre-test for DXA and treatment **DOUBLE STRATIFICATION** | Immediate access to DXA and (ii) no screening. | To determine the cost effectiveness of osteoporosis screening with QUS as a pre-test for DXA and treatment with alendronate | Cost-effectiveness Analysis |

| Author | Country | Osteoporosis Treatment | Population | Test Type | Comparators | Purpose | Type of Economic Evaluation |
|---|---|---|---|---|---|---|---|
| Nagata-Kobayashi et al. (2002) | Japan | Hormone replacement therapy | Postmenopausal women aged 50 and over without specific risk of breast cancer. | DXA | No screening or treatment; treatment for osteopenia and osteoporosis after screening; and universal treatment | To identify the cost-effectiveness of screening for osteoporosis in postmenopausal Japanese women | Cost-effectiveness Analysis |
| Nayak et al. (2011) | U.S | Oral bisphosphonate therapy | Postmenopausal U.S. women aged 55 years or older. | Comparison of various screening and treatment strategies.<br>Three different tests; Central DXA; QUS; SCORE tool.<br><br>Also use of **DOUBLE STRATIFICATION** - SCORE tool to determine whether patients receive DXA. | | To identify the cost-effectiveness of various screening strategies for osteoporosis in postmenopausal American women | Cost-effectiveness Analysis |
| Nayak et al. (2012) | U.S | Alendronate | Community-dwelling women aged 65 and over | Screening with dual-energy x-ray absorptiometry (DXA) of the femoral neck and lumbar spine. | No screening with treatment only after fracture | Assess the impact of Generic Alendronate Cost on the Cost-Effectiveness of Osteoporosis Screening and Treatment | Cost-effectiveness Analysis |
| Schott et al. (2007) | France | Risedronate or Alendronate | Postmenopausal women aged 70 years and over | Measuring BMD, using DXA, of those having at least one risk factor (based upon ANAES guidance).<br>Double stratification.<br>**DOUBLE STRATIFICATION** | Measuring BMD of all women;<br>No screening | Identify the cost-effectiveness of screening strategies applied to elderly women aged 70 years and older | Cost-effectiveness Analysis |

| Author | Country | Osteoporosis Treatment | Population | Test Type | Comparators | Purpose | Type of Economic Evaluation |
|---|---|---|---|---|---|---|---|
| Stevenson et al. (2007) | UK | Strontium ranelate | Postmenopausal women without a prior fracture aged 50-84. All patients in the model have CRF's to make them candidates for treatment, and therefore the model does not represent stratification by CRF. | Measuring BMD, using DXA | Neither treatment nor a BMD scan;<br><br>Treatment without a BMD scan. | To assess the impact of alternative identification approaches on the cost-effectiveness of the screening strategies for receipt of strontium ranelate for prevention of osteoporotic fractures | Cost-effectiveness Analysis |
| Abbreviations; BMD (Bone mineral density); CRF (Clinical risk factors); DXA (dual energy x-ray absorptiometry); OST (Osteoporosis Self-Assessment Tool); QUS (Calcaneal quantitative ultrasonography); SCORE (Simple Calculated Osteoporosis Risk Estimation). | | | | | | | |

### 5.3.2 Modelling Structure

Of the eight models which specified the time horizon of their analysis, six were lifetime, and two (Schott et al., 2007; Stevenson et al., 2007) used ten-year time horizons.

In terms of model type, six models used a Markov cohort model (Borgstrom et al., 2010; Ito et al., 2009; Ito et al., 2014; Mueller and Gandjour, 2006; Nagata-Kobayashi et al., 2002; Schott et al., 2007), and three an individual state transition model (Nayak et al., 2011; Nayak et al., 2012; Stevenson et al., 2007). Four models featured a schematic with a decision tree prior to their main model to represent the stratification process, two models placed their decision tree prior to Markov models (Ito et al., 2014; Schott et al., 2007) and two models placed their decision tree prior to individual simulation models (Nayak et al., 2012; Stevenson et al., 2007). Models which did not provide such a schematic (n = 5) used the same arithmetic calculations as if they had used a decision tree (Borgstrom et al., 2010; Ito et al., 2009; Mueller and Gandjour, 2006; Nagata-Kobayashi et al., 2002; Nayak et al., 2011).

The approach to the structuring of all six Markov models was similar and can be best highlighted by considering the schematic provided by Ito et al. (2014), shown in figure 5.2. In the model, the decision tree shows that all women were assigned to either the watchful waiting or stratified by a bone mineral density (BMD) or clinical risk factor (CRF) based strategy which would determine whether or not they would receive treatment.

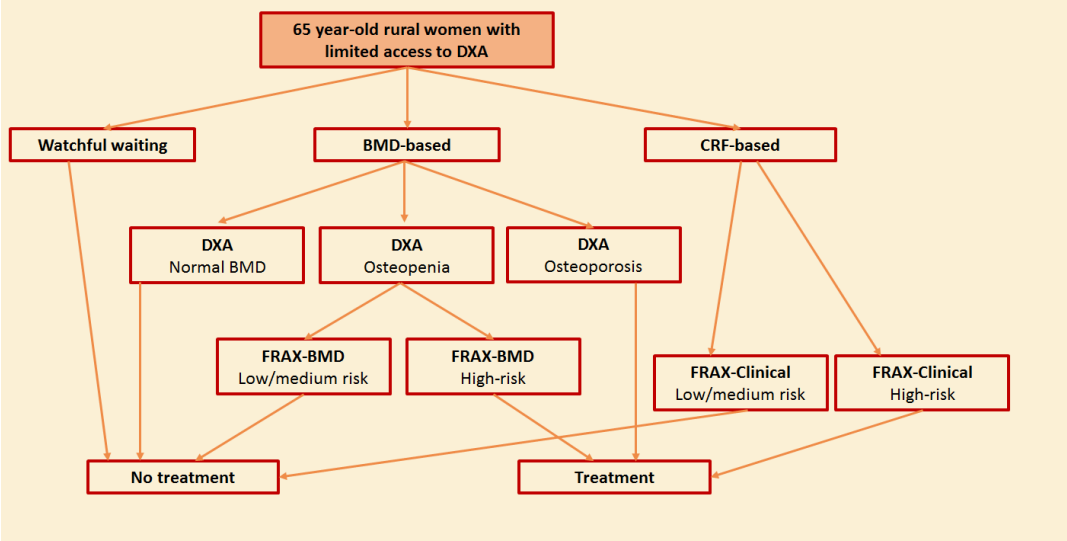



**Figure 5-2 Decision Tree showing how each strategy impacts upon whether or not a patient receives treatment (Ito et al., 2014)**

Those under watchful waiting did not receive treatment until experiencing a fracture. Those women undergoing a BMD-based strategy were stratified by the World Health Organization Fracture Risk Assessment Tool calculated with BMD (FRAX-BMD) after undergoing dual-energy X-ray absorptiometry (DXA) screening, with those considered to have osteoporosis and high-risk osteopenia receiving treatment. Meanwhile, those in the CRF-based strategy, patients were stratified according to the World Health Organization Fracture Risk Assessment Tool calculated without BMD (FRAX-Clinical), and again, high-risk patients received treatment. With the proportions of patients who would receive treatment in each strategy determined, patients were then entered into a Markov model of which the health states are essentially concerned with fractures, see Figure 5.3.

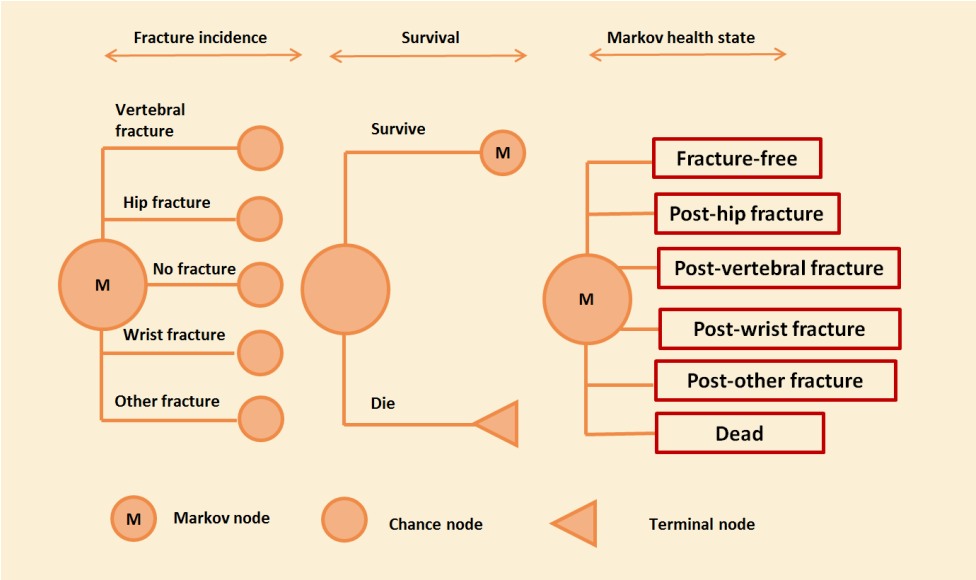



**Figure 5-3 Markov model structure, Ito et al. (2014)**

In the Markov model, all women were entered in the fracture-free state. Each year women were at risk of sustaining a fracture (hip, vertebral, wrist, or other) or death from other causes. Whether or not they experienced an event determined whether they remained in "fracture-free" or proceed to any of the "post-fracture" states, or became absorbed into the death state. The probability of experiencing a fracture was determined as a function of the baseline characteristics of the cohort (e.g. age, BMD, history of fracture) but also of whether or not they had received the treatment. In other words, each cohort faced an identically structured Markov model, with differential fracture risks based upon the characteristics of their cohort, as well as whether or not they received treatment. The other cohort models (Borgstrom et al., 2010; Ito et al., 2009; Mueller and Gandjour et al., 2008; Nagata-Kobayashi et al., 2002) all used the same principles and schematics.

The decision tree and Markov model by Schott et al. (2007), see figure 5.4, demonstrates how such a model is generated within the specific software used in the modelling chapters of this thesis, TreeAge for Healthcare (TreeAge Software Inc., Williamstown, MA, USA).

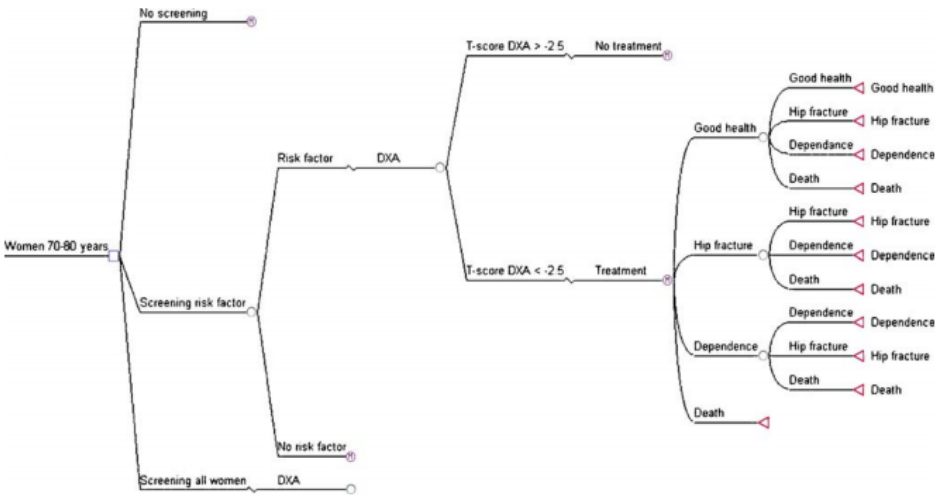



**Figure 5-4 Decision tree and Markov model by Schott et al. (2007)**

This model is one of the types with a double stratification. The first stratification is based upon risk factors, with a second subsequently based upon BMA test for those who received one. The fracture risks in this model are dependent upon age and dependence, as well as whether the patient received treatment or not.

The simulation models in this review have similarities to the cohort state transition models, with the exception that these models simulate individual women through the model one at a time. Of the three models which used individual simulation models, two were by the same author, Nayak et al. (2011; 2012). The two models by Nayak and colleagues are schematically similar to the Markov models which have a decision tree prior to them. Essentially, the authors evaluated seven testing strategies alongside no screening, although their model allowed for the strategy to be repeated every five years and also every ten years, resulting in 22 alternative strategies. Again, the same principles are adopted to that of the Markov models, i.e. with a positive screening result, the individual is offered treatment which lowers fracture risk. Key difference with this class of models is that patients can experience a series of events over time, including a new osteoporotic fracture, death, moving to a nursing home, or recovery, as well as adverse medication events. These events subsequently determine future progression through the model. The use of the

simulation model allows an *individual* patient to have their fracture risk based upon the women's characteristics, such as age, femoral neck or lumbar spine BMD, and presence or absence of a history of fracture.

Finally, Stevenson et al. (2007) also run a simulation model in their HTA monograph. Their analysis differs most from the analytical models in other studies. This is because their analysis ran patients through the following simulation model first to answer their first research question related to the use of strontium ranelate as a treatment for osteoporosis, it is only subsequent that this model is used as the basis for identifying the optimum screening strategy.

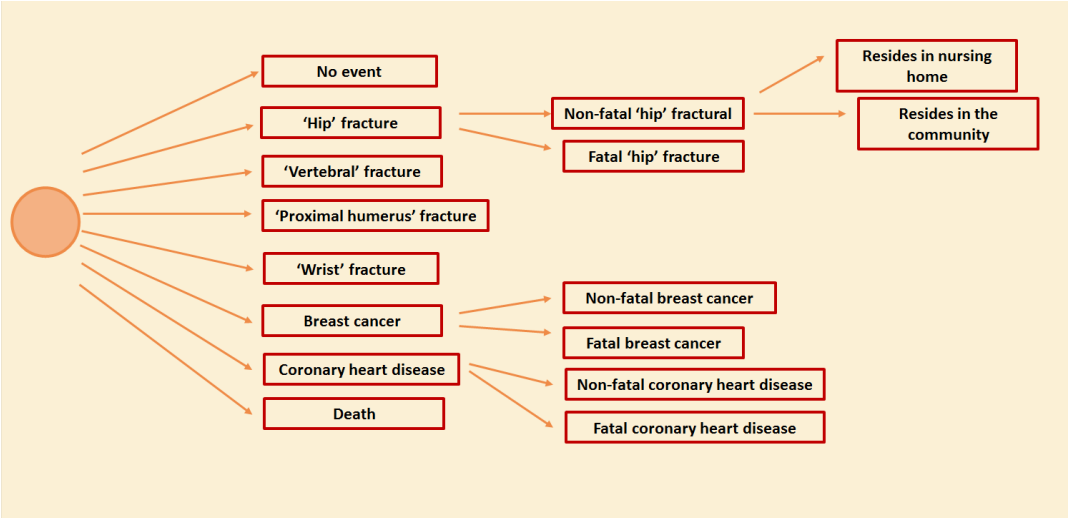



**Figure 5-5 Simulation model by Stevenson et al. (2007)**

In their model, full patient history is recorded, meaning that events such as prior fractures and current residential status can, therefore, be used to determine the likelihood of events in the next period. Again, treatment lowers fracture risk but incurs treatment costs.

Having established the cost-effectiveness of using treatment in various groups of patients using a simulation model, Stevenson et al. (2007) then move to identifying the cost-effectiveness of identification strategies in various groups. They do this essentially by

arithmetic calculations on the principle of monetary net benefit, with the optimal strategy being the strategy with the highest monetary net benefit.

They evaluate three strategies for different ages, and CRFs, (a) offer neither treatment nor a BMD scan; (b) offer treatment without a BMD scan; and (c) offer BMD scans to all and treatment to those whose T-score shows that they can be treated cost-effectively. These strategies were compared to current standard practice. For option a) the net benefit is assumed to be zero minus the costs of identification, which would include the costs of asking the initial questions. For option (b) the net benefit is the number of women in each T-score band who can be treated cost-effectively multiplied by the appropriate net benefit from treatment, minus the costs of identification and BMD scanning. The net benefit for option (c) is the number of women multiplied by the appropriate net benefit of treatment minus the cost of identification.

**Table 5.2 Modelling Characteristics**

| Author | Model Type | Name of States | Time horizon | Method of comparing strategies |
|---|---|---|---|---|
| Borgstrom et al. (2010) | Markov cohort | Wrist fx, Well, Other Osteo fx, Vertebral fx, Hip fx, Post Vertebral fx, Post hip fx, Dead | Time horizon death or 100 years of age. | The same model structure was used for each stratified risk group. Each subgroup group faces a unique set of model parameters based upon the fracture risk. BMD testing determines how many osteoporotic patients are identified and thus receive treatment. Results are presented separately for whether the cohort receives BMD test or not. |
| Ito et al. (2009) | Markov cohort | Death from nonfracture, adverse event from alendronate therapy, wrist fracture, subclinical vertebral fracture, clinical vertebral fracture, hip fracture, or second hip fracture. Long terms states "death, nursing home, community-dwelling but dependent, or community-dwelling and well" | Time horizon death or 100 years of age. | The same model structure was used for each stratified risk group. The number of patients identified and subsequently treated is dependent upon testing strategy. Each cohort essentially faces a fracture risk dependent upon receipt of treatment, as well as age and pre-fracture functional status of the patient. Results are shown in a table showing the ICER for testing strategies. |
| Ito et al. (2014) | Decision tree prior to Markov cohort | Fracture-free, post-hip fracture, post-vertebral fracture, post-wrist fracture, post-other fractrure, dead. | Time horizon death or 100 years of age. | The same model structure was used for each stratified risk group. Decision tree shows that two types of patients enter their Markov model, those who receive treatment on the basis of the screening and those who do not. Whether or not the patient had received treatment, as well as patient characteristics, then determines fracture risk within the model. Results are shown in a table showing the ICER for testing strategies. |
| Mueller and Gandjour (2006) | Markov model | Stratification unspecified. Clinical stages in Markov model with eight stages, no fracture, hip fracture, vertebral fracture, forearm fracture, post-fracture hip, post-fracture vertebral and forearm) | Time horizon death or 100 years of age. | The same model structure was used for each stratified risk group. The difference between the two strategies being the number of women selected for treatment, not graphically modelled in the study. Fracture risks based upon receipt of treatment, as well as risk factors and BMD, Results are shown in a table showing the ICER for testing strategies. |

| Author | Model Type | Name of States | Time horizon | Method of comparing strategies |
|---|---|---|---|---|
| Nagata-Kobayashi et al. (2002) | Markov cohort | Healthy, Acute hip fracture, good prognosis, poor prognosis, death. | | The same model structure was used for each stratified risk group. Fracture risk dependent upon risk level, as well as receipt of treatment. The hypothetical cohort classified into 3 groups according level of risk of hip fracture: 41.7% of people at low risk of hip fracture; 31.0% with osteopenia, and 27.3% with osteoporosis. All patients in cohort run through the model for the four different strategies. Post-fracture outcomes are standardised regardless of initial risk group. |
| Nayak et al. (2011) | Individual state transition model | Screening strategies (8); Treatment (2); Fracture state (5); Outcomes (11) | Lifetime of the patient. | The same model structure was used for each stratified risk group. Patient were sent through the individual sampling model such that their screening strategy influenced whether they received bisphosphonate, which in turn, along with individual characteristics, influenced the degree to which they obtained fractures or not. Results are shown in a table showing the ICER for testing strategies. |
| Nayak et al. (2012) | Decision-Tree prior to individual state transition model | Decision Tree (4): DXA screening and no screening (into usual care). ISTM(6): (I) 5 fracture types or death. (II) Survive or Death. (III) Community dwelling or nursing home. (IV) No new fracture, new wrist fracture, new vertebral fracture, new hip fracture, hospital death. (V) Alendronate Adverse Event or no adverse event. | Lifetime of the patient | The same model structure was used for each stratified risk group. Patient were sent through the ISTM, with their screening strategy influencing whether they receive alendronate, which in turn, along with their individual characteristics, then influences fracture risk. Results are shown in a table showing the ICER for testing strategies. |
| Schott et al. (2007) | Decision tree prior to Markov model | Three branches of a decision tree - Screening strategies (3); DXA screening or not (2); treatment or not (2). All stratifications used decision tree.<br><br>Markov model for all patients - good health, hip fracture, dependence, or death (4) | Ten year time horizon | The same model structure was used for each stratified risk group. Patients are stratified according to the risk factors, and then those with the presence of risk factors have a subsequent stratification according to DXA score. In the model, the fracture risks are dependent upon presence of risk factors and whether they had treatment or not. This process is shown in a decision tree. Results are shown in a table showing the ICER for testing strategies. |

| Author | Model Type | Name of States | Time horizon | Method of comparing strategies |
|---|---|---|---|---|
| Stevenson et al. (2007) | Decision tree then individual state transition | Then ISM for health states - No event; hip fracture (to non-fatal hip fracture (to nursing home or resides in the community) or fatal hip fracture); vertebral fracture; proximinal humerus fracture; wrist fractures; breast cancer (to non-fatal breast cancer or fatal breast cancer); coronary heart disease (to non-fatal coronary heart disease); death. | Ten-year time horizon | The same model structure was used for each stratified risk group. Patient outcomes are simulated in an ISM, using fracture risks and death rates dependent upon whether the patient received treatment as well as individual risk factors. The net benefit for each testing strategy reflects an arithmetic function of costs of identification and the variation in costs and outcomes for those receiving treatment. |
| Abbreviations; BMD (Bone mineral density); CRF (Clinical risk factors); DXA (dual energy x-ray absorptiometry); fx (Fracture); ICER (Incremental cost-effectiveness ratio); ITSM (Individual state transition model); OST (Osteoporosis Self-Assessment Tool). | | | | |

### 5.3.3 Sensitivity analyses

Given this review is about the stratification process, this section concerns only sensitivity analyses directly related to the stratification.

The predominant form of sensitivity analyses in this study which related to the stratification process, concerned the cost-effectiveness implications of varying the sensitivity and specificity of the test. Deterministic analysis appeared in five studies (Ito et al., 2009; Ito et al., 2014; Mueller and Gandjour, 2008; Nayek et al., 2011), these studies also included uncertainty over their screening rate in their PSA.

Nagata-Kobayashi et al. (2002) included a deterministic analysis on the proportion of screened patients in each risk group. Whilst Stevenson et al. (2007) performed a deterministic analysis on GP time to perform initial risk assessment and time to discuss DXA scan, as well as the admin costs of DXA scan.

**Table 5.3 Sensitivity Analysis**

| Author | Sensitivity Analyses |
|---|---|
| Borgstrom et al. (2010) | None |
| Ito et al. (2009) | Deterministic analysis around the age which the patient receives screening, as well as the sensitivity and specificity of the OST tool |
| Ito et al. (2014) | Deterministic and probabilistic sensitivity analyses were performed upon varying screening rates |
| Mueller and Gandjour (2008) | Sensitivity analysis conducted for the sensitivity and specificity of the tests |
| Nagata-Kobayashi et al. (2002) | Deterministic sensitivity analysis on the proportion screened in each risk group. |
| Navak et al. (2011) | Deterministic analyses varying the duration of rescreening, and the ages of patients entering the analysis.<br>PSA reviews the uncertainty regarding the test sensitivity and specificity. |
| Nayak et al. (2012) | None |
| Schott et al. (2007) | Many variables were tested, and only those with the greatest effect were disclosed. |
| Stevenson et al. (2007) | Deterministic analysis performed on GP time to perform initial risk assessment, time to discuss DXA scan, and the admin costs of DXA scan. |
| Abbreviations; DXA (dual energy x-ray absorptiometry); GP (General Practitioner); OST (Osteoporosis Self-Assessment Tool); PSA (Probabilistic sensitivity analysis). | |

### 5.3.4 Treatment Efficacy

As previously noted, with few exceptions, the health states included in this review reflect some form of fracture. The treatments in these analyses impact the cost-effectiveness by reducing the risk of fracture. Therefore fracture risk, as well as treatment efficacy (reduction in fracture risk), are the key parameters in these models.

A key question for modelling in stratified care therefore must be, does the progression through the model relate to the stratification. In relation to fracture risk, as noted, all of the eight models follow a similar pattern with models using fracture risks which are for the most part independent of the stratification of patients. Whilst fracture risks are not based upon the stratification, many of the risk factors which are used for the stratification of patients are also used to parameterise fracture risk. In other words, progression through the model is not determined by the stratification, but rather based upon similar characteristics used to stratify. For example, five models which stratify using BMD score also allow BMD to influence fracture risks (Ito et al., 2014; Mueller and Gandjour, 2008; Nagata-kobayashi et al., 2002; Nayak et al., 2012; Nayak et al., 2011). Meanwhile, Ito et al. (2009) use the OST tool, which is used for stratification also to impact upon fracture risk.

The six models which use cohort modelling use cohort characteristics to derive their fracture risks, whilst the three which use individual simulation models have the benefit of basing their fracture risks upon individual characteristics and previous events experienced in the model.

In terms of how the stratification impacts upon treatment efficacies within the model, treatments, commonly alendronate, reduce the fracture risk for patients, which were modelled as a relative risk reduction in all models. In all but one of the models, treatment efficacy was set as entirely independent from the stratification mechanism, with the

exception of Nayak et al. (2011) who set their treatment efficacies as partially dependent upon T-score.

**Table 5.4 Treatment Efficacy**

| Author | Sources | Method of deriving parameters for fracture risks | Fracture risk independent of stratification | Method of deriving parameters for treatment efficacy | Treatment efficacy independent of stratification |
|---|---|---|---|---|---|
| Borgstrom et al. (2014) | Baseline risks come from prospective study by Singer et al. (1998) and systematic review by Stevenson et al. (2005). Fracture risks calculated using the FRAX® tool.<br><br>Effects of risedronate on fracture risk from meta-analysis | Fracture risks in model cohorts were set as relative risks to the 'normal population'.<br><br>**Risks reflected cohort characteristics;** low BMI, prior fragility fracture, parental history of hip fracture, long-term use of oral glucocorticoids, rheumatoid arthritis, current smoking status, and alcohol consumption. | ✓ | Treatment using risedronate impacts upon relative risk of fracture, independent of stratification.<br><br>Fracture reduction benefit returns to null in a linear fashion over 5 years post-completion of risedronate therapy. | ✓ |
| Ito et al. (2009) | Baseline fracture incidence from a population-based survey<br><br>Fracture relative risks due to the presence of osteoporosis, fractures, and OST score.<br><br>Relative risk reduction of fractures owing to alendronate therapy based on meta-analysis | Fracture risks in model cohorts set as relative risks relative to incidence rates in the 'nonosteoporotic population'.<br><br>**Risks reflected cohort characteristics;** OST score (age and low body weight), presence of osteoperosis, vertebral fracture.<br>Nursing home placement also impacts fracture risk. | ✓ | Treatment using alendronate impacts upon relative risk of fracture, independent of stratification.<br><br>Treatment benefit assumed to start in year three of therapy.<br>Fracture reduction benefit returns to null in a linear fashion over 5 years post-completion of risedronate therapy. | ✓ |
| Ito et al. (2014) | Baseline fracture incidence from the US National Inpatient Sample database. | Fracture risks in model cohorts were set as relative risks to the general population.<br><br>**Risks reflected cohort characteristics;** age, BMD Z-score, and history of prior fractures. | ✓ | Treatment using alendronate impacts upon relative risk of fracture, independent of stratification. | ✓ |

| Author | Sources | Method of deriving parameters for fracture risks | Fracture risk independent of stratification | Method of deriving parameters for treatment efficacy | Treatment efficacy independent of stratification |
|---|---|---|---|---|---|
| Ito et al. (2014) (contd) | Treatment effect, and relative risks of fracture from metaanalyses. | | | Fracture reduction benefit returns to null in a linear fashion over 5 years post-completion of alendronate therapy. | |
| Mueller and Gandjour (2008) | Fracture risk in baseline, and risk groups determined by various methods of integrating sources with assumptions.<br><br>Effectiveness of alendronate was derived from meta-anaylsis of studies identified via literature search. | Fracture risks in model cohorts were set as relative risks to the general population.<br><br>**Risks reflected cohort characteristics;** age, prior fracture, body weight, smoking status, mobility, number of falls, and family history. High risk patients determined by BMD T-score. | ✓ | Treatment using alendronate impacts upon relative risk of fracture, independent of stratification. | ✓ |
| Nagata-Kobayashi et al. (2002) | Risk group cohorts obtained from epidemiological study (Fujiwara et al., 1995).<br><br>HRT treatment effect from a review of observational studies by Grady et al. (1992). Weiss et al., [1980] reported that the risk of hip fracture decreased with prolonged use of HRT. | Model uses three defined risk groups (fracture risk) according to BMD T-score, 41.7% of people at low risk; 31.0% with osteopenia; 27.3% with osteoporosis.<br><br>**Risks reflected cohort characteristics;** age, BMD T-score.<br>Nursing home placement also impacts fracture risk. | ✓ | Treatment using HRT impacts upon relative risk of fracture, independent of stratification. | ✓ |
| Nayak et al. (2012) | Fracture rates for women not on alendronate treatment were based on Study of Osteoporotic Fractures (SOF) data. | Fracture risks in model simulations were set as relative risks to population without risk factors. | ✓ | Treatment using alendronate impacts upon relative risk of fracture, independent of stratification. | ✓ |

| Author | Sources | Method of deriving parameters for fracture risks | Fracture risk independent of stratification | Method of deriving parameters for treatment efficacy | Treatment efficacy independent of stratification |
|---|---|---|---|---|---|
| | Relative risk reduction for alendronate from a study by Liberman et al. (1995), meta-analysis by Karpf et al. (1997), and 3 trials | Risks reflected individual characteristics; age, femoral neck or lumbar spine BMD, and history of fracture. | | | |
| Nayak et al. (2011) | Fracture risks probabilities calculated used logistic regression equations developed from Study of Osteoporotic Fractures data.<br><br>Treatment effects of oral bisphosphonate from meta-analysis. | Fracture risks in model simulations were set as relative risks to population without risk factors.<br><br>**Risks reflected individual characteristics;** age, femoral neck or lumbar spine BMD, and presence or absence of a history of fracture. | ✓ | Treatment using alendronate impacts upon relative risk of fracture, dependent upon T-score used for stratification.<br><br>Fracture reduction benefit returns to null in a linear fashion over 5 years post-completion of alendronate therapy | χ |
| Schott et al. (2007) | Hip fracture rates (Baudoin et al., 1996; 1997).<br><br>35% reduction of hip fracture incidence over 10 years after a treatment 5 years based upon three Kanis et al. (2002a; 2002b; 2004) studies. | Fracture risks seem to be the same across the population, with annual hip fracture probability.<br><br>**Risks reflected cohort characteristics;** Unstated.<br><br>Hip fractures were modelled at a higher rate in nursing home resident | ✓ | Treatment using alendronate impacts upon relative risk of fracture, independent of stratification.<br>The treatment effect over the time horizon took into account efficacy, offset, compliance, and discontinuation rate. | ✓ |

| Author | Sources | Method of deriving parameters for fracture risks | Fracture risk independent of stratification | Method of deriving parameters for treatment efficacy | Treatment efficacy independent of stratification |
|---|---|---|---|---|---|
| Stevenson et al. (2007) | Stevenson et al. (2007) | Individual fracture risks in model simulations set as function of risk factors<br><br>**Risks reflected individual characteristics;** age, previous fracture at each site, and residential status. | ✓ | Treatment using alendronate impacts upon relative risk of fracture, independent of stratification.<br><br>Fracture reduction benefit returns to null in a linear fashion over 5 years post-completion of alendronate therapy | ✓ |

Abbreviations; BMI (Body Mass Index); BMD (Bone Mineral Density); DXA (dual energy x-ray absorptiometry); FRAX ( Fracture Risk Assessment Tool); GP (General Practitioner); OST (Osteoporosis Self-Assessment Tool); PSA (Probabilistic sensitivity analysis);

### 5.3.5 Utility Values

An important issue for assessment within the models is how utility values are calculated for each stratified comparator, and none of the models appear to use risk-specific utility values. In Table 5.5, the column on the far-right shows that of the eight models which use utility values in their analysis, all of them used utility values which were independent of their initial stratification type, and were based upon defined health states. All models had utility values associated with fractures, most had specific post-fracture states with unique utility values (Ito et al., 2014; Mueller and Gandjour, 2008; Nayak et al., 2011; Stevenson et al., 2007), some had utility values associated with mode of residence (Ito et al., 2009; Nayak et al., 2012; Nayak et al., 2011; Stevenson et al., 2007) as well as disutilities associated with adverse events (Ito et al., 2009; Nayak et al., 2012; Nayak et al., 2011; Stevenson et al., 2007). One model included a utility value based upon ‘prognosis’ after hip fracture (Nagata-Kobayashi et al., 2002).

**Table 5.5 Utility Values**

| Author | Sources | Method of deriving parameters for individual risk groups | Utility values independent of stratification |
|---|---|---|---|
| Borgstrom et al. (2014) | Two prospective observational cohort studies (Borgstrom et al., 2006; Strom et al., 2008). Supplemented by expert opinion in a published study, a study containing empirical observations and the population tariff values for the UK. | Population tariff values for the UK are used. Fracture states carry specific utility losses. Utility losses in the first year after a fracture at the hip, spine, or forearm were based on empirical estimates. | ✓ |
| Ito et al. (2009) | Baseline utility for well state was collected from nationally representative survey (Hanmer et al., 2006)<br>Utility multipliers for each fracture type were obtained from systematic review (Brazier et al., 2002). | Utility values are set for each health state, with specific values calculated for the fracture states, as well as dwelling type, and esophageal ulcer. | ✓ |
| Ito and Leslie (2014) | "Nationally representative samples" come from a report by Hanmer et al. (2006). Cohort study by Kanis et al. (2004). | Utility values are set for each health state, with specific values calculated for the fracture states, as well as post-fracture state. | ✓ |
| Mueller and Gandjour (2008) | For the no-fracture state (QOL) data from the general population was used, estimated by a<br>time trade-off questionnaire by Brazier et al. (2002). Utility values for forearm and clinical vertebral fractures were calculated using the EQ-5D, as reported in Kanis et al. (2004). Quality-adjusted life expectancy associated with hip fractures was modelled in a previous paper (Gandjour et al., 2006).<br>For the post-fracture state QOL improved, although not to the prior level, authors reference (Schousboe et al., 2005). | Utility values are set for each health state, with specific values calculated for the fracture and post-fracture states. | ✓ |
| Nagata-Kobayashi et al. (2002) | Time-trade off study by Salkeld et al. (n=194) | Utility values are set for health state, and prognosis after hip fracture. | ✓ |
| Nayak et al. (2012) | Baseline values from Hammer et al. (2006).<br>Disutilities come from six multiple published studies on various fractures. | This analysis uses a baseline utility, then subtracts disutilities associated with fractures, living in a nursing home, and medication adverse events. | ✓ |

| Author | Sources | Method of deriving parameters for individual risk groups | Utility values independent of stratification |
|---|---|---|---|
| Nayak et al. (2011) | Baseline values come from Hanmer et al. (2006), a report of nationally representative values for the noninstitutionalized US adult population.<br>Other utility values are derived from numerous other sources. | Utilities are given for each health state and are also set as dependent upon age. Post-fracture states, nursing home residence, and esophagitis also have utility values. | ✓ |
| Schott et al. (2006) | *Effectiveness was estimated as the average number of years without hip fracture over 10 years* | | N/A |
| Stevenson et al. (2007) | Utility values for fractures were calculated using the EQ-5D, as reported in Kanis et al. (2004). | Utility values are calculated before the strategy calculations are performed. | ✓ |
| Abbreviations; BMI (Body Mass Index); BMD (Bone Mineral Density); DXA (dual energy x-ray absorptiometry); FRAX ( Fracture Risk Assessment Tool); GP (General Practitioner); OST (Osteoporosis Self-Assessment Tool); PSA (Probabilistic sensitivity analysis). | | | |

### 5.3.6 Costing stratification and individual risk groups

A further key issue for this review was how the stratification process impacted upon the costs included in the model, and how the costs were derived.

Firstly, as was the case with utility values, costs were associated with the health states, which were independent of the stratification process. Therefore the models in this review did not make any significant alterations to the costing process. In terms of costs associated with stratified treatments, all studies included the costs associated with screening and testing patients, but it was not clear how this was included in the model, e.g. which states or which time points were used for these costs.

Some studies included some unique aspects of the testing process. Firstly, Ito et al. (2014) included in their societal analysis not only the costs of administering the test, but also travelling to the test site, as well as lost earnings from the trip. Second, Mueller and Gandjour et al. (2008) included the costs of false positives and follow up of patients. Third, Schott et al. (2007) included the costs of a screening campaign. Finally, Stevenson et al. (2007) used a rigorous and thorough method for costing the identification process. They cost the GP's initial risk assessment and subsequent GP consultation of the algorithm in minutes, treatment without a DXA scan is given a further 10-minute appointment to discuss and initiate treatment. They also cost booking and execution of the DXA scan, followed by a 10- minute appointment to discuss the DXA results.

All studies included the direct costs of the treatments as well as the costs associated with managing fractures, and as these are not related to stratification these are not listed in table 5.6.

**Table 5.6 Costs**

| Author | Sources | Costs associated with risk groups | How was cost of the test incorporated into the analysis |
|---|---|---|---|
| Borgstrom et al. (2014) | Cost analysis by Stevenson et al. (2006) | Costs were independent of stratification, other than testing costs.<br>**Other costs included;** Nursing home costs for hip fracture | Unclear.<br>The cost of the test was included for those who had the test and for those who didn't. |
| Ito et al. (2009) | Direct medical costs taken from a population-based cost analysis (Gabriel et al., 2002).<br>Home healthcare or nursing home care costs from a cost study (Harrow et al., 1995).<br>Costs of screening for osteoporosis Medicare Physician Fee Schedule Look-up Tool.<br>The price of generic alendronate was taken from an online retail pharmacy. | Costs were independent of stratification, other than testing costs.<br>**Other costs included;** Adverse events, home healthcare and nursing home care costs. | Cost of the test included.<br>Unclear where test cost was included. |
| Ito et al. (2014) | Cost of DXA and physician visit from 2014 Medicare National Average Rates (ACoR, 2014). Median weekly earnings from the US Bureau of Labor Statistics to estimate the opportunity cost of travel time to complete DXA (USDoL, 2014). Treatment costs included cost of generic alendronate reported in online retail pharmacy (Drugstore.com, 2014). Fracture costs from a population-based cost analysis in Olmsted County, MN (Gabriel et al., 2002) | Costs were independent of stratification, other than testing costs.<br>**Other costs included;** societal costs included patient **t**ravelling to the test site as well as lost earnings also. | Cost of the test included.<br>Unclear where test cost was included. |

| Author | Sources | Costs associated with risk groups | How was cost of the test incorporated into the analysis |
|---|---|---|---|
| Mueller and Gandjour (2008) | Costs of DXA based on the German price list for outpatient treatment. QUS costs based on the German medical fee schedule.<br>The price for alendronate was taken from a public database at the Federal Statistical Office Germany. | Costs were independent of stratification, other than testing costs.<br>**Other costs included;** costs of false positive results and follow-up. | Costs of test included.<br>Unclear where test cost was included. |
| Mueller and Gandjour (2008) cont'd | Costs of Treatment for Fractures For hip fractures, all inpatient, using weighted average diagnosis-related group rates (InEK, 2006) | | |
| Nagata-Kobayashi et al. (2002) | All costs except for drug costs were based upon a cost of illness study for osteoporosis [Ogawa et al., 1996) | Costs were independent of stratification, other than testing costs. | Cost of the test included.<br>Unclear where test cost was included. |
| Nayak et al. (2012) | Medicare database - Centers for Medicare and Medicaid Services national physician fee schedule website (2010)<br><br>Drug Topics Red Book.: Physician's Desk Reference | Costs were independent of stratification, other than testing costs. | Cost of the test included.<br>Unclear where test cost was included. |
| Nayak et al. (2011) | Fracture-related resource use from Medicare reimbursement rates (CMMS, 2011).<br>Nursing home costs based upon a national nursing home insurance survey (GE, 2003). Over-the-counter omeprazole price from a low-cost pharmacy (Walmart, 2010). | Costs were independent of stratification, other than testing costs. | Cost of the test included.<br>Unclear where test cost was included.. |

| Author | Sources | Costs associated with risk groups | How was cost of the test incorporated into the analysis |
|---|---|---|---|
| Schott et al. (2007) | In France, the price of DXA a cost of 75 euros based on the average costs observed in practices.<br>Costs of the screening campaign were derived from French screening campaign for breast cancer [Watt, 2003]. Costs for preventive treatments based on risedronate or alendronate (Kanis and Johnson, 2002). Costs of hospitalization after a hip fracture were based on three previous French data. | Costs were independent of stratification, other than testing costs.<br>**Other costs included;** Costs of a screening campaign. | Cost of the test included.<br>Unclear where test cost was included. |
| Stevenson et al. (2007) | This report uses the costs reported in a systematic review by Kanis and colleagues (Kanis et al., 2002) having inflated, where applicable, to 2003/04 prices (Curtis et al., 2004).<br>NICE appraisal (2005).<br>BMD scan cost (Stevenson et al., 2005). | Costs were independent of stratification, other than testing costs.<br>**Other costs included;** Costing of GP consultation times in minutes for the entire testing process. | Cost of the test included.<br>Unclear where test cost was included. |
| Abbreviations; BMD (Bone Mineral Density); DXA (dual energy x-ray absorptiometry; GP (General Practitioner); OST (Osteoporosis Self-Assessment Tool); | | | |

## 5.4 Discussion of results

### 5.4.1 Statement of principal findings

The review demonstrated that the decision models built to assess stratified care in osteoporosis had used consistent methods. In what follows, the particular aspects of the modelling process will be discussed, as well as suggestions of how these approaches might inform model construction in subsequent chapters.

### 5.4.2 Modelling Structure

Modelling guidelines stipulate that the analyst should favour the simplest model able to appropriately represent the study objectives, natural history of the disease, and the treatment pathways (e.g. Philips et al., 2006). A simple model has the advantage of transparency and clarity, which facilitates easier validation (Barton et al., 2004). Due to the long time horizons considered in the analyses and natural history of the condition, none of the models used the simplest of model structures, the decision tree, with studies using either Markov or ISM models. Six of the models selected a Markov cohort model, which at a superficial level at least appears appropriate given in all these models the decision problem requires patients to transition in and out of transient and potentially recurrent health states over time, such as osteoporotic hip fractures.

However, generally Markov models are used for homogenous cohorts, so the crucial consideration is; can Markov models appropriately model a heterogeneous cohort which requires subdivision according to risk stratification? The answer appears to be yes to a degree. The Markov models in this review commonly modelled the stratification with the inclusion of a decision tree prior to the Markov analysis, where the branches of the decision tree reflect the different groups within the stratification. The different model parameters associated with different risk groups and treatments were used to parameterise unique Markov models for each branch. The overall costs and effects of stratification by screening risk factor, therefore, becomes a function (weighted by the probabilities on the branches) of the costs and effects associated with each of the Markov models associated

with the branching of the cohort in the decision tree. Given six of the models appear to have adopted such an approach; it seems these authors consider it as a reasonable means with which to approach modelling a stratified pathway.

Clearly, there are advantages of using the decision-tree prior to the Markov model in representing a stratified cohort, namely the simplicity of the approach. However, in more complex stratifications, the number of branches in the decision tree will increase, and where there are many states in the Markov model, the model may become unmanageable. Indeed, three models used an individual sampling model, albeit with some similarities in structure. The rationale given by these papers for the use of the ISM was the ability of the individual patient approach to provide more accuracy and flexibility than the cohort approach. Obviously, where stratified treatment pathways are being modelled numerous individual parameters are likely to influence subsequent model events than non-stratified treatments. For example, in the simulation model by Nayak et al. (2011), fracture risk is dependent upon age, femoral neck or lumbar spine BMD, and history of fracture. Given that seven strategies are being evaluated, with two treatment types, and fracture risks based upon four individual patient characteristics, and a Markov model with 14 states, it is clear that a Markov structure, in this case, would be impractical.

Therefore the critical issue for the analyst selecting the modelling approach, is can all relevant information relevant to the cost-effectiveness of a stratified care approach be captured within the decision tree and Markov structure? Or is an ISM model using individual parameters required to capture all information relevant to the cost-effectiveness of the approach? This is likely to depend upon the number of states in the model, as well as the importance of individual parameters deemed necessary in order to parameterise future events in the model.

### 5.4.3 Parameter Values

In terms of how the stratification affected the derivation of model parameters, in certain studies the fracture incidences were partially reflective of the same risk factors used to perform the stratification. However, this is conceptually different from basing those values solely upon which group patients were stratified to. For osteoporosis, data on fracture risks according to particular characteristics are readily available, and therefore obtaining such data posed no significant difficulties. In some case, models were able to base fracture risks upon a complex array of individual or cohort characteristics. This may not be the case for other conditions which make use of stratified treatments.

It was however noted that data were lacking on the long-term outcomes for patients who were stratified to receive either alendronate or risedronate. As a consequence, five studies assumed that the fracture reduction benefit returned to zero in a linear relationship over five years after completion of either therapy.

In all bar one study, the stratification process had no direct relationship with the treatment effect, with treatment effect set as a relative risk reduction for patients identified as having an osteoporotic fracture. Costs and utility values for the models were also set as reflective of health state, and independent of which group the patient was stratified or which treatment path the patient followed. The concept of having health states set independently of the treatment, and reflective of the natural biology of the condition, was also endorsed in the previous review of modelling in low back pain and sciatica in chapter 4. Part of the justification for doing so, which also appears relevant here, is reducing the number of parameters required and avoids problems such as that associated with the HTA model in the previous chapter, where the authors assumed the utility values associated various different treatments were all the same as successful surgery. Where a model is structured in such a way, the primary concern then becomes deriving differential parameters for movements between the states, in this case, fracture risks.

As expected the costs of stratification testing were included in the model to different degrees of complexity, this was the only way in which the stratification impacted upon the model parameters. Unfortunately, it is not clear in which states or stages the costs were inputted into any of the models.

### 5.4.4 Sensitivity Analyses

It was considered important by several authors to include the impact of varying the sensitivity and specificity of the stratification tool in various forms of sensitivity analysis. The commonality of this approach suggests that this ought to be included within any analyses where the sensitivity and specificity of the test are likely to impact upon its cost-effectiveness. Indeed, two of the studies included the uncertainty regarding the value of the test specificity and sensitivity in their PSA.

Screening rates, as well as age cohorts for screening, were also included in sensitivity analyses; these were shown as sub-group and scenario analyses. Clearly running the model for different cohorts will provide decision makers with richer information upon which to base a decision, and these analyses ought to reflect important sub-groups relevant to the decision.

### 5.4.5 Strengths and weaknesses of the review

The major strength of the review relates to the breadth of the search criteria as well as the variety of databases used to identify studies. Moreover, this is the first systematic review to identify, document and classify model-based economic evaluations of stratified treatment pathways for osteoporosis.

The major limitation of this review could relate to the search criteria. It is possible that model-based economic evaluations were excluded because they did not contain both ‘economic’ and ‘modelling’ terms used by the search strategy. In order to overcome this, the review did use as broad a search strategy as possible, using terms such as ‘economic’ or ‘model’ which in other contexts could be said to lack specificity.

Relating this back to the purpose of the thesis, to construct decision analytic models for stratified care in LBP and sciatica; whilst there are similarities between these two conditions and osteoporosis, there are also important differences between the two. For example, back pain is highly recurrent, whereas movement into and out of osteoporotic fractures are not likely to be as frequent as future flare-ups of back pain. Mean pain levels associated with back pain is also more constant over time, whereas osteoporotic fractures are mostly isolated events. These differences could limit the applicability of the approaches described in this review.

### 5.4.6 Implications

The review identified a consensus around the use of particular methodologies in the area of modelling in osteoporosis. It may, therefore, be possible to advocate the use of the techniques within this review for other stratified treatment approaches which share common features.

- The use of decision trees to subdivide the cohort, followed by a Markov structure can be considered appropriate. In some instances it may be necessary to use an ISM model in order to handle increasingly complex decision problems. This is a decision which ought to be taken in consultation with clinicians and experts in the field.
- States are best set as reflective of the underlying condition, instead of the treatment itself. This limits the number of parameters required in order to populate the model.
- Associated model utility values and transition probability have commonalities with the means of stratification, but are not directly associated with stratification.
- Sensitivity analyses may be required to consider the sensitivity and specificity of the test where this impacts the cost-effectiveness of the approach. It could also be important to consider the cost-effectiveness of the approach according to different sub-groups; this includes although is not limited to age cohorts and screening rates.

## 5.5 Conclusion

This review has highlighted commonalities in approaches towards the modelling of stratified treatments in the condition of osteoporosis. Namely, all models are initially sub-divided by a decision tree, following which each of the risk groups has its own Markov model (with same model structure regardless of stratification). Each Markov model is then differentially parameterised on the basis of cohort or individual characteristics. Treatment efficacies, utilities and costs, are almost entirely unrelated to the stratification process, with fracture risks only partially dependent upon the same data used to stratify patients.

These approaches were considered and in some cases applied, in the construction of the LBP and sciatica models in the following two chapters (Chapters 6 and 7 respectively).

# Chapter 6: MARKOV MODEL-BASED COST-EFFECTIVENESS ANALYSIS OF STRATIFIED CARE VERSUS USUAL CARE FOR LOW BACK PAIN

## 6.1 Introduction and Objective

The previous chapters have explored the theoretical and practical underpinnings of decision analytic modelling in health economic analysis; the presentations and clinical course of low back pain (LBP) and sciatica; and systematically reviewed the use of decision analytic modelling in both conditions and stratified care. In this chapter, all the above are considered together, beginning by describing the methodological approaches adopted for the Markov state transition model for LBP, outlining the justifications and assumptions which underpin model structure, data inputs and methods of analysis, reporting the model findings, finishing with a discussion of the implications of these results.

Chapter 1.5 described the current management of LBP patients including the stratified care model for LBP (STarT Back approach) which is clinically and cost-effective compared to non-stratified care. However, the long-term cost-effectiveness of the STarTBack stratified care model for LBP, is currently unknown, this could be addressed by decision modelling.

In the systematic review of published studies with decision model-based economic evaluations in LBP in Chapter 4, these models were shown to have several shortcomings (Hall et al., 2019). These include failing to adequately characterise the condition in health states and the absence of modelling the long-term pathway due to the lack of data and difficulty of modelling symptom flare-ups. Moreover, there are currently no decision-analytic models of a stratified care approach to managing LBP. Therefore, the work described in this chapter aims to conceptualise the first decision model of a stratified care approach for the management of LBP.

## 6.2 Methods

A Markov model was constructed with two-monthly cycles, and analysis performed over a ten-year time horizon in order to assess the cost-effectiveness of stratified care versus usual care from the NHS perspective, with a secondary analysis also performed from the societal perspective. In what follows, each aspect of the methodological approach is discussed in turn.

### 6.2.1 Consultation with experts

The lack of adequate modelling in both LBP and sciatica meant it was imperative to ensure that at all stages of the model development process consultation took place with clinicians as well as experts drawn from epidemiological and health economic research backgrounds (full details in appendix 4).

### 6.2.2 Choice of Model

In determining which model to select, the simplest model ought to be chosen to adequately capture the complexity of the decision problem (Barton, 2004). A Markov model was determined as the most appropriate for this condition, given that patients move between different levels of functioning and can experience periods of recurring poor function due to LBP.

A decision-tree model cannot easily facilitate representation of adequate time horizon necessary for the clinical course of LBP and was therefore ruled out. An individual simulation model (ISM) had initially been proposed, in Chapter 5, ISM models were shown to be suitable to model stratified treatments, owing to their ability to track and update patient characteristics, as well as allow the probability of future events to be dependent upon the aforementioned characteristics. After consultation with experts, and analysis of relevant available data from cohort studies (BaRNS, see 6.3.10; BeBack see 6.3.11), it became evident that movement of patients between health states over the long-

term (between twelve months and seven years) would be best dependent upon their health state at twelve months. In a Markov model, the number of states needed to incorporate the dependency could cause the model to grow infinitely complex, whereas simulation modelling via trackers allow for the incorporation of these phenomena whilst avoiding unnecessary complexity. However, the use of matrix algebra to generate the transition probabilities (Chhatwal et al., 2016) determined that transition probabilities from 12-months to 7-years assumed a linear form. Accordingly, during this period, patients would transition as a mathematical function of patient function at 12-months (full explanation in 6.3.7).

ISMs can also be appropriate where the analysis is of interventions with heterogeneous populations, such as LBP. However, given the data available for analysis, the use of an ISM would have provided little informational gain relative to the loss of statistical power that occurs from having low numbers of observations to derive estimates for each patient's profiles. Information presented in Appendix 5, shows that even the introduction of one additional patient characteristic, function at baseline, would lead to very low sample sizes for transition estimates. Therefore, given the use of individual trial-based sampling would have reduced statistical power and therefore widened the uncertainty around model estimates, and with only treatment and function at twelve month determining progression through the a cohort model was settled upon. Heterogeneity was explored to a degree in sub-group analysis, where each risk-group had an individual unique analysis.

The model was constructed in TreeAge Pro 2017 version. All statistical analyses were performed in STATA 15 / MP.

### 6.2.3 Model Population

Whilst the treatment effect, costs, and utility values used in the base case analysis were derived from the STarT Back trial (Hill et al., 2011; Whitehurst et al., 2012), the consensus

of the experts involved in consultations was that the IMPaCT Back study (Foster et al., 2014) population would most closely resemble the consulting population who would likely receive the intervention. As a consequence, the relevant study population was drawn from the IMPaCT Back study (full methods are reported, Foster et al., 2014). Briefly, IMPaCT Back was a prospective, primary care–based, quality improvement study in England with a before-and-after design, including adults aged over 18 years old, consulting with non-specific LBP of any duration, as identified using standardised Read codes (Hassey et al. 2001). Baseline model characteristics were taken directly from the before phase of this study, where participants had a mean age of 53 years at the start of the model, 55% were female, 37% were classified as low-risk on the STarT Back tool, 41% medium-risk, and 22% were high-risk. Baseline characteristics of patients who consented to participate in the study are presented in Table 6.1 below.

**Table 6.1 Baseline characteristics of IMPaCT Back study patients**

| Characteristic of participating patients | Before | After |
|---|---|---|
| Age, mean (SD), y | 53.0 (15.0) | 54.1 (14.8) |
| Sex, female, No. (%) | 202 (55) | 330 (60) |
| Currently in paid employment, No. (%) | 227 (62) | 323 (59) |
| Time off work for back pain in past 12 months, No. (%) | 109 (49) | 133 (42) |
| Disability: RMDQ score, mean (SD) | 8.7 (5.9) | 8.4 (5.7) |
| Pain intensity: NRS rating, mean (SD) | 5.3 (2.4) | 5.0 (2.6) |
| Duration of back pain episode, No. (%) | | |
| <1 month | 75 (20) | 94 (17) |
| 1–3 months | 62 (17) | 102 (18) |
| 3–6 months | 75 (20) | 111 (20) |
| 6 months to 3 years | 82 (22) | 130 (24) |
| >3 years | 74 (20) | 117 (21) |
| Risk group, No. (%) | | |
| Low | 136 (37) | 214 (39) |
| Medium | 151 (41) | 232 (42) |
| High | 81 (22) | 108 (20) |

Abbreviations: NRS (Numerical Rating Scale); RMDQ (Roland Morris Disability Questionnaire); SD (Standard deviation)

### 6.2.4 Definition of the intervention for non-specific LBP

Details of the intervention (including the stratification tool used) are reported in 1.5.1, and described in full elsewhere (Hay et al., 2008; Hill et al., 2011). Briefly, in the intervention group, patients' risk group was calculated using the STarT Back Screening Tool during the baseline clinical assessment, and patients were matched to treatments recommended for their risk group. Patients in the low-risk group had one off session where they received advice on appropriate levels of activity and return to work, explanation of the condition and reassurance on expected good outcome/prognosis. Medium-risk patients were referred for further physiotherapy sessions to address symptoms and function. High-risk patients were referred for further physiotherapy with emphasis on addressing psychosocial obstacles to recovery in addition to symptoms and function.

In the control group, usual care, during baseline clinical assessment, further management decisions were made in accordance with the assessing physiotherapist's clinical judgment, without the use of the STarT Back Tool, and physiotherapy treatments received were part of normal NHS physiotherapy care.

### 6.2.5 Model health states and structure

The model had seven different health states, shown in figure 6.1.

Initially, for the first twelve months, a patient's transition through the model states of risk reflect transitions derived from the STarT Back trial directly, imposed upon underlying baseline patient risk profiles drawn from the IMPaCT Back cohort. Accordingly, patients begin in one of three risk groups, low-, medium-, or high-risk of persistent disability, and may move between these states during the first year; patients could also exit the model at any time into death. These states were considered reasonable by all experts involved in the consultation phase.

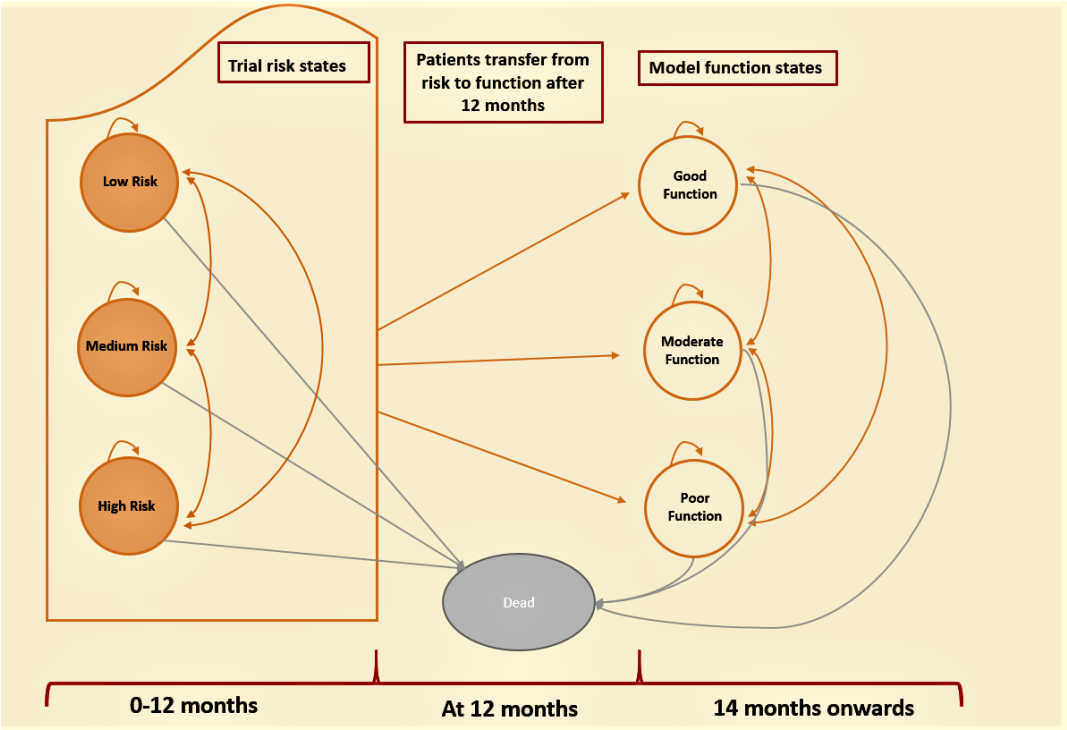



**Figure 6-1 State transition model schematic**

As indicated by the arrows, patients were able to move between the risk groups throughout the first-year in each two-month time cycle and had utility values associated with their risk group. However, it was not possible to calculate the individual contribution of each two-monthly cycle to overall first-year cost owing to healthcare resource use data only being available for patient healthcare use in the "past twelve months". Accordingly, costs for the first year were a one-off annual cost based upon baseline risk group and attached at cycle 1.

At twelve months, as indicated by the three arrows projecting from risk states towards function states, patients moved into one of three possible states dependent upon their function as measured by their RMDQ (Roland Morris Disability Questionnaire) score (Roland and Fairbank, 2000) at the end of the trial. The RMDQ is a measure of back pain-related disability, where a higher score reflects higher levels of disability on a 24-point scale.

Having initially proposed using pain or trajectory grouping (see 2.2.3) as health state, the decision to use patient function as the long-term health state choice was initially suggested during the consultations with the clinicians and experts in the field, on the grounds that improving patient function was the main aim of treatments provided in the STarTBack trial. Pain as a health state is included as a sensitivity analysis. Further expert consultations endorsed this approach. Function data (RMDQ scores) from the BaRNS cohort study (Dunn et al., 2013) were also available across seven years, to allow estimates of patient prognosis based upon function. In The BaRNS prospective cohort study (Dunn et al., 2013), 228 people consulting their GP with LBP, aged 30-59 years, on whom data had previously been collected during 2001 and 2003, were surveyed again in 2009. One hundred fifty-five participants responded and provided sufficient longer-term data for the model's primary analyses. To calculate model transitions, RMDQ scores taken at twelve months were used as a baseline, with ratings collected at the start of the seventh year of follow up used to determine how patients had moved between function states throughout that period in the model.

There was, however, an absence of literature on the definition of RMDQ score cut-offs for LBP patients. As a result, opinions were elicited from experts and augmented by analysis of the distribution of RMDQ scores in the risk groups in the STarT Back trial and from evidence on RMDQ cut-offs in sciatica patients (Patrick et al., 1995). The following categories were used; RMDQ score of 11 and over was classified as poor function, 5 to 10 inclusive was considered moderate function, and 0-4 was considered good function.

### 6.2.6 Model time horizon and cycle length

In this model, a time horizon of ten years was adopted, which was considered adequate to capture meaningful differences between stratified care and usual care. This view was endorsed by researchers in stratified care at Keele, as well as the GPs and physiotherapists

advising on these issues. In terms of cycle length, because of the time points in the STarT Back trial, it was considered appropriate to use two-monthly cycles to facilitate calculation of the transition probabilities used.

### 6.2.7 Transition probabilities

In the first year of the Markov model, patients transitioned every two months between states representing risk groups from the STarT Back trial (low, medium, high). At the twelve-month endpoint of the trial, patients then transitioned into health states based upon their allocated function category, as measured by RMDQ in the STarT Back data. Upon assignment to the initial functional category, patients subsequently transitioned every two months between states (or remained in the same state) for the remaining nine years of the model. Their transitions for this period were based upon observed transitions in patients with the same function at twelve months to seven years follow-up from the BaRNS cohort study.

### 6.2.8 Transition probabilities for the first year

In order to derive the transition probabilities for the first year in both stratified and usual care, matrix multiplication was used to transform the four-month (representing baseline to four months) and eight-month (representing four to 12 months) transition probabilities available in the STarT Back data into two-monthly probabilities. The matrices require the simplifying assumption that transition probabilities are linear across the first four months, and then assume a different linear function across four to 12 months. The use of matrix multiplication transforms the probability into an underlying rate, and where there is movement between various states over time; it has been shown to produce more accurate estimates than the traditional method (Chhatwal et al., 2016).

Using the transitions of patients within the first four months of the STarT Back trial allows the model to represent the initial improvement achieved in that first four months on both

stratified care and usual care. Using separate transitions for zero to four months, and four-12 months is preferred to using zero-12 month transitions where the rate of transition would have been the linear across the entire first year.

Full details of the process of Matrix multiplication can be found in Appendix 6. Tables 6.2 and 6.3 show how this method of minimising the sum of errors between the observed data and the numbers predicted by the matrix, using the Microsoft Excel solver add-on, generate closely matching estimates between predicted patients and transitions observed directly in STarT Back data.

**Table 6.2 Observed number of patients moving in risk groups from baseline to four months on stratified care (from observed data – STarT Back trial)**

| From/ to | n | n | n | n |
|---|---|---|---|---|
| | low | med | high | Total |
| low | 99 | 6 | 2 | 107 |
| med | 150 | 38 | 6 | 194 |
| high | 81 | 21 | 9 | 111 |
| | | | | 412 |

**Table 6.3 Predicted numbers of patients moving in risk groups from baseline to four months on stratified care (from Matrix multiplication)**

| From / To | n | n | n | n |
|---|---|---|---|---|
| | low | Med | high | Total |
| low | 99 | 7 | 1 | 107 |
| med | 150 | 37 | 7 | 194 |
| high | 81 | 21 | 9 | 111 |
| | | | | 412 |

As the transition probabilities were calculated for each risk group separately, no standardisation was applied as it was assumed that these probabilities could be directly applied to the risk groups in the IMPaCT Back study population, although it is acknowledged that there could be some small differences between the characteristics of risk groups in both studies.

**Table 6.4 Model transition probabilities and distributions.**

| | Stratified Care | | | Usual Care | | |
|---|---|---|---|---|---|---|
| Movement | Mean (SE) | Dist | α,β or n | Mean (SE) | Dist | α,β or n |
| Zero to Four Months | | | | | | |
| Low to Medium | 0.053 (0.022) | Beta | 5.64, 100.4 | 0.105 (0.043) | Beta | 5.35,45.65 |
| Low to High | 0 | N/A | - | 0 | N/A | - |
| Medium to Low | 0.588 (0.035) | Dirichlet | 114.1 | 0.764 (0.044) | Dirichlet | 71 |
| Medium to Med | 0.279 (0.032) | Dirichlet | 54.1 | 0.035 (0.019) | Dirichlet | 3.3 |
| Medium to High | 0.133 (0.024) | Dirichlet | 25.8 | 0.201 (0.042) | Dirichlet | 18.7 |
| High to Low | 0.386 (0.046) | Beta | 42.5, 67.5 | 0 | N/A | |
| High to Medium | 0.614 (0.046) | Beta | 67.5,42.5 | 0.639 (0.067) | Beta | 31.97,45.65 |
| Four months to Twelve Months | | | | | | |
| Low to Low | 0.965 (0.011) | Beta | 259.67,9.33 | 0.966 (0.017) | Dirichlet | 105.3 |
| Low to Medium | 0.035 (0.011) | Beta | 9.33,259.67 | 0.030 (0.016) | Dirichlet | 3.3 |
| Low to High | 0 | N/A | - | 0.004 (0.006) | Dirichlet | 0.5 |
| Medium to Low | 0.146 (0.049) | Dirichlet | 7.6 | 0.166 (0.066) | Dirichlet | 5.3 |
| Medium to Medium | 0.784 (0.057) | Dirichlet | 40.8 | 0.787 (0.072) | Dirichlet | 25.2 |
| Medium to High | 0.069 (0.035) | Dirichlet | 3.6 | 0.047 (0.037) | Dirichlet | 1.5 |
| High to Low | 0.065 (0.071) | Dirichlet | 0.2 | 0 | N/A | - |
| High to Medium | 0.198 (0.115) | Dirichlet | 2.38 | 0.223 (0.120) | Beta | 8.54,2.45 |
| High to High | 0.737 (0.127) | Dirichlet | 9.48 | 0.777 (0.120) | Beta | 2.45,8.54 |
| Initial Distributions of patients into function states | | | | | | |
| Movement | Mean (SE) | Dist | α,β or n | Mean (SE) | Dist | α,β or n |
| Low-risk to Poor Function | 0.071 (0.026) | Dirichlet | 21 | 0.038 (0.017) | Dirichlet | 5 |
| Low-risk to Moderate Function | 0.203 (0.023) | Dirichlet | 60 | 0.269 (0.039) | Dirichlet | 35 |
| Low-risk to Good Function | 0.726 (0.015) | Dirichlet | 215 | 0.692 (0.041) | Dirichlet | 90 |
| Medium-risk to Poor Function | 0.595 (0.062) | Dirichlet | 38 | 0.765 (0.074) | Dirichlet | 26 |
| Medium-risk to Moderate Function | 0.343 (0.060) | Dirichlet | 22 | 0.269 (0.039) | Dirichlet | 5 |

| Movement | Mean (SE) | Dist | α,β or n | Mean (SE) | Dist | α,β or n |
|---|---|---|---|---|---|---|
| Medium-risk to Good Function | 0.062 (0.031) | Dirichlet | 4 | 0.088 (0.049) | Dirichlet | 3 |
| High-risk to Poor Function | 0.857 (0.078) | Beta | 16.28,2.71 | 0.90 (0.100) | Beta | 7.2, 0.8 |
| High-risk to Moderate Function | 0.143 (0.078) | Beta | 2.71,16.28 | 0.100 (0.100) | Beta | 0.8, 7.2 |
| High-risk to Good Function | 0 | N/A | - | 0 | N/A | - |

| Transitions between function states, one year to ten years | | | |
|---|---|---|---|
| | Mean (SE) | Distribution | n |
| Poor to Poor | 0.986 (0.017) | Dirichlet | 48.3 |
| Poor to Moderate | 0.005 (0.011) | Dirichlet | 0.3 |
| Poor to Good | 0.008 (0.013) | Dirichlet | 0.4 |
| Moderate to Poor | 0.009 (0.013) | Dirichlet | 0.4 |
| Mod to Mod | 0.959 (0.029) | Dirichlet | 46 |
| Moderate to Good | 0.032 (0.025) | Dirichlet | 1.6 |
| Good to Poor | 0.001 (0.003) | Dirichlet | 0.1 |
| Good to Moderate | 0.007 (0.008) | Dirichlet | 0.8 |
| Good to Good | 0.991 (0.009) | Dirichlet | 110 |

For probabilistic analyses, Dirichlet distributions, a multivariate form of the beta distribution, were used for transition probabilities, unless one of the movements out of the state was known to be zero, in which case beta distributions were used.

Transitions to death were calculated by using 2015/16 ONS life tables (ONS, 2016). Annual probability of death for males and females between 55 and 65 was calculated by firstly by weighting annual mortality risk towards the 57% of the IMPaCT Back sample which was female. Weighted annual mortality risk was converted to annual probability using $p = 1 - e^{-rt}$. This annual probability was converted into a two-monthly rate using $r = \frac{1}{t}\ln(1-p)$, and then converted to a two-monthly probability again using $p = 1 - e^{-rt}$. Transitions to death used in the model are shown in Table 6.5

**Table 6.5 Two monthly model transition to death**

| Age | Weighted two-monthly probability of death |
|---|---|
| 55 | 0.000652 |
| 56 | 0.000704 |
| 57 | 0.000767 |
| 58 | 0.000842 |
| 59 | 0.000937 |
| 60 | 0.001044 |
| 61 | 0.001131 |
| 62 | 0.001241 |
| 63 | 0.001365 |
| 64 | 0.001481 |
| 65 | 0.001589 |

### 6.2.9 Moving from the risk groups to the function health states

At twelve months, patients moved into health states representing function, the values of which are shown in Table 6.4 under the heading "Initial distributions of patients into function states". These transition probabilities were based upon the proportion of patients in each function state in each of the risk groups at twelve months in the STarT Back data. As the model is calibrated with STarT Back data, proportion of patients in each function state at twelve months in the model is almost identical to function group at twelve months in the STarT Back data. Probabilities were calculated separately for both stratified care and usual care, in order to reflect the lower proportion of stratified care patients in poor function at twelve months. In the model, costs and rewards associated with each of the function states are first attached in cycle seven, at fourteen months, based upon the proportion of patients in each of the function states at twelve months.

### 6.2.10 Transition probabilities for twelve months to seven years.

From 14 months onwards, the transition probabilities between function states were exactly the same in both the stratified care and usual care arms. However, the differential

distribution of patients into function states at twelve months, with equal transitions from one year onwards, means that fewer stratified care patients remained in poor function for the duration of the model. This assumption that transitions between function states occur at the same rate in both arms was explored further in sensitivity analysis. Whilst patient movement between the states is linear over time, they are able to move between any of the states, although only a small proportion of patients make movements from good to poor function.

Again, matrix algebra was used to derive the transition of patients from one year to seven years using data from the BaRNS prospective cohort study (Dunn et al., 2013). To determine whether it was appropriate to calculate transition for the model population from BaRNS data, mean RMDQ scores for each function state in the STarT Back trial patients at twelve months were compared with the mean RMDQ scores of the patients in the BaRNS sample at twelve months and seven years. These comparisons are shown in Table 6.6.

**Table 6.6 Mean RMDQ scores of patients in STarT Back trial and BaRNS cohort**

| Function | STarT Back at twelve months | BaRNS at twelve months | BaRNS at seven years |
|---|---|---|---|
| Poor | 15.021 | 16.227 | 15.591 |
| Moderate | 7.228 | 6.691 | 6.667 |
| Good | 1.177 | 1.237 | 1.374 |

Whilst the overall scores were slightly lower in the STarT Back trial sample at twelve months relative to the BaRNS sample at the same time, neither of the samples had statistically significant differences either in population or function state RMDQ ($p<0.05$).

With a gap of six years between the twelve-month and seven-year follow questionnaires in the BaRNS study, it was again necessary to use the matrix algebraic methods described above to transform six-year transition probabilities into two-monthly transitions. One set of

matrix derived transition probabilities were used for all movements in the model from twelve months onwards, equal in both stratified care and usual care; these are shown in Table 6.7.

**Table 6.7 Matrix derived two-month transition probabilities used for extrapolation**

| Two months | function | | |
|---|---|---|---|
| From / To | Poor | Moderate | Good |
| Poor | 0.98640 | 0.00544 | 0.00816 |
| Moderate | 0.008725 | 0.959074 | 0.032201 |
| Good | 0.00130 | 0.00739 | 0.99131 |

Whilst these transition probabilities were calculated reflecting patient movement from twelve months to seven years, it was further assumed that the derived probabilities would hold over the next three years. Accordingly, these probabilities were used to move patients between states from seven to ten years in the model. In all cases the standard errors were calculated by using SE (p) = $\sqrt{(p(1-p)/n)}$, where p was the probability of movement between the states as generated by the matrices, and n reflects the number of patients in each of the function/risk states at the beginning of the period the matrix is reflecting, directly obtained from the BaRNS cohort data.

### 6.2.11 Costs

The base-case economic evaluation was performed from the NHS perspective, which takes into account costs solely incurred by the NHS and excludes the value of private healthcare costs. Costs associated with the first year reflected the resource usage from the Start Back trial (Whitehurst et al., 2012). The data used were the same as that presented in the published within-trial analysis, which performed multiple imputation and included costs directly associated with the trial (Whitehurst et al., 2012). The resource use data for the first year originates from STarT Back data reflecting responses to self-report

questionnaires and included consultations with GPs and nurses, other healthcare professionals, prescriptions, hospital procedures, as well as prescribed medications and over the counter treatments, the costs of the latter were included only in the societal perspective.

As the modelling for the first year is predicated upon the treatment effect associated with the trial, the costs related to delivering the stratified care model in the trial were included, including the costs associated with the trial-specific protocol. The mean costs for low, medium and high-risk patients (classified at baseline) were calculated separately for stratified care and usual care. Resource usage estimates are similar, although not identical, to those reported by the cost-utility analysis of the STarT Back trial (Whitehurst et al., 2012). The difference in the value of the costs used for the first year in this study originates primarily from the differential increase in unit costs since the initial trial evaluation used 2008 costs.

Beyond the first year, data from the BeBack study (Foster et al., 2008) were used to derive the costs associated with one year of healthcare for LBP patients in each of the function states. The BeBack study was a prospective cohort study of patients aged 18 to 60 years, consulting their GP for LBP between September 2004 and April 2006. One thousand five hundred and ninety-one patients participated in the cohort at the baseline stage. The BeBack study data reflect healthcare usage in what would be considered “usual care”, and therefore for the base case analysis, it was assumed that stratified care does not impact upon the long-term healthcare usage of patients beyond 12 months. Given this represents one of the most significant sources of uncertainty in the model, the impact of varying this assumption is explored later.

Unit costs in the model were based upon 2015/2016 prices, and are shown in Table 6.8. Total costs were discounted at 3.5% as reflective of NICE guidance (NICE, 2018). It is

noted that all costs are derived from questionnaires, which stipulate costs directly related to back pain only. Physiotherapy, nurse and GP visits were costed from the Unit Costs of Health and Social Care 2016 (PSSRU, 2016). The PSSRU includes the capital overheads associated with the receipt of care, whilst costs-per-visit were calculated by using face-to-face patient multipliers from the 2009 version of the document (PSSRU, 2009). X-Ray, blood test, and epidural injection costs were not available in the National Schedule of Reference Costs; therefore the costs were taken from the original Whitehurst et al. (2012) study and inflated using the Health Service Cost Index (HCSA, 2016). The costs associated with over-the-counter and prescribed medications in all cycles were taken from STarT Back trial data, owing to the greater detail in the STarT Back trial than the BeBack cohort study. These costs were also inflated using the Health Service Cost Index (HCSA, 2016).

Resource usage data derived from STarT Back and BeBack studies were multiplied by the unit prices to create annual costs associated with each state; these are shown in Table 6.9. As noted above, in the first year, these estimates were entered into the model at cycle zero, to reflect the initial risk group. In cycles seven onwards, the annual estimates of the costs associated with each of the function states are divided by six to create a two-monthly cost.

**Table 6.8 Unit costs assigned to healthcare resource use data**

| Healthcare Resource | Unit cost | Source |
|---|---|---|
| Primary Care: | | |
| GP consultation | £34.00 | Unit costs of Health and Social Care 2016* |
| Nurse consultation | £15.93 | Unit costs of Health and Social Care 2016* |
| Physiotherapy: Initial session | £41.88 | Unit costs of Health and Social Care 2016* |
| Physiotherapy: initial 1-hour assessment (high-risk) | £55.84 | Unit costs of Health and Social Care 2016* |
| Physiotherapy: Follow up sessions (1/2 hour) | £27.92 | Unit costs of Health and Social Care 2016* |
| Hospital-Based Care: | | |
| Consultant LBP first attendance | £136 | NHS National Schedule of Reference Costs |
| Consultant LBP follow up | £83 | NHS National Schedule of Reference Costs |
| Admission to A and E | £147 | NHS National Schedule of Reference Costs |
| X-Ray | £34.23 | NHS National Schedule of Reference Costs |
| CT Scan | £99 | NHS National Schedule of Reference Costs |
| MRI Scan | £145 | NHS National Schedule of Reference Costs |
| Blood test | £18.49 | Whitehurst et al. (2012)# |
| Epidural injections | £218.89 | Whitehurst et al. (2012)# |
| First consultation with other HC professionals | £57.00 | NHS National Schedule of Reference Costs |
| Follow up consultation with other HC professionals | £47.00 | NHS National Schedule of Reference Costs |
| Over Counter medication | Patient specific | STarT Back data |
| Prescribed medication | Patient specific | STarT Back data |

*Ratio of patient face-to-face contact is taken from Unit Costs of Health and Social Care 2009. All visits assumed to be to surgery, no home visits.

#Costs inflated to 2016 prices using the Health Service Cost Index

**Table 6.9 Total annual costs per state**

| | | NHS annual | | Annual private costs | |
|---|---|---|---|---|---|
| | Distributions | Costs(£) mean (SD) | α, λ | Costs(£) mean (SD) | α, λ |
| Costs year 1 | | | | | |
| Low-risk_UC | Gamma | 136.47 (147.20) | 0.86, 0.006 | 23.57 (51.27) | 0.211, 009 |
| Medium-risk_UC | Gamma | 312.72 (313.09) | 0.998,0.003 | 42.71 (114.21) | 0.14, 0.003 |
| High-risk_UC | Gamma | 372.35 (399.93) | 0.867,0.002 | 35.35 (69.12) | 0.262, 0.007 |
| Low-risk_SC | Gamma | 90.90 (169.27) | 0.288,0.003 | 22.70 (72.50) | 0.098,0.004 |
| Medium-risk_SC | Gamma | 248.81 (239.96) | 3.436,0.014 | 37.09 (120.67) | 0.094,0.003 |
| High-risk_SC | Gamma | 375.26 (228.09) | 2.708,0.007 | 30.82 (85.62) | 0.130,0.004 |
| Costs 2-10 years | | | | | |
| Poor function | Gamma | 503.53 (703.66) | 0.512,0.001 | 174.09 (380.65) | 0.209,0.001 |
| Moderate function | Gamma | 235.45 (340.11) | 0.479,0.002 | 153.26 (300.34) | 0.260,0.002 |
| Good function | Gamma | 85.10 (245.73) | 0.120,0.001 | 112.02 (249.01) | 0.202,0.002 |
| Abbreviations: SC (Stratified care); UC (Usual care) | | | | | |

For the societal analysis, due to lack of nationally representative unit cost estimates for private health care, unit costs associated with private healthcare were assumed to be the same as that of the NHS equivalent. Resource use was taken from the BeBack and STarT Back datasets, which provide detail on the number of visits to healthcare practitioners, alongside whether those visits were to an NHS professional or a private one. In relation to healthcare practitioner visits, very few patients received both NHS and private care, and where patients indicated they had received both NHS and private healthcare, the number of visits was divided by two, with half of the total cost allocated to NHS and half to private. The first-year cost breakdown is presented in Table 6.10.

**Table 6.10 Healthcare costs, £ (2016), first year, for stratified care and usual care, by risk group**

| | Stratified care | | | Usual care | | |
|---|---|---|---|---|---|---|
| | Low | Medium | High | Low | Medium | High |
| Clinic and physiotherapy | 40.18 | 141.56 | 213.76 | 98.97 | 114.98 | 122.11 |
| GP | 15.05 | 33.51 | 61.41 | 20.62 | 47.48 | 61.13 |
| Nurse | 0.00 | 0.98 | 5.09 | 0.34 | 1.65 | 3.63 |
| Other consultations | 0.00 | 4.41 | 1.76 | 4.52 | 10.61 | 4.11 |
| Hospitalisation costs | 6.85 | 12.30 | 23.80 | 3.08 | 21.76 | 23.36 |
| Non-study NHS physio | 23.59 | 21.95 | 43.92 | 14.37 | 57.29 | 93.95 |
| Non-study private physio | 0.59 | 2.06 | 6.22 | 4.71 | 15.92 | 6.83 |
| NHS other | 1.76 | 12.44 | 25.41 | 3.60 | 19.97 | 30.11 |
| Private other | 12.63 | 21.33 | 11.07 | 9.04 | 9.64 | 8.64 |
| Prescriptions | 3.30 | 17.26 | 22.68 | 1.08 | 28.32 | 12.31 |
| OTC medication | 9.49 | 13.65 | 13.61 | 17.74 | 17.15 | 19.89 |
| Total NHS costs | 90.90 | 248.81 | 375.26 | 136.47 | 312.72 | 372.35 |
| Total private costs | 22.70 | 37.09 | 30.82 | 23.57 | 42.71 | 35.35 |
| Total healthcare costs | 113.60 | 286.43 | 407.23 | 160.04 | 355.44 | 407.80 |
| OTC: Over the counter | | | | | | |

The breakdown of annualised costs associated with each of the function states, is presented in Table 6.11 below.

**Table 6.11 Costs, £ (2016) associated with function states beyond 12-months**

| | Poor function | Moderate function | Good function |
|---|---|---|---|
| GP NHS | 121.91 | 55.75 | 23.1 |
| GP private | 0.81 | 0 | 0.25 |
| Nurse NHS | 3.77 | 3.91 | 0 |
| Nurse private | 0 | 0.28 | 0.94 |
| Hospital Attendance* NHS | 70.51 | 28.05 | 6.31 |
| Hospital Attendance* private | 15.13 | 14.39 | 1.52 |
| NHS physio | 207.62 | 108.75 | 31.02 |
| private physio | 36.64 | 61.52 | 51.27 |
| NHS other | 63.39 | 27.17 | 31.55 |
| Private other | 19.65 | 9.29 | 4.51 |
| Prescriptions | 26.69 | 10.66 | 0.97 |
| OTC medication | 17.39 | 12.83 | 7.16 |
| Total NHS costs | 503.53 | 235.45 | 85.10 |
| Total private costs | 174.09 | 153.26 | 112.02 |
| Total healthcare costs | 677.62 | 388.71 | 197.12 |

OTC: Over the counter

*Hospital attendance includes the cost of imaging

To assess the impact of stratified care upon work participation, self-reported days of work absence owing to back pain were estimated from the STarT Back and BeBack datasets, with associated costs assigned using the HCM, whereby productivity losses are estimated by multiplying some time loss from work absence by a wage rate. In this particular model, during the first year work absence in days was estimated from the STarT Back study, and multiplied by a daily mean IMPaCT Back population wage (derived from respondent-specific wage estimates based upon social class), and inflated to 2015/16 equivalent using the ONS Wages and salaries annual growth rate (ONS, 2019). For the remainder of the

model, length of work absence was derived from the BeBack data and multiplied by mean wage identified in IMPaCT Back estimates and adjusted using ONS inflators. In both cases, mean work absence in days has been adjusted to account for the 40.35% not in employment in the IMPaCT Back study.

Table 6.12 presents the mean absence due to back pain in the STarT Back trial and BeBack cohort study, the wage used, and the percentage of participants employed in the model population, the costs of work absence, and total societal costs used in the analysis.

**Table 6.12 Societal costs of stratified care vs usual care**

| | Stratified care | Usual care |
|---|---|---|
| Mean days absence due to back pain (STarT Back data) | Low: 0.37 | Low: 3 |
| | Medium: 4.07 | Medium: 18.44 |
| | High: 9.85 | High: 10.57 |
| Mean daily wage (IMPaCT Back data) | £95.24 | |
| % employed in IMPaCT Back study | 59.65% | |
| Mean cost of back pain related work absence* | Low: £23.56 | Low: £176.30 |
| | Medium: £179.59 | Medium: £968.82 |
| | High: £444.45 | High: £449.86 |
| Total societal costs (HC costs plus absenteeism costs) Mean (SD) | Low: £137.17 (£233.01) | Low: £340.17 (£975.86) |
| | Medium: £471.31 (£1125.12) | Medium: £1324.26 (£3588.46) |
| | High: £867.55 (£2499.55) | High: £867.43 (£1494.77) |
| By function | | |
| Mean days absence due to back pain (BeBack data) | Poor: 27.43 | |
| | Moderate: 13.20 | |
| | Good: 5.60 | |
| Mean wage (IMPaCT Back data) | £95.24 | |
| Mean cost of back-pain related work absence* | Poor: 727.26 (1677.50) | |
| | Moderate: £527.45 (1336.14) | |
| | Good: £232.85 (721.23) | |
| Total societal costs (HC costs plus LOP costs) Mean (SD) | Poor: 1276.78 (149.57) | |
| | Moderate: 900.41 (131.38) | |
| | Good: 429.56 (82.39) | |
| * Patients who do not work accrue zero days off and £0 societal cost, numbers are adjusted for 0.5965 of model population in employment<br>LOP Loss of productivity | | |

### 6.2.12 Quality adjusted life years (QALYs)

A QALY is a function of quality and quantity of life and is calculated simply by the multiplication of a quality of life (QoL) score, by the number of years lived. For the model, EQ-5D 3L scores for the first year were taken from STarT Back data, with longer-term values originating from BeBack. Both sets of responses were converted to utility scores based upon the York tariff (Dolan et al., 1995). In the model, QALYs were discounted at 3.5% per year, as per current NICE guidelines (NICE, 2018). In the PSA, the model uses EQ-5D scores sampled from their distribution, and subsequently divides by six in order to obtain two-monthly QALYs.

There were a number of calculations used to obtain the EQ-5D scores used in the first year of this model. The patient population reflected standardised mean baseline EQ-5D score in the IMPaCT Back study and applied a treatment effect upon EQ-5D score for each risk group derived from the STarT Back trial data. Therefore, it was assumed that the absolute changes in EQ-5D for each of the risk groups in STarT Back trial are transferrable to the IMPaCT Back study population.

These steps were necessary because the IMPaCT back population generally had higher EQ-5D scores at baseline compared to the STarT Back study population. These are shown in Table 6.13.

**Table 6.13 EQ-5D values at baseline in IMPaCT Back study**

| Risk Group | Baseline EQ-5D IMPaCT Back | Baseline EQ-5D STarT Back |
|---|---|---|
| Low | 0.798 | 0.730 |
| Medium | 0.677 | 0.551 |
| High | 0.389 | 0.299 |

Stratified care and usual care each provided different changes in EQ-5D score in each risk group across the first twelve months. To calculate the EQ-5D treatment effect across the

first year, regression analysis would usually have been employed to estimate the change in EQ-5D score for each intervention, controlling for baseline EQ-5D in each risk group at the three different time points. However, because patients move risk groups in each of the time intervals, this was not possible.

Instead, EQ-5D scores were calculated at three time points; baseline, four, and twelve months, for each risk group, across both stratified care and usual care, using simple descriptive statistics, shown in Table 6.14. The impact of the treatment upon mean EQ-5D score in each risk group was then calculated at four months, and again at twelve months, to produce a value of the treatment effect for each risk group. For example, low risk on stratified care, improved by 0.057 from baseline to four months.

**Table 6.14 Treatment effect upon EQ-5D, stratified care vs usual care**

| Risk Group | Baseline | 4 months | 12 months | Treatment Effect 0-4 months | Treatment Effect 4-12 months |
|---|---|---|---|---|---|
| Low_ SC | 0.728 | 0.785 | 0.779 | +0.057 | -0.006 |
| Medium_SC | 0.541 | 0.480 | 0.414 | -0.061 | -0.066 |
| High_SC | 0.325 | 0.108 | 0.156 | -0.217 | +0.048 |
| Low_UC | 0.733 | 0.798 | 0.750 | +0.065 | +0.017 |
| Medium_UC | 0.571 | 0.530 | 0.425 | -0.041 | -0.105 |
| High_UC | 0.245 | 0.176 | 0.235 | -0.069 | -0.010 |
| Abbreviations: SC Stratified Care; UC Usual Care | | | | | |

Next, EQ-5D scores in each of the risk groups for each arm of the IMPaCT Back study were standardised to the population mean, low-risk (0.798), median (0.667) and high (0.389), shown in Table 6.15 as model baseline utility. The treatment effects calculated (Table 6.14), were now added to each of the standardised model baseline utility scores to create standardised four and twelve-month utility scores with treatment effect.

**Table 6.15 Standardised baseline, four and twelve month utility values**

| Risk Group | Model baseline utility | Standardised 4-month utility with TE | α,β or n | Standardised 12-month utility with TE | α,β or n | Distribution |
|---|---|---|---|---|---|---|
| Low_SC | 0.798 | 0.855 | 1655,281 | 0.849 | 1700,302 | Beta |
| Medium_SC | 0.677 | 0.616 | 204,127 | 0.550 | 190,156 | Beta |
| High_SC | 0.389 | 0.172 | 9,41 | 0.220 | 13,47 | Beta |
| Low_UC | 0.798 | 0.863 | 1593,253 | 0.880 | 1919,436 | Beta |
| Medium_UC | 0.677 | 0.636 | 206,118 | 0.531 | 185,163 | Beta |
| High_UC | 0.389 | 0.320 | 24,52 | 0.310 | 31,51 | Beta |
| Abbreviations: SC Stratified Care; TE Treatment Effect; UC Usual Care | | | | | | |

As these scores were artificially created, standard errors were taken from EQ-5D scores at equivalent time points in the STarT Back trial. To generate the utility values at 2, 6, 8, and 10 months, values were assumed to take a linear function between observed periods; these appear in the columns entitled "generated x months" in table 6.16. Standard errors originate from the STarT Back trial baseline and four-month data.

**Table 6.16 Estimating EQ-5D scores at unobserved time points**

| Risk Group | **Model baseline utility** | Generated 2-month utility | **Standardised 4-month utility with TE** | Generated 6-months | Gen 8 months | Gen 10 months | **Standardised 12-month utility with TE** |
|---|---|---|---|---|---|---|---|
| Low_SC | **0.798 (0.011)** | 0.827 | **0.855 (0.008)** | 0.854 | 0.852 | 0.851 | **0.849 (0.008)** |
| Medium_SC | **0.677 (0.014)** | 0.647 | **0.616 (0.027)** | 0.600 | 0.583 | 0.567 | **0.550 (0.027)** |
| High_SC | **0.389 (0.022)** | 0.281 | **0.172 (0.053)** | 0.184 | 0.196 | 0.208 | **0.220 (0.053)** |
| Low_UC | **0.798 (0.011)** | 0.831 | **0.863 (0.008)** | 0.867 | 0.872 | 0.876 | **0.880 (0.008)** |
| Medium_UC | **0.677 (0.014)** | 0.657 | **0.636 (0.027)** | 0.610 | 0.584 | 0.557 | **0.531 (0.027)** |
| High_UC | **0.389 (0.022)** | 0355 | **0.320 (0.053)** | 0.318 | 0.315 | 0.313 | **0.310 (0.053)** |
| Abbreviations; SC( Stratified Care); TE (Treatment Effect); UC(Usual care) | | | | | | | |

EQ-5D scores for the function states (beyond 12-months) originated from patients in the BeBack study sample who were categorised by function based upon the RMDQ thresholds. BeBack data were used, as there were no EQ-5D scores available in the BaRNS cohort study. However, as there were no EQ-5D scores in the BeBack 5-year follow up study, twelve-month follow-up values were used under the assumption that EQ-5D scores for each of the function states at twelve months were stable over the next nine years. EQ-5D scores used from fourteen months onwards in each of the function states are shown in Table 6.17.

**Table 6.17 EQ-5D scores for function states, years 1-10 of model**

| Risk Group | Mean | Standard Error | n | Distributions | α,β |
|---|---|---|---|---|---|
| Poor Function | 0.371 | 0.033 | 101 | Beta | 79.13, 134.16 |
| Moderate Function | 0.686 | 0.023 | 88 | Beta | 278.65, 127.54 |
| Good Function | 0.886 | 0.009 | 231 | Beta | 1103.92, 142.04 |

The validity of assuming stable EQ-5D scores over time is evidenced by the statistically indistinguishable RMDQs score at twelve-months and five-years in the BeBack study data, shown in appendix 7.

### 6.2.13 Methods of Analysis

6.2.13.1 Base case analysis

The base case is a cost-utility analysis of stratified care versus usual care for LBP, performed from the NHS perspective. The results are presented in the form of a cost-per-additional QALY gained. In order to obtain the base case estimates Monte Carlo simulation was utilised to perform 10,000 iterations of the model, the purpose of this was to capture the extent to which uncertainty over the parameter values impacts upon the cost-effectiveness. Accordingly, base case QALYs and costs are not point estimates, but means of 10,000 replications. The distributions, standard errors, and standard deviations around

the point estimates required to perform this analysis are detailed in Tables 6.4, 6.9, and 6.16.

Monte Carlo simulation also facilitates probabilistic sensitivity analysis (PSA). As noted in chapter 3, the PSA aims to quantify the degree of confidence in the analytical output with reference to the uncertainty associated with parameter values. The results from the PSA are represented by cost-effectiveness planes, showing all of the 10,000 model simulations, graphically plotting the difference in costs against the difference in QALYs. The range of values presented is a visual representation of the uncertainty over differences in costs and QALYs between the two treatments. The plane has four quadrants: in the north-east (NE) quadrant, a new intervention will offer more health gains but be more expensive; in the NW quadrant a new intervention offers worse health outcomes and is more expensive; in the SW quadrant an intervention will be cheaper and less effective, and in the SE quadrant the intervention will be more effective and cheaper. In the SE quadrant, therefore, the intervention is said to be dominant over the comparator; whereas in the NW usual care is dominant.

The uncertainty associated with the adoption decision is commonly represented using a cost-effectiveness acceptability curve (CEAC) (Briggs et al., 2006), which is a graphical plot of a range of cost-effectiveness thresholds against the probability that the intervention is cost-effective at said thresholds. The CEAC allows the decision-maker to assess the uncertainty associated with making a particular adoption decision. In the results, Willingness to pay (WTP) thresholds of between £20,000 and £30,000 per QALY gained are commonly displayed, as these are the values considered acceptable by NICE guidelines (NICE, 2008).

#### 6.2.13.2 Secondary analyses

Methodological uncertainty can be defined as the uncertainty in the results of an economic evaluation which arise from the choice of analytic methods (Briggs and Gray, 1999). Whilst all results are presented discounted, methodological uncertainty (6.3.2) is considered by representing subsidiary analysis showing base case results undiscounted. Next owing to the high private healthcare costs and volume of time taken off work associated with LBP (detailed in Chapter 2), analysis was performed from the societal perspective.

Sub-groups analysis (6.3.3) was also performed on the discounted base case, with the model run separately for the three STarT Back risk groups, as the cost-effectiveness implications are likely to be different in each risk group.

#### 6.2.13.3 Deterministic sensitivity analyses

The various assumptions, simplifications and scientific judgments made when constructing a model, can be referred to as structural uncertainty (Bojke et al., 2009). Given the uncertainty associated with the long-term parameters in the model, a number of sensitivity analyses were performed to assess the robustness of the results to both these structural uncertainties as well as parameter uncertainties (6.3.4), namely changes in modelling assumptions or temporal assumptions over input parameters used in the base case. All scenarios considered were pre-specified in conjunction with the advice and views of the group of experts.

There were a number of analyses assessing structural and parameter uncertainties.

Subsidiary analysis explored the cost-effectiveness implications of the different expert views regarding the impact of stratified care upon long-term treatment costs (6.3.4.1). Analysis explored a raising or lowering of the long-term cost of stratified care by 2%, 5%, and 10%.

As baseline model utility values originated from the IMPaCT Back study population, analysis also explored the impact of using the STarT Back utility values throughout the first year (6.3.4.2).

Experts agreed that the most significant source of uncertainty in the model was the long-term treatment effect for patients who received stratified care. Accordingly, two analyses consider the implications of pessimistic (6.3.4.3) and optimistic (6.3.4.4) long-term treatment effects by varying the numbers of patients in each function states over time on stratified care. A related issue was that the IMPaCT Back treatment effect may better represent the true treatment effect possible in this population. An analysis for the 6-month treatment effect achieved in the IMPaCT Back study is therefore presented (6.3.4.5). A worst-case scenario is also explored (6.3.4.6) whereby, costs are 10% higher long-term on stratified care, whilst patients achieve no overall improvement in function beyond two years.

There were some differences in the private cost estimates derived from BeBack and STarT Back datasets, and consequently, an alternate deterministic costing scenario (6.3.4.7) was evaluated for first-year costs. This analysis applied mean healthcare cost-per-patient observed in the STarT Back trial (e.g. £318.26 on usual care) but split those costs across the public and private sector using the distribution of NHS vs private costs observed in the BeBack study (66% vs 34%).

Finally, since pain is acknowledged to be an important outcome for LBP patients, it was also suggested that a sensitivity analysis (6.3.4.8) assesses the impact of modelling pain states instead of function states. In this analysis, transitions were taken from the BeBack data, and patients were categorised according to their pain intensity, as measured in the STarT Back study, e.g. average across three questions (average LBP pain in last 2 weeks, least LBP pain in past 2 weeks, and most painful LBP pain in last 2 weeks). Patients were then categorised as in the original BeBack study (Dunn et al., 2006), where >=5 was

classified as severe pain; >=1 and <5 as moderate pain, and <1 as no pain. Transitions, EQ-5D values and costs were calculated from the BeBack data for each of the pain states. In order to transition patients from twelve months into pain states, patients were also allocated a pain state at the end of the STarT Back trial (by 12 months).

#### 6.2.13.4 Value of Information Analysis

As noted in chapter 3, the purpose of value of information analysis (VOI) is to quantify the value of conducting further research. The "value" as such, derives from the fact that decision-makers want to implement the most cost-effective treatment, and where there is uncertainty regarding which is the best option, an incorrect adoption decision could follow. In this case, two VOI analyses were performed, an expected value of perfect information (EVPI) analysis, and an expected value of perfect parameter information (EVPPI) analysis on the base case model. The EVPI uses the quantification of uncertainty from the PSA output, and calculates the net value of eliminating that uncertainty, such that the best treatment option could be selected in each iteration. This analysis will provide a quantification of the value of further research to the NHS. The EVPPI can be estimated by assessing the impact of reducing the standard error of a particular parameter to zero on the reduction in standard error of overall INB. In other words, the EVPPI is the (expected) reduction in expected loss from the reduction in overall decision uncertainty attributable to eliminating uncertainty in a particular parameter.

Using the PSA output an expected value of perfect information (EVPI) per person is calculated. To obtain the overall value of removing decision uncertainty, the individual estimate is then multiplied by the population expected to benefit from the intervention. In order to consider the total value of removing decision uncertainty, it is essential to account for, not only the population impacted by this decision annually, but also the duration that the comparison holds relevancy. This comparison is assumed to hold relevance for the next ten

years, and this time frame is adopted in this analysis. Accordingly, population EVPI can be denoted by the following equation:

$$\text{EVPI}^* \sum_{t=1}^{T} \frac{I_t}{(1+r)^t}$$

Where EVPI is the individual EVPI estimate, $I_t$ is incidence in the *t*th year, *T* is total number of years the research would be relevant, and *r* is the discount rate.

In this study, per person EVPI was estimated using the Sheffield Accelerated Value of Information (SAVI) software (Strong et al., 2014). Information about the chosen willingness to pay threshold, and units of cost and effect measures, were entered into the software. Iterations of parameters, costs and effects from the PSA were saved as .csv files and inputted into the software.

In order to calculate the population expected to benefit, to ensure there is no double counting, ten-year consultation prevalence is required. Such a calculation was not found in the literature for LBP, so based upon Jordan et al. (2012) who estimated a cumulative seven-year consultation prevalence of 21.13% it is assumed that the ten-year prevalence is 25%, and as a consequence the prevalence of "new" annual consulters is 2.5%. This assumption was checked with Kelvin Jordan author of the original paper.

As the SAVI software does not incorporate a discount rate, the estimated total 10-year prevalence was discounted at 3.5%. In order to do this, the 2.5% new consultation rate, was multiplied by the UK 2016 population over 18 years old, 51,863,500 (ONS, 2018) to derive an annual population likely to receive the intervention, of 1,296,588. This was then discounted over 10-years using the formula, $\sum_{t=1}^{T} \frac{I_t}{(1+r)^t}$, which gives a discounted 10-year consultation prevalence of 11,160, 526.

In order to calculate the single and group parameter EVPPI, the Sheffield Accelerated Value of Information (SAVI) software (Strong et al., 2014) was used. Given that there are 75 parameters in the model, it was expected that the initial contribution of each parameter to the overall uncertainty would be minor, and therefore EVPPI was computed for groups of

associated parameters. Subsets used in analysis were first year transition probabilities, utility values, and costs.

### 6.2.14 Validity Checks

Internal model validity was assessed by running the model in Excel, and comparing proportions of the sample in each of the states and one year, and ten years.

External validity will be assessed by comparing the proportions in each state, as well as costs and QALYs at one-year with that of the STarT Back trial. In terms of the longer-term validity the model will be compared with the proportions in each state in the BaRNS cohort study.

## 6.3 Model Results

The presentation of the results appears in four components. Firstly, base case analysis is presented, with accompanying CEAC and cost-effectiveness planes. Secondly, subsidiary analyses on the base case are performed, namely undiscounted results, a societal analysis, and sub-group analysis for each of the risk groups. Thirdly, deterministic sensitivity analysis addresses the structural uncertainty over the long-term costs and treatment effect. Finally, a value of information analysis considers the potential value of further information.

All results presented are obtained from 10,000 replications using Monte Carlo simulation, and all results are discounted aside from those in the undiscounted analysis. Results are presented with willingness to pay thresholds of £20,000 per QALY gained.

### 6.3.1 Base case analysis

The results of the base case analysis are presented in Table 6.18, showing the QALYs and NHS costs associated with stratified care and usual care for the model population.

**Table 6.18 Base case analysis stratified care vs usual care**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified Care | 1596.83 | 6.35 |
| Usual Care | 1732.02 | 6.21 |
| Difference (Stratified care-usual care) | -135.19 | +0.14 |

The ten-year LBP related healthcare costs of stratified care were estimated to be £1596.83 per patient, with mean QALYs experienced by a patient during the time of 6.35. Usual care treatment was more expensive at £1732.02 per patient, and produced 6.21 QALYs. Therefore stratified care yielded 0.14 more QALYs and saved £135.19 per patient. Stratified care dominates usual care.

The Monte Carlo simulations demonstrate the variability in these results, shown in the cost-effectiveness plane in figure 6.2, where each point on the scatterplot represents one iteration of the simulated model, with associated costs on the y-axis and effects on the x-axis.

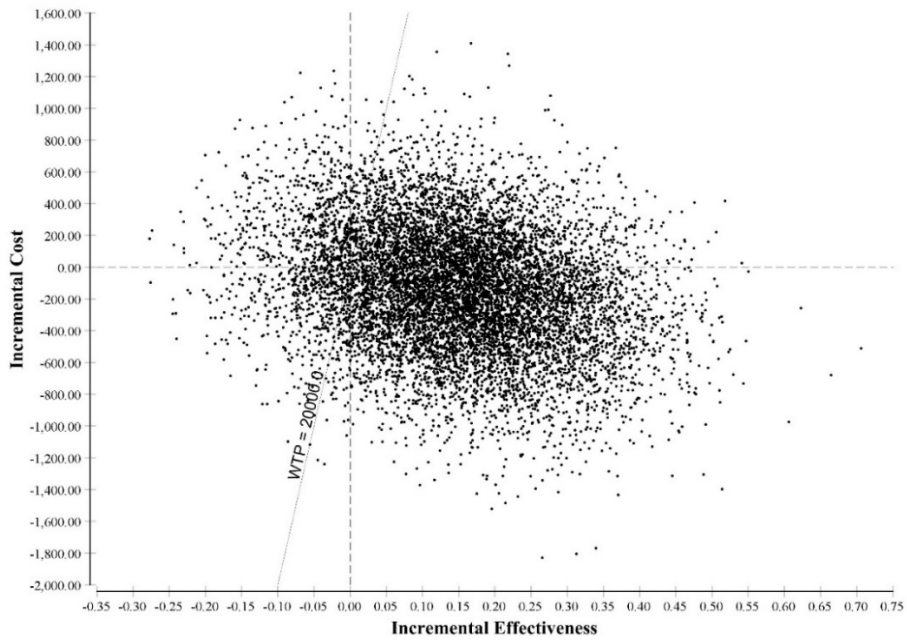



**Figure 6-2 Cost effectiveness plane, stratified care vs. Usual Care, base case**

In the majority of the 10,000 simulations, stratified care was more effective, indicated by most of the scatter points lying to the right of the y-axis, in SE and NE quadrants. In most iterations, stratified care was also less costly with the majority of points below the x-axis in the SE and SW quadrants. However, a few of the replications are also in the NW quadrant where usual care is cheaper and more effective, indicating that there is still some minor uncertainty over the cost-effectiveness of stratified care.

When these results are plotted onto a cost-effectiveness acceptability curve (see figure 6.3) in order to understand how this uncertainty impacts the likelihood of stratified care being cost-effective at different WTP thresholds, it can be seen that it is very likely that stratified care is cost-effective relative to usual care given the variability in the results.

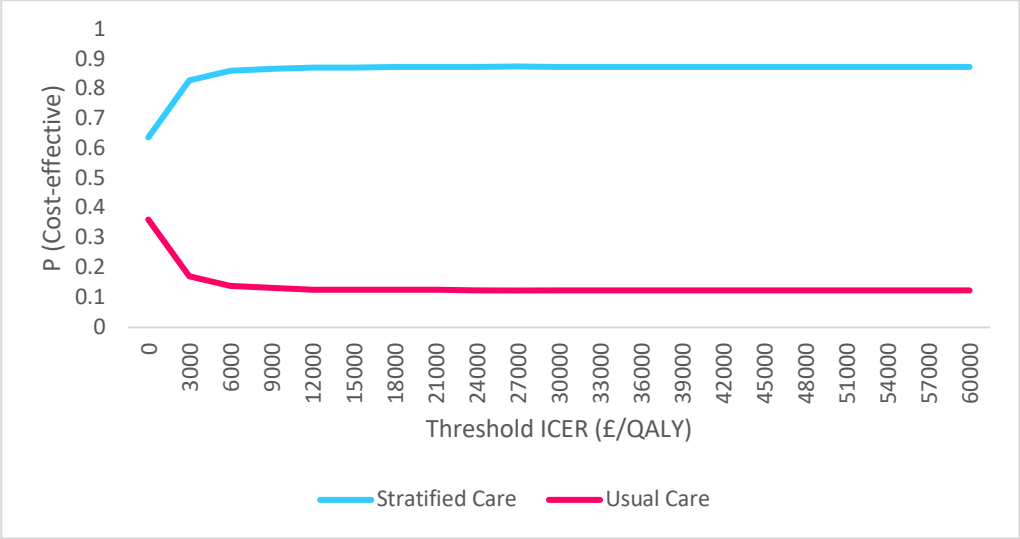



**Figure 6-3 Cost-Effectiveness Acceptability Curve, stratified care vs usual Care, base case**

At the critical threshold of £20,000 per QALY, stratified care is 87.4% likely to be cost-effective, rising slightly to 87.5% at £30,000 per QALY.

To understand this result it is important to observe how the model estimates patient prognosis (in terms of function state) on stratified care and usual care. Table 6.19 shows model estimates of prognosis over time in each function group. At ten years after treatment

there are a greater proportion of patients in good function on stratified care (63.1%) vs usual care (61.4%) with associated lower costs and higher QALYs.

**Table 6.19 Patient function, stratified care vs usual care, over 10 years**

| | End of One-year | End of Five years | End of Ten years |
|---|---|---|---|
| Stratified care* | | | |
| Good Function | 58.2% | 62.8% | 63.1% |
| Moderate Function | 22.3% | 16.5% | 14.2% |
| Poor Function | 19.1% | 18.4% | 16.9% |
| Usual Care* | | | |
| Good Function | 53% | 59.7% | 61.4% |
| Moderate Function | 23.4% | 16.8% | 14.2% |
| Poor Function | 23.2% | 21.3% | 18.7% |

*Death not shown, same in both arms

### 6.3.2 Sensitivity analyses, methodological uncertainty

Accordingly, this component of the results section assesses the impact of methodological uncertainty, e.g. the uncertainty caused by analytical choices, by assessing how the inclusion of societal costs in the analysis, as well as not discounting costs and outcomes, impact the cost-effectiveness of the approach.

#### 6.3.2.1 Undiscounted analysis

The base case model was rerun with no discounting performed; results are shown in Table 6.20.

**Table 6.20 Stratified care vs usual care, no discounting**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified care | 1876.13 | 7.50 |
| Usual Care | 2029.60 | 7.33 |
| Difference (Stratified care-usual care) | -153.47 | +0.17 |

When discounting is not performed, the incremental cost savings of the intervention are £18.28 higher than in the base case (-£153.47 vs -£135.19), and produce higher incremental QALYs (+0.17 vs +0.14). Accordingly, not discounting does not significantly impact the probability the intervention is cost-effective; at £20,000 WTP threshold it is 87.5% likely to be cost-effective, and dominant over usual care.

#### 6.3.2.2 Societal Analysis

The results of the societal analysis, showing QALYs and total societal costs associated with stratified care and usual care for the IMPaCT Back study population are shown in Table 6.21.

**Table 6.21 Societal analysis stratified care vs usual care**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified care | 8146.90 | 6.35 |
| Usual Care | 8836.89 | 6.21 |
| Difference (Stratified care-usual care) | -689.99 | +0.14 |

The ten-year societal cost of LBP-related expenses was £8146.90 per patient with the mean QALYs experienced by a patient during the time reaching 6.35. Treatment with usual care was more expensive at £8836.89 and produced 6.21 QALYs. Stratified care contributed 0.14 more QALYs and costed £689.99 less. Therefore, from the societal perspective usual care is also dominated by stratified care. Compared with the base case results, mean cost savings from stratified care are now much higher when societal costs are included (£689.99 vs £135.19).

### 6.3.3 Heterogeneity, sub-group analysis

The base case model was run separately for each of the risk subgroups. QALYs and NHS costs associated with stratified care and usual care for each risk subgroup are shown in Table 6.22.

**Table 6.22 Stratified care vs usual care, by risk group**

| Low-Risk | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified care | 1396.07 | 6.50 |
| Usual Care | 1465.35 | 6.46 |
| Difference (Stratified care-usual care) | -69.28 | +0.04 |
| Medium-Risk | | |
| Stratified care | 1659.12 | 6.30 |
| Usual Care | 1788.11 | 6.21 |
| Difference (Stratified care-usual care) | -128.99 | +0.09 |
| High-Risk | Mean Cost (£) | Mean QALYs |
| STarT Back | 1821.10 | 6.20 |
| Usual Care | 2088.80 | 5.78 |
| Difference (Stratified care-usual care) | -267.70 | +0.42 |

Stratified care provides a small QALY gain to patients in the low-risk group (+0.04 QALY), a good benefit (+0.09 QALY) in medium-risk patients and greater benefits for high-risk patients (+0.42 QALY). Stratified care is cheaper in all of the risk groups, but most cost-saving in the high-risk group (-£267.70). Stratified care dominates usual care in

all risk groups, and is most cost-effective for high-risk patients, saving £267.70 and providing an additional 0.42 QALYS.

The probabilistic simulations show considerable variability in results for low-risk patients, demonstrated in figure 6.4

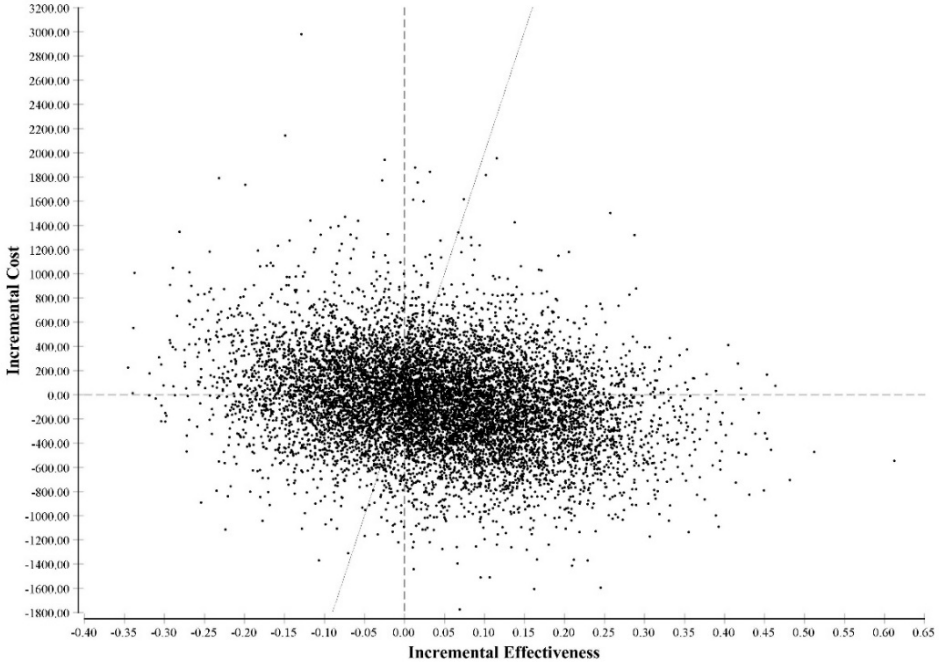



**Figure 6-4 Cost effectiveness plane, stratified care vs usual care, low-risk patients**

This uncertainty over whether or not stratified care is effective for low-risk patients is demonstrated in the Cost-Effectiveness Acceptability Curve in figure 6.5 which shows that for low-risk patients at £20,000 per QALY, stratified care is 63.9% likely to be cost-effective, and 63.7% likely to be cost-effective at £30,000 per QALY.

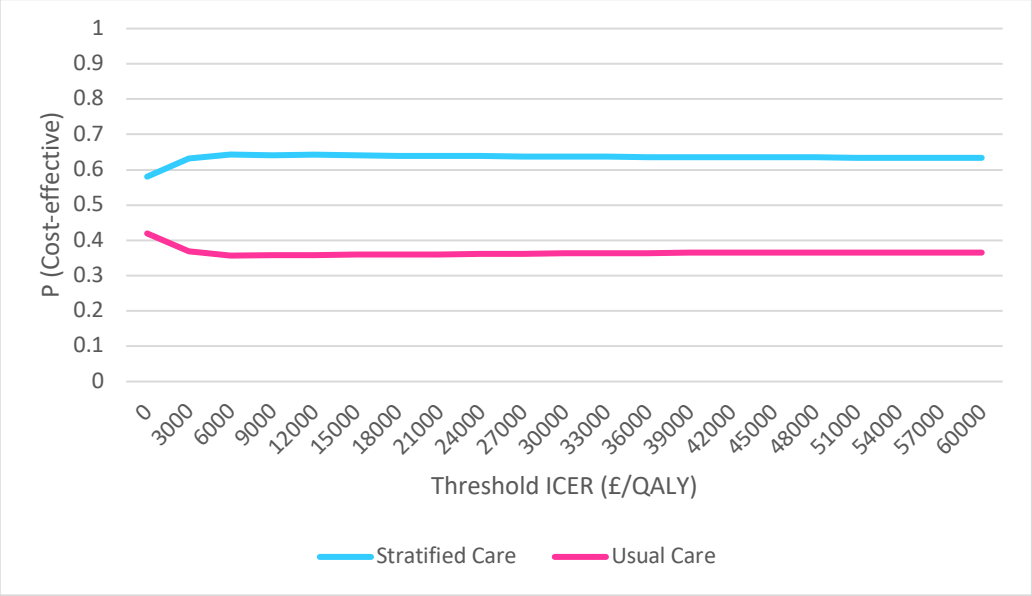



**Figure 6-5 Cost-Effectiveness Acceptability Curve, stratified care vs usual care, low-risk patients**

Stratified care was shown to be £128.99 cheaper (£1659.12) than usual care (£1788.11) for medium-risk patients, and provided 0.09 more QALYs (Table 6.22). As there are mean cost savings, and only a small treatment benefit, the cost-effectiveness plane in figure 6.6 shows more of the points of the scatterplot in the southern quadrants (NE and SE).

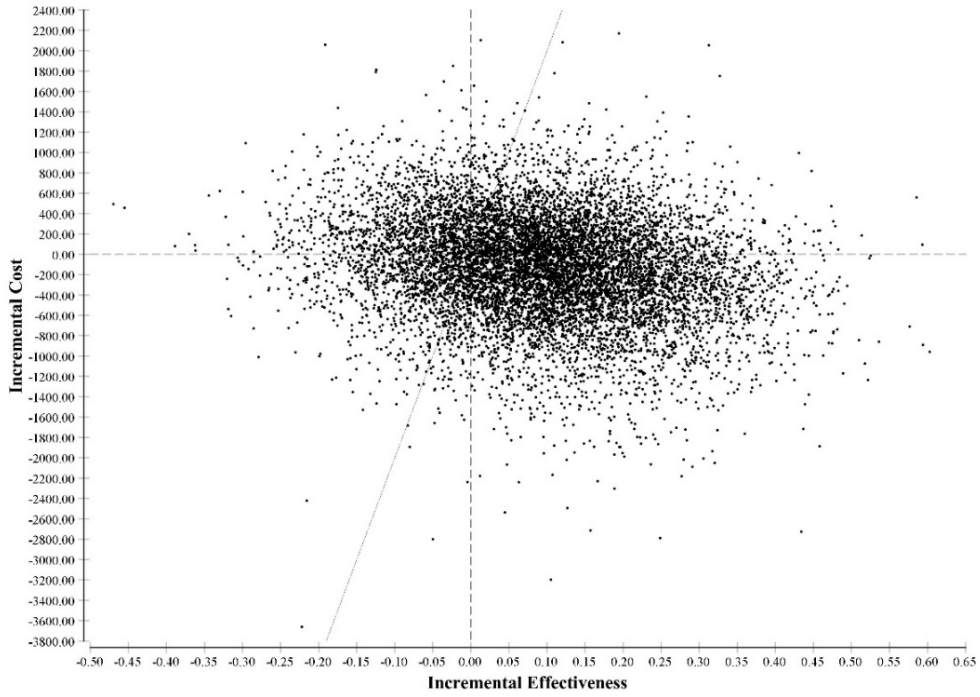



**Figure 6-6 Cost-effectiveness plane, stratified care vs usual care, medium-risk patients.**

The cost-effectiveness curve in figure 6.7 shows that for medium-risk patients, at the £20,000 and £30,000 WTP thresholds, there is a 75.8% chance that stratified care is cost-effective.

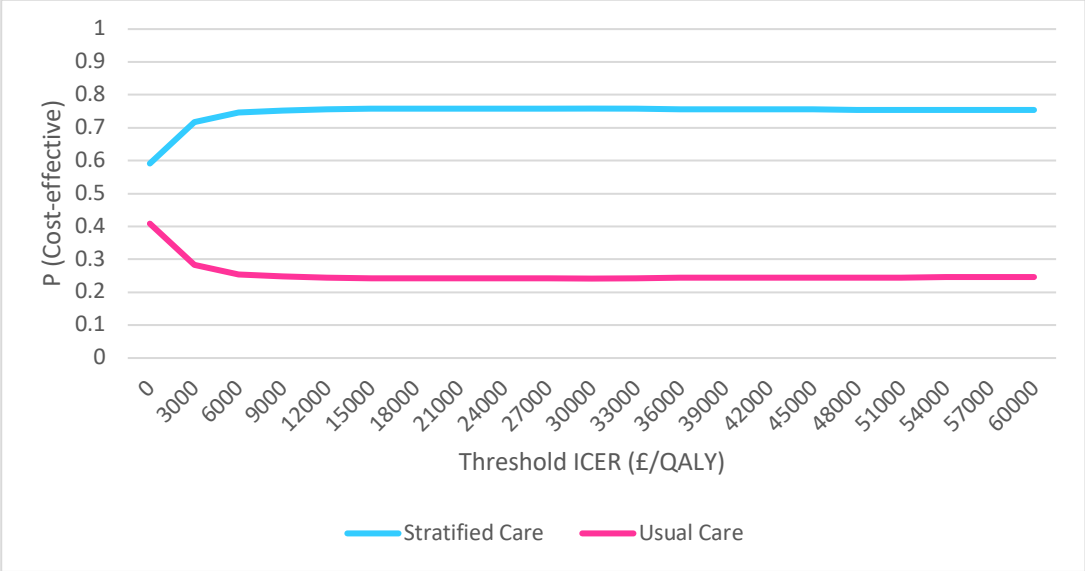



**Figure 6-7 Cost-Effectiveness Acceptability Curve, stratified care vs usual care , medium-risk patients**

For high-risk patients, stratified care was £267.70 cheaper than usual care, with a QALY gain of 0.42 over ten years (Table 6.22). As figure 6.8 shows, almost all model iterations

for high-risk patients lie in the eastern quadrants, indicative that stratified care provides much more benefit.

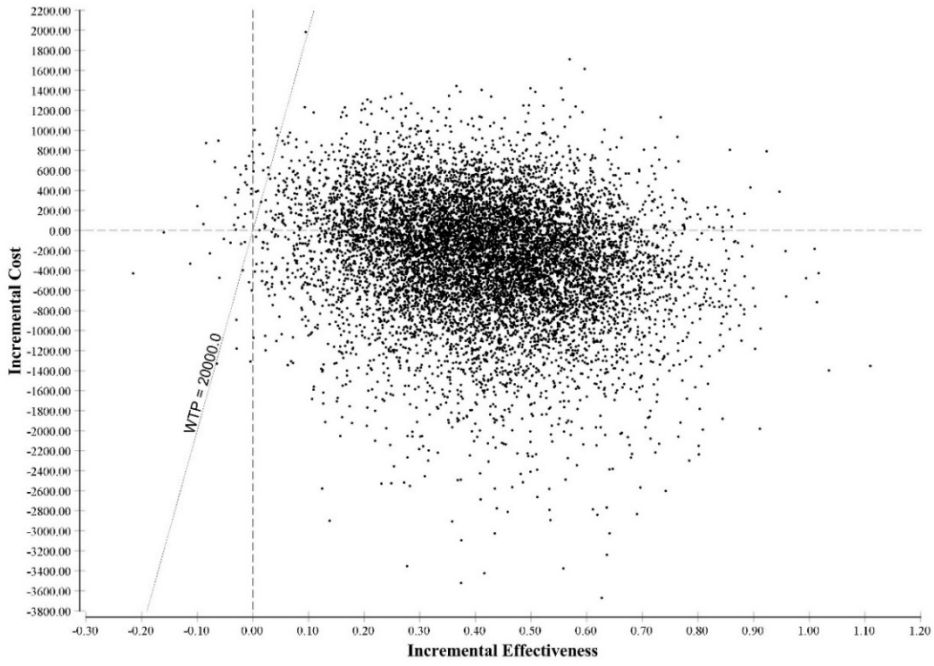



**Figure 6-8 Cost-effectiveness plane, stratified care vs usual care, high-risk patients.**

The CEAC in figure 6.9 shows that at the WTP threshold of £20,000, there is a 99.5% chance that stratified care is cost-effective. At £30, 000 per-QALY that rises to a 99.6% likelihood.

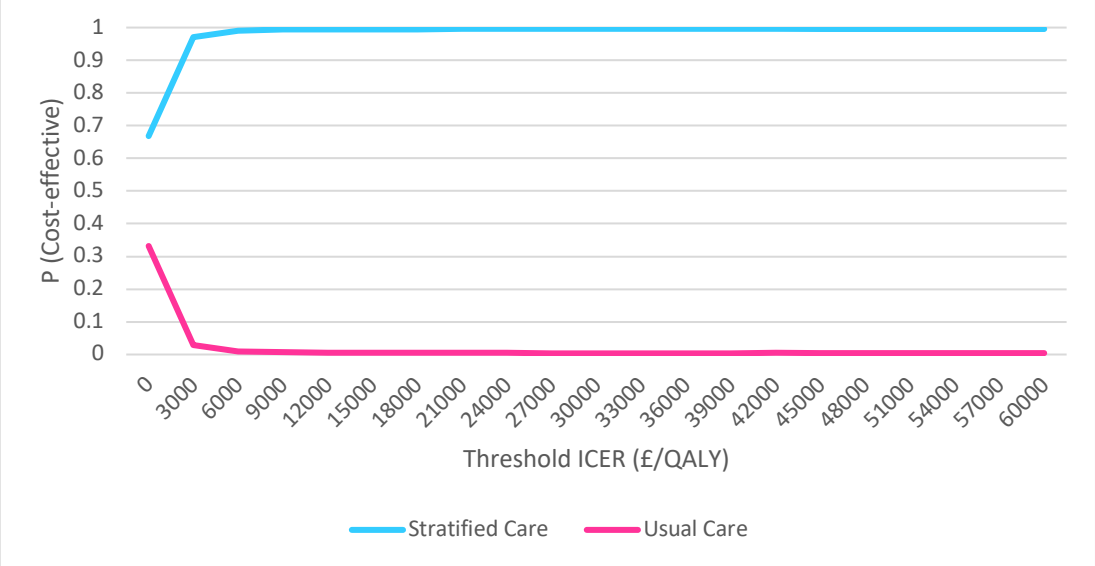



**Figure 6-9 Cost-Effectiveness Acceptability Curve, stratified care vs usual care, high-risk patients**

### 6.3.4 Sensitivity analyses, structural and parameter uncertainty

In what follows this section assesses the impact of differing assumptions regarding the long-term costs and effectiveness of stratified care, as well as using the implementation study data on treatment effect, EQ-5D values from the STarT Back trial, and running the model using pain states instead of function, beyond one-year.

#### 6.3.4.1 Temporal uncertainty over treatment cost

Table 6.23 shows the impact upon cost-effectiveness of different costing assumptions regarding the effect of long-term treatment. The table also presents the probability that stratified care is cost-effective at a £20,000 WTP threshold given the different costing assumptions.

**Table 6.23 Cost-effectiveness of stratified care versus usual care, in different cost scenarios**

| Cost-variation | Per patient cost, SC | Incremental Cost, SC vs Usual care | % cost-effective at £20,000 WTP |
|---|---|---|---|
| 10% lower costs on SC | 1459.74 | -271.98 | 89.2% |
| 5% lower costs on SC | 1524.19 | -211.00 | 88.5% |
| 2% lower costs on SC | 1570.83 | -159.82 | 87.9% |
| Base case | 1596.83 | -135.19 | 87.4% |
| 2% higher costs on SC | 1623.17 | -115.37 | 87.1% |
| 5% higher costs on SC | 1665.16 | -71.76 | 87.0% |
| 10% higher costs on SC | 1732.05 | -1.60 | 86.5% |

Abbreviations: ICER (Incremental cost-effectiveness ratio); SC (Stratified Care)

It can be seen that the costs of treatment change the incremental cost of the stratified care intervention, ranging from saving £271.98 to £1.60, but the likely cost-effectiveness of the approach is relatively unchanged. Stratified care is dominant in all scenarios.

#### 6.3.4.2 STarT Back trial utility values

Using the baseline utility values from the STarT Back trial, results in a fall in total QALYs of 0.1 from 6.35 to 6.25 for stratified care, and of 0.09 from 6.21 to 6.12 for usual care. Accordingly, incremental QALYs for the intervention fall slightly to 0.13 QALYs compared with 0.14 in the base case. Stratified care is once again dominant, and likely cost-effectiveness remains high at 86%. Results are shown in Table 6.24.

**Table 6.24 Stratified care vs usual care, STarT Back study utility values**

| | Baseline EQ-5D used | Mean QALYs |
|---|---|---|
| Stratified care | 0.73; 0.551; 0.299 | 6.25 |
| Usual Care | 0.73; 0.551; 0.299 | 6.12 |
| Difference (Stratified care-usual care) | | 0.13 |

#### 6.3.4.3 Temporal uncertainty over long-term treatment effect, pt.1

The impact upon cost-effectiveness of more pessimistic assumptions regarding patient function on stratified care, are shown in Table 6.25. In the table, the column "convergence at year" represents scenarios where, for example, in the case of the row titled "2 years" at the end of two-years the distribution of the patient cohort in each of the function states is identical on stratified care and usual care, and remains so for the remainder of the model. In this sense there is a convergence in patient function at each time point instead of the base case assumption that there are always more patients are in good function on stratified care throughout the model.

**Table 6.25 Impact of assumptions over long-term treatment effect upon cost-effectiveness of stratified care**

| Convergence at year | Incremental cost | Incremental QALY | % cost-effective at £20,000 WTP threshold |
|---|---|---|---|
| 2-years | -51.25 | +0.05 | 65.2% |
| 3 years | -74.45 | +0.08 | 70.2% |
| 4 years | -86.07 | +0.09 | 73.9% |
| 5 years | -94.98 | +0.09 | 76.6% |
| 6 years | -111.58 | +0.11 | 81.4% |
| 7 years | -119.82 | +0.12 | 83.9% |
| 8 years | -126.45 | +0.13 | 84.6% |
| 9 years | -129.90 | +0.13 | 85.2% |
| Base case – no convergence | -135.19 | +0.14 | 87.5% |

Whilst stratified care was dominant in all scenarios, the impact of varying assumptions over treatment effect changed the probability of cost-effectiveness. Where the same proportion of patients are in function groups at 2 years, the probability that stratified care is cost-effective falls to 65.2% from 87.5% in the base case.

6.3.4.4 Temporal uncertainty over long-term treatment effect, pt.2

More optimistic scenarios were modelled, where stratified care is assumed to have a continued positive impact, transitioning more patients into good function over time. Results for stratified care achieving 2.5% and 7.5% more patients in good function at ten years relative to usual care are shown in Table 6.26.

**Table 6.26 Stratified care vs usual care, additional treatment benefit on stratified care**

| Additional patients in good function at ten years | Incremental cost | Incremental QALY | % cost-effective at £20,000 WTP threshold |
|---|---|---|---|
| +2.5% | -157.17 | +0.17 | 91.4% |

| | | | |
|---|---|---|---|
| +7.5% | -173.72 | +0.21 | 93.8% |

It can be seen that the cost-effectiveness of stratified care is influenced by increasing numbers of patients in good function, reaching 93.8% chance to be cost-effective if 7.5% more patients were to achieve good function with this management, at ten years.

#### 6.3.4.5 IMPaCT Back study treatment effect

A sensitivity analysis was conducted using the treatment effect obtained in the IMPaCT Back implementation study. Results are shown in Table 6.27.

**Table 6.27 Costs and effects associated with the implementation study treatment effect**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified care | 1632.81 | 6.28 |
| Usual Care | 1783.15 | 6.16 |
| Difference (Stratified care-usual care) | -150.34 | +0.12 |

Despite a reduced improvement in function in the first year relative to usual care, stratified care remains dominant, saving £150.34 with a 0.12 QALY gain, and an 83.1% chance of being cost-effective at a £20,000/QALY threshold.

#### 6.3.4.6 Worst Case Scenario Analysis

A worst-case scenario was performed, where patient function was equalised on both comparators at the end of the second year, and future costs were 10% higher for patients in stratified care, shown in Table 6.28.

**Table 6.28 Stratified care vs usual care, Worst case scenario**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified care | 1814.87 | 6.26 |
| Usual Care | 1734.38 | 6.21 |

| | | |
|---|---|---|
| Difference (Stratified care-usual care) | +80.51 | +0.05/ ICER= £1610/QALY |

In the worst case scenario, stratified care resulted in 0.05 additional QALYs but costed £80.51 more, giving an ICER of £1610/QALY. Given the variability in the results, probabilistic output showed that stratified care was still 63.9% likely to be cost-effective at a £20,000 WTP threshold.

6.3.4.7 Alternate NHS vs private cost distribution, first year

On the basis that the STarT Back trial may overstate NHS costs relative to private costs, if total healthcare costs taken from the STarT Back trial are assigned to the NHS costs using the NHS vs private split (66% vs 34%) found in the BeBack cohort study, mean NHS costs are lower in both stratified and usual care arms than the base case, as shown in table 6.29. Accordingly, the cost savings associated with the approach are slightly less from the NHS perspective than in the base case (-£97.83 vs £135.19), but stratified care is still dominant.

**Table 6.29 Stratified care vs usual care, NHS costs using BeBack cost distribution**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified care | £1532.35 | 6.35 |
| Usual Care | £1630.18 | 6.21 |
| Difference (Stratified care-usual care) | -97.83 | +0.14 |

Probabilistic analysis in this scenario showed that the intervention is 87.1% likely to be cost-effective at the £20,000 WTP threshold, slightly below the 87.4% in the base case.

#### 6.3.4.8 Use of pain states instead of function states

Where patients transition into pain states the cost-effectiveness of the stratified care was considerably lower. Results are shown below in Table 6.30.

**Table 6.30 Stratified care vs usual care, pain level to determine health states**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified care | £2020.53 | 6.11 |
| Usual Care | £2058. 53 | 6.07 |
| Difference (Stratified care-usual care) | -£38.00 | +0.04 |

The use of pain health states to perform the extrapolation reduces the likely cost-effectiveness of the approach, to 65% at the £20,000 WTP threshold.

It is important to note that this result arises because the analysis uses equivalent EQ-5D values for pain states on stratified care and usual care beyond 12-months. However, as Table 6.31 shows, in the STarT Back trial, at 12-months mean EQ-5D was higher in each of the pain states on the stratified care arm.

**Table 6.31 EQ-5D scores on stratified care vs usual care, at twelve-months**

| | STarT Back | | Usual care | |
|---|---|---|---|---|
| Pain state | n | EQ-5D | n | EQ-5D |
| No pain | 95 | 0.8948 | 44 | 0.8726 |
| Moderate pain | 176 | 0.743 | 86 | 0.7036 |
| Severe pain | 99 | 0.3786 | 47 | 0.3682 |

As a consequence as patients accrue their QALYs at 14-months the model using pain states provides a lower QALY benefit for stratified care than the data would suggest.

### 6.3.5 Value of information analysis

A VOI analysis was performed for uncertainty relating to all parameterised components of the model. Results in Table 6.32 show an individual level EVPI of £262.90 per patient, which can be extrapolated to a population EVPI using the consultation population expected to benefit (11,160, 526), this is shown in the right-hand column.

**Table 6.32 Per-Person and population EVPI**

| Scenario | Annual Per Person EVPI at £20,000 WTP threshold | Ten-year population EVPI at the £20,000 WTP threshold |
|---|---|---|
| Base case | £262.90 | £2,934,102,285 |
| Abbreviations; EVPI (Expected value of perfect information); WTP (Willingness to pay) | | |

When conducting VOI analysis, estimating which parameters are the major contributors to the decision uncertainty can be of particular interest, as can the potential value of reducing that uncertainty. As noted in 6.2.13.4, this forms the basis of expected value of partial perfect information (EVPPI) analysis, estimates for single parameters are shown in Table 6.33.

**Table 6.33 Single parameter EVPPI, Per Person and population**

| Parameters | Annual Per Person EVPPI at £20,000 WTP threshold | Approximate Standard Error | Ten-year population EVPPI at the £20,000 WTP threshold |
|---|---|---|---|
| TP: SC, low to low risk, months 4-12 | 1.22 | 1.06 | 13,570,000 |
| TP: SC: low risk to function, 12 months | 0.04 | 0.12 | 482,500 |

| | | | |
|---|---|---|---|
| TP: UC: medium to high risk, months 4-12 | 10.17 | 2.64 | 113,500,000 |
| TP: UC: medium to medium risk, months 4-12 | 0.02 | 0.17 | 277,300 |
| *Parameters with zero EVPPI are not shown<br>Abbreviations; EVPPI (Expected value of partial perfect information); SC (Stratified Care); TP (Transition probabilities); UC (Usual care); WTP (Willingness to pay) | | | |

It can be seen that the parameter with the most influence on the uncertainty in the model is the transition of medium risk patients on usual care from 4 to 12 months, with a ten-year EVPPI of £113,500,000. The transition of low risk patients on the stratified care from 4 to 12 months has some role in contributing to the uncertainty, with a ten-year EVPPI of £13,570,000. However, most other parameters in the table play a small role, and the majority of the parameters in the model are not given in the table, because the single parameter EVPPI estimates are returned at zero.

Often EVPPI estimates are more informative when computed for groupings of related parameters, for example all transition probabilities. Group EVPPI estimates therefore are the maximum expected value of further research which informs this set of parameters (Strong et al. 2014). Table 6.34 shows EVPPI estimates related to transition probabilities, utility values, and costs.

**Table 6.34 EVPPI parameter groups, per person and population**

| Parameter(s) | Annual Per Person EVPPI at £20,000 WTP threshold | Approximate Standard Error | Population EVPPI at the £20,000 WTP threshold |
|---|---|---|---|
| Transition probabilities first year | 261.07 | 7.14 | £2,913,680,442 |
| Transition probabilities year 2-10 | 0.40 | 0.96 | £449,249 |
| All Utilities | 7.72 | 10.033 | £86,121,677 |
| All Costs | 0.00 | 0.69 | 0.00 |

| Abbreviations; EVPPI (Expected value of partial perfect information); WTP (Willingness to pay) |
|---|

Given the constitution of this model structure, first year transition probabilities create the largest source of uncertainty in this decision model; there is high benefit to resolving that uncertainty, of £2,913,680,442. However, interpretation of this result requires logical and statistical thought, and is further elucidated in the discussion in 6.4.3.

### 6.3.6 Validity

In order to check the internal and external validity of the model, checks between model output and observed data were undertaken, and reported as follows.

In Table 6.35, it can be seen that the proportion of patients in each of the risk groups at twelve months is almost identical in the model as in the observed data from the STarT Back trial. To assess the internal validity of the model, the observed data was weighted to the standardised model population.

**Table 6.35 Proportion of patients in each risk group at 12 months, modelled estimates versus observed data**

| | Model Output | Observed Data* |
|---|---|---|
| Usual Care | | |
| Low Risk | 74.5% | 74.5% |
| Medium Risk | 18.5% | 18.9% |
| High Risk | 6.90% | 6.60% |
| Stratified Care | | |
| Low Risk | 78.8% | 78.8% |
| Medium Risk | 16.3% | 16.7% |

| | | |
|---|---|---|
| High Risk | 4.8% | 4.40% |
| *Observed data weighted for model population. | | |

The same is also true of function at seven years, where patient function on usual care is very similar to that within the BaRNS usual care cohort study, shown in Table 6.36.

**Table 6.36 Proportion of patients in each function state at 7 years, modelled estimates versus observed data**

| | Function at 7-years BaRNS data | Function at 7-years model output UC | Function at 7-years model output SC |
|---|---|---|---|
| Poor function | 21.2% | 20.8%% | 18.5% |
| Moderate function | 15.9% | 16.2% | 15.9% |
| Good function | 62.9% | 63.0% | 65.6% |
| Abbreviations; SC (Stratified care); UC (Usual care) | | | |

## 6.4 Discussion

The discussion includes three principal components, the statement of principal findings, followed by the strengths and weaknesses of the study, and finally the implications of the results.

### 6.4.1 Principal findings

The body of work contained within this chapter reflects the design and realisation of a state transition model to perform a cost-effectiveness analysis of the potential application of a stratified care model (STarT Back approach) for the management of LBP. The Monte Carlo simulations performed for the base case analysis, from the NHS perspective, including the assumption of no additive treatment effect from stratified care beyond one-year, showed that the intervention is very likely to be cost-effective and cost-saving, on average.

There are however, important caveats. Firstly, it is important to remember it is impossible in this analysis, to isolate the impact of the stratification of care, with the matched treatments available. It is plausible it is the psychologically informed physiotherapy that drives the QALY gain in high risk patients for example. Moreover, as has been mentioned

throughout this thesis, the long-term treatment effect for LBP patients managed according to this stratified care model is unknown. Conceptually, this is important. Whilst the impact upon cost-effectiveness when altering the long-term costs of the stratified care even in the two extreme cost scenarios was minimal, the impact of different assumptions over how patient function improves on stratified care over the duration of the model, had far more impact upon model results.

Indeed, during consultation with experts, there was considerable disagreement regarding the impact of treatment on patient function. It is worth noting that both the modelled optimistic and pessimistic scenarios were based on the suggestions of experts. Many experts believed that there would be some degree of convergence of patients in terms of function over the longer-term. Model results show that if patient function is identical at two-years the likely cost-effectiveness falls to only 65%. Whilst advocates of stratified care argued that there would be additional gain in function above and beyond modelled in the base case, if 7.5% more were in good function at ten years on stratified care the cost-effectiveness rises as high as 94%. The significant variation in cost-effectiveness according to long-term patient exemplifies why this long-term uncertainty is a caveat over the result, although the treatment approach is still likely cost-effective even in the pessimistic scenario.  The relevance of this sensitivity to patient prognosis is explored in more detail in 8.2.1

Considering the cost differentials between the function states (shown in Table 6.9), and the proportions of patients in each function state over time for stratified care and usual care (Table 6.19), it is evident why the long-term treatment effect is such a significant influence upon model results. A patient in good function has costs of only £85.10, whereas a patient in poor function has annual costs of £503.53. A treatment that can keep patients in better function over time will prove to be highly cost-effective.

The use of pain as a state instead of function could also provide caveats over the results. Where patients transitioned into pain states at twelve-months, this stratified care model was much less likely to be cost-effective. This analysis does, however, highlight the importance of the suggestion contained within the systematic review (Hall et al., 2019) that choice of state is likely to have a considerable effect upon cost-effectiveness results. Moreover, assuming that modelled transitions between states will fully capture the fundamental impact of treatment upon HRQoL could be a flawed assumption. For example, building a model using pain states (with equal extrapolated EQ-5D scores per state), as in 6.3.4.8, is not sensitive enough to capture the improvements in EQ-5D on stratified care at 12-months. This issue was also raised in the systematic review in Chapter 4, where it was suggested that using equal EQ-5D scores for "success" and "failure" for all treatments will not adequately represent the quality of life gain. To rectify or supplement the pain state model in 6.3.4.8, either one could assume continued EQ-5D gain over some period, or different assumptions explored in scenario analyses. This was not necessary with the use of function states, because the differences between the EQ-5D scores for the function states were minimal between stratified care and usual care at 12 months.

Leaving aside these methodological debates, there are other reasons to think that this model of stratified care is cost-effective in the long term. The first is that the stratified care trial (STarT Back) showed lower mean costs associated with low-risk and medium-risk patients on stratified care, with these two groups accounting for most patients. However this long-term reduction on costs on stratified care is not modelled in the base case analysis, it is likely, therefore, some further reduction in the longer term costs associated with the implementation of stratified care are plausible. A 10% fall in treatment costs did raise incremental cost-savings and likely cost-effectiveness slightly. Second, as was the case with other modelled cost-effectiveness analysis on LBP (Kim et al., 2010), the inclusion of societal costs in the analysis, if the intervention has a favourable impact upon

function, will improve the desirability of the intervention. The results here are very promising; the societal analysis shows that stratified care produces a cost saving of £689.11 per person over model duration.

Risk sub-group analysis also reveals interesting results. Where high-risk patients only were entered into the model, stratified care was 95.1% likely to be cost-effective, saving a significant £267.70 per patient, and affording a 0.42 incremental QALY gain over model duration. The reason this is so cost-effective in this group is that stratified care seems very effective at moving more patients out of the high-risk group at zero to four months (see Table 6.5). The transition from high-risk to low-risk is 0.386 on stratified care but zero on usual care. In addition, there is a higher rate of transition from high-risk to poor function on usual care at 12 months (0.90 versus 0.86 on stratified care). However, clinicians noted, it was plausible that patients in poor function on stratified care will experience some deterioration of function over and above that which was captured amongst high-risk patients in the BeBack study. Of course this is likely to be the case, but these patients may, if they wish, revisit their GP and access the effective matched treatments. With any further effect of psychologically informed physiotherapy not included in the base case analysis beyond one-year, there could be some offsetting between deterioration and availability of future effective treatments.

### 6.4.2 Strengths and weaknesses of the study

There are three major strengths of this analysis. Firstly, it is the first decision model produced for stratified care for LBP; in that regard it is a novel development. Secondly, and related, it offers solutions to the methodological problems identified in chapter 4, the review of previous LBP models. Through extensive consultations with a number of experts from clinical practice, health economics and academic research in the field of LBP and stratified care, the model is an attempt to adequately represent the condition to meet the

needs of modelling guidance. The use of states of function to represent the patient pathway is a methodological advance and allows the modeller to sidestep somewhat the issue of modelling the frequent recurrence of LBP symptoms. Mean EQ-5D at each time point is in essence an average of those who have, and those who do not have recurrence of symptoms.

Third, as identified in the systematic review (Hall et al., 2019); there is a hesitance to produce modelled cost-effectiveness analyses over an extended time horizon. By meeting with clinicians and academic experts, this analysis was able to make assumptions about the long-term treatment effect, and then test the significance of these assumptions through a series of sensitivity analyses on a number of expert informed likely scenarios. The probabilistic output is also robust, using 10,000 replications, and offering probabilistic output for all scenario analyses.

The study has a number of potential weaknesses which need to be considered. Whilst the PSA samples transitions of patients between function states beyond one year, and is in that sense probabilistic, the assigning of equivalent transitions to patients in both treatments (stratified and usual care) does not capture the "true" uncertainty over long-term transitions. The "true" probabilistic uncertainty over the long-term treatment effect, therefore, could only be accounted for by attaching a probability distribution to the differential transitions of patients on stratified care versus usual care, over time. This information is currently unknown, and therefore the modeller could consider options such as (i) try and parameterise this distribution possibly through elicitation and/or Bayesian methods or (ii) attempt statistical extrapolation of the treatment effect (Mahon, 2013), or make a base case assumption with robust scenario analyses included (NICE, 2014). The latter was the approach taken here, and the study has attempted to investigate all possible scenarios and explain how this would impact the likely long-term cost-effectiveness.

It is worth proposing caution over the generalisability of the results. This model is predicated upon the treatment benefits obtained in the STarT Back trial, a particular patient population with healthcare practices specific to Staffordshire, United Kingdom. Moreover, as Whitehurst and colleagues (2015) also noted, policymakers must interpret the findings alongside the need to support both GPs and physiotherapists to use this stratified care model.

In relation to data; transitions beyond one-year were only available at seven years, and as a consequence, the transitions do not accurately reflect the likely fluctuations in function state for patients between twelve months and seven years. Although the assumption of linear transition may be unrealistic, it does impact both stratified and usual care, in the same manner. Moreover, EQ-5D scores were not available at follow up points in the cohort studies used in this analysis, although RMDQ scores in BaRNS were stable over time. Whilst it is acknowledged that RMDQ and EQ-5D are conceptually distinct, stability in RMDQ implies that assuming stability in EQ-5D scores is likely to be a reasonable assumption.

There are some points to consider regarding costings. Resource use was taken from the first year of the BeBack cohort study, and therefore may over represent likely long-term costs. The cohort study was undertaken in 2006 and perhaps does not represent current treatment pathways for usual care. The fact that all healthcare resource use came from self-report, can be considered a limitation, on the grounds of recall bias (Petrou et al., 2002). Nonetheless, self-report resource use questions provide an efficient means of collecting information where routine data sources are not available, and are used extensively in economic analyses (Whitehurst et al., 2015).

It is also noted that, despite very similar NHS costs, the private healthcare cost estimates obtained from the STarT Back dataset are lower (£35.83) than those observed in the

BeBack observational cohort study (£282.44) (Table 6.8). However, where different absolute costings and distributions of those costs were analysed (6.3.4.7), the likely cost-effectiveness differed very little from the base case.

### 6.4.3 Implications for researchers, clinicians and policymakers

The findings in these analyses strengthen the economic case for the cost-effectiveness of this model of stratified care for the management of patients consulting with non-specific LBP, which includes systematic identification of future risk of back pain-related disability with matched treatments according to risk level. The results are important because they indicate not only that the approach can produce tangible treatment benefits, but that it can also contribute towards the ongoing rationalisation of efficient healthcare.

When considering the effect of applying treatment estimates from other stratified care implementation studies investigating the same approach, although there is some reduction in the likely cost-effectiveness, stratified care remains cost-effective and dominant. This result does support the assertion in the economic evaluation of IMPaCT Back study (Whitehurst et al., 2015), that the degree of cost-effectiveness of the approach will correspond to the level of implementation. The analysis contained here within supports the assertion that improving the utilisation of stratified care could help improve clinical outcomes and cost-savings.

The results of the risk sub-groups analysis show the large treatment effect the stratified care management (provided in STarT Back trial) afforded to high-risk patients in the long-term. These results are different to the one-year STarT Back trial, and demonstrate the importance of engaging in longer-term extrapolation, where the treatments are likely to improve long-term patient outcomes, however, it should be remembered that these results are dependent on specific, underlying assumptions.

The VOI analysis showed that removing some of the uncertainty detailed in this study could be valuable. The considerable size of the estimate reflects the fact that the condition can incur considerable costs, whilst affecting a significantly large population. The EVPPI estimates suggest that the first-year transition probabilities are the largest source of uncertainty in the model. However, given that patients on stratified care and usual care move at identical rates between function states beyond 12 months, the first year transitions become in essence a proxy for the entire patient trajectory during the entire model. A related point, it is not surprising that the value of reducing uncertainty around cost and utility is minimal given that the uncertainty over cost and utility estimates at all stages of the model are not as significant as differences in cost and utility estimates between each stages. It is logical that parameters that determine movements into the stages would drive the parameter uncertainty. As a consequence, the logical interpretation of the EVPPI result is to say that further investigation of the transitions of patients on stratified care and usual care across a time period of ten years could reduce the uncertainty in the model.

## 6.5 Conclusion

In this chapter, the design of and analysis of a model-based cost-effectiveness analysis of stratified care versus usual care over a long 10-year-time horizon, were presented, from the healthcare perspective in patients consulting in primary care with non-specific LBP. The analyses conclude that this stratified care model was cost-saving and provided incremental QALY gains in nearly all scenarios modelled. Base case results showed it is likely to be cost-effective at £20,000 WTP threshold. Sensitivity analyses and value of information analysis stress the importance of investigation of the long-term treatment effect upon the likely cost-effectiveness of the approach.

# Chapter 7: A COST-EFFECTIVENESS ANALYSIS OF STRATIFIED CARE VERSUS BEST CARE AND USUAL CARE FOR SCIATICA

Adopting the same structure as Chapter 6, this chapter begins by describing the methodological approaches adopted for an individual simulation model for sciatica, and moves on to outline the justifications and assumptions which underpin the model structure, data inputs and methods of analysis. The chapter then reports the model findings, and the implications of these results are discussed.

## 7.1 Background and objectives

When compared to patients with NSLBP, sciatica patients suffer more severe pain, higher levels of disability, higher absence from work, and require more health resources

(Konstantinou et al. 2013). The SCOPiC trial (Sciatica Outcomes in Primary Care) (section 1.6.1) was a multi-center, pragmatic, assessor-blind, two-arm, randomised controlled trial. The trial tested whether stratified care leads to faster recovery and overall better outcomes for sciatica patients compared to usual, non-stratified care, and as noted in 1.6.1, the trial found stratified care was not more clinically or cost-effective relative to usual care provided in the trial.

Nonetheless, given that there are currently no decision-analytic models of a stratified care approach to managing sciatica, and concerns were expressed with the methods used in existing model-based sciatica analyses, this chapter, therefore, aims to explore these challenges by conceptualising the first decision model of a stratified care approach for management of patients with sciatica.

Moreover, in expert consultations it was suggested that the outcomes obtained in the usual care arm of the SCOPiC trial were better than had been expected, and therefore experts suggested labelling this "best usual care" and including usual care from a cohort study into the model as a further comparator, "usual care". In what follows, stratified care will be compared with best usual care and usual care.

## 7.2 Methods

In order to assess the cost-utility of stratified care versus best usual care versus usual care in sciatica patients, from the NHS perspective, an individual sampling model was constructed with two-monthly cycles, and analysis was performed over a ten year time horizon. Each aspect of the methodological approach is now discussed in turn.

### 7.2.1 Consultation with experts

The lack of adequate modelling in sciatica meant it was imperative to ensure that at all stages of the model development process, consultation took place with clinical

practitioners as well as experts from epidemiological and health economic research backgrounds (full details in appendix 8).

### 7.2.2 Choice of Model

When selecting the model type, the simplest model ought to be chosen to adequately capture the complexity of the decision problem (Barton, 2004). Following discussion with experts, it was determined that an individual sampling model (ISM) would be most appropriate for this condition. In chapter 5, ISMs were shown to be suitable for modelling stratified treatments, owing to their ability to track and update patient characteristics, as well as easily allowing the probability of future events to be dependent upon the aforementioned characteristics.

After consultation with experts, and analysis of the BeBack cohort study (see 6.3.11), it became evident that movement of patients between health states over the long-term (between twelve months and ten years) should reflect patient function at baseline (measured by the RMDQ score), symptom resolution at 12-months, as well as whether or not the patient had undergone spinal surgery. A decision-tree model cannot easily facilitate longer time horizons, or the dependency necessary for the clinical course of sciatica, and was therefore ruled out. Whereas in a Markov model, the number of states needed to incorporate dependency alongside patient stratification would cause the model to grow infinitely complex. Simulation modelling allows the incorporation of these phenomena easily, via trackers.

The model was constructed in TreeAge Pro 2017 version. All statistical analyses were performed in STATA 15 / MP.

### 7.2.3 Model Population

The consensus of the experts involved in consultations, was that the SCOPiC study population would most closely resemble those who would receive stratified care. The

SCOPiC trial recruited adults (≥ 18 years old), consulting with suspected sciatica. Baseline characteristics in this model were taken directly from this study, patients had a mean age of 52 at the start of the model, 54.6% were female. 54% were classified as having poor function, 34% moderate function, and 12% good function (as per RMDQ cut-offs determined in the previous chapter 6). Table 7.1 shows baseline characteristics of patients who consented to participate in the study.

**Table 7.1 Baseline characteristics of SCOPiC patients**

| Characteristic of participating patients | Stratified Care n=232 | Best usual care n=230 | Population n=462 |
|---|---|---|---|
| Age, mean, y (SD) | 50.86 (0.96) | 53.36 (0.88) | 52.10 (0.66) |
| Sex, female, n (%) | 129 (55.36) | 124 (53.91) | 253 (54.64) |
| Currently in paid employment, n (%) | 162 (70.13) | 155 (67.98) | 317 (69.06) |
| Disability: RMDQ score, mean (SD) | 11.17 (0.35) | 11.20 (0.35) | 11.18 (0.25) |
| Pain intensity: NRS rating, mean (SD) | 5.88 (0.18) | 5.75 (0.201) | 5.81 (0.14) |
| Symptom duration, n (%) | | | |
| 0-3 months | 167 (72.99) | 172 (74.78) | 339 (73.38) |
| 3-6 months | 31 (13.36) | 28 (12.17) | 59 (12.77) |
| 6-12 months | 10 (4.31) | 8 (3.48) | 18 (3.9) |

| Characteristic of participating patients | Stratified Care n=232 | Best usual care n=230 | Population n=462 |
|---|---|---|---|
| >12 months | 24 (10.34) | 22 (9.57) | 46 (9.96) |
| Function state* | | | |
| Poor, n (%) | 125 (53.88) | 126 (54.78) | 251 (54.33) |
| Moderate, n (%) | 76 (32.76) | 80 (34.78) | 156 (33.77) |
| Good, n (%) | 31 (13.36) | 24 (10.44) | 66 (11.90) |

Abbreviations: NRS (Numerical rating scale); RMDQ (Roland Morris Disability Questionnaire)
*Good function (RMDQ 0-4); Moderate function (RMDQ 5-10): Poor function (RMDQ 11-24)

At the beginning of the model, the sampling model defined patient age, gender, and function derived from this distribution of patients in the SCOPiC trial. These characteristics, as well as an occurrence of surgery, are then tracked throughout the duration of the time horizon.

Ages were defined in direct proportion to their relationship with gender in the model population, e.g. an age of under 30 was selected in 4.3% of the simulations for males and 8.3% for female. Summary table of age category by gender is provided below in Table 7.2.

**Table 7.2 Age by gender, composition of model population**

| Age group | Male, n (%) | Female, n (%) |
|---|---|---|
| Under 30 | 9 (4.29) | 21 (8.30) |
| 31-40 | 36 (17.14) | 42 (16.60) |
| 41-50 | 44 (20.95) | 58 (22.92) |
| 51-60 | 56 (26.67) | 55 (21.74) |
| 61-70 | 48 (22.86) | 44 (17.39) |
| 71-80 | 16 (7.62) | 29 (11.46) |
| 81-90 | 1 (0.48) | 4 (1.58) |
| Total | 210 | 253 |

Within the model, age was updated by 2 months every cycle, and age and gender both determined probability of death. However, age and gender were not used to determine transitions between resolved (meaning recovery from sciatica symptoms), unresolved and having surgery, owing to lack of statistical power in both SCOPiC and BeBack datasets. The issue related primarily to subdivision of patients in the BeBack dataset, meaning some transitions and EQ-5D scores were already based upon <10 observations and subdivision by age and gender would have reduced this even further.

### 7.2.4 Definition of the intervention and comparator

Details of the interventions were reported in 1.6.1, and described in full elsewhere (Foster et al. 2017). In the stratified care arm, an algorithm was used to allocate patients in one of three groups with matched care pathways; group 1: brief physiotherapy input in one or two sessions to support self-management, group 2: a course of physiotherapy of up to 6 sessions, group 3: fast-track referral to MRI test and specialist spinal services, and (Konstantinou et al 2019).

The SCOPiC usual care arm, 'best usual care' in this model, was based on non-stratified usual care, delivered by physiotherapists not involved in the stratified care arm of the trial. All usual care patients received their first treatment in the SCOPiC research clinic (a one-off session of assessment, advice and education), where the treating physiotherapist could arrange referral to NHS community physiotherapy or specialist spinal services as required. Most usual care patients were referred for a course of physiotherapy.

Experts advised that the most appropriate data source to derive usual care parameters would be the BeBack cohort study (see 6.3.11). As BeBack includes patients with LBP only as well as LBP and leg pain, only those reporting pain below the knee at baseline were considered to be representative of patients with sciatica (Konstantinou et al. 2012;

Dionne et al. 2008), and were subsequently included in this analysis. Baseline characteristics of the BeBack sample are shown in Table 7.3.

**Table 7.3 Baseline characteristics of BeBack sciatica patients**

| Characteristic of participating patients | Usual care n=1591 |
|---|---|
| Age, mean, y (SD) | 43.88 (0.259) |
| Sex, female, n (%) | 922 (58.50) |
| Currently in paid employment, n (%) | 1177 (75.06) |
| Disability: RMDQ score, mean (SD) | 8.64 (0.151) |
| Pain intensity: NRS rating, mean (SD) | 4.64 (0.068) |
| Symptom Duration, n (%) | |
| 0-3 months | 1023 (66.86) |
| 3-6 months | 148 (9.67) |
| Over 6 months | 359 (23.47) |
| State of function*, n (%) | |
| Poor, | 504 (31.70) |
| Moderate, | 624 (39.25) |
| Good, | 462 (29.06) |

Abbreviations: NRS (Numerical rating scale); RMDQ (Roland Morris Disability Questionnaire)
*Good function (RMDQ 0-4); Moderate function (RMDQ 5-10): Poor function (RMDQ 11-24)

The BeBack cohort was younger than the SCOPiC population, with lower mean disability and pain intensity, although there were more patients with longer symptoms duration in BeBack. There were also more patients in good function in the BeBack sample, although differences in function by category were controlled for in the modelled analysis by applying a model population standardised by baseline function group.

### 7.2.5 Model health states and structure

The model had seven different health states, shown in figure 7.1, with the model in TreeAge form shown in Figure 7.2.

Initially, for the first twelve months, patient movement through the model directly reflected the SCOPiC data, and BeBack data in the case of the usual care patients. Accordingly, all patients began in a symptomatic state, and as indicated by the solid dark orange arrows, patients either stayed symptomatic, moved to surgery, or recovered. Movements between these health states in the model were dependent upon patient function at baseline. Resolution of symptoms was calculated using the global measure of change (ordinal scale for global perceived change (GPC)) which is measured on a six-point ordinal scale. The definition of resolution used in this model was 'completely recovered', 'much better', and 'better' in terms of sciatica symptoms. This differed from the primary analysis in the SCOPiC trial, which used a definition of symptom resolution as 'completely recovered' or 'much better'., with patients considered symptomatic if they reported 'better', 'same/no change', 'worse' or 'much worse' on the same scale. These states were settled upon as a consequence of discussions with researchers and clinicians, the difference with the SCOPiC trial definition made on the basis that the SCOPiC definition was too strict. State definitions were considered reasonable by all experts involved in the consultation phase. All comparators used the same core model structure.

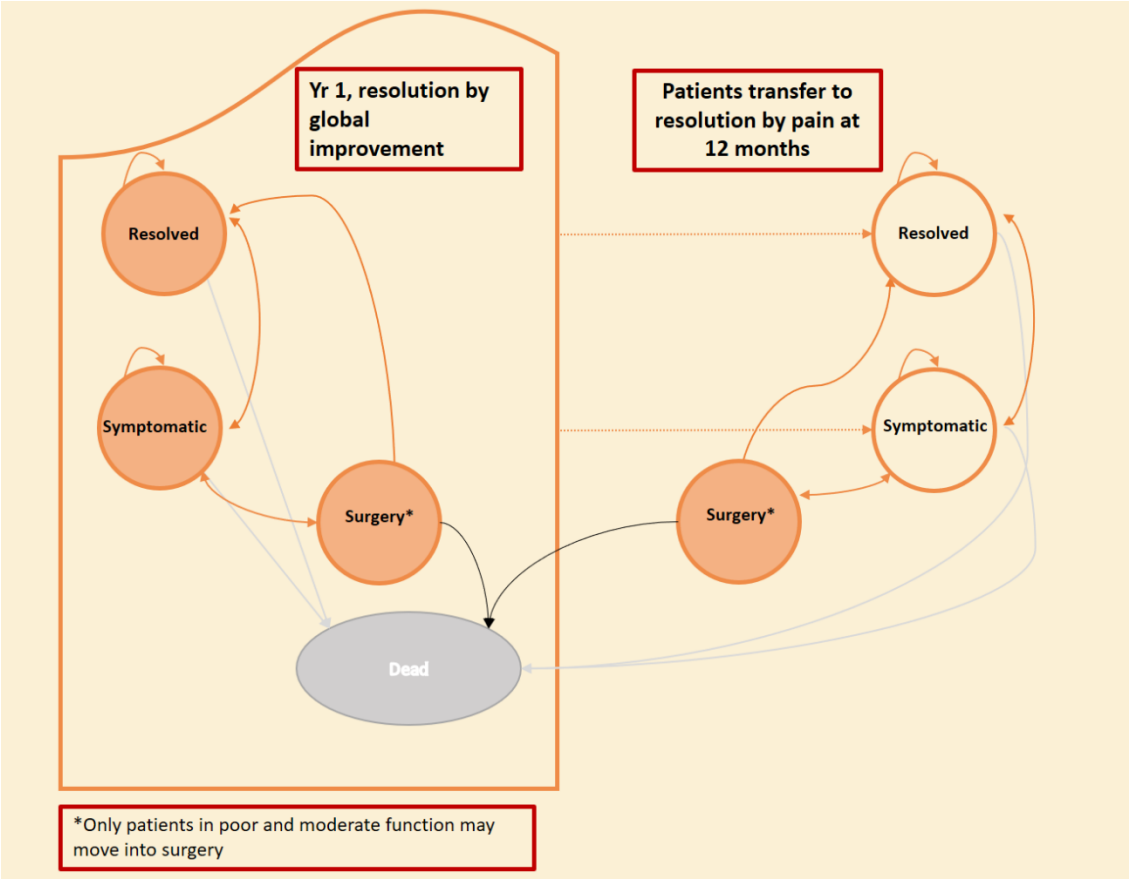



**Figure 7-1 State transition model schematic**

Patient's utility values, and transitions between states, were associated not only with their health state of symptomatic, resolved or surgery, but also their function at baseline. The decision to use patient function to stratify patients in the model so as to determine patient progression was suggested during expert consultation. Prognostic risk groups could not be used in this case, as in the LBP model, as this was not relevant for groups 2 and 3 in the SCOPiC population. Function groups were the same as those used in the LBP model, and based upon the process described in 6.2.5.

As in the LBP model (6.2.11), it was not possible to calculate the individual contribution of each two-monthly cycle to the overall first-year cost, owing to healthcare resource use data only being available for patient healthcare used in the "past twelve months". Accordingly, costs for the first year were a one-off annual cost based upon baseline function group, and resolution, and attached at cycle 6 (month 12).

Surgery was incorporated into the model, however only for patients in moderate or poor function at baseline. Patients, spend one cycle in surgery, and have the utility values of patients with poor function with unresolved symptoms. Once patients have received surgery, they initially move back to "resolved" in the subsequent time cycle. However, once in the resolved health state, they do have a possibility of re-operation, according to a probability assigned from the literature.

At twelve months, as indicated by the two dashed arrows projecting from the year 1 states towards resolution by pain states, patients moved into one of two possible states dependent upon their resolution of symptoms, as measured by numerical rating scale (NRS) pain intensity at twelve months. A score of 5 and below was categorised as resolved, whereas above 5 was categorised as symptomatic. The use of pain scores was necessary, given that the global measure of change (ordinal scale for global perceived change (GPC)) was not available in the BeBack study beyond 12-months.

As can be seen from the Table 7.4, using NRS pain scores to denote resolution appears to be valid as most patients reporting symptom improvement on the GPC scale also report low levels of pain according to the NRS. A Spearman rank correlation coefficient test found a correlation of 0.69 between resolution by pain and global resolution at 12 months, and 0.72 between NRS pain score and global perceived change in symptoms at 12 months.

**Table 7.4 Resolution by global perceived change vs NRS pain scale**

| | Resolved (using NRS pain scores) | Unresolved (using NRS pain scores) | Total |
|---|---|---|---|
| Resolved (using GPC scale) | 126 | 11 | 137 |
| Unresolved (using GPC change) | 4 | 32 | 36 |
| Total | 130 | 43 | |
| NRS: Numerical rating scale, (0-5 Resolved; >5 Unresolved)<br>GPC: Global perceived change, (Resolved – 'completely recovered', 'much better', and 'better'; Unresolved - 'same/no change', 'worse' or 'much worse'). | | | |

### 7.2.6 Model time horizon and cycle length

In this model, a time horizon of ten years was adopted, which was considered adequate to capture meaningful differences between treatment options. This view was endorsed by the experts. In terms of cycle length, because of the 4-month and 12-month follow up in the SCOPiC trial, it was considered appropriate to use two-monthly cycles to facilitate calculation of our transition probabilities. These two-monthly cycles were used throughout the model horizon.

### 7.2.7 Transition probabilities

In the first year of the individual sampling model, patients transitioned every two months between symptomatic, resolved, and surgery, with movements determined also by their function at baseline. These movements were based upon the SCOPiC trial for stratified care and best usual care, and on BeBack data for usual care. At the endpoint of the SCOPiC trial (twelve months), patients then transitioned into model health states, of symptomatic, resolved, and surgery, determined by their pain level (measured by NRS pain scores) and function at baseline. Upon assignment to symptomatic, resolved, and surgery states based upon NRS pain scores, patients subsequently moved between these states (or remained in the same state) for the remaining nine years of the model. Patient transitions for this period were based upon observed transitions in patients from the BeBack cohort study. Transitions used in the model are shown in Table 7.5, and now discussed in turn.

**Table 7.5 Model transition probabilities and distributions**

| | Stratified Care | | Best Usual Care | | Usual care | |
|---|---|---|---|---|---|---|
| Transition to resolved/ unresolved by patient function | Mean (SE) | Dist (α,β) | Mean (SE) | Dist (α,β) | Mean (SE) | Dist (α,β) |
| **Zero to Four Months** | | | | | | |
| Poor to resolved | 0.386 (0.049) | Beta (37.4,59.6) | 0.500 (0.05) | Beta (49.5,49.5) | 0.163 (0.036) | Beta (16.8,86.2) |
| Moderate to resolved | 0.527 (0.066) | Beta (30,27) | 0.500 (0.061) | Beta (33.5,33.5) | 0.256 (0.036) | Beta (37.6,109.4) |
| Good to resolved | 0.423 (0.095) | Beta (11,15) | 0.541 (0.114) | Beta (9.7,8.3) | 0.246 (0.081) | Beta (6.6,20.4) |
| **Four months to Twelve Months** | | | | | | |
| Poor resolved to unresolved | 0.077 (0.047) | Beta (2.4,28.6) | 0.089 (0.036) | Beta (5.4,55.6) | 0 | - |
| Poor unresolved to resolved | 0.128 (0.047) | Beta (6.27,42.7) | 0.163 (0.092) | Beta (2.4,12.6) | 0.024 (0.029) | Beta (0.7,27.3) |
| Moderate resolved to unresolved | 0.022 (0.023) | Beta (0.8,38.2) | 0.043 (0.030) | Beta (2,44) | 0.120 (0.046) | Beta (5.8,42.2) |
| Moderate unresolved to resolved | 0.182 (0.129) | Beta (1.5,6.5) | 0.174 (0.101) | Beta (2.3,10.7) | 0.067 (0.047) | Beta (1.9,26.1) |
| Good resolved to unresolved | 0.032 (0.043) | Beta (0.5,15.5) | 0.152 (0.059) | Beta (0.6,11.4) | 0.075 (0.152) | Beta (0.2,1.8) |
| Good unresolved to resolved | 0.118 (0.132) | Beta (0.6,4.4) | 0.105 (0.153) | Beta (0.3,2.7) | 0.184 (0.173) | Beta (0.7,3.3) |
| **Initial distributions of patients into resolved/unresolved by NRS** | | | | | | |
| Poor resolved to unresolved | 0.063 (0.031) | Beta (3.9,58.1) | 0.068 (0.033) | Beta (3.9,54.1) | 0.188 (0.098) | Beta (2.8,12.2) |
| Poor unresolved to resolved | 0.154 (0.071) | Beta (3.9,21.2) | 0.428 (0.094) | Beta (11.6,15.4) | 0.324 (0.077) | Beta (11.7,24.3) |
| Moderate resolved to unresolved | 0 | - | 0.02 (0.020) | Beta (1,48) | 0.04 (0.028) | Beta (2,47) |
| Moderate unresolved to resolved | 0.50 (0.158) | Beta (4.5,4.5) | 0.308 (0.128) | Beta (3.7,8.3) | 0.688 (0.082) | Beta (21.3,9.7) |
| Good resolved to unresolved | 0 | - | 0.063 (0.061) | Beta (0.9,14.1) | 0 | - |

| | Mean (SE) | Dist | Mean (SE) | Dist | Mean (SE) | Dist (α,β) |
|---|---|---|---|---|---|---|
| Good unresolved to resolved | 0.333 (0.192) | Beta (1.7,3.3) | 0.333 (0.272) | Beta (0.7,1.3) | 0.50 (0.354) | Beta (0.5,0.5) |
| **Surgical transitions (poor & moderate only), by time** | | | | | | |
| Time | Mean | | SE | | Distribution (α,β) | |
| First year | 0.004 | | 0.003 | | Beta (1.8,439.9) | |
| Second year | 0.016 | | 0.006 | | Beta (7,429) | |
| Third year | 0.026 | | 0.007 | | Beta (13,502) | |
| Fourth year | 0.032 | | 0.008 | | Beta (15,468) | |
| Fifth year | 0.037 | | 0.009 | | Beta (19,464) | |
| Sixth year | 0.041 | | 0.009 | | Beta (20,465) | |
| Seventh year | 0.045 | | 0.010 | | Beta (19,409) | |
| Eighty year | 0.048 | | 0.010 | | Beta (22,434) | |
| Ninth year | 0.051 | | 0.010 | | Beta (24,458) | |
| Tenth year | 0.053 | | 0.010 | | Beta (27,474) | |
| Recurrence (any year) | 0.002 | | 0.004 | | Beta (0.25,123) | |
| **Transitions between resolved states, one year to ten years** | | | | | | |
| | Mean | | SE | | Distribution (α,β) | |
| Resolved to unresolved | 0.006 | | (0.016) | | Beta (0.3,44.7) | |
| Unresolved to resolved | 0.018 | | (0.043) | | Beta (0.3,15.7) | |

*Good function (RMDQ 0-4); Moderate function (RMDQ 5-10): Poor function (RMDQ 11-24)

#### 7.2.7.1 Transition probabilities for first year

In order to derive the transition probabilities for the first year in both stratified and best usual care, 4-month (representing baseline to 4 months) and 8-month (representing 4 to 12 months) transition probabilities available in the SCOPiC data, were transformed into two-monthly probabilities. Firstly, probabilities (p) from the data were transformed into a two-monthly rate using the formula; $r = \frac{1}{t}\ln(1-p)$, where r is the rate per time unit t. Two monthly rates were then transformed to a two-monthly probability using$p = 1 - e^{-rt}$. For the usual care patients, observations were available at baseline, six months, and 12 months in the BeBack data, and therefore similar calculations were undertaken to transform six monthly transition probabilities into two-monthly probabilities.

As the transition probabilities were calculated for each function group, no standardisation was applied as it was assumed these probabilities reflect movements of patients of similar function, although it is acknowledged that there could be some small differences between the characteristics of function groups in SCOPiC and BeBack populations.

Derivation of surgical probabilities was taken from SCOPiC trial data whereby 9 patients had spinal surgery in one-year. This corresponds to an annual probability of surgery of 0.019 or 1.9% for any patient in the SCOPiC trial, and is the probability used in this model. However, because only poor or moderate function symptomatic patients could receive those surgeries within the model, the probability of surgery, in symptomatic patients in poor and moderate function, throughout the model had to be weighted towards the proportion of patients in those two states at each cycle. In order to establish how many patients were in symptomatic poor and moderate function health states over time, a SCOPiC model population was simulated in Excel combining stratified and usual care patients.

For the first six cycles, unique probabilities were set for each cycle, reflecting the rapid improvement in patients within that trial. For the first cycle, 88% of patients began the model in poor and moderate function symptomatic states and could receive surgery in the model; therefore 415 of the total 472 patients in SCOPiC trial were eligible to receive those 9 surgeries in the model. Therefore the annualised probability of surgery in poor and moderate function patients was 9/415, equal to 2.12% or 0.0212. Converting this annual probability to a two-monthly rate; $r = \frac{1}{t}\ln(1-p)$, where r is the rate per time unit t, the two monthly rate is 0.00357. Translating this rate back to a probability using; $p = 1 - e^{-rt}$, yields a two monthly probability of 0.0036. By cycle 6 however, only 21% of the sample were symptomatic in poor and moderate function, and therefore the annualised probability of surgery in those states that year was 9/99, 9.1% or 0.091, this yields a two-monthly rate of 0.0159, equating to a two-monthly probability of 0.0158 (1.58%).

From 12-months onwards an annual rate of surgery is set, which changes every six cycles as the proportion of patients in those states changes. By cycle 54, the two-monthly probability of surgery in those states reaches 5.3% as few patients remain in the symptomatic states. This assumes a constant annual probability of surgery across time, and possibly overstates the number of surgeries. Given the cost implications of surgery, and the fact that the SCOPiC trial was not designed to detect differences in rates of surgery between comparators, it is assumed that surgery is equivalent on all three comparators throughout the model horizon.

The recurrence rate for surgery is 6% over five years, as found in Lequin et al. (2013), this was transformed to a two-monthly probability to provide a two-monthly recurrence probability of 0.002, SD (0.004). This is kept constant throughout the model, and is the same for all management strategies.

Note that all transition probabilities are assigned Beta distributions, this reflects the fact that at each node in the TreeAge model there is a dichotomous choice, death/survive, surgery/no surgery, resolved/unresolved, see Figure 7.2. Where a state transition (p) was assigned a Beta distribution, the probability of movement into the other state was set in TreeAge as # a notation representing (1-p), to prevent the probabilities of moving out of the state exceeding one.

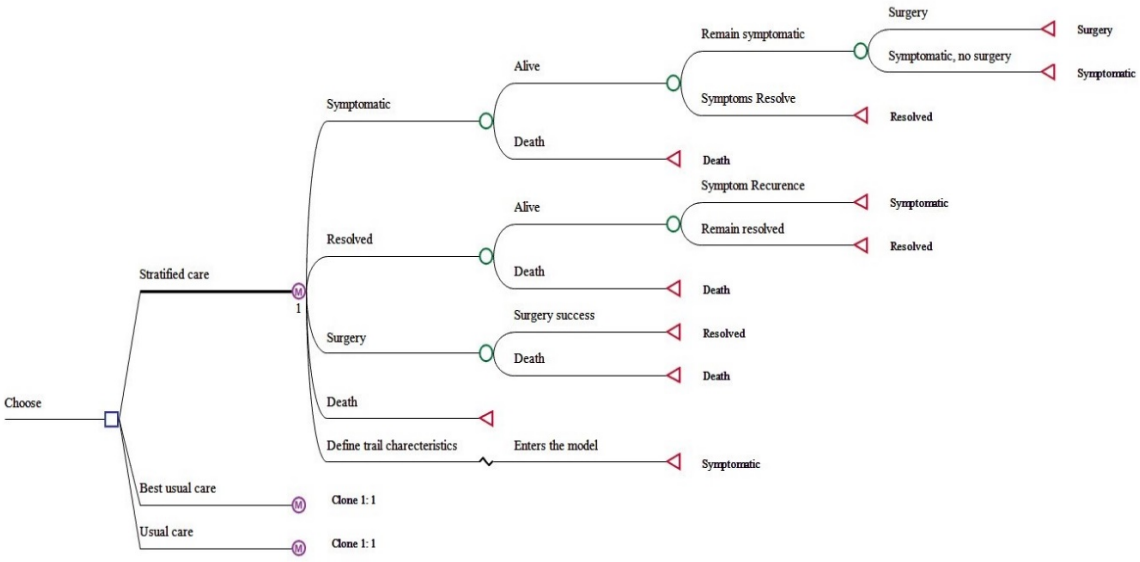



**Figure 7-2 Use of dichotomies and Beta distributions.**

Transitions to death were calculated by using 2016/17 ONS life tables (ONS, 2018). The annual mortality rate for males and females between ages 18 and 100 were taken from the life tables, and converted to annual probabilities using $p = 1 - e^{-rt}$. This annual probability was converted into a two-monthly rate using, $r = \frac{1}{t}\ln(1-p)$, and then converted to a two-monthly probability again using $p = 1 - e^{-rt}$.

### 7.2.8 Moving from resolved / symptomatic by GPC to resolved / symptomatic by NRS pain scores

At twelve months, patients moved into the same health states, e.g. resolved / symptomatic / surgery, but in this case, these state definitions reflect pain change instead of global perceived change. The values of these transitions are shown in Table 7.5 under the heading "Initial distributions of patients into resolved/unresolved by pain NRS". These transition probabilities were based upon the proportion of patients in each state of resolution in each of the function groups at twelve months. Probabilities were calculated separately for stratified care, best usual care, and usual care, in order to reflect the lower proportion of patients who achieved resolution (NRS pain score) at twelve months, on usual care. In the model, costs and QALYs associated with each of the resolution by pain states were first attached in cycle seven, at fourteen months, based upon the proportion of patients in each of the resolved by pain states at twelve months.

### 7.2.9 Transition probabilities for twelve months to ten years.

In order to perform the extrapolation, the assumption that patients transition at equivalent rates over time on all comparators is again used. Accordingly, from fourteen months onwards, the transition probabilities between resolution states, shown in Table 7.5 under the heading 'Transitions between resolved states, one year to ten years', were assumed to be exactly the same in stratified care, best usual care, and usual care. However, the differential distribution of patients into resolved by pain states at twelve months, with equal transitions from one year onwards, ensures that more patients in best usual care achieve resolution for the duration of the model.

Natural logs were used to transform the four-year transition of patients using data from the BeBack cohort study, (6.3.11) into two-monthly transitions for the model. In the analysis, NRS pain scores were then assessed at twelve months with scores 5 and below taken to be

symptom resolution, and above 5 taken to be unresolved. For each resolution state, NRS pain scores were then assessed at five years, in order to determine how patients have transitioned across those four years. Whilst these transition probabilities were calculated to reflect patient movement from twelve months to five years, it was further assumed that the probabilities calculated from the linear matrix would hold over the next five years. Accordingly, these probabilities were used to move patients between states from five to ten years in the model. In all cases the standard errors were calculated by using $\frac{p(1-p)}{\sqrt{n}}$ where p was the probability of movement between the states as generated by natural logs, and n reflects the number of patients in each of the states at the beginning of the period the matrix is reflecting, directly obtained from BeBack data.

To understand whether five-year data from the BeBack study were appropriate to calculate transitions for the model patient population, mean RMDQ scores for each function and resolution state in the SCOPiC patients at twelve months, were compared with the RMDQ scores of the patients in the BeBack sample at twelve months, shown in Table 7.6

**Table 7.6 Mean RMDQ scores in SCOPiC and BeBack patients**

| Function and resolution yes/no | SCOPiC at twelve months | BeBack at twelve months |
|---|---|---|
| Poor Resolved | 4.61 | 7.44 |
| Poor Unresolved | 14.37 | 15.60 |
| Moderate Resolved | 2.4 | 3.98 |
| Moderate Unresolved | 11.06 | 10.54 |
| Good Resolved | 1.54 | 1.58 |
| Good Unresolved | 5.86 | 8.00 |
| Population | 5.17 | 5.83 |

*Good function (RMDQ 0-4); Moderate function (RMDQ 5-10): Poor function (RMDQ 11-24)

Whilst there were some small differences between the three different comparators across the periods, most of the groups did not have statistically significant differences ($p<0.10$), aside from patients who began in poor function who had resolution of symptoms, who were in better function in the SCOPiC dataset. Patients in the BeBack sample were in slightly worse function overall (5.83 vs 5.17).

### 7.2.10 Costs

The base-case economic evaluation was performed from the NHS perspective which takes into account costs solely incurred by the NHS and excludes the value of private healthcare costs. Costs associated with the first year reflected the resource usage from the SCOPiC trial in the case of stratified care and best usual care, and the BeBack cohort study for usual care. The SCOPiC data used was from the within-trial economic analysis (submitted for publication), which performed multiple imputation and included costs directly associated with the trial. The resource use data reflected responses to self-report questionnaires and included consultations with GPs and nurses, other healthcare professionals, prescriptions, hospital procedures, as well as prescribed medications and over the counter treatments. The costs of the latter were included only in a secondary analysis from a societal perspective.

As the modelling for the first year is predicated upon the treatment effect associated with the trial, the costs related to delivering the stratified care model in the trial were included, including the costs associated with the trial-specific protocol for treatment. The mean costs for patients in poor, moderate and good function (classified at baseline), were calculated separately for stratified, best usual care and usual care.

Beyond the first year, data from the BeBack cohort (Foster et al. 2008) were used to derive the costs associated with one year of healthcare for sciatica patients dependent upon their function and pain at the beginning of the cycle. Costs derived from BeBack were annualised, and therefore the costs per 2-month time cycle in the model reflect 1/6 of the

total first year costs in BeBack. BeBack reflects healthcare usage in what would be considered “usual care”, and therefore for the base case analysis it was assumed that neither stratified care nor best usual care impact upon long-term healthcare usage of patients. Given this represents one of the most significant sources of uncertainty in the model, the impact of varying this assumption is discussed later. Unit costs in the model were based upon 2017 prices, and are shown in Table 7.7. Total costs were discounted at 3.5% as per NICE guidance (NICE, 2018), and were derived from questionnaires which stipulate costs directly related to back pain and sciatica only.

**Table 7.7 Unit prices for healthcare**

| Health care resource | Unit cost (£) | Unit Cost Source |
|---|---|---|
| Primary care contacts: | | |
| General Practitioner: surgery consultation | 37 | Unit costs of Health and Social Care 2017 |
| Practice Nurse: surgery consultation | 11 | Unit costs of Health and Social Care 2017 |
| Practice Nurse: home visit | 94 | Unit costs of Health and Social Care 2017 |
| Hospital-based care | | |
| Consultant: Sciatica pain first attendance | 167 | DOH, 2017 |
| Consultant: Sciatica pain follow-up | 141 | DOH, 2017 |
| Consultant: Pain management first attendance | 177 | DOH, 2017 |
| Consultant: Pain management follow-up | 101 | DOH, 2017 |
| Physiotherapist: First attendance | 65 | DOH, 2017 |
| Physiotherapist: Follow-up attendance | 49 | DOH, 2017 |
| Consultation: A&E | 180 | DOH, 2017 |
| Hospital nurse | 89 | DOH, 2017 |
| Diagnostic tests: x-ray | 31 | DOH, 2017 |
| Diagnostic tests: CT scan | 103 | DOH, 2017 |
| Diagnostic tests: MRI scan | 169 | DOH, 2017 |
| Diagnostic tests: Blood test | 6 | DOH, 2017 |
| Spinal epidural injection | 575 | DOH, 2017 |
| Surgery (Discectomy) | 5,298 | DOH, 2017 |
| Out-of-pocket treatments | Participant reported costs | |
| Prescribed medication | Participant-specific | BNF, 2017 |
| Work absence/reduced productivity | Participant - specific | ONS, 2017 |

Resource use data derived from SCOPiC and BeBack were multiplied by unit prices to create annual costs associated with each state, shown in Table 7.8. As noted above, in the first year, these estimates were entered into the model at cycle six to reflect the initial function group and resolution. In cycles seven onwards, the annual estimates of the costs associated with each of the resolution by pain states were divided by six to create a two-monthly cost. The cost of surgery was a one-off non-variable cost of £5298.

**Table 7.8 Total annual costs and gamma parameters, per health state**

| | | NHS annual costs | | | Total annual societal costs | | |
|---|---|---|---|---|---|---|---|
| First year costs | | | | | | | |
| | | **SC** | **BUC** | **UC** | **SC** | **BUC** | **UC** |
| oor | £ | 628.91 | 587.43 | 497.72 | 1016.14 | 1222.86 | 1845.45 |
| esolved | SD | (525.55) | (583.53) | (557.86) | (1142.63) | (2079.02) | (5162.20) |
| | α, λ | 1.43,0.002 | 1.24,0.002 | 0.80,0.002 | 0.791,0.001 | 0.346,0.0002 | 0.128,0.0001 |
| oor | £ | 808.04 | 456.73 | 791.58 | 1129.48 | 695.67 | 2559.05 |
| nresolved | SD | (618.74) | (457.32) | (1039.17) | (836.37) | (1048.97) | (4834.75) |
| | α, λ | 1.71,0.002 | 0.545,0.001 | 0.58,0.001 | 1.824,0.002 | 0.44,0.00 | 0.28,0.0001 |
| oderate | £ | 575.72 | 598.38 | 252 | 1334.40 | 1040.17 | 1999.73 |
| esolved | SD | (703.05) | (834.74) | (446.77) | (2767.79) | (1749) | (1434.07) |
| | α, λ | 0.67,0.001 | 0.724,0.001 | 0.318,0.001 | 0.232,0.001 | 0.354,0.0003 | 1.944,0.001 |
| oderate | £ | 567.6 | 536.80 | 409.54 | 1969.41 | 760.63 | 1080.69 |
| nresolved | SD | (880.96) | (494.78) | (559.40) | (4230.57) | (478.48) | (4551.04) |
| | α, λ | 0.415,0.001 | 0.371,0.001 | 0.536,0.001 | 0.217,0.001 | 2.527,0.003 | 0.056,0.0001 |
| ood | £ | 441.99 | 278.9 | 69.30 | 730.35 | 439.96 | 366.25 |
| esolved | SD | (657.58) | (496.85) | (129.30) | (974.51) | (733.19) | (430) |
| | α, λ | 0.451,0.001 | 0.18,0.001 | 0.287,0.004 | 0.562,0.001 | 0.360,0.001 | 0.725,0.002 |
| ood | £ | 505.05 | 216.08 | 8.97 | 1228.96 | 1655.41 | 28.13 |
| resolved | SD | (518.84) | (166.02) | (3.06) | (664.85) | (2360.18) | (13.06) |
| | α, λ | 0.948,0.002 | 0.173,0.001 | 8.59,0.958 | 3.417,0.003 | 0.492,0.0002 | 4.639,0.165 |
| Annual costs years 2-10 | | | | | | | |
| esolved | £, SD | 289.00 (466.74) | | | 1383.42(3322.53) | | |
| | α, λ | 0.383,0.001 | | | 0.173,0.0001 | | |
| nresolved | £, SD | 599. 79 (863.85) | | | 1492(4425.16) | | |
| | α, λ | 0.482,0.001 | | | 0.114,0.0001 | | |

obreviations; BUC (Best Usual care); SC (Stratified care); UC (Usual care)
λ are gamma parameters

For the secondary analysis, using a societal perspective, the costs associated with private healthcare were assumed to be the same as that of the NHS equivalent. This is due to lack of nationally representative unit cost estimates for private healthcare. Resource use was

taken from the BeBack and SCOPiC datasets, which provide detail on the number of visits to a healthcare practitioner, alongside whether those visits were to an NHS professional or private healthcare. Very few patients received both NHS and private care, and where patients indicated they had received both NHS and private healthcare, the number of visits was divided by two, with half of the total cost allocated to NHS and half to private.

To assess the impact of stratified care upon work participation, self-reported days of work absence owing to sciatica were estimated from SCOPiC and BeBack datasets, with associated costs calculated using the human capital approach. For the first year of the model, for stratified care and best usual care, work absence in days was estimated from the SCOPiC study, and multiplied by the UK population mean wage in 2017. For usual care in the first twelve months, work absence was derived from the BeBack data, and multiplied by the UK population mean wage in 2017. In all cases, mean work absence in days has been adjusted to account for the 30.94% not in employment in the SCOPiC study.

Table 7.9 presents the work absence due to back pain and sciatica in the SCOPiC and BeBack studies, the wage used, and the percentage of participants employed in the model population.

**Table 7.9 Work absence, stratified care, best usual care, and usual care**

| | Stratified care | Best usual care | Usual care |
|---|---|---|---|
| Mean days absence due to back pain | Poor: 1.13 | Poor: 3.47 | P:12.66 |
| | Moderate: 3.80 | Moderate: 1.1 | M:7.56 |
| | Good: 1.12 | Good: 2.83 | G:3.10 |
| | Population: 2.12 | Population: 2.57 | Population: 7.86 |
| Mean daily wage (SCOPiC) | £107.74 | | |
| % employed in SCOPiC | 69.06% | | |
| | By pain resolution | | |
| Mean days absence due to back pain (BeBack) | Resolved: 7.73 | | |
| | Unresolved: 7.99 | | |

### 7.2.11 Quality adjusted life years (QALYs)

EQ-5D 5L scores for the first year were taken from SCOPiC data for stratified care and best usual care, and converted to their 3L equivalent using the EQ-5D crosswalk algorithm (van Hout et al. 2012).

Usual care values were derived from the BeBack study, which used the EQ-5D 3L. Both sets of responses were converted to utility scores based upon the York tariff (Dolan et al. 1996). In the model, QALYs were discounted at 3.5% per year, as per current NICE guidelines (NICE, 2018). In the PSA, the model uses annual EQ-5D scores sampled from their distribution and then divided by six in order to obtain two-monthly weighted QALYs.

There were a number of calculations used to obtain the EQ-5D scores used in the first year of this model. Baseline EQ-5D for each function group in the model reflected standardised mean baseline EQ-5D scores of patients in the SCOPiC study. In order to obtain utility

values for each treatment, function group, and whether or not resolution was achieved, a treatment effect upon EQ-5D score for resolved and unresolved states was calculated for each function group. Treatment effect upon EQ-5D was calculated using linear regression analysis controlling for baseline EQ-5D score. In the case of the SCOPiC data, regression models were calculated separately for stratified and best usual care patients, results are shown in Table 7.10. These treatment effects were then added or subtracted to a baseline model population EQ-5D for each patient function. This was necessary because there were differences in baseline EQ-5D amongst the three comparators.

**Table 7.10 EQ-5D treatment effect by time point and comparator**

| Function, by resolution (Yes/No) | Model Baseline EQ-5D | EQ-5D treatment effect at four months | | | EQ-5D treatment effect at twelve months | | |
|---|---|---|---|---|---|---|---|
| | | SC | BUC | UC | SC | BUC | UC |
| Poor function Resolved | 0.3995 | +0.319 | +0.285 | +0.249 | +0.332 | +0.361 | +0.324 |
| Poor function Unresolved | | -0.039 | +0.010 | -0.016 | -0.023 | +0.186 | -0.102 |
| Moderate function Resolved | 0.5906 | +0.203 | +0.180 | +0.095 | +0.231 | +0.231 | +0.198 |
| Moderate function Unresolved | | +0.055 | +0.061 | -0.044 | +0.021 | +0.043 | -0.087 |
| Good function Resolved | 0.7360 | +0.109 | +0.129 | +0.028 | +0.129 | +0.199 | +0.046 |
| Good function Unresolved | | -0.014 | -0.083 | -0.043 | -0.198 | -0.133 | +0.058 |

Abbreviations; BUC (Best Usual care); SC (Stratified care); UC (Usual care)

*Good function (RMDQ 0-4); Moderate function (RMDQ 5-10): Poor function (RMDQ 11-24)

As these scores were artificially created, standard errors were taken from actual EQ-5D scores in SCOPiC data, at baseline, four months, and 12-months, and BeBack data at baseline, six and 12 months.

To generate the utility values between time-points, values were assumed to take a linear function between observed periods; these are shown only for stratified care, in the columns titled “generated x months” in Table 7.11. Calculations for other treatments were identical.

**Table 7.11 Estimating EQ-5D scores at unobserved time points, stratified care**

| Function / Resolution group | **Model baseline utility Mean (SE)** | Generated 2-month utility | **Standardised 4-month utility Mean (SE)** α, λ | Generated 6-months | Gen 8 months | Gen 10 months | **Standardised 12-month utility Mean (SE)** α, λ |
|---|---|---|---|---|---|---|---|
| Poor function Resolved | 0.399 (0.019) | 0.559 | 0.718 (0.022) 230,118 | 0.721 | 0.725 | 0.728 | 0.732 (0.025) 222,81 |
| Poor function Unresolved | 0.399 (0.019) | 0.395 | 0.392 (0.039) 62,96 | 0.388 | 0.384 | 0.381 | 0.377 (0.063) 22,36 |
| Moderate function Resolved | 0.591 (0.014) | 0.691 | 0.794 (0.018) 409,106 | 0.801 | 0.808 | 0.814 | 0.821 (0.02) 307,67 |
| Moderate function Unresolved | 0.591 (0.014) | 0.618 | 0.648 (0.055) 49,26 | 0.639 | 0.630 | 0.621 | 0.611 (0.071) 28,18 |
| Good function Resolved | 0.736 (0.017) | 0.800 | 0.845 (0.026) 161,30 | 0.850 | 0.855 | 0.860 | 0.865 (0.032) 98,15 |
| Good function Unresolved | 0.736 (0.017) | 0.729 | 0.722 (0.035) 62,96 | 0.676 | 0.630 | 0.584 | 0.538 (0.052) 49,43 |
| *Good function (RMDQ 0-4); Moderate function (RMDQ 5-10): Poor function (RMDQ 11-24) | | | | | | | |

EQ-5D scores for the resolution states (beyond 12-months) originated from patients in the SCOPiC population, and were determined in the model by resolution (pain score) at 12 months and baseline function, shown in Table 7.12. As with the LBP model, the

assumption is made that twelve-month follow-up values were stable over the next nine years.

**Table 7.12 EQ-5D scores for symptom resolution, years 2-10 of model**

| Function / Resolution group | Mean | Standard Error | n | Distributions | α,β |
|---|---|---|---|---|---|
| Poor function Resolved | 0.744 | (0.016) | 134 | Beta | 553,190 |
| Poor function Unresolved | 0.447 | (0.040) | 46 | Beta | 69,85 |
| Moderate function Resolved | 0.808 | (0.013) | 100 | Beta | 741,176 |
| Moderate function Unresolved | 0.596 | (0.046) | 19 | Beta | 67,46 |
| Good function Resolved | 0.833 | (0.025) | 35 | Beta | 185,37 |
| Good function Unresolved | 0.671 | (0.067) | 11 | Beta | 32,16 |
| *Good function (RMDQ 0-4); Moderate function (RMDQ 5-10): Poor function (RMDQ 11-24) | | | | | |

## 7.3 Methods of Analysis

### 7.3.1 Base case analysis

The base case is a cost-utility analysis of stratified care versus best usual care versus usual care for sciatica patients, performed from the NHS perspective. In order to obtain the base case estimates, Monte Carlo simulation was utilised to perform 25,000 patient microsimulations. The results are presented in the form of a cost-per-additional QALY gained. As the interest of the analysis is the cost-effectiveness of stratified care, results will be presented in tables as two-way comparisons between stratified care and best usual care, and stratified care and usual care.

Probabilistic sensitivity analyses were also performed to explore the uncertainty arising from parameter uncertainty. Second-order Monte Carlo simulation was repeated 1000 times using all parameters with distributions attached, transitions, utility values, and costs. Given the presence of three strategies, probabilistic analyses are presented by cost-effectiveness acceptability curves (CEAC) in the form of three-way comparisons between all comparators.

#### 7.3.1.1 Secondary analyses

Three main secondary/supplementary analyses were undertaken as follows:

1. Whilst all results are presented discounted, methodological uncertainty is considered by representing undiscounted results.
2. Due to the high private healthcare costs and volume of time taken off work associated with sciatica (detailed in Chapter 2), a secondary analysis was performed from the societal perspective.
3. Sub-group analysis was performed on the discounted base case, with the model run separately for patients in each of the three SCOPiC groups. This analysis was performed by assigning each of the SCOPiC groups a different function profile; Group 1 (8% poor, 51% moderate, and 41% good function), Group 2 (58% poor, 38% moderate, and 5% good function), and Group 3 (82% poor, 17% moderate, 1% good function).

### 7.3.2 Deterministic sensitivity analyses

Given the uncertainty associated with the long-term parameters in the model, a number of sensitivity analyses were performed to assess the robustness of the results in light of these structural and parameter uncertainties, namely changes in assumptions or input parameters used in the base case. All scenarios considered were pre-specified in conjunction with the views of the group of experts.

1. With stratified care likely to lead to an altered treatment pathway, analyses explored the opinion of experts that stratified care may lower long-term treatment costs.
2. As utility values for each state beyond 12-months were assumed to be the same on each comparator, an analysis explored the impact of using the higher EQ-5D scores

for patients on best usual care for equivalent states (as found in the SCOPiC trial) lasting one, two, five, and ten years.

3. Finally, given the concerns regarding the impact of health-state choice upon likely cost-effectiveness of the approach, alongside the importance of function as an outcome for sciatica patients, it was also suggested that a sensitivity analysis assess the impact of modelling function states instead of resolution by symptom/pain. In this analysis, the model proceeds in a similar structure but patient's transition between function states, accruing costs and QALYs related to their function instead of resolution of symptoms.

The results of all sensitivity analyses are presented in the form of a cost-per-additional QALY gained, based upon 25,000 first order simulation trials and 1000 second-order Monte Carlo samples, results are displayed either as CEACs, or in numeric form in tables reflecting the % likelihood of cost-effectiveness at WTP thresholds of £20,000 per QALY.

### 7.3.3 Value of information Analysis

This analysis will provide a quantification of the value of further research to the NHS. Using the PSA output, an expected value of perfect information (EVPI) per person is calculated. To obtain the overall value of removing decision uncertainty, the individual estimate is then multiplied by the population expected to benefit from the intervention.

To consider the total value of removing decision uncertainty it is essential to account for, not only the population impacted by this decision annually, but also the duration that the comparison holds relevancy. This comparison is assumed to hold relevance for the next 10 years, and this time frame is adopted in this analysis.

Per person EVPI was estimated using the Sheffield Accelerated Value of Information (SAVI) software (Strong et al., 2014). In order to calculate the population expected to benefit and to ensure there is no double counting, 10-year consultation prevalence is required. Such

a calculation was not found in the literature for sciatica, so based upon the assumed 10-year consultation prevalence used in the LBP model (derived from the estimates of Jordan et al. (2012)) and that between 20-35% of LBP patients suffer from sciatica (Laroche & Perrot, 2013), it was assumed that the consulting population was 27.5% of the LBP population, giving a 10-year discounted consultation prevalence of 3,885,026.

The Sheffield Accelerated Value of Information (SAVI) software (Strong et al., 2014) was also used in order to calculate the single and group parameter EVPPI. Given that there are over 100 parameters in this model, it is expected that the initial contribution of each parameter to the overall uncertainty will be minor, and therefore EVPPI was computed for groups of associated parameters. Subsets used in analysis will be first year transition probabilities, transition probabilities years 2-10, utility values, and costs.

## 7.4 Model Results

The presentation of the results is in four components. Firstly, the base case analysis is presented, with accompanying CEACs reflecting probabilistic analyses on second-order uncertainty. Secondly, sensitivity analyses on the base case are performed, namely undiscounted results, a societal analysis, and subgroup analysis for each of the sciatica subgroups. Third, deterministic sensitivity analysis addresses the structural uncertainty over the long-term costs. Finally, a value of information analysis (using the base case PSA output), considers the potential value of further information. All results, except those in the undiscounted analysis, are discounted. Results are presented with willingness to pay thresholds of £20,000 per QALY gained.

### 7.4.1 Base case analysis

The results of the base case analysis are presented in Table 7.13, showing the QALYs and NHS costs associated with stratified care, best usual care and usual care for the model population.

**Table 7.13 Base case analysis stratified care vs usual care vs best usual care**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified Care | 4781.84 | 5.16 |
| Best Usual Care | 4474.01 | 5.19 |
| Difference (SC-BUC) | +307.83 | -0.03 |
| Usual Care | 5038.62 | 4.85 |
| Difference (SC-UC) | -256.78 | +0.31 |
| Abbreviations: BUC (Best usual care); (Quality-adjusted life year); SC (Stratified care), UC (Usual care) | | |

The ten-year sciatica-related healthcare costs of stratified care were estimated to be £4781.84 per patient, with mean QALYs experienced per-patient of 5.16 during the 10-year time horizon. Treatment with best usual care was cheaper at £4474.01 per patient, and produced 5.19 QALYs. Therefore, stratified care yielded 0.03 less QALYs for an additional cost of £307.83 per patient over 10 years, and was dominated by best usual care. Stratified care is however dominant relative to usual care, offering 0.31 more QALYs at £256.78 cheaper.

The second-order Monte Carlo PSA demonstrate the variability in these results, shown in the cost-effectiveness acceptability curve in Figure 7.3, indicate that at £20,000 per QALY, best usual care is 63% likely to be cost-effective, stratified care 37%, both more likely to be cost-effective than usual care. At willingness-to-pay thresholds above around £2,500 per QALY likely cost-effectiveness of usual care falls to near zero.

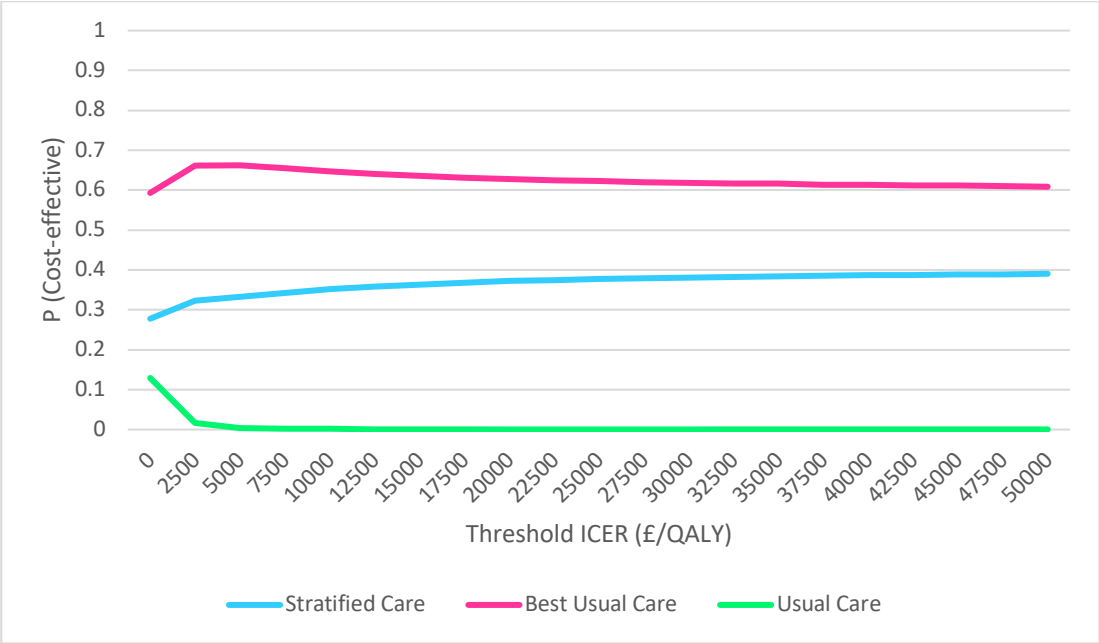



**Figure 7-3 Cost-Effectiveness Acceptability Curve, Stratified care vs best usual care vs usual care, base case**

To understand why the model estimates the cost-effectiveness to be so uncertain despite the inferiority of stratified care in the trial data, Table 7.14 shows EQ-5D scores by pain resolution at 12 months obtained from trial data compared with the model population score used for the extrapolation.

**Table 7.14 EQ-5D scores for symptom resolution at 12 months, stratified care vs best usual care**

| | Stratified Care | Best Usual Care | Model Population |
|---|---|---|---|
| Poor Resolved | 0.736 | 0.765 | 0.744 |
| Moderate Resolved | 0.813 | 0.806 | 0.808 |
| Good Resolved | 0.820 | 0.854 | 0.833 |
| Poor Unresolved | 0.364 | 0.50 | 0.447 |
| Moderate Unresolved | 0.508 | 0.624 | 0.596 |
| Good unresolved | 0.437 | 0.72 | 0.671 |

As can be seen, similar to the LBP model, in all but moderate resolved, best usual care had EQ-5D scores higher than stratified care at 12 months. Yet the modelled extrapolation

beginning at 12 months assumes that EQ-5D scores were equal for each state on all model treatments. The consequence of this assumption is that the first extrapolation cycle overestimates stratified care and underestimates best usual care by 0.003 QALYs per cycle, the combined effect of which would be 0.02 QALYs annualised. As this was the pre-specified analysis this was not changed for the base case, instead the impact of using higher EQ-5D scores for best usual care is explored in sensitivity analysis.

### 7.4.2 Secondary analyses for methodological uncertainty and heterogeneity

This component of the results section assesses the impact of analytic choices, by assessing how the inclusion of societal costs, as well as not discounting costs and outcomes, impact on the cost-effectiveness of the approach.

#### 7.4.2.1 Undiscounted analysis

The base case model was rerun with no discounting performed; results are shown in Table 7.15.

**Table 7.15 Stratified care vs best usual care vs usual care, no discounting**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified Care | £5584.13 | 6.12 |
| Best Usual Care | £5263.50 | 6.14 |
| Difference (SC-BUC) | +320.63 | -0.02 |
| Usual Care | £5906.71 | 5.78 |
| Difference (SC-UC) | -322.58 | +0.34 |

Abbreviations: BUC (Best usual care), QALY (Quality-adjusted life year); SC (Stratified care), UC (Usual care)

When discounting is not performed, the incremental cost of stratified care vs best usual care is higher than in the base case (£320.63 Vs £307.83), and produces a slightly lower incremental QALY gain for best usual care (+0.02 Vs +0.03). Accordingly, not discounting, slightly improves the probability that stratified care is cost-effective; at the £20,000/QALY WTP threshold it is 40% likely to be cost-effective, see below.

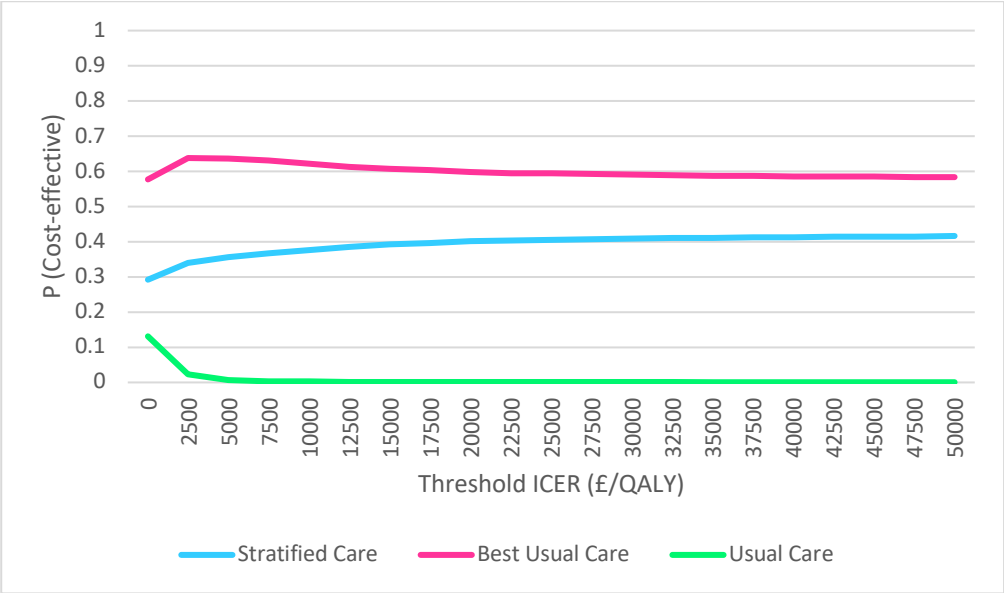



**Figure 7-4 Cost-Effectiveness Acceptability Curve, Stratified care vs best usual care vs usual care, undiscounted**

7.4.2.2 Societal Analysis

The results of the societal analysis, showing QALYs and total societal costs are shown in Table 7.16.

**Table 7.16 Societal analysis for stratified care versus best usual care versus usual care**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified Care | £8555.62 | 5.16 |
| Best Usual Care | £8134.09 | 5.19 |
| Difference (SC-BUC) | +£421.53 | -0.03 |
| Usual Care | £8998.78 | 4.85 |
| Difference (SC-UC) | -443.16 | +0.31 |
| Abbreviations: BUC (Best usual care); SC (Stratified care), UC (Usual care) | | |

The ten-year societal cost of sciatica related expenses on stratified care was £8555.62 per patient with a mean QALY of 5.16 experienced per-patient during that time. Treatment with best usual care was cheaper at £8134.09 and produced 5.19 QALYs. Stratified care costed £421.53 more compared to best usual care, this was more than the incremental cost

in the base case (£307.83), reflecting the additional societal costs accruing to stratified care in the first year, and slight advantage of best usual care over the longer-term.

#### 7.4.2.3 Subgroup analysis

The results of the base case analysis ran separately for each of the SCOPiC sciatica subgroups, estimating QALYs and NHS costs associated with all three comparators, results are shown in Table 7.17.

**Table 7.17 Stratified care vs best usual care vs usual care in each SCOPiC subgroup**

| Group 1 | Mean Cost (£) | Mean QALYs | Cost-effectiveness result |
|---|---|---|---|
| Stratified care | £3568.90 | 5.95 | |
| Best usual care | £3541.93 | 5.84 | |
| Difference (SC-BUC) | +26.97 | +0.11 | £245.18/ QALY |
| Usual Care | £4086.39 | 5.61 | |
| Difference (SC-UC) | £517.49 | +0.24 | Dominant |
| Group 2 | | | |
| Stratified care | £4976.27 | 5.10 | |
| Best usual care | £4655.02 | 5.16 | |
| Difference (SC-BUC) | +£321.25 | -0.06 | Dominated |
| Usual Care | £5228.83 | 4.81 | |
| Difference (SC-UC) | -£252.56 | +0.35 | Dominant |
| Group 3 | | | |
| Stratified care | £5339.22 | 4.70 | |
| Best usual care | £4880.10 | 4.78 | |
| Difference (SC-BUC) | +459.12 | -0.08 | Dominated |
| Usual Care | £5432.28 | 4.39 | |
| Difference (SC-UC) | -93.06 | +0.31 | Dominant |
| Abbreviations: BUC (Best usual care); QALY (Quality-adjusted life year); SC (Stratified care), UC (Usual care) | | | |

Relative to best usual care stratified care provides significant additional benefit to patients in Group 1 (+0.11 QALY), and dis-benefit for patients in Group 2 (-0.06 QALY) and Group 3 (-0.08 QALY). Stratified care is more expensive than best usual care for all patients, but considerably more expensive for patients with the most severe symptoms in Group 3 (+459. 12). Stratified care is cost-effective relative to usual care for Group 1 patients only (£245/QALY), and dominated by best usual care in Group 2 and 3.

The probabilistic simulations show considerable likelihood that for patients in Group 1 stratified care is cost-effective relative to best usual care, as demonstrated in Figure 7.5. For patients in Group 1 at £20,000 per QALY, stratified care is 78% likely to be cost-effective, with best usual care only 22% likely.

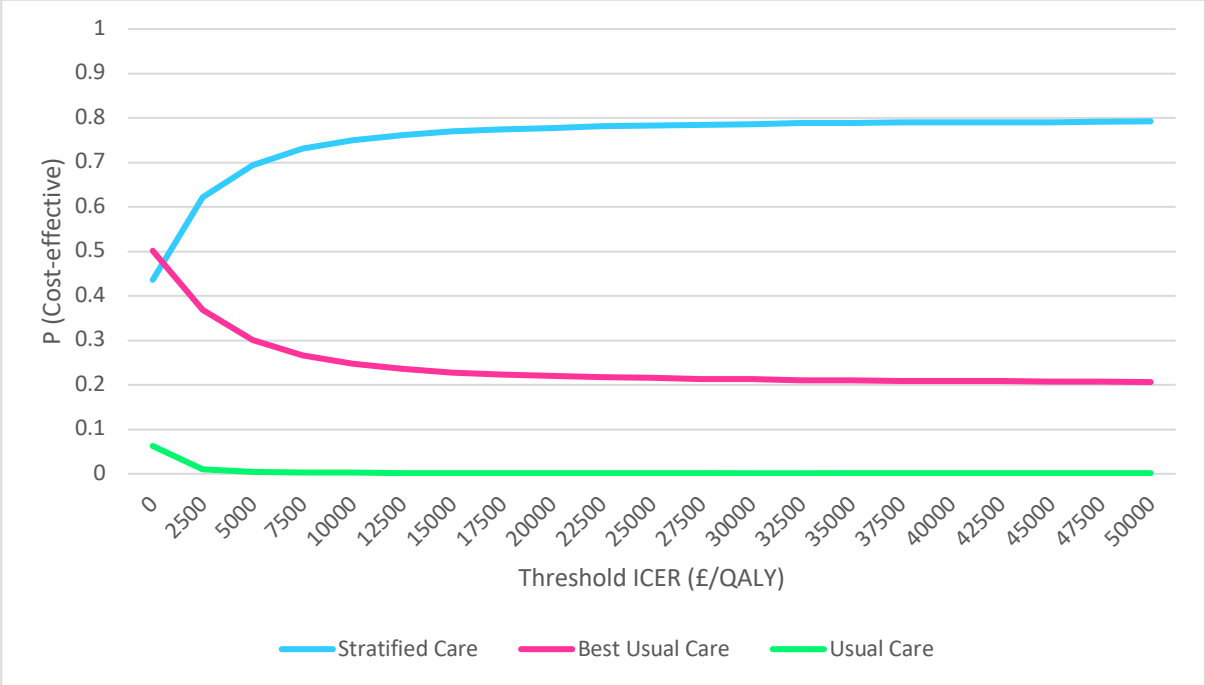



**Figure 7-5 Cost-Effectiveness Acceptability Curve, Stratified care vs best usual care vs usual care, Group 1 patients**

In Table 7.17 stratified care was shown to be £321.25 more expensive for patients in Group 2, providing 0.06 less QALYs than best usual care, which dominates stratified care. The CEAC in Figure 7.6 shows that for patients in Group 2, at the £20,000/QALY WTP

threshold, there is only a 28% chance that stratified care is cost-effective with best usual care most likely (72%) to be cost-effective).

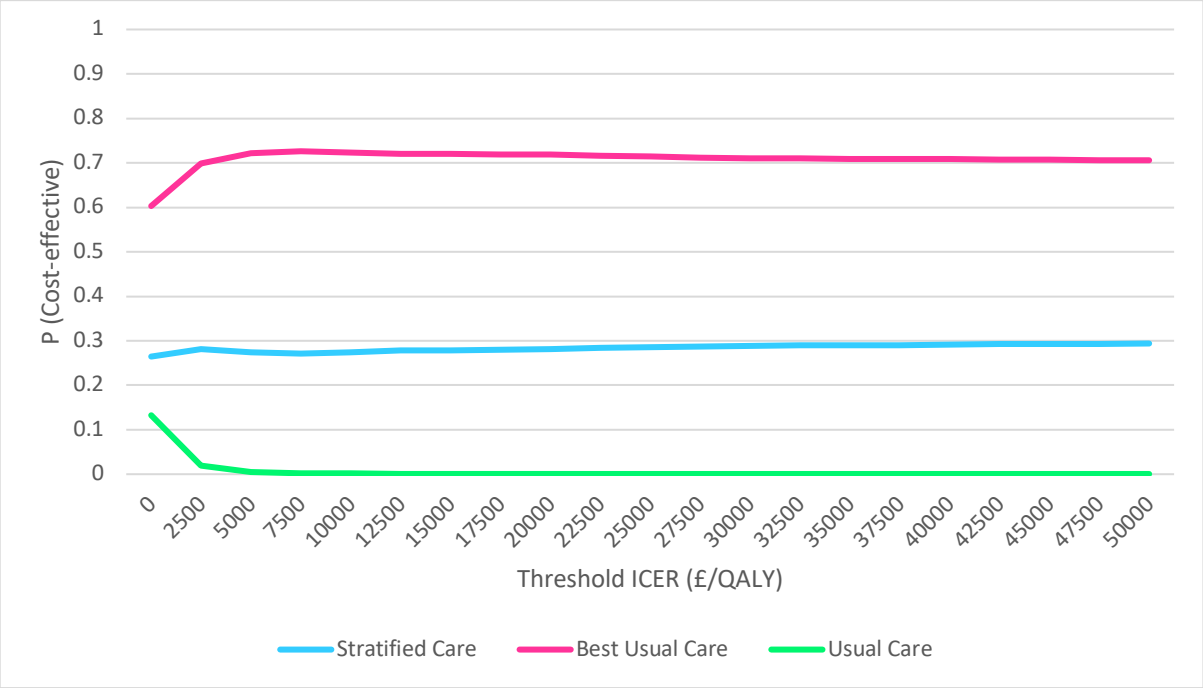



**Figure 7-6 Cost-Effectiveness Acceptability Curve, Stratified care vs best usual care vs usual care, Group 2 patients**

In Table 7.17 it was shown that for patients in Group 3, stratified care was dominated by best usual care, costing £459.12 more and producing 0.08 less QALYs over ten years. Accordingly, as Figure 7.7 shows, at the WTP threshold of £20,000, there is only a 25% chance that stratified care is cost-effective for patients in poor function with best usual care more likely (75%) to be cost-effective).

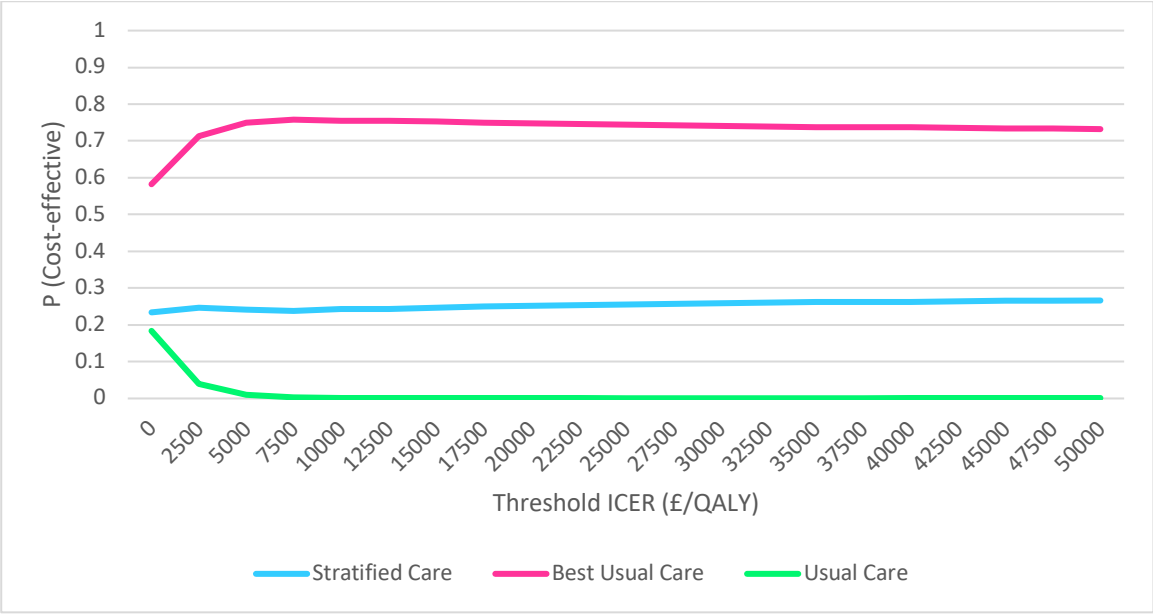



**Figure 7-7 Cost-Effectiveness Acceptability Curve, Stratified care vs best usual care vs usual care, Group 3 patients**

### 7.4.3 Structural uncertainty

This section assesses the impact upon cost-effectiveness of differing assumptions regarding the long-term costs of stratified care, assumptions over long-term EQ-5D values, and running the model using function states instead of symptom resolution beyond one-year. Whilst there is some overlap with parameter uncertainty these are included here as structural uncertainties as they are based upon long-term structural assumptions.

#### 7.4.3.1 Temporal uncertainty over treatment cost

Table 7.18 shows the impact upon the cost-effectiveness of different assumptions regarding the long-term treatment impact on the costs of treatment. Note, the cost of surgery is not changed in these analyses. The table also presents the probability that stratified care is cost-effective at a £20,000/QALY WTP threshold.

**Table 7.18 Cost-effectiveness of stratified care versus best usual care, in different cost scenarios**

| Cost-variation | Per patient cost, SC | Incremental Cost, SC Vs BUC | Likely cost-effectiveness SC |
|---|---|---|---|
| 10% lower costs on SC | £4529.96 | +£55.95 | 41% |
| 5% lower costs on SC | £4630.35 | +156.34 | 39% |
| 2% lower costs on SC | £4684.89 | +£210.88 | 38% |
| Base case | £4781.84 | +307.83 | 37% |
| Abbreviations: BUC (Best usual care); QALY (Quality adjusted life year); SC (Stratified Care) | | | |

It can be seen that whilst the incremental costs of treatment change considerably, falling to £55.95 in the 10% lower cost scenario, stratified care is always more expensive, and the likely cost-effectiveness of the approach is relatively unchanged by increased potential cost savings.

7.4.3.2 Alternate utility values

Using the initial assumption that EQ-5D values would be equal on stratified care versus usual care in the base case, was inconsistent with findings from the EQ-5D scores at 12 months in the SCOPiC trial. Results shown in Table 7.19 suggest what happens if the higher EQ-5D scores achieved on best usual care were maintained for another 1, 2, 5 and 10 years. Only stratified care and best usual care are shown here, since best usual care is more cost-effective than usual care even using the conservative base case assumptions regarding EQ-5D.

**Table 7.19 Stratified care vs best usual care, trial EQ-5D scores beyond 12 months**

| Duration | Incremental QALY SC | Likely cost-effectiveness SC |
|---|---|---|
| 10 years | +0.16 | 9% |
| 5 years | +0.09 | 24% |
| 2 years | +0.06 | 30% |
| 1 year | +0.04 | 33% |
| Base case | +0.03 | 37% |
| Abbreviations: QALY (Quality adjusted life year); SC (Stratified Care) | | |

Under the assumption of higher EQ-5D scores for best usual care, the likely cost-effectiveness of stratified care falls for each additional year the assumption is held. As Table 7.14 suggests these results arise because best usual care patients had a mean EQ-5D score 0.03 higher than stratified care patients at twelve months, despite stratified care patients having marginally more favourable outcomes according to time to symptoms resolution.

#### 7.4.3.3 Use of function states instead of pain

Where patient's transition into states of function instead of resolution defined by pain state at twelve months, stratified care is even less favourable than in the base case, and therefore still dominated by best usual care. Results are shown below in Table 7.20.

**Table 7.20 Stratified care Vs best usual care, function used as a state.**

| | Mean Cost (£) | Mean QALYs |
|---|---|---|
| Stratified Care | £5380.39 | 5.25 |
| Best Usual Care | £4960.28 | 5.32 |
| Difference (SC-BUC) | +£420.11 | -0.07 |
| Abbreviations: BUC (Best usual care); QALY (Quality adjusted life year); SC (Stratified Care) | | |

The use of function states alters the likely cost-effectiveness implications, with stratified care now costing more and producing less QALYs than in the base case, and only 27% likely to be cost-effective at the £20,000/QALY WTP threshold.

### 7.4.4 Internal and external validity

The internal validity of the model was checked by a series of logic checks, including dummy runs of all model parameters. A patient cohort was also simulated in Excel in order to check the validity of the model, results showed no more than 1% difference in resolution at both 5 and 10 years against the output in TreeAge.

In order to assess the external validity of the model, outcomes were compared at five years, with that of patients in the BeBack cohort study, shown in Table 7.21

**Table 7.21 Model output vs cohort study observations.**

| | SC | BUC | UC | BeBack |
|---|---|---|---|---|
| Symptom resolution in base case | 84.7% | 84% | 79.1% | 76.2% |
| Abbreviations: BUC (Best usual care); SC (Stratified Care); UC (Best Usual care) | | | | |

It can be seen that patients in usual care performed slightly better than the BeBack cohort at five years. The higher proportions in SC and BUC reflect the significantly improved symptom resolution achieved in SCOPiC trial relative to the BeBack cohort study.

### 7.4.5 Value of Information analysis

A value of information analysis was performed for uncertainty relating to all parameterised components of the model. Results in Table 7.22 show individual level EVPI, £343.16 per patient, which can be extrapolated to a population using the consultation population expected to benefit (3,885,026), this is shown in the right-hand column at £1,333,185,522.

**Table 7.22 Per-person and population Value of information**

| Scenario | Per Person EVPI at £20,000 WTP threshold (£) | Population EVPI at the £20,000 WTP threshold (£) |
|---|---|---|
| Base case | 343.16 | 1,333,185,522 |
| Abbreviations; EVPI (Expected value of perfect information); WTP (Willingness to pay) | | |

As we can see from Table 7.23, there are many parameters which can be considered to have mae minor contributions to the ovveral decision uncertainty. However, those parameters which make the largest contributions are predominantly transition probabilities from 4 to 12 months for stratified and best usual care.

**Table 7.23 Single parameter EVPPI, Per Person and population**

| Parameters | Annual Per Person EVPPI at £20,000 WTP threshold | Approximate Standard Error | Ten-year population EVPPI at the £20,000 WTP threshold |
|---|---|---|---|
| TP: Stratified care, H1 | £0.05 | 0.25 | £179,200 |
| TP: Stratified care, poor function to resolved, months 4-12 | £13.64 | 3.15 | £52,980,000 |
| TP: Stratified care, poor function to unresolved, months 4-12 | £0.93 | 0.97 | £3,617,000 |
| TP: Stratified care, moderate function to resolved, months 4-12 | £0.86 | 1.24 | £3,353,000 |
| TP: Best usual care, poor function to unresolved, months 4-12 | £11.31 | 2.63 | £43,930,000 |
| TP: Best usual care, moderate function to unresolved, months 4-12 | £0.76 | 0.77 | £2,941,000 |

| | | | |
|---|---|---|---|
| TP: Best usual care, good function to unresolved, months 4-12 | £8.32 | 1.67 | £32, 320,000 |
| TP: BUC, good resolved to resolved, 12 months | £0.06 | 0.11 | £237,700 |
| TP: BUC, moderate resolved to resolved | £0.53 | 0.30 | £2,054,000 |
| TP: BUC, poor function, resolved to resolved | £0.16 | 0.19 | £632, 200 |
| Utility, poor function, resolved, years 2-10 | £0.01 | 0.14 | £23,160 |
| Utility, good function, unresolved, years 2-10 | £0.08 | 0.34 | £323,200 |
| Costs, BUC, year 1, poor function resolved | £3.13 | 0.98 | £12.160,000 |
| Costs, BUC, year 1, poor function unresolved | £0.04 | 0.12 | £148, 900 |
| Costs, BUC, year 1, moderate function resolved | £2.61 | 0.88 | £10,130,000 |
| Costs, usual care, poor resolved | £0.02 | 0.08 | £62,190 |

*Parameters with zero EVPPI are not shown

Abbreviations; BUC EVPPI (Expected value of partial perfect information); SC (Stratified Care); TP (Transition probabilities); UC (Usual care); WTP (Willingness to pay)

When these parameters are grouped, in Table 7.24, it becomes evident just how much the transition probabilities contribute to the decision uncertainty.

**Table 7.24 EVPPI parameter groups, per person and population**

| Parameter(s) | Annual Per Person EVPPI at £20,000 WTP threshold | Approximate Standard Error | Population EVPPI at the £20,000 WTP threshold |
|---|---|---|---|
| Transition probabilities first year | £266.87 | 7.55 | £1,036,807,900 |
| Transition probabilities year 2-10 | £0.00 | 0.00 | 0.00 |
| All Utilities | £55.99 | 9.98 | £217,527,002 |
| All Costs | £12.22 | 3.47 | £47,467,204 |
| Abbreivations; EVPPI (Expected value of partial perfect information); WTP (Willingness to pay) | | | |

## 7.5 Discussion

### 7.5.1 Principal findings

The body of work contained within this chapter reflects the design and realisation of an individual simulation model to perform a cost-effectiveness analysis of the potential application of stratified care for the management of sciatica in primary care. The simulations performed for the base case analysis, from the NHS perspective, containing the assumption of no additive treatment from stratified care beyond one-year, equivalent EQ-5D scores for each state beyond 12-months, showed that the stratified care intervention is not cost-effective relative to best usual care, although dominant relative to usual care.

However, the relative uncertainty of the result must be taken within the context of the results of the subsidiary and sensitivity analyses, namely that, as shown in Table 7.1, patients on best usual care have higher EQ-5D scores within the same states compared with stratified care patients. Currently, there are no means of determining how long the EQ-5D benefits of best usual care could last (or indeed if they are statistically significant). Yet, as Table 7.19 shows, even assuming they last two years pushes the likely cost-effectiveness even further towards best usual care. Table 7.19 provides a clear indication of the sensitivity of model results towards this structural assumption, a vivid illustration of the concerns raised in Chapter 4 about assuming equivalent EQ-5D scores for identical states across different treatments. This highlights it is imperative to explore the implications of different structural assumptions in future modelled analyses in sciatica.

A similar issue arises when function states are used instead of symptom resolution states, showing improved QALY gain, higher incremental cost savings, and higher likely cost-effectiveness for best usual care relative to stratified care. This result arises from the fact that at twelve months, patients were marginally more likely to be in favourable states of function on best usual care vs stratified care compared with symptom resolution by pain. Moreover, this analysis does not include the slightly more favourable EQ-5D scores achieved on best usual care vs stratified care found in the SCOPiC trial. Taken together, the base result, results from the EQ-5D analysis, and the use of function as a state, suggest that it is highly likely that stratified care is not cost-effective relative to best usual care.

Interestingly, the results from the subgroup analysis demonstrated that stratified care produces beneficial outcomes and saves significant amounts of money for patients in good function, but is more costly and less effective for patients in poor function. This result suggests that the approach seems effective for patients who were in better condition, but

the overall cost-effectiveness of this stratified care model is compromised by the care pathway for patients in poor function.

In relation to the results achieved on stratified care and best usual care versus usual care, it is clear that both best usual care and stratified care were superior to results obtained by patients in the usual care cohort. The significance of the difference between these results is interesting, however difficult to interpret. It may be that despite usual care patients being younger and having lower levels of disability and pain, the higher proportion of patients with longer symptom duration in BeBack may have contributed towards this differential. It should also be noted that the BeBack cohort study was from 2006; expert consultations suggested that since then, care for sciatica patients has changed to some degree with GPs referring sooner, and directing more patients for treatment such as physiotherapy, and to specialist services. Since 2006, physiotherapists also have improved in terms of knowledge and training regarding the biopsychological model on which stratified care leans, and it is plausible that the lack of impact of stratified care found in the trial and this secondary analysis reflect overall underlying improvements in treating patients rather than treatment protocols related to the SCOPiC trial. As mentioned above, the SCOPiC stratified care model for patients in poor function seemed to drive the overall lack of cost-effectiveness of the approach. This is possibly due to the fact that the stratification algorithm used to identify patients for fast-track to MRI and spinal specialist assessment, was not adequately specific or discriminant, and although patients in Group 3 had the higher levels of pain and disability, not all of them needed to be fast-tracked to specialists.

Nonetheless, the results obtained here possibly lend some support to the idea that providing some good quality of care (mainly conservative management) early on in the presentation, over and above what might be expected in true usual care, whether it be stratified care or best usual care, can improve patient outcomes in the short term, and potentially result in

partial repayment of initial outlay across time if the improvements in symptoms are maintained.

### 7.5.2 Strengths and weaknesses

There are three major strengths of this analysis. Firstly, it is the first decision model produced for stratified care in sciatica; in that regard it is a novel development. Here by importing the simulation modelling techniques used in osteoporosis stratification modelling, the model offers a different approach to the representation of patient stratification than that of the LBP model in Chapter 6. This allows the model to track patient function at baseline and experience of surgery, and subsequent transitions dependent upon those characteristics. This was especially useful when tailoring different configurations of patient function for the subgroup analyses on the SCOPiC sciatica groups. Moreover, on the face of it, the model itself has a very simple structure, although the underling logic and commands are more complex, it is therefore intuitively straightforward to interpret and explain to others, and prevents the number of Markov states becoming unmanageable.

Second, and related, whilst not finding solutions to all of the methodological problems identified in chapter 4, the analysis has demonstrated how important structural and temporal assumptions over EQ-5D scores across time and choice of health state can be for the likely cost-effectiveness of comparisons between interventions with similar outcomes at twelve months.

Third, as identified in the systematic review in chapter 4, there is a hesitance to produce modelled cost-effectiveness analyses over an extended time horizon. By meeting with clinicians and other experts, this analysis was able to make assumptions about long-term treatment outcomes being dependent upon resolution by NRS pain scores at twelve months and five years.

The study has some weaknesses. Whilst the PSA samples transitions of patients between states of resolution beyond one year, and it is in that sense probabilistic, as was the case in the LBP model, the assigning of equivalent transitions to patients in all three treatments does not capture the "true" uncertainty over long-term transitions. The "true" probabilistic uncertainty over the long-term treatment effect, therefore, could only be accounted for by attaching a probability distribution to the differential transitions of patients on stratified care versus best usual care versus usual care over time. This information is currently unknown, and in this case the impact of having best usual care achieve better outcomes for patients, was addressed by looking at improved EQ-5D scores for patients on best usual care at various time points. The same is true for the uncertainty over long-term EQ-5D scores, and costs, not being parameterised between the comparators. These could have been assessed more formally and empirically parameterised using more sophisticated methods (see Bojke et al. 2009 for a review), such as model averaging and parameterisation, and whilst lying outside the scope of this work, could be a future endeavour in modelling in both areas.

In relation to the cost-effectiveness result in the usual care comparator, there were more patients in the BeBack sample with symptoms of longer duration, which was not controlled for in the analysis of transitions or EQ-5D scores. However, by using EQ-5D scores controlling for baseline EQ-5D, and applying a model population standardised by baseline function category, it was hoped to control for important differences between cohorts.

In relation to data, the transition probabilities beyond one-year were only available at five years, and as a consequence this raises two issues. Firstly, the assumption of linear transitions between twelve months and five years in unrealistic and does not accurately reflect the likely fluctuations in function experienced by patients, although at least impacts each treatment in the same manner. Moreover, assuming equivalent transitions from years

five to ten are similar to one to five years could be unrealistic. As could the assumption of stable EQ-5D scores of stability of time.

Validity checks showed the model predicts higher symptom resolution that would be expected from the BeBack data. This possibly reflects the fact that surgical patients are placed into symptom resolution at twelve months, and is a possible limitation of the model. Further analysis of outcomes for surgical patients could help improve this issue.

Moreover, the superior results achieved for stratified care and best usual care in the model could reflect the superior improvements experienced by patients in the SCOPiC trial. However, it is possible that this result may require a tempering of the base case assumption that patients who already improved significantly during the first year on the SCOPiC trial, would continue to improve at the same rate as patients in the BeBack sample over time, given those patients in BeBack hadn't experienced that same initial improvement.

There were a number of issues regarding costings. Resource use is taken from the first year of the BeBack cohort study, and therefore may over-represent likely long-term costs. This study was also undertaken in 2006 and possibly does not represent the nature of the treatment pathway today. As previously noted, that all healthcare resource use came from self-report can be considered a limitation, on the grounds of recall bias (Petrou et al. 2002). Additionally, there is a significant difference in the number of days off work experienced by patients in the SCOPiC trial, and patients in the BeBack cohort study, this could reflect underlying changes in work absence in this population.

### 7.5.3 Implications for researchers, clinicians and policymakers

The findings in these analyses provide strong evidence that either stratified care or best usual care, is almost certainly likely to be cost-effective relative to usual care for the management of sciatica. Lacking the evidence to determine whether or not, stratified care improves symptoms, versus best usual care, or whether best usual care improves EQ-5D

scores versus stratified care, future research could provide benefits given the potential cost savings likely to follow from a cost-effective management approach for sciatica.

The results of the subgroup analysis show how the more costly, and less effective outcomes for patients in poor function, compromise the cost-effectiveness of the approach, and perhaps this provides evidence that researchers ought to target their future work towards better identification of those sciatica patients that truly need to see a specialist early on, and also towards improving the interventions for patients with greater disability.

The value of information analysis showed that removing the uncertainty detailed in this study could hold significant value. This very high value reflects not only the degree of uncertainty over which is the best treatment, but the large potential net benefit of identifying timely and effective treatments for sciatica. In terms of interpretation of this result it is to be concluded that there is undoubtedly value in identifying cost-effective treatment approaches for this patient population going forward. However, whether or not further investigation of stratified care (or different models of stratified care) vs best usual care is useful is likely to be a clinical judgment reflecting whether or not clinicians believe that this approach has value in this patient population. Subgroup results in this analysis do at least help by suggesting where this approach provided clear benefit, e.g. SCOPiC Group 1.

Results of single parameter and parameter groupings can help suggest further avenues for research, namely collection of data on the first year transition probabilities on stratified and best usual care. However, further logical consideration must be given to the nature of the model structure, which assumes equivalent transitions from years 2-10. The EVPPI output from such a structure will not adequately capture the value of information associated with identifying these longer-term parameters, with first-year transition probabilities acting as a catch-all for the uncertainty associated with all transitions across the model horizon.

Therefore, research on the likely long-term treatment effect would also help reduce the uncertainty in the model, as well as being able to differentiate accurately between sources of uncertainty.

## 7.6 Conclusion

This chapter presented the design and analysis of a model-based cost-effectiveness analysis of stratified care versus best usual care versus usual care from the healthcare perspective, in patients consulting in primary care with sciatica. The analyses conclude that stratified care is not cost-effective, although providing patients with care above or beyond what they might reasonably expect in usual care, best available care is likely to be cost-effective over the long-term. Sensitivity analyses reveal, once again, that certain structural uncertainties impact upon the model results.

# Chapter 8: DISCUSSION AND RECOMMENDATIONS

This chapter comprises a discussion of the overall findings and their relevance as a valuable source of information to policy makers, as well as wider implications for decision modelling in low back pain and sciatica.

## 8.1 Context

The longer-term cost effectiveness of the Keele stratified care model was unknown in both LBP and sciatica. The ultimate objective of the thesis was the production of long-term cost-effectiveness estimates of stratified care approaches in both conditions, by extrapolating beyond trial results. In order to provide these estimates, in addition to data analysis, a process of learning was undertaken to consider how modellers had approached modelling in both conditions as well as representing stratified care pathways in models. As a consequence, it is possible to comment upon methodological issues relating to modelling in these conditions, as well as the specific cost-effectiveness implications of the two treatment approaches. Therefore, this thesis concludes by offering recommendations on decision analytic modelling in both conditions.

The purpose of this chapter is to unite the methodological and empirical work within this thesis. The first section provides an overview of key findings from the thesis. The next section then considers the overall strengths and limitations of the thesis. Third, the findings for policy and future research are presented. The chapter concludes with potential modelling guidance in these conditions.

## 8.2 Research objectives in this thesis

The following section answers four of the five specific research objectives detailed in 1.8.1. The final objective concerning guidance is explored in Section 1.3

### 8.2.1 What are the lessons to be learnt regarding current modelling approaches in low back pain and sciatica

Both decision models designed for this thesis are de-novo models and include modelling methodologies unique to this body of work. The LBP model takes a unique model structure, using patient functional status achieved at twelve months in order to extrapolate trial results over ten years. Stratification in this model is reflected by the use of three separate Markov models for each risk subgroup. The sciatica model uses an individual sampling model, a novel development within this condition, and suggested to be the most appropriate means to model sciatica given that outcomes are dependent upon patient characteristics, and surgery is an infrequent but nonetheless available treatment for this population.

The cost-effectiveness result amongst high-risk patients in the LBP model highlights the importance of modelled analyses in both conditions. In the modelled analysis, stratification dominated usual care over ten-years, yet in the one-year trial was more expensive than usual care. The problem with one-year trial analyses is they do not capture the potential long-term benefits and cost savings, in this case treating the high-risk patients with initially costly stratified care with cost savings expected later.

Given the value in the production of high-quality economic evidence based upon decision modelling it makes it all the more surprising that that existing model-based economic evidence is either non-existent or of poor quality (Chapter 4, Section 4.4; Hall et al. 2019). The review identified critical flaws in existing models of interventions for LBP and sciatica. Issues related to studies not modelling across adequate time horizons, inappropriate use of utility data, calculation errors, a lack of transparency regarding methodologies, and the failure to consider the extent to which uncertainty and assumptions

limit the applicability and generalisability of the results. It is the argument of this thesis, that all these issues can be addressed, through rigor and adherence to modelling guidance.

Extrapolation using decision analytic modelling ought to now be considered a research imperative, given that many treatments for both LBP and sciatica now aim at a) encouraging the patient to self-manage, and b) treat patients early and appropriately, in the hope of better outcomes and overall improvement in prognosis, and/or saving future treatment costs. This should begin at the trial design stage, so thought ought to go into a potential modelling strategy, which preserves the integrity and reporting transparency of the working methodologies.

Ultimately, health economists and modellers developing models in both conditions need to be more willing to explore the implications of extrapolation of treatment effect over an appropriate time horizon. Guidance is available on how to capture the associated uncertainty relating to extrapolation of unobserved treatment parameters in sensitivity analyses. The National Institute for Health and Care Excellence (NICE, 2013) methods guidance advocate scenario analyses with (1) nil treatment effect over the unobserved period; (2) treatment effect during the unobserved period is set equal to the observed period; and (3) treatment effect diminishes over time. The analysis presented in this thesis used a variant on (1), with sensitivity analyses (see Chapter 6, section 6.4.4) attempting to represent both (2) and (3).

Yet, notably, none of the existing studies in the systematic review using extrapolation undertook sensitivity analyses with sufficient rigour to capture the uncertainty over the long-term treatment pathway that their assumptions demanded. Both models in this thesis, attempt to extrapolate using data in conjunction with expert input as regards assumptions, and capture, to the best degree possible, the cost-effectiveness implications of different extrapolation assumptions. The results of these analyses show the importance of

undertaking these sensitivity analyses, given the impact that different structural and temporal assumptions can have on the cost-effectiveness of different approaches.

Moreover, further research needs to explore the implications of using different structural assumption, in modelled analysis of both conditions. In both models, whilst it did not affect the overall cost-effectiveness result, the choice of health state significantly affected the degree of cost-effectiveness decision. In studies with closer cost-effectiveness result, this could be extremely significant.

In relation to the question of which health state ought to be selected, established guidance (Philips et al. 2006), states that disease states should reflect the underlying biological disease process and intervention impact. In this case, using function, symptom resolution, or pain, as health state, could all fall within the scope of this guidance. An analysis could choose a health state classification in order to improve the cost-effectiveness of the intervention, or conversely a team may not capture the full benefits or cost savings accruing to their intervention, both scenarios are equivalently likely to lead to incorrect adoption decisions. In order to minimise the impact of uncertainty over appropriate health state, analyses should be performed using different health states, it is suggested here that function and pain (possibly dichotomised to indicate symptom resolution) be used as health states in LBP or sciatica models.

Moreover, calibration of utility values should be undertaken in order to avoid a scenario (as in Chapter 6, section 6.4.4.9) where patients in each of the states of pain, had higher EQ-5D scores on usual care than on stratified care. This situation also arose in the sciatica model, where EQ-5D scores for “success” or “resolution” were set as equal across the length of the extrapolated analysis, despite evidence that this is not the case at twelve months, this is likely to cause bias in the model results. The solution, as performed in 7.4.3.2, was to run the model for alternate assumptions regarding EQ-5D score

differentials. With regard to the otherwise excellent Lewis et al. (2011) model, and various other models that used similar methodology, the evidence presented here suggests that assuming equivalent EQ-5D scores for a "success" or "failure" on different treatments could compromise the overall cost-effectiveness result.

Future models should endeavour to pay attention to the methodological challenges raised in Chapter 4, and Chapters 6-8, and summarised in Hall et al. 2019, to ultimately help advance this field, and enable more useful comparisons between treatments, and a better standard of cost-effectiveness evidence. Until modellers produce more high-quality modelling studies, consistent with modelling guidelines, the standard of discourse necessary to stimulate methodological improvements in these areas, is likely to be restricted.

Finally, a note on extrapolation, the fundamental principles underpinning the extrapolation used in this analysis, were reflective of the work on trajectories (Dunn et al. 2013). Whilst the statistical methodologies differ, the principle of extrapolation used in this thesis reflects the idea that groups of patients have trajectories, which are stable over time, and therefore it is reasonable to perform linear extrapolation between observed periods. Even where a longer-term dataset is not available, it may be possible to use information from the trajectories literature, to perform extrapolation.

### 8.2.2 What are the lessons to be learnt regarding current modelling approaches to stratified care?

The stratification in these analyses were modelled in accordance with how modellers had constructed their models previously, presented in the review in Chapter 5. Using different model forms, both analyses presented here demonstrated how a Markov state-transition, or an individual sampling model, could adequately carry out this form of analysis.

Clearly, the principal lesson to be learnt from the data analysis performed in this thesis relates to the problem of data. When deriving parameters for a model-based analysis, having the stratification presents another means by which the statistical power of the estimates is reduced. In the LBP model, the BaRNS cohort study had sufficient data with which to derive the longer-term transitions as dependent upon patient function, although not risk group. Patient transitions ultimately became a function of the composition of patient function states at twelve-month, with results in each risk-group reflective of patient function within that risk-group at twelve months. Similarly, in the sciatica model, patient's subdivision could only account for function and symptom resolution at twelve months, given the limitations of data in the BeBack study. This somewhat diminishes the value of using individual simulation modelling in this thesis, and also compromises the accuracy of the sub-group analyses.

### 8.2.3 Is stratified care for low back pain likely to be cost-effective?

A de-novo state transition model was constructed to perform a cost-effectiveness analysis of the potential application of a stratified care model (STarT Back approach) for the management of LBP. The Monte Carlo simulations performed for the base case analysis, from the NHS perspective, including the assumption of no additive treatment effect from stratified care beyond one-year, showed that the intervention is very likely to be cost-effective and cost-saving, on average, with cost-effectiveness result robust to sensitivity analyses.

There are some caveats about the effectiveness of this stratified care model for LBP. Notably, the clinical results have not been replicated by others, in other areas/countries (e.g. Cherkin et al. 2018; Morso et al. 2018), and therefore there are questions about the generalisability of these results beyond the UK, or indeed beyond Staffordshire and surrounding regions. This model could provide a theoretical platform for authors in those

countries researching this model of care, to perform their own analyses, as there may be reasons to assume that there could be future cost savings accruing to stratified care patients that are not captured within their trial-based analyses to-date. If the results of the trials provide non-inferiority but are likely to lead to short-term and long-term cost savings, then modelling can be undertaken and may show that on balance it is likely that stratified care could be cost-effective even where it is not likely to lead to clearly superior clinical outcomes.

### 8.2.4 Is stratified care for sciatica likely to be cost-effective?

The base case result for the de-novo individual sampling model performed for sciatica was unambiguous about the likely cost-effectiveness of either of the trial management options relative to usual care assumed in the cohort study. However, stratified care was not cost-effective in comparison to best usual care. Results could be used to hypothesise that the provision of effective early care, received either in the stratified care arm of the trial, as well as best usual care, that are likely to be cost-effective. Although, the analysis may partially be reflecting underlying changes in the sciatica treatment pathway over the past 15 years.

The results in the sub-groups were informative, given that stratified care provided strong benefits and cost-savings for patients in good function, but inferior and more costly outcomes for patients in poor function, providing some evidence to motivate an improved future model of care. Clearly there are dimensions to the stratified care model which can be cost-effective, and the evidence suggests it should be considered a work in progress.

## 8.3 Guidelines for modelling in both conditions

In what follows, the implications of this work are considered with respect to future modelling endeavours.

- A clear need for those conducting economic evaluations to consider the use of decision modelling

Given the evolution of the understanding of back pain as a chronic condition, with frequent recurrence over time, there is a clear need for economic evaluations to consider the use of decision modelling to capture the long-term impacts of treatments for both conditions, a need made more urgent by the rising socio-economic burden of both conditions.

- More engagement with economic modelling to strengthen methodologies.

This thesis has identified, and attempted to resolve the methodological issues related to decision modelling in both conditions. However, the answers provided here are intended to be the beginning of conversations, there will need to be more engagement with economic modelling to strengthen methodologies.

- Suitable, high quality data is needed to underpin the modelling process.

As is evident, there are shortages of data with which to perform the extrapolation. The data used here for extrapolation of transition probabilities, was of high quality, but was low on statistical power, which contributes to substantial decision uncertainty. Moreover, costs were taken from older cost sources, and quality of life values used in the extrapolation were derived at 12 month follow-up. Researchers need to produce high-quality research on patient function over time, but also quality of life, and healthcare costs. Direct evidence on the impact of stratified care would improve the validity of the modelling. Considering the data needs of decision models at the trial-design stage could be of real utility.

- Health economists and modellers developing models in both conditions need to be more willing to explore the implications of extrapolation of treatment effect over an

appropriate time horizon. In expert consultations, the consensus appeared to be extrapolation over ten-years should be the absolute minimum time horizon in both conditions.

This is because extrapolation better captures the extent of longer-term benefits and cost-savings associated with treatments, and modelling could plausibly alter the cost-effectiveness implications of treatments in certain scenarios. This is an imperative, given the proliferation of treatment approaches in both LBP and sciatica, which have the explicit aim of treating early to save money and improve patient outcomes further down the line. It is also likely that encouraging patients to self-manage in the longer-term, will potentially lower treatment costs – again the full extent of the potential savings accruing to the health service are not likely to be captured within a twelve-month evaluation. As described in Chapter 4, methods exist to handle temporal parameter uncertainty.

- Where data is limited in engaging in extrapolation, modellers should follow as a minimum the NICE guidance.

Modellers should ensure their base case and sensitivity analyses cover the three scenarios advised by NICE (2013); (1) nil treatment effect over the unobserved period; (2) treatment effect during the unobserved period is set equal to the observed period; and (3) treatment effect diminishes over time. This thesis provides a starting point for how this could be done in these conditions (Chapter 6, section 6.4.4.3 – Temporal uncertainty over long-term treatment effect).

- Parameter and structural uncertainty need to be addressed in accordance with best practice.

There was under representation of probabilistic analyses in the systematic review (Hall et al. 2019). Whilst the sensitivity analyses in these studies account for some of the

uncertainty in model parameters, studies rarely undertook probabilistic analyses or the rigorous sensitivity analyses required to capture the uncertainty over the long-term treatment pathway that their assumptions demanded. The two modelled analyses in this thesis demonstrate how sensitive the cost-effectiveness outcomes were to assumptions regarding long-term patients' outcomes, EQ-5D values, and choice of health state. It could be argued that this thesis would have benefited from inclusion of a further sensitivity analysis to review the impact of alternative cut-offs for function in the LBP model, and resolution in the sciatica model. Modellers in both conditions should, as a minimum follow the Philips et al. (2004) checklists related to sensitivity analysis.

- Research needs to explore the implications of using different health states in both conditions.

In both modelled analyses it was shown that changing the health states used in the analyses altered the cost-effectiveness results, and whilst not changed the implied adoption decision in both models, given the degree of sensitivity could likely influence the cost-effectiveness implications where the comparators were more similar in cost-effectiveness. In order to avoid incorrect adoption decisions, two health states ought to be modelled as a minimum. In a related matter, work in both conditions should continue to advance the dialogue within this thesis related to consideration of what the most appropriate health states might be it is suggested function and pain are appropriate.

- The importance of the calibration of utility values.

As was shown by the systematic review, it was common to use identical EQ-5D scores for various states on both treatments, over the long-term. The significance of this issue was highlighted in the sciatica analysis, where altering the EQ-5D scores for each state, significantly changed the QALY gain. The implication of alternative assumptions should be modelled in sensitivity analysis.

- Future modelling studies should pay attention to the methodological challenges raised in Chapter 4, and Chapters 6-8, and summarised in (Hall et al. 2019.

Strengthening methodologies will ultimately enable a better standard of cost-effectiveness evidence to compare between treatments. The potential value of improving research in this area was evidence by the value of information analyses indicating the potential benefit of further information is extremely high.

- EVPPI should be produced to guide future research.

Given the need for better quality of economic evidence in this field, it is imperative that data needs are continually highlighted; EVPPI is a key tool in highlighting where to guide future research. Yet none of the models included in the review included EVPPI, with only one including an EVPI estimate.

## 8.4 Strengths and limitations of the research

Taken as a body of work, this thesis has provided a high standard of cost-effectiveness evidence on the long-term cost-effectiveness of stratified management options for both conditions. However, having engaged with extensive issues related to the modelling of treatments and management approaches, this thesis has provided far more than that.

Referring to the issues raised in the review, the work has shown with some fairly simple assumptions, that models in both conditions can extrapolate trial results over ten years, and the extensive body of sensitivity analyses provides suggestions as to a means of quantifying the uncertainty around those assumptions. The inputs used in the models, therefore not only represent suggestions for methods that could be employed in order to undertake extrapolation, but also provide estimates that could be used in alternate models to undertake similar approaches. The presentation of a range of reasonably powered and appropriate utility values especially, could also aid decision modelling in this field.

The work has identified the key points of sensitivity in models addressing both conditions. Namely, EQ-5D scores on treatments, assumptions around long-term patient outcomes, and choice of health states. EQ-5D scores and long-term patient outcomes can be resolved by improved data availability. However, debates around how to represent the condition in the form of health states, as the sensitivity analysis results have shown is not inconsequential. Selection of health states for these models could perhaps be an empirical question, perhaps reflecting whether pain or function do better predict long-term patient outcomes. Whilst not issuing a definitive solution, engaging with these debates is a key strength of this thesis.

Clearly, the absence of data on long-term outcomes and healthcare costs, limit the applicability of the study results. It is strongly advised that readers interpret base case results in conjunction with the sensitivity analysis, and not rely solely upon the base case estimates, which lean heavily upon assumptions and involve possibly problematic data. Findings must also be interpreted with reference to the data limitations considered within each of the modelling chapter discussions (6.4.2 and 7.5.2). Cost estimates used in the extrapolations are a particular concern, given they are from the first year of a cohort study from 2006, there is reason therefore to consider that these results do not represent current resource usage associated with each condition. However, sensitivity analyses undertaken in both analyses reveal that results are fairly robust to even major changes in assumptions regarding costs. In this case, it is sensible to uphold concerns regarding the estimates, whilst accepting the cost-effectiveness implications of the analyses using these estimates are likely to be robust.

## 8.5 Implications for future research

The results of the EVPI analyses highlight the potential value from further research. Whilst the stratified care approach for LBP is likely to be cost-effective, the high value of future

research arises from the large numbers consulting in this population. The value in sciatica arises from the uncertainty of the cost-effectiveness relative to best usual care. Meanwhile, the EVPPI analyses show that improving upon the data limitations acknowledged in the study regarding long term transition probabilities, could lead to a reduction in overall uncertainty, and can act as a motivation for future research. Other uncertain structural assumptions could be resolved by the facilitation of data collection, e.g. longer-term follow up data for EQ-5D scores and healthcare usage which are powered to detect differences in stratification groups.

In relation to stratified care for sciatica patients, the sub-group analyses produced in this thesis highlight the value of stratified care for patients in good function, but perhaps serve to motivate improvements in stratification approaches for patients experiencing more severe symptoms.

## 8.6 Conclusion

The purpose of this research was to investigate the methodologies that had been used in decision analytic modelling in low back pain and sciatica as well as stratified care, to develop de-novo models to perform cost-effectiveness analyses for stratified care for low back pain, and sciatica.

The modelled analyses show that stratified care for low back pain is likely to be cost-effective, whereas stratified treatment of sciatica is unlikely to be cost-effective. Findings were robust to a number of scenario analyses

However, the main theoretical message of the thesis is that the state of decision modelling in both conditions is still in its infancy, and long-term cost-effectiveness evidence to date, is sparse. This thesis calls for not only more attempts at modelling, but for further discussion of issues considered within this body of work, and lays out the importance of

doing so. The need for the production of better quality data to place into decision models was also discussed, and EVPPI analyses show long-term patient transitions as a priority.

This work, produces guidance and suggestions on how decision modelling might proceed, produces some parameter estimates which may be of use going forward, and highlights the key drivers of uncertainty within these models, the ultimate subsequent aim being the production of high-quality long-term economic evidence in both conditions.

## References

Abdi, S., Datta, S., Trescot, A. M., Schultz, D. M., Adlaka, R., Atluri, S. L., Smith, H. S. and Manchikanti, L. (2007). Epidural steroids in the management of chronic spinal pain: a systematic review. *Pain physician,* 10, pp. 185-212.

Ades, A.E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N. and Lu, G. (2006). Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics,* 24(1), pp. 1–19.

Al-Janabi, H., Flynn, T.N. and Coast, J. (2012). Development of a self-report measure of capability wellbeing for adults: the ICECAP-A, *Quality of Life Research*, 21(1), pp. 167–176

Allegretti, A., Borkan, J., Reis, S. and Griffiths, F. (2010). Paired interviews of shared experiences around chronic low back pain: classic mismatch between patients and their doctors. *Family practice,* 27(6), pp. 676-683.

Anand, P. (2010). Welfare Economics and Social Choice, in Simonetti, R et al. [eds] (2010). *Doing Economics: People, Markets and Policy,* Milton Keynes, Open University.

Andersson, G. B. (1999). Epidemiological features of chronic low-back pain. *Lancet,* 354(9178)**,** pp. 581-585.

Andronis, L., Kinghorn, P., Qiao, S., Whitehurst, D. G., Durrell, S. and McLeod, H. (2017). Cost-Effectiveness of Non-Invasive and Non-Pharmacological Interventions for Low Back Pain: a Systematic Literature Review, *Applied Health Econmoics and Health Policy,* 15(2), pp. 173-201

Appleby, J. (2016). *How does NHS spending compare with health spending internationally?* [online] The Kings Fund. Available at:

https://www.kingsfund.org.uk/blog/2016/01/how-does-nhs-spending-compare-health-spending-internationally [Accessed 2 Jun. 2017].

Ashworth, J., Konstantinou, K. and Dunn, K. (2011) Prognostic factors in non-surgically treated sciatica: a systematic review. *BMC Musculoskeletal Disorders*, 12(208).

Assendelft, W. J. J., Morton, S. C., Yu, E. I., Suttorp, M. J. and Shekelle, P. G. (2004) Spinal manipulative therapy for low-back pain. *Cochrane Database of Systematic Reviews*, 1.

Australian Department of Health and Ageing. (2010) *Medicare Benefits Schedule Book, operating from 1 May 2010*. Canberra. DoHA.

Axen, I., Bodin, L., Bergstrom, G., Halasz, L., Lange, F., Lovgren, P. W., Rosenbaum, A., Leboeuf-Yde, C. and Jensen, I. (2011). Clustering patients on the basis of their individual course of low back pain over a six month period. *BMC musculoskeletal disorders,* 12 (99).

Axen, I. and Leboeuf-Yde, C. (2013). Trajectories of low back pain. *Best Practice and Research Clinical rheumatology,* 27(5), pp. 601-612.

Babad, H., Sanderson, C., Naidoo, B., White, I., and Wang, D. (2002) The development of a simulation model for modelling primary prevention strategies for coronary heart disease. *Health Care Management Science*, 5(4), pp. 269–274.

Bakhtiary, A. H., Safavi-Farokhi, Z. and Rezasoltani, A. (2005). Lumbar stabilizing exercises improve activities of daily living in patients with lumbar disc herniation. *Journal of Back and musculoskeletal Rehabilitation,* 18 (3-4), pp. 55-60.

Balague, F., Nordin, M., Sheikhzadeh, A., Echegoyen, A. C., Brisby, H., Hoogewoud, H.M., Fredman, P. and Skovron, M.L. (1999). Recovery of severe sciatica. *Spine,* 24(23), pp.2516–2524.

Balagué, F., Mannion, A. F., Pellisé, F. and Cedraschi, C. (2012) Non-specific low back pain. *The Lancet***,** 379(9814), pp. 482-491.

Barton, P., Bryan, S., and Robinson, S. (2004) Modelling in the economic evaluation of health care: selecting the appropriate approach**,** *Journal of Health Services Research and Policy***,** 9(2), pp. 110 – 118

Becker, A., Held, H., Redaelli, M., Strauch, K., Chenot, J.F., Leonhardt, C., Keller, S., Baum, E., Pfingsten, M., Hildebrandt, J., Basler, H.D., Kochen, M.M., and Donner-Banzhoff, N. (2010). Low back pain in primary care. Costs of care and prediction of future health care utilization. *Spine*, 35(18), pp. 1714 – 1720.

Beneciuk, J. M., Hill, J. C., Campbell, P., Afolabi, E., George, S. Z., Dunn, K. M. and Foster, N. E. (2017). Identifying Treatment Effect Modifiers in the STarT Back Trial: A Secondary Analysis. *The Journal of Pain,* 18(1)**,** pp. 54-65.

Berg, S., Tullberg, T., Branth, B., Olerud, C. and Tropp, H. (2009). Total disc replacementcompared to lumbar fusion: a randomised controlled trial with 2-year follow-up. *European Spine Journal,* 18(10), pp. 1512–9

Bhattacharya, J., Hyde, T., and Tu, P. (2014) *Health Economics*. New York, Palgrave.

Biering-Sorensen, F., Thomsen, C. E. and Hilden, J. (1989). Risk indicators for low back trouble. *Scandinavian Journal of Rehabilitation Medicine,* 21(3)**,** pp. 151-157.

Birch, S. and Donaldson, C. (2003). Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism? *Social Science and Medicine,* 56(5), pp. 1121–1133

Bojke, L., Claxton, K., Sculpher, M. and Palmer, S. (2009). Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. *Value in Health*, 2009, 12(5), pp. 739-749.

Borgstrom, F., Strom, O., Kleman, M., McCloskey, E. Oden, A., Johansson, H. and Kanis, J. A. (2010). FRAX® and its applications in health economics—Cost-effectiveness and intervention thresholds using bazedoxifene in a Swedish setting as an example. *Bone,* 47, pp. 430-437.

Boswell, M. V., Hansen, H. C., Trescot, A. M. and Hirsch, J. A. 2003. Epidural steroids in the management of chronic spinal pain and radiculopathy. *Pain Physician,* 6(3)**,** pp. 319-334.

Brennan, A., Akehurst, R. (2000) Modelling in health economics evaluation. What is its place? What is its value? *Pharmacoeconomics,* 17(5), pp. 445–459.

Brennan, A., Chick, S. E. and Davies, R. (2006). A taxonomy of model structures for economic evaluation of health technologies. *Health Economics,* 15(12), pp.1295–1310.

Bridges S. (2011) Chronic pain. In: Craig R, Mindell J, eds. (2011) *Health Survey for England 2011 Health, social care and lifestyles.* NHS Digital. Available at (https://files.digital.nhs.uk/publicationimport/pub09xxx/pub09300/hse2011-ch9-chronic-pain.pdf) [Accessed 17th July 2017].

Briggs, A. H. and O'Brien, B.J. (2001) The death of cost-minimization analysis. Health Economics, 10(2), pp. 179-184.

Briggs, A., Claxton, K., Sculpher, M. (2006) *Decision modelling for health economic evaluation***.** New York: Oxford University Press.

Brouwer, W.B.F., Culyer, A.J., Van Exel, N.J. A., and Rutten, F.F.H. (2008) Welfarism vs. extra-welfarism, *Journal of Health Economics,* 27(2): pp. 325–338

Buchbinder, R., Blyth, F. M., March, L. M., Brooks, P., Woolf, A. D. and Hoy, D. G. (2013). Placing the global burden of low back pain in context. *Best Practice and Research Clinical Rheumatology,* 27(5), pp. 575-589.

Buchbinder, R., van Tulder, M., Öberg, B., Menezes Costa, L.M., Woolf, A., Schoene, M. and Croft, P. (2018). Low back pain: a call for action, *Lancet,* 391(10137) pp. 2384-2388.

Burton, A. K., Waddell, G., Tillotson, K. M. and Summerton, N. (1999). Information and advice to patients with back pain can have a positive effect. A randomized controlled trial of a novel educational booklet in primary care. *Spine,* 24(23)**,** pp. 2484-2491.

Büssing, A., Ostermann, T., Lüdtke, R. and Michalsen, A. (2012). Effects of yoga interventions on pain and pain-associated disability: a meta-analysis. *The Journal of Pain,* 13(1)**,** pp. 1-9.

M.J. Buxton., Drummond, M.F., Van Houdt, B.A., Prince, R. L., Sheldon, T.A., Szucs, T. and Vray, M (1998) Modelling in economic evaluation: an unavoidable fact of life, *Health Economics*, 6(3), pp. 217-227.

Bydon, M., Macki, M., Abt, N.B., Witham, T.F., Wolinsky, J.P., Gokaslan, Z.L., Bydon, A., and Sciubba, D.M. (2015) The cost-effectiveness of interbody fusions versus posterolateral fusions in 137 patients with lumbar spondylolisthesis. *Spine*, 15(3), pp. 492-498.

Calvo-Muñoz, I., Gómez-Conesa, A. and Sánchez-Meca, J. (2013). Prevalence of low back pain in children and adolescents: a meta-analysis. *BMC pediatrics,* 13(14).

Campbell, P., Foster, N. E., Thomas, E. and Dunn, K. M. (2013). Prognostic Indicators of Low Back Pain in Primary Care: Five-Year Prospective Study. *The Journal of Pain,* 14(8)**,** pp. 873-883.

Campello, M. A., Weiser, S. R., Nordin, M. and Hiebert, R. (2006). Work retention and nonspecific low back pain. *Spine,* 31(16)**,** pp. 1850-1857.

Carey, T. S., Garrett, J. M., Jackman, A. and Hadler, N. (1999). Recurrence and care seeking after acute back pain: Results of a long-term follow-up study. *Medical Care,* 37(2)**,** pp. 157-164.

Cassidy, J.D., Cote, P., Carroll, L.J. and Kristman, V. (2005). Incidence and course of low back pain episodes in the general population. *Spine,* 30(24), pp. 2817-2823.

Centers For Disease Control and and Prevention (2001). Prevalence of disabilities and associated health conditions among adults--United States, 1999. *MMWR Morbidity and mortality weekly report,* 50 (07), pp. 120-125.

Centers for Medicare and Medicaid Services. MEDPAR Limited Data Set (LDS) - Hospital (National). Available at: http://cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/MEDPARLDSHospitalNational.html [Authors accessed January 29, 2013].

Charlesworth, A. and Johnson, P. (eds) (2018). *Securing the future: funding the health and social care to the 2030s.* [Online] Institute for Fiscal Studies, Health Foundation. Available at http://www.ifs.org.uk/publications/12994 [Accessed 16 July 2019].

Cherkin, D., Kovacs, F.M., Croft, P., Borkan, J., Foster, N., Oberg, P., Urrutia, G. and Zamora, J. (2009). The ninth international forum for primary care research on low back pain. *Spine,* 34(3), pp. 304–307.

Cherkin, D., Balderson, B., Wellman, R., Hsu, C., Sherman, K.J., Evers, S.C., Hawkes, R., Cook, A., Levine, M.D., Piekara, D., Rock, P., Estlin, K.T., Brewer, G., Jensen, M. and LaPorte, A-M. (2018). Effect of Low Back Pain Risk-Stratification Strategy on Patient Outcomes and Care Processes: the MATCH Randomized Trial in Primary Care. *Journal of General Internal Medicine,* 33(8), pp. 1324-1336.

Cheung, K. M., Karppinen, J., Chan, D., Ho, D. W., Song, Y. Q., Sham, P., Cheah, K. S., Leong, J. C. and Luk, K. D. (2009). Prevalence and pattern of lumbar magnetic resonance imaging changes in a population study of one thousand forty-three individuals. *Spine,* 34(9)**,** pp. 934-940.

Chhatwal, J, Jayasuriya, S., Elbasha, E.H. (2016) Changing Cycle Lengths in State-Transition Models: Challenges and Solutions. *Medical Decision Making,* 36(8), pp. 952-64.

Chou, R., Qaseem, A., Snow, V., Casey, D., Cross, J. T., Jr., Shekelle, P., and Owens, D. K. (2007). Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Annals of Internal Medicine,* 147(7)**,** pp. 478-491.

Chou, R. and Shekelle, P. (2010). Will this patient develop persistent disabling low back pain? *Jama,* 303(13)**,** pp. 1295-1302.

Chou, R. (2014). Low Back Pain. *Annals of Internal Medicine,* 160(11), ITC6-1.

Clark, W., Jobanputra, P., Barton, P., Burls, A. (2004). The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic analysis. *Health Technology Assessment,* 8(18), pp. 1–118.

Coast, J., Smith, R.D., and Lorgelly, P. (2008) Welfarism, extra-welfarism and capability: The spread of ideas in health economics. Social Science and Medicine, 67(7), pp. 1190-1198

Coeuret-Pellicer, M., Descatha, A., Leclerc, A. and Zins, M. (2010). Are tall people at higher risk of low back pain surgery? A discussion on the results of a multipurpose cohort. *Arthritis care and research,* 62(1), pp. 125-127.

Cook, C. E., Taylor, J., Wright, A., Milosavljevic, S., Goode, A. and Whitford, M. (2014). Risk factors for first time incidence sciatica: a systematic review. *Physiotherapy Research International,* 19(2)**,** pp. 65-78.

Cooper, K., Davies, R., Roderick, P., Chase, D. and Raftery, J. (2002). The development of a simulation model for the treatment of coronary artery disease. *Health Care Management Science,* 5(4), pp. 259–267.

Cooper, K., Brailsford, S.C. and Davies, R. (2007). Choice of modelling technique for evaluating health care interventions, *Journal of the Operational Research Society*, 58(2), pp. 168-176.

Costa Lda, C., Maher, C. G., Mcauley, J. H., Hancock, M. J., Herbert, R. D., Refshauge, K. M. and Henschke, N. (2009). Prognosis for patients with chronic low back pain: inception cohort study. *BMJ,* 339(b3829).

Cramer, H., Lauche, R., Haller, H. and Dobos, G. (2013). A systematic review and meta-analysis of yoga for low back pain. *The Clinical journal of pain,* 29(5)**,** pp. 450-460.

Croft, P. R., Macfarlane, G. J., Papageorgiou, A. C., Thomas, E. and Silman, A. J. (1998). Outcome of low back pain in general practice: a prospective study. *BMJ;* 316, pp. 1356-1359.

Culyer, A. (1991). The normative economics of health care finance and provision. In A. McGuire, P. Fenn, and K.Mayhew (Eds.), *Providing health care: The economics of alternative systems of finance and delivery.* Oxford: Oxford University Press, pp. 65–98.

Culyer, A., and Evans, R. (1996). Normative rabbits from positive hats: Mark Pauly on welfare economics. *Journal of Health Economics*, 15(2), pp. 243–251.

Da C Menezes Costa, L., Maher, C. G., Hancock, M. J., Mcauley, J. H., Herbert, R. D. and Costa, L. O. (2012). The prognosis of acute and persistent low-back pain: a meta-analysis. *Canadian Medical Association Journal,* 184(11)**,** E613-24.

Dahm, K. T., Brurberg, K. G., Jamtvedt, G. and Hagen, K. B. (2010). Advice to rest in bed versus advice to stay active for acute low-back pain and sciatica. *The Cochrane Database of Systematic Reviews, 6.*

Dangerfield, B.C. (1999). System dynamics applications to European health care issues. *Journal of the Operational Research Society*, 50(4), pp. 345–353.

Dart, R. C., Green, J. L. and Bogdan, G. M. (2005). The safety profile of sustained release paracetamol during therapeutic use and following overdose. *Drug safety,* 28(11)**,** pp. 1045-1056.

Davis, R.J., Errico, T.J., Bae, H. and Auerbach, J.D. (2013). Decompression and Coflex interlaminar stabilization compared with decompression and instrumented spinal fusion for spinal stenosis and low-grade degenera¬tive spondylolisthesis: two-year results from the prospective, random¬ized, multicenter, Food and Drug Administration Investigational Device Exemption trial. *Spine,* 38(18), pp. 1529–1539.

De Schepper, E. I., Damen, J., Van Meurs, J. B., Ginai, A. Z., Popham, M., Hofman, A., Koes, B. W. and Bierma-Zeinstra, S. M. (2010). The association between lumbar disc

degeneration and low back pain: the influence of age, gender, and individual radiographic features. *Spine,* 35(5)**,** pp. 531-536.

Deaton, A. (2002) Policy Implications of The Gradient Of Health And Wealth, *Health Affairs,* 21(2), pp. 13-30.

Degeling, K., Koffijberg, H., IJzerman, M.J. (2017) A systematic review and checklist presenting the main challenges for health economic modeling in personalized medicine: towards implementing patient-level models. Expert Review of Pharmacoeconomics & Outcomes Research, 17(1), pp. 17-25.

Department of Health (UK) (2006) *The musculoskeletal services framework. A joint responsibility: doing it differently.* Leeds, Department of Health.

Deyo, R.A. and Weinstein, J.N. (2001) "Low back pain," *The New England Journal of Medicine***,** 344(5), pp. 363–370.

Deyo, R.A., Mirza, S.K., Martin, B.I., Kreuter, W., Goodman, D. C. and Jarvik, J.G (2003) Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA,* 303(13), pp. 1259–1265

Dionne, C. E., Von Korff, M., Koepsell, T. D., Deyo, R. A., Barlow, W. E. and Checkoway, H. (2001). Formal education and back pain: a review. *Journal of epidemiology and community health,* 55(7)**,** pp. 455-468.

Dionne, C. E., Dunn, K. M. and Croft, P. R. (2006). Does back pain prevalence really decrease with increasing age? A systematic review. *Age and Ageing,* 35(3)**,** p. 229-234.

Dionne, C.E., Dunn, K.M. and Croft, P.R. (2008). A consensus approach toward the standardization of back pain definitions for use in prevalence studies. *Spine,* 33(10, pp. 95–103.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1995). *"A social tariff for EuroQol: results from a UK general population survey,"* Working Papers 138chedp, Centre for Health Economics, University of York.

Dolan, P., Shaw, R., Tsuchiya, A., and Williams, A. (2005). QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics,* 14(2), pp. 197–208.

Dolan, P. and Kahneman, D. (2008) Interpretations of Utility and their implications for the valuation of health, *The Economic Journal,* 118(525), pp. 215–234.

Donaldson, C., Mitton, C. (2009) *Priority Setting Toolkit.* [pdf]. BMJ Books. Available at: http://www.myilibrary.com?ID=21394 [Accessed 27 April 2017]

Drummond, M.F., Schulper, M.J., Claxton, K., Stoddart, G.L. and Torrance, G.W. (2015) *Methods for the Economic Evaluation of Health Care Programme.* Oxford: University Press.

Dunn, K. M., Jordan, K. and Croft, P. R. (2006). Characterizing the course of low back pain: a latent class analysis. *American Journal of Epidemiology,* 163(8), pp. 754-761.

Dunn, K. M., Jordan, K. P., Mancl, L., Drangsholt, M. T. and Le Resche, L. (2011). Trajectories of pain in adolescents: a prospective cohort study. *Pain,* 152(1), pp. 66-73.

Dunn, K. M., Campbell, P. and Jordan, K. P. (2013a). Long-term trajectories of back pain: cohort study with 7-year follow-up. *BMJ Open,* 3(e003838).

Dunn, K. M., Hestbaek, L. and Cassidy, J. D. (2013b). Low back pain across the life course. *Best Practice and Research Clinical Rheumatology,* 27(5), pp. 591-600.

Duthey, B (2013) *Background Paper 6.24 Low back pain,* in: World Health Organisation (WHO) (ed.) Priority Medicines for Europe and the World "A Public Health Approach to Innovation". Geneva: WHO.

Economist (2015) Mind stretching: Fine words on mental health should be matched with money for research—but not just from the state, Mental-health research, *The Economist* Oct 24th 2015 [online] Available at http://www.economist.com/news/leaders/21676771-fine-words-mental-health-should-be-matched-money-researchbut-not-just [Accessed 22/11/2015].

M. Ekman., O. Johnell, and Lidgren, L. (2005). The economic cost of low back pain in Sweden in 2001, *Acta Orthopaedica,* 76(2), pp. 275-284

Endean, A., Palmer, K. T. and Coggon, D. (2011). Potential of magnetic resonance imaging findings to refine case definition for mechanical low back pain in epidemiological studies: a systematic review. *Spine,* 36(2), pp. 160-169.

Engers, A. J., Jellema, P., Wensing, M., Van Der Windt, D. A. W. M., Grol, R. and Van Tulder, M. W. (2008). Individual patient education for low back pain. *The Cochrane Database of Systematic Reviews, 23(1).*

Enthoven P, Skargren E, Oberg B. (2004). Clinical course in patients seeking primary care for back or neck pain: a prospective 5-year follow-up of outcome and health care, *Spine,* 29(21), pp. 2458-2465

Fairbank, J. C. (2007) Sciatic: An archaic term. *BMJ,* 335(7611), p. 112.

Fejer, R. and Leboeuf-Yde, C. (2012). Does back and neck pain become more common as you get older? A systematic literature review. *Chiropractic & Manual Therapies, 20(*24).

Fitzsimmons, D., Philips, C. J., Bennett, H., Jones, M., Williams, N., Lewis, R., Sutton, A., Matar, H. E., Din, N. and Burton, K. (2014) Cost-effectiveness of different strategies to manage patients with sciatica. *PAIN,* 155 (7), pp. 1318-1327

Foster, N.E., Bishop, A., Thomas, E., Main, C., Horne, R., Weinman, J. and Hay, E. (2008) Illness perceptions of low back pain patients in primary care: what are they, do they change and are they associated with outcome? *PAIN,* 136(1-2), pp. 177–87.

Foster, N.E, Mullis, R., Hill, J.C., Lewis, M., Whitehurst, D.G., Doyle, C., Konstantinou, K., Main, C., Somerville, S., Sowden, G., Wathall, S., Young, J., and Hay, E.M. (2014) Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. *Annals of Family Medicine*, 12(2), pp. 102-11.

Foster, N.E., Konstantinou, K., Lewis, M., Ogollah, R., Dunn, K.M., van der Windt, D., Beardmore, R., Artus, M., Bartlam, B., Hill, J.C., Jowett, S., Kigozi, J., Mallen, C., Saunders, B., and Hay, E.M. (2017). The clinical and cost-effectiveness of stratified care for patients with sciatica: the SCOPiC randomised controlled trial protocol (ISRCTN75449581) *BMC Musculoskelet Disorders,* 18(1), p. 172.

Foster, N.E., Anema, J.R., Cherkin, D., Chou, R., Cohen, S. P., Gross, D. P., Ferreira, H., Fritz, J. M., Koes, B. W., Peul, W., Turner, J. A. and Maher, C. G. (2018) Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet,* 391(10137), pp. 2368-2383

Fritz, J.M., Delitto, A. and Erhard, R.E. (2003). Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine***,** 28(13), pp. 1363–1371

Frymoyer, J. W. And Cats-Baril, W. L. (1991). An overview of the incidences and costs of low back pain. *The Orthopedic clinics of North America,* 22(2), pp. 263-271.

Fryback, D.G., Dasbach, E.J., Klein, R., Klein, B. E., Dorn, N., Peterson, K. and Martin, P. A (1993) The Beaver Dam Health Outcomes study: Initial Catalog of Health-state Quality Factors. *Medical Decision Making,* 13(2), pp. 89-102.

Furlan, A. D., Pennick, V., Bombardier, C. and Van Tulder, M. (2009) updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine,* 34 (18), pp. 1929–41.

Ghabri, S., Hamers, F. and Josselin, J-M. (2016). Exploring uncertainty in economic evaluations of new drugs and medical devices: lessons from the first review of pharmaceutical companies' submissions to the French National Authority for Health. *Pharmacoeconomics,* 34, pp. 617–24.

Gibson, J. N. and Waddell, G. (2007). Surgical interventions for lumbar disc prolapse. *Spine,* 32(16), pp. 1735-1747.

Glassman, S.D., Polly, D.W., Dimar, J.R. and Carreon, L.Y. (2012). The cost effectiveness of single-level instrumented posterolateral lumbar fusion at 5 years after surgery. *Spine,* 37(9), pp. 769–774.

Glick, H, A., Doshi, J.A., Sonnad, S. S. and Polsky, D. (2014). *Economic Evaluation in Clinical Trials.* Oxford: University Press

Goodacre, S. and McCabe, C. (2002) An introduction to economic evaluation. *Emergency Medicine Journal,* 19, pp. 198-201

Goode, A., Cook, C., Brown, C., Isaacs, R., Roman, M and Richardson, W. (2011) Differences in comorbidities on low back pain and low back related leg pain. *Pain Practice,* 11(1), pp. 42-47.

Government Accountability Office. (2008). *Medicare Part B imaging services: rapid spending growth and shift to physician offices indicate need for CMA to consider additional management practices*. Washington DC: Government Accountability Office.

Gray, A.M., Clarke, P.M., Wolstenholme, J.L. and Wordsworth, S., (2011) *Applied Methods of Cost-effectiveness Analysis in Health Care***.** Oxford: Oxford University Press.

Grotle M, Brox JI, Veierod MB, et al., Clinical course and prognostic factors in acute low back pain: patients consulting primary care for the first time. Spine; 2005:30(8):976-982

Guzman, J., Esmail, R., Karjalainen, K., Malmivaara, A., Irvin, E. and Bombardier, C. (2001). Multidisciplinary rehabilitation for chronic low back pain: systematic review. *BMJ,* 322 (7301), pp. 1511-1516.

Haines, T. P., Hill, A. M., Hill, K.D., Brauer, S. G., Hoffman, T., Etherton-Beer, C. and McPhail, S. M. *(2013).* Cost effectiveness of patient education for the prevention of falls in hospital: economic evaluation from a randomized controlled trial, *BMC Med*, 11: 135.

Hall, J.A., Konstantinou, K., Lewis, M., Oppong, R., Ogollah, R. and Jowett, S. (2019) Systematic Review of Decision Analytic Modelling in Economic Evaluations of Low Back Pain and Sciatica, *Applied Health Economics and Health Policy,* 17(4), pp. 467-491

Hartvigsen, J., Hancock, M.J., Kongsted, A., Louw, Q., Ferreira, M. L., Genevay, S., Hoy, D., Karppinen, J., Pransky, G., Sieper, J., Smeets, R. J. and Underwood, M. (2018). Low back pain series: what low back pain is and why we need to pay attention, *Lancet,* 391(10137) pp. 2356-2367

Hassey, A., Gerrett D. and Wilson, A. (2001) A survey of validity and utility of electronic patient records in general practice. *BMJ,* 322:1401

Haugen, A.J., Grøvle, L., Brox, J.I., Natvig, B., Keller, A., Soldal, Grotle, M. (2011) Estimates of success in patients with sciatica due to lumbar disc herniation depend upon outcome measure. *European Spine Journal,* 20: 1669.

Hay, E.M., Dunn, K.M., Hill, J.C., Lewis, M., Mason, E. E., Konstantinou, K., Sowden, G., Somerville, S., Vohora, K., Whitehurst, D. and Main, C. J. (2008). A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. *BMC Musculoskelet Disorders,* 9: 58.

Hayden, J. A., Van Tulder, M. W., Malmivaara, A. V. and Koes, B. W. (2005). Meta-analysis: exercise therapy for nonspecific low back pain. *Annals of Internal Medicine,* 142(9), pp. 765-775.

Health Service Cost Index (HCSA) (2016) Health Service Cost Index, Department of Health and Social Care. [online]. Available at http://www.info.doh.gov.uk/doh/finman.nsf/af3d43e36a4c8f8500256722005b77f8/bb8108f0fa6e031980257fef005352d7?OpenDocument [Accessed 17/04/2017].

Heliövaara, M., Knekt, P. and Aromaa, A. (1987). Incidence and risk factors of herniated lumbar intervertebral disc or sciatica leading to hospitalization. *Journal of chronic diseases,* 40(3), pp. 251-258.

Heneweer, H., Vanhees, L. and Picavet, H. S. (2009). Physical activity and low back pain: a U-shaped relation? *Pain,* 143(1-2), pp. 21-25.

Henschke, N., Maher, C. G., Refshauge, K. M., Herbert, R. D., Cumming, R. G., Bleasel, J., York, J., Das, A. and Mcauley, J. H. (2008). Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ;* 337.

Henschke, N., Ostelo, R. W. J. G., Van Tulder, M. W., Vlaeyen, J. W. S., Morley, S., Assendelft, W. J. J. and Main, C. J. (2010). Behavioural treatment for chronic low-back pain. *The Cochrane Database Systematic Reviews,* 7(7).

Hernández, C. P., Sánchez, N., Navarro-Siguero, A. and Saldaña, M. T. 2013. What are the causes of sciatica and radicular pain? in F. Laroche and S. Perrot (eds.), *Managing Sciatica and Radicular Pain in Primary Care Practice, pp. 17-31.* Springer.

Hestbaek, L., Leboeuf, Y, C. and Manniche, C. (2003). Low back pain: what is the long-term course? A review of studies of general patient populations, *European Spine Journal,* 12(2), pp. 149–165.

Hestbaek, L., Korsholm, L., Leboeuf-Yde, C. and Kyvik, K. O. (2008). Does socioeconomic status in adolescence predict low back pain in adulthood? A repeated cross-sectional study of 4,771 Danish adolescents. *European Spine Journal,* 17(12)**,** pp. 1727-1734

Hicks, J. (1939). "The Foundations of Welfare Economics". *The Economic Journal,* 49(196), pp. 696–712.

Hicks, G. E., Gaines, J. M., Shardell, M. and Simonsick, E. M. (2008). Associations of back and leg pain with health status and functional capacity of older adults: findings from the retirement community back pain study. *Arthritis and Rheumatism,* 59(9)**,** pp. 1306-1313.

Hill, J.C., Dunn, K.M., Lewis, M., Mullis, R., Main, C.J., Foster, N.E. and Hay, E. M. (2008) A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis and Rheumatism,* 59(5), pp. 632–41.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E, Konstantinou, K., Main, C.J., Mason, E., Somerville, S., Sowden, G., Vohora, K., Hay, E.M (2011)

Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet,* 378(9802), pp. 1560-1571

Hillman, M., Wright, A., Rajaratnam, G., Tennant, A. and Chamberlain, M. A. (1996). Prevalence of low back pain in the community: implications for service provision in Bradford, UK. *Journal of Epidemiology and Community Health,* 50, pp. 347–52

Hingorani A.D., Windt D.A., Riley R.D., Abrams, K. A., Moons, K.G., Steyerberg, E.W., Schroter, S., Sauerbrei, W., Altman, D. G. and Hemingway, H. for PROGRESS Group (2013). Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ.* 2013;346:e5793

Hoffman, B. M., Papas, R. K., Chatkoff, D. K. and Kerns, R. D. (2007). Meta-analysis of psychological interventions for chronic low back pain, *Health psychology,* 26(1), pp. 1-9

Hofstee, D. J., Gijtenbeek, J. M. M., Hoogland, P. H., Van Houwelingen, H. C., Kloet, A., Lötters, F. and Tans, J. T. J. (2002). Westeinde sciatica trial: randomized controlled study of bed rest and physiotherapy for acute sciatica. *Journal of Neurosurgery,* 96(1)**,** 45-49.

Hollingworth, W., Todd, C. J., King, H., Males, T., Dixon, A. K., Karia, K. R. and Kinmonth, A. L. (2002). Primary care referrals for lumbar spine radiography: diagnostic yield and clinical guidelines. *The British journal of general practice,* 52(479), pp. 475-480.

Hoy, D., Brooks, P., Blyth, F. and Buchbinder, R. (2010). The epidemiology of low back pain. *Best practice and research Clinical rheumatology,* 24(6)**,** pp. 769-781.

Hoy, D., Bain, C., Williams, G., March, L., Brooks, P., Blyth, F., Woolf, A., Vos, T. and Buchbinder, R. (2012). A systematic review of the global prevalence of low back pain. *Arthritis and Rheumatism,* 64(6), pp. 2028-2037.

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., Barendregt, J., Murray, C., Burstein, R. and Buchbinder, R. (2014) The global burden of low back pain: estimates from the Global Burden of Disease 2010 study, *Annals of the Rheumatic Diseases*, 73(6), pp. 968-974.

Hurley, J. (1998). Welfarism, extra-welfarism and evaluative economic analysis in the health sector. In M. Barer, T. Getzen, and G. Stoddart (Eds.), *Health, health care and health economics: Perspectives on distribution* (pp. 373–395). Chichester: John Wiley.

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A.H., Mauskopf, J. and Loder, E (2013) Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ,* 346 :f1049

Igarashi, A., Akazawa, M., Murata, T., Taguchi, T., Sadosky, A., Ebata, N., Willke, R., Fujii, K., Doherty, J. and Kobayashi, M. (2015) Cost-effectiveness analysis of pregabalin for treatment of chronic low back pain in patients with accompanying lower limb pain (neuropathic component) in Japan. *Clinicoeconomics and Outcomes Research,* 7(7), pp. 505-520.

IMS-Brogan Inc. (2011) Provincial Formulary Database . Kirkland, Quebec : Delta PATM

Institute for Clinical and Economic Review. (2011) *Management options for patients with low back disorders*. Available from: https://icer-review.org/wp-content/uploads/2016/02/LBP-Final-Appraisal-6-24-11.pdf [Accessed 15th May 2017].

Ito, K., Hollenberg, J. P., Charlson, M. E. (2009). Using the Osteoporosis Self-Assessment Tool for Referring Older Men for Bone Densitometry: A Decision Analysis. *Journal of the American Geriatrics Society,* 57(2), pp. 218-224.

Ito, K. and Leslie, W. D. (2015). Cost-effectiveness of fracture prevention in rural women with limited access to dual-energy X-ray absorptiometry. Osteoporosis International, 26(8), pp. 2111-2119.

Jaffé, G. (1974). A doubie-blind, between-patient comparison of alclofenac ('Prinalgin') and indomethacin in the treatment of low back pain and sciatica. *Current medical research and opinion,* 2(7)**,** pp. 424-429.

Jaime-Caro, J., Briggs, A.H., Siebert, U. and Kuntz, K.M. (2012). Modeling Good Research Practices-Overview: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value Health*, 32(5), pp. 796-803.

Jeffries, L. J., Milanese, S. F. and Grimmer-Somers, K. A. (2007). Epidemiology of adolescent spinal pain: A systematic overview of the research literature. *Spine,* 32(23)**,** pp. 2630-2637.

Jespersen, T., Jørgensen, M. B., Hansen, J. V., Holtermann, A. and Søgaard, K. (2012). The relationship between low back pain and leisure time physical activity in a working population of cleaners-a study with weekly follow-ups for 1 year. *BMC musculoskeletal disorders,* 13: 28.

Jones, G. T. and Macfarlane, G. J. (2005). Epidemiology of low back pain in children and adolescents. *Archives of Disease in Childhood,* 90(3)**,** pp. 312-316.

Jordan, K. P., Kadam, U, T., Hayward, R., Porcheret, M., Young, C. and Croft, P. (2010) Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC musculoskeletal disorders,* 11, (144).

Jordan, K.P., Jöud, A., Bergknut, C., Croft, P., Edwards, J.J., Peat, G., Petersson, I.F., Turkiewicz, A., Wilkie, R. and Englund, M. (2014) International comparisons of the

consultation prevalence of musculoskeletal conditions using population-based healthcare data from England and Sweden, *Annals of the Rheumatic Diseases,* 73(1), pp. 212-218.

Kääriäl, S., Leino-Arjasl, P., Rahkonenl, O., Lahtil, J., Lahelmal, E. and Laaksonenl, M. (2011) Risk factors of sciatic pain: A prospective study among middle—aged employees. *European Journal of Pain,* 15(60, pp. 584-590.

Kaldor, Nicholas (1939). "Welfare Propositions in Economics and Interpersonal Comparisons of Utility". *The Economic Journal,* 49(195), pp. 549–552.

Kalauokalani, D., Cherkin, D. C., Sherman, K. J., Koepsell, T. D. and Deyo, R. A. (2001). Lessons from a trial of acupuncture and massage for low back pain: patient expectations and treatment effects. *Spine,* 26(13), pp. 1418-1424.

Kamper, S. J., Apeldoorn, A. T., Chiarotto, A., Smeets, R. J. E. M., Ostelo, R. W. J. G., Guzman, J. and Van Tulder, M. W. (2014). Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. *The Cochrane Database Systematic Reviews,* 2(9).

Karacan, I., Aydin, T., Sahin, Z., Cidem, M., Koyuncu, H., Aktas, I. And Uludag, M. (2004). Facet angles in lumbar disc herniation: their relation to anthropometric features. *Spine,* 29**,** 1132-1136.

Karjalainen, K., Malmivaara, A., Van Tulder, M., Roine, R., Jauhiainen, M., Hurri, H. and Koes, B. (2001). Multidisciplinary biopsychosocial rehabilitation for subacute low back pain in working-age adults: a systematic review within the framework of the Cochrane Collaboration Back Review Group. *Spine,* 26(3)**,** pp. 262-269.

Karnon, J. and Brown, J. (1998). Selecting a decision model for economic evaluation: a case study and review. *Health Care Management Science*, 1(2), pp. 133-140

Karnon J. (2003). Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Economics,* 12(10), pp. 837-848.

Katz, J.N., Lipson, S.J., Lew, R.A., Grobler, L. J., Weistein, J. N., Brick, G. W., Fossel, A. H. and Liang, M. H. (1997) Lumbar laminectomy alone or with instrumented or noninstrumented arthrodesis in degenerative lumbar spinal stenosis: Patient selection, costs, and surgical outcomes. *Spine*, 22(10), pp. 1123–31.

Kearney, P. M., Baigent, C., Godwin, J., Halls, H., Emberson, J. R. and Patrono, C. (2006). Do selective cyclo-oxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? Meta-analysis of randomised trials. *BMJ,* 332(7553)**,** pp. 1302-1308.

Keefe, F. J., Brown, G. K., Wallston, K. A. and Caldwell, D. S. (1989). Coping with rheumatoid arthritis pain: catastrophizing as a maladaptive strategy. *Pain,* 37(1), pp. 51-56.

Keele University (2015) STarT Back / Commissioners / Impact [online] Available at: https://www.keele.ac.uk/sbst/commissioners/impact/ [Accessed 20 April 2017].

Keele University (undated) SCOPiC trial - SCiatica Outcomes in Primary Care [online] Available at: https://www.keele.ac.uk/pchs/research/musculoskeletalpainandstratifiedcare/scopic/ [Accessed 26th April 2017].

Khoromi, S., Cui, L., Nackers, L. and Max, M. B. (2007). Morphine, nortriptyline and their combination vs. placebo in patients with chronic lumbar root pain. *Pain,* 130**,** 66-75.

Kigozi, J., Konstantinou, K., Ogollah, R., Dunn, K., Lewis, M. and Jowett, S. (2019). Factors associated with costs and health outcomes in patients with Back and leg pain in primary care: a prospective cohort analysis. *BMC Health Services Research,* 19: 406

Kim, N., Yang, B., Lee, T. and Kwon, S. (2010) An economic analysis of usual care and acupuncture collaborative treatment on chronic low back pain: a Markov model decision analysis. *BMC Complementary Alternative Medicine*, 10 (74).

Kim, S., Hedjri, S., Coyte, P.C., and YR, R. (2012) Cost-utility of lumbar decompression with or without fusion for patients with symptomatic degenerative lumbar spondylolisthesis. *Spine*, 12(1), pp. 44-54.

Klaber Moffett, J., Richardson, G., Sheldon, T. and Maynard, A. (1995). *"Back pain: its management and costs to society",* Working Papers 129chedp. York: Centre for Health Economics, University of York.

Kluge, E. H. (2007). Resource Allocation in Healthcare: Implications of Models of Medicine as a Profession. *Medscape General Medicine, 9*(1), p, 57.

Kobayashi, S., Baba, H., Uchida, K., Kokubo, Y., Kubota, C., Yamada, S., Suzuki, Y. and Yoshizawa, H. (2005). Effect of mechanical compression on the lumbar nerve root: localisation and changes of intraradicular inflammatory cytokines, nitric oxide, and cyclooxygenase. *Spine*, 30(15), pp. 1699–705.

Koenig L, Dall, T.M., Gu, Q., Saavoss, J. and Schafer, M.F. (2014) How Does Accounting for Worker Productivity Affect the Measured Cost-Effectiveness of Lumbar Discectomy? *Clinical Orthopaedics and related research,* 472(4), pp. 1069-1079

Koes, B., van Tulder, M. and Peul, W. C. (2007). Diagnosis and treatment of sciatica. *BMJ,* 334(7607), pp. 1313–7.

Kolb, E., Canjuga, M., Bauer, G. F. and Läubli, T. (2011). Course of back pain across 5 years: A retrospective cohort study in the general population of Switzerland. *Spine,* 36(4), E268-E273.

Kongsted, A. and Leboeuf-Yde, C. (2010). The Nordic back pain subpopulation program: Can low back pain patterns be predicted from the first consultation with a chiropractor? A longitudinal pilot study. *Chiropractic and osteopathy,* 18: 8.

Kongsted, A., Kent, P., Hestbaek, L. and Vach, W. (2015). Patients with low back pain had distinct clinical course patterns that were typically neither complete recovery nor constant pain. A latent class analysis of longitudinal data. *The spine journal,* 15(5)**,** pp. 885-894.

Kongsted, A., Kent, P., Axen, I., Downie, A. S. and Dunn, K. M. 2016. What have we learned from ten years of trajectory research in low back pain? *BMC musculoskeletal disorders,* 17: 220.

Konstantinou, K. and Dunn, K. M. (2008) Sciatica: review of epidemiological studies and prevalence estimates. *Spine,* 33(22), pp. 2464-2472.

Konstantinou, K., Lewis, M. and Dunn, K. M. (2012). Agreement of self-reported items and clinically assessed nerve root involvement (or sciatica) in a primary care setting. *European Spine Journal,* 21(11), pp. 2306-2315.

Konstantinou K, Hider S, L., Jordan J, L., Lewis, M. and Hay, E. M. (2013). The Impact of Low Back-related Leg Pain on Outcomes as Compared With Low Back Pain Alone: A Systematic Review of the Literature**.** *Clinical journal of pain,* 29(7), pp. 644-654

Konstantinou K, Dunn K, Ogollah R, Lewis M, van der Windt D, Hay E. (2018). Prognosis of sciatica and back-related leg pain in primary care: the ATLAS cohort. *Spine, 18(6),* pp. 1030-1040.

Konstantinou, K., Dunn, K. M., van der Windt, D., Ogollah, R., Jasani, V. and Foster, N. (2019). Subgrouping patients with sciatica in primary care for matched care pathways: development of a subgrouping algorithm. *BMC Musculoskeletal Disorders,* 20:313.

Koopman, J.S., Chick, S.E., Riolo, C.S., Simon, C.P. and Jacquez, G. (2002). Stochastic effects on endemic infection levels of disseminating versus local contacts. *Mathematical Biosciences*, 180(1-2), pp. 49–71.

Koopman, J.S., Jacquez, G. and Chick, S.E. (2006). Integrating discrete and continuous population modeling strategies in population health and aging: strengthening the dialog between epidemiology and demography. In Weinstein M, Hermalin A, Stoto MA (eds), *Annals of the New York Academy of Sciences,* vol. 954. New York, New York Academy of Science, pp. 268–294.

Koopmanschap, M.A., Rutten, F.F., van Ineveld, B.M. and van Roijen, L. (1995). The friction cost method for measuring indirect costs of disease. *Journal of Health Economics,* 14(2), pp. 171-189.

Korean National Health and Nutrition survey. (2007). Available at https://knhanes.cdc.go.kr/knhanes/main.do [Accessed 08/03/2019].

Kuntz, K. M., Snider, R.K., Weinstein, J.N., Pope, M.H. and Katz, J.N. (2000). Cost-effectiveness of fusion with and without instrumentation for patients with degenerative spondylolisthesis and spinal stenosis. *Spine,* 25(9), pp. 1132-1139.

Kuntz, K. and Weinstein, M. (2001). Modelling in economic evaluation. In: Drummond M, McGuire A, editors. *Economic Evaluation in Health Care: Merging Theory with Practice.* Oxford: Oxford University Press.

Lamb, S.E., Lall, R., Hansen, Z., Castelnuovo, E., Withers, E. J., Nichols, V., Griffiths, F., Potter, R., Szczepura, A., and Underwood M (2010) A multicentred randomised controlled trial of a primary carebased cognitive behavioural programme for low back pain. The Back Skills Training (BeST) trial. *Health Technology Assessment*, 14(41).

Lambeek, L.C., van Tulder, M.W., Swinkels, I.C., Koppes, L.L., Anema, J.R. and van Mechelen, W. (2011). The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. *Spine,* 36(13), pp. 1050–58.

Laroche, F. And Perrot, S. (2013). *Managing Sciatica and Radicular Pain in Primary Care Practice,* London: Springer Healthcare.

Launois, R., Henry, B., Marty, J.R., Gersberg, M., Lassale, C., Benoist, M. and Goehrs, J. M. (1994). Chemonucleolysis versus surgical discectomy for sciatica secondary to lumbar disc herniation. A cost and quality-of-life evaluation. *Pharmacoeconomics,* 6(5), pp. 453–463.

Lawrence, J. S. (2016). *Rheumatism in populations*, London: Elsevier.

Leclerc, A., Tubach, F., Landre, M. F. and Ozguler, A. (2003). Personal and occupational predictors of sciatica in the GAZEL cohort. *Occupational medicine* 53(6), pp. 384-391.

Leino-Arjas, P., Solovieva, S., Riihimaki, H., Kirjonen, J. and Telama, R. (2004). Leisure time physical activity and strenuousness of work as predictors of physical functioning: a 28 year follow up of a cohort of industrial employees. *Occupational and environmental medicine,* 61**,** pp. 1032-1038.

Lewis, R., Wiliams, N., Matar, H. E., Din, N., Fitzsimmons, D., Phillips, C., Jones, M., Sutton, A., Burton, K., Nafees, S., Hendry, M., Rickard, I., Chakraverty, R. and Wilkinson, C. (2011). The clinical effectiveness and cost-effectiveness of management strategies for sciatica: systematic review and economic model. *Health Technology Assessment, 1*5(39).

Lloyd, A., Scott, D.A., Akehurst, R.L., Lurie-Luke, E., Jessen, G. (2004). Cost-effectiveness of low-level heat wrap therapy for low back pain. *Value in Health,* 7(4), pp. 413-422.

Lofland, J.H., Pizzi, L. and Frick, K.D. (2004) A review of health-related workplace productivity loss instruments. *Pharmacoeconomics,* 22(3), pp. 165–184

Luijsterburg, P. A. J., Verhagen, A. P., Ostelo, R. W. J. G., Van Os, T. A. G., Peul, W. C. And Koes, B. W. (2007). Effectiveness of conservative treatments for the lumbosacral radicular syndrome: a systematic review. *European Spine Journal,* 16(7), pp. 881-899.

Lyons, R.A., Lo, S.V. and Littlepage, B.N.C. (1994) Comparative health status of patients with 11 common illnesses in Wales. Journal of epidemiology and community health, 48(4), pp. 388–90.

Ma, V. Y., Chan, L. and Carruthers, K. J. (2014). Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. *Archives of Physical Medicine and Rehabilitation,* 95(5), pp. 986-995. e1.

Mahon, R. (2014). *Temporal uncertainty in cost-effectiveness decision models.* PhD thesis. 2014.

Malter, A.D., Larson, E.B., Urban, N. and Deyo, R.A. (1996) Cost-effectiveness of lumbar discectomy for the treatment of herniated intervertebral disc. *Spine,* 21(9), pp. 1048-1055.

Manchikanti, L., Singh, V., Falco, F. J. E., Benyamin, R. M. and Hirsch, J. A. (2014). Epidemiology of low back pain in adults. *Neuromodulation: Technology at the Neural Interface,* 17(2), pp. 3-10.

Maniadakis, N. and Gray, A. (2000) The economic burden of back pain in the UK. *Pain,* 84(1), pp. 95-103.

Manninen, P., Riihimak, H. And Heliovaara, M. (1995). Incidence and risk factors of low-back pain in middle-aged farmers. *Occupational medicine,* 45(3), pp. 141-146.

Martell, B. A., O'connor, P. G., Kerns, R. D., Becker, W. C., Morales, K. H., Kosten, T. R. and Fiellin, D. A. (2007). Systematic review: opioid treatment for chronic back pain: prevalence, efficacy, and association with addiction. *Annals of Internal Medicine,* 146(2), pp. 116-127.

Mccabe C, Dixon S. (2000) Testing the Validity of Cost-Effectiveness Models. *Pharmacoeconomics,* 17(5), pp. 501–513

McCabe, C, Claxton, K. and Culyer, AJ. (2008) The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics*, 26(9), pp. 733-44.

McGrail, M.P. Jr, Lohman, W.H. and Gorman, R. (2001). Disability prevention principles in the primary care office. American Family Physician, 63(4), pp. 679–84.

Ministry of Employment and Labor. Available at http://laborstat.molab.go.kr [Accessed 13th March 2019]. .

Miranda, H., Viikari-Juntura, E., Martikainen, R., Takala, E. P. and Riihimaki, H. (2001). Physical exercise and musculoskeletal pain among forest industry workers. *Scandinavian Journal of Medicine and Science in Sports,* 11**,** 239-246.

Miranda, H., Viikari-Juntura, E., Martikainen, R., Takala, E. P. and Riihimaki, H. (2002). Individual factors, occupational loading, and physical exercise as predictors of sciatic pain. *Spine,* 27(1)**,** pp. 1102-1109.

Mitchell, M. D., Mannino, D. M., Steinke, D. T., Kryscio, R. J., Bush, H. M. and Crofford, L. J. (2011). Association of smoking and chronic pain syndromes in Kentucky women. *The Journal of Pain,* 12(8)**,** pp. 892-899.

Morris, S., Devlin, N., Parkin, D., Spencer, A. (2012). *Economic Analysis in Health Care, Second Edition,* Chichester: John Wiley.

Morso, L., Schiøttz-Christensen, B., Søndergaard, J., de Vos Andersen, N., Pedersen, F., Olsen, K.R., Sall Jensen, M., Hill, J. and Høyrup Christiansen, D. (2018). The effectiveness of a stratified care model for non-specific low back pain in Danish primary care compared to current practice: study protocol of a randomised controlled trial. *Trials,* 315.

Mortimer, M., Pernold, G. and Wiktorin. C. (2006). Low back pain in a general population. Natural course and influence of physical exercises—a 5-year follow-up of the Musculoskeletal Intervention Center— Norrtalje Study. *Spine,* 31(26), pp. 3045–51.

Mueller, D, and Gandjour, A. (2008) Cost Effectiveness of Ultrasound and Bone Densitometry for Osteoporosis Screening in Post-menopausal Women, Applied Health Economics and Health Policy, 6(2-3), pp. 113-135.

Nadler, S.F., Steiner, D.J., Erasala, G.N., Hengehold, D.A., Hinkle, R.T., Beth Goodale, M., Abeln, S.B. and Weingand, K.W. (2002) Continuous low-level heat wrap therapy provides more efficacy than ibuprofen and acetaminophen for acute low back pain. *Spine,* 27(10), pp. 1012–7.

Nagata-Kobayashi, S., Shimbo, T. and Fukui, T. (2002). Cost-effectiveness analysis of screening for osteoporosis in postmenopausal Japanese women. *Journal of Bone & Mineral Metabolism, 20(6), pp. 350-357.*

National Health Service (2013) *Commissioning spinal services – getting the service back on track, a guide for commissioners of spinal services January 2013* [online] Available at http://www.nationalspinaltaskforce.co.uk/downloads/ [Accessed 26/04/2017]

National Institute for Health and Care Excellence (NICE). (2009). *Low back pain in adults: early management, Clinical Guidance.* National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence (NICE). (2013) *Guide to the methods of technology appraisal 2013* [online] Available at https://www.nice.org.uk/process/pmg9/chapter/the-reference-case [Accessed 19/04/2017]

National Institute for Health and Care Excellence (NICE). (2016) *Low back pain and sciatica in over 16s: assessment and management (NICE guideline NG59). Full guideline 2016.* [online] Available at www.nice.org.uk/guidance/ng59/evidence [Accessed 26/04/2017]

National Institute for Health and Care Excellence (NICE). (2009) *Low back pain in adults: early management, Clinical Guidance.* National Institute for Health and Care Excellence.

National Institute for Health and Care Excellence (NICE). (2008). *Guide to the methods of technology appraisal*. National Institute for Health and Care Excellence.

Navarro-Siguero, A., Hernández, C. P. and Saldaña, M. T. (2013) How are sciatica and spinal radicular pain classified?. In: Laroche F., Perrot S. (eds) *Managing Sciatica and Radicular Pain in Primary Care Practice.* Tarporley: Springer Healthcare.

Nayak, S., Roberts, M. S. and Greenspan, S. L. (2011). Cost-effectiveness of different screening strategies for osteoporosis in postmenopausal women. *Annals of Internal Medicine,* 155(11), pp. 751-761.

Nayak, S., Roberts, M. S. and Greenspan, S. L. (2012). Impact of generic alendronate cost on the cost-effectiveness of osteoporosis screening and treatment. PLoS ONE, 7(3), e32879.

Neumann, P.J., Greenberg, D., Olchanski, N.V., Stone, P.W. and Rosen, A.B. (2005). Growth and quality of the cost-utility literature, 1976-2001. *Value in Health,* 8(1), pp. 3-9.

Nordin, M., Skovron, M. L., Hiebert, R., Weiser, S., Brisson, P. M., Campello, M., Harwood, K., Crane, M. and Lewis, S. (1997). Early predictors of delayed return to work in patients with low back pain. *Journal of Musculoskeletal Pain,* 5(2), pp. 5-27

Norton G, McDonough CM, Cabral H, et al. (2015) Cost-utility of cognitive behavioral therapy for low back pain from the commercial payer perspective. *Spine,* 40(10), pp. 725–33.

Oakley J. (2002). *Value of information for complex costeffectiveness models.* Research Report No. 533/02, Department of Probability and Statistics, University of Sheffield.

Office for National Statistics (ONS). (2017). *Total of 137 million working days lost to sickness and injury in 2016 - Office for National Statistics.* [online] Available at: https://www.ons.gov.uk/news/news/totalof137millionworkingdayslosttosicknessandinjuryin2016 [Accessed 13 Jun. 2019].

Office for National Statistics (ONS). (2018). *Age groups – Office for National Statistics.* [online] Available at: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest [Accessed 16 June 2019].

Ontario Case Costing Initiative. (2010). *OCCI Costing Analysis Tool 2010*. Available at: http://www.occp.com. [Accessed August 24 2018].

Ostelo, R. W. J. G., Van Tulder, M. W., Vlaeyen, J. W. S., Linton, S. J., Morley, S. and Assendelft, W. J. J. (2005). Behavioural treatment for chronic low-back pain. *The Cochrane Database of Systematic Reviews,* 25(1).

Osterman, H., Seitsalo, S., Karppinen, J. and Malmivaara, A. (2006). Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. *Spine,* 31(21), pp. 2409-2414.

Owens, D.K., Lohr, K.N., Atkins, D., Treadwell, J.R., Reston, J.T., Bass, E.B., Chang, S. and Helfand, M. (2010). Grading the strength of a body of evidence when comparing medical interventions – AHRQ and the effective health care program. *Journal of Clinical Epidemiology,* 63(5), pp. 513–23

Ozguler, A., Leclerc, A., Landre, M. F., Pietri-Taleb, F. and Niedhammer, I. (2000). Individual and occupational determinants of low back pain according to various definitions of low back pain. *Journal of epidemiology and community health,* 54(3), pp. 215-220.

Palmer, K.T., Griffin, M.J., Syddall, H.E., Pannett, B., Cooper, C. and Coggon, D. (2003) The relative importance of whole body vibration and occupational lifting as risk factors for low-back pain. *Occupational and Environmental Medicine,* 60(10), pp. 715–21.

Parker, S.L., Godil, S.S., Mendenhall, S.K. Zuckerman, S.L., Shau, D.N. and McGirt, M.J. (2014). Two-year comprehensive medical management of degenerative lumbar spine disease (lumbar spondylolisthesis, stenosis, or disc herniation): a value analysis of cost, pain, disability, and quality of life. Journal of Neurosurgery. *Spine,* 21(2), pp. 143-149.

Parker, S.L., Anderson, L.H., Nelson, T. and Patel, V.V. (2015). Cost-effectiveness of three treatment strategies for lumbar spinal stenosis: Conservative care, laminectomy, and the Superion interspinous spacer. *International Journal of Spine Surgery,* 9(28), pp. 1-11.

Parkinson, B., Goodall, S. and Thavaneswaran, P. (2013). Cost-effectiveness of lumbar artificial intervertebral disc replacement: driven by the choice of comparator. *ANZ Journal of Surgery*, 83(9), pp. 669-675.

Parson, S., Ingram, M., Clarke-Cornwell, A.M. and Symmons, D.P.M. (2011) *A Heavy Burden The occurrence and impact of musculoskeletal conditions in the United Kingdom today*. Arthritis Research UK Epidemiology Unit [online] Available at (www.escholar.manchester.ac.uk/uk-ac-man-scw:123774) [Accessed 15th June 2017].

Patel, V.V., Whang, P.G., Haley, T.R., Bradley, W.D., Nunley, P.D., Miller, L.E., Block, J.E. and Geisler, F.H. (2014). Two-year clinical outcomes of a multicenter randomized controlled trial comparing two interspinous spacers for treatment of moderate lumbar spinal stenosis. *BMC Musculoskelet Disorders,* 15(221).

Patrick, D.L., Deyo, R.A., Atlas, S.J., Singer, D.E., Chapin, A. and Keller, R.B. (1995). Assessing health related quality of life in patients with sciatica. Spine, 20(17), pp. 1899-909.

Peat, G., Thomas, E., Wilkie, R. and Croft, P. (2006). Multiple joint pain and lower extremity disability in middle and old age. *Disability and rehabilitation,* 28(24), pp. 1543-1549.

Peñaloza Ramos, M. C., Barton, P., Jowett, S. snd Sutton, A. J. (2015). A systematic review of research guidelines in decision-analytic modelling**,** *Value in Health*, 18(4), pp. 512-529.

Personal Social Services Research Unit (PSSRU). (2016) Unit costs of Health and Social Care 2016. [online]. PSSRU. Available at https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2016/ [Accessed 20/09/2017]. Health Service cost index

Petrou, S. and Gray, A (2011) Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. Collecting economic data at the same time as evidence of effectiveness maximises the information available for analysis but requires proper consideration at the design stage, *BMJ,* 342:d1548.

Petrou S, Gray A. (2011) Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ*, 342: d1766

Petrou, S., Murray, L., Cooper, P. and Davidson, L. L. (2002). The accuracy of self-reported health care resource utilization in health economic studies. *International Journal of Technology Assessment in Health Care,* 18(3), pp. 705–40.

Peul, W., Brand, R., Thomeer, R. and Koes, B. (2008). Improving prediction of 'inevitable' surgery during non-surgical treatment of sciatica. *Pain,* 138(3), pp. 571–6.

Philips, Z., Ginnelly, L., Sculpher, M., Claxton, K., Golder, S., Riemsma, R., Woolacoot, N. and Glanville, J. (2006). Good practice guidelines for decision-analytic modelling in health technology assessment. *Pharmacoeconomics,* 24(4), pp. 355–71.

Picavet, H. S., Struijs, J. N. and Westert, G. P. (2008). Utilization of health resources due to low back pain: survey and registered data compared. *Spine,* 33(4), pp. 436-444.

Pietri-Taleb, F., Riihimaki, H., Viikari-Juntura, E., Lindstrom, K. and Moneta, G. B. (1995). The role of psychological distress and personality in the incidence of sciatic pain among working men. *American Journal of Public Health,* 85(4), pp. 541-545.

Plenet, A., Gourmelen, J., Chastang, J. F., Ozguler, A., Lanoe, J. L. and Leclerc, A. (2010). Seeking care for lower back pain in the French population aged from 30 to 69: the results of the 2002-2003 Decennale Sante survey. *Annals of physical and rehabilitation medicine,* 53(4), pp. 224-31.

Poetscher, A. W., Gentil, A. F., Lenza, M. and Ferretti, M. (2014). Radiofrequency denervation for facet joint low back pain: a systematic review. *Spine,* 39(14), pp. E842-9.

Puts, M. T., Deeg, D. J., Hoeymans, N., Nusselder, W. J. and Schellevis, F. G. (2008). Changes in the prevalence of chronic disease and the association with disability in the older Dutch population between 1987 and 2001. *Age and Ageing,* 37(2), pp. 187-193.

Qiao, Q., Tervahauta, M., Nissinen, A. and Tuomilehto, J. (2000). Mortality from all causes and from coronary heart disease related to smoking and changes in smoking during a 35-year follow-up of middle-aged Finnish men. *European Heart Journal,* 21(19), pp. 1621-1626

Qin, Z., Liu, X., Wu, J., Zhai, Y. and Liu, Z. (2015) Effectiveness of Acupuncture for Treating Sciatica: A Systematic Review and Meta-Analysis. *Evidence-based complementary and alternative medicine***,** 2015 (425108).

Raiffa, H. (1968) *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* New York: Addison-Wesley.

Ramond, A., Bouton, C., Richard, I., Roquelaure, Y., Baufreton, C., Legrand, E. and Huez, J. F. (2011). Psychosocial risk factors for chronic low back pain in primary care--a systematic review. *Family practice,* 28(1), pp. 12-21.

Ramond-Roquin, A., Bodin, J., Serazin, C., Parot-Schinkel, E., Ha, C., Richard, I., Le Manach, A. P., Fouquet, N. and Roquelaure, Y. (2015a). Biomechanical constraints remain major risk factors for low back pain. Results from a prospective cohort study in French male employees. *The Spine Journal,* 15(4), pp. 559-569.

Ramond-Roquin, A., Bouton, C., Begue, C., Petit, A., Roquelaure, Y. and Huez, J. F. (2015b). Psychosocial Risk Factors, Interventions, and Comorbidity in Patients with Non-Specific Low Back Pain in Primary Care: Need for Comprehensive and Patient-Centered Care. *Frontiers in medicine,* 2:73.

Ramsey, S.D., Willke, R.J., Glick, H., Reed, S.D., Augustovski, F., Jonsson, B., Briggs, A. and Sullivan, S.D. (2015). Cost-effectiveness analysis alongside clinical trials II—an ISPOR Good Research Practices Task Force report. *Value in Health,* 18(2), pp. 161–72.

Rice, D,P. (1994) Cost-of-illness studies: fact or fiction? *Lancet,* 344, pp. 1519–1520.

Riihimäki, H., Viikari-Juntura, E., Moneta, G., Kuha, J., Videman, T. and Tola, S. (1994). Incidence of Sciatic Pain Among Men in Machine Operating, Dynamic Physical Work, and Sedentary Work: A Three-Year Follow-Up. *Spine,* 19(2), pp. 138-142.

Rivinoja, A. E., Paananen, M. V., Taimela, S. P., Solovieva, S., Okuloff, A., Zitting, P., Jarvelin, M. R., Leino-Arjas, P. and Karppinen, J. I. (2011). Sports, smoking, and overweight during adolescence as predictors of sciatica in adulthood: a 28-year follow-up study of a birth cohort. *American Journal of Epidemiology,* 173(8), pp. 890-897.

Rizzo, J. A., Abbott, T. A. and Berger, M. L. (1998) The labor productivity effects of chronic backache in the United States. Medical Care, 36 (10), pp. 1471-1488

Roberts, M., Russell, L.B., Paltiel, A.D., Chambers, M., McEwan, P. and Krahn, M. (2012). Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. Medical Decision Making. 32(5), pp. 678-689.

Roberts, T.E., Robinson, S., Barton, P., Bryan, S. and Low, N. (2006). Screening for Chlamydia trachomatis: a systematic review of the economic evaluations and modelling. Sexually Transmitted Infections, 82: pp. 193–200

Roland, M., and Fairbank, J. (2000) The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire, Spine, 25(24), pp. 3115–3124.

Rubenstein, S, M., van Middelkoop, M., Assendelft, W. J., de Boer, M.R. and van Tulder, M.W. (2011) Spinal manipulative therapy for chronic low-back pain. Spine, 16(2), CD008112

Russell, L.B., Gold, M.R., Siegel, J.E., Daniels, N. and Weinstein, M.C. (1996). The role of cost-effectiveness analysis in health and medicine. JAMA, 276(14), pp.1172-1177

Sackett, D.L (1996) Evidence-Based Medicine. London. Churchill

Samuelson, P.A and Nordhaus, W.D (2004). Economics, 18th edition. New York: Tata McGraw Hill

Savettieri, G., Salemi, G., Rocca, W. A., Meneghini, F., D'arpa, A., Morgante, L., Coraci, M. A., Reggio, A., Grigoletto, F. and Di Perri, R. (1996). Prevalence of lumbosacral radiculopathy in two Sicilian municipalities. Sicilian Neuro-Epidemiologic Study (SNES) Group. Acta Neurologica Scandinavica, 93(6), pp. 464-469.

Savigny P, Kuntze S, Watson P, et al. (2009) Low back pain: early management of persistent non-specific low back pain. London: National Institute of Clinical Evidence, 2009.

Schaafsma, F. G., Whelan, K., Van Der Beek, A. J., Van Der Es-Lambeek, L. C., Ojajärvi, A. and Verbeek, J. H. (2013). Physical conditioning as part of a return to work strategy to reduce sickness absence for workers with back pain. Cochrane Database of Systematic Reviews, 30(8).

Schmier, J., Halevi, M., Maislin, G. and Ong, K. (2014). Comparative cost effectiveness of Coflex interlaminar stabilization versus instrumented posterolateral lumbar fusion for the treatment of lumbar spinal stenosis and spondylolisthesis. ClinicoEconomics and Outcomes Research, 6(1), pp. 125-131.

Schonstein, E., Kenny, D., Keating, J., Koes, B. and Herbert, R. D. (2003). Physical conditioning programs for workers with back and neck pain: a cochrane systematic review. Spine, 28(19), E391-5.

Schott, A. M., Ganne, C., Hans, D., Monnier, G., Gauchoux, R., Krieg, M. A., Delmas, P. D., Meunier, P. J. and Colin, C. (2007). Which screening strategy using BMD measurements would be most cost effective for hip fracture prevention in elderly women?

A decision analysis based on a Markov model. *Osteoporosis International, 18(2), pp. 143-151.*

Sculpher, M., Fenwick, E. and Claxton, K. (2000). Assessing quality in decision analytic cost-effectiveness models. Pharmacoeconomics, 17(5), pp. 461-477.

Sculpher, M.J., Pang, F.S., Manca, A., Drummond, M.F., Golder, S., Urdahl, H., Davies, L.m. and Eastwood, A. (2004). Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technology Assessment, 8(49), pp. 1–192.

Sculpher, M. and Drummond, M. (2005) Common methodological flaws in economic evaluations. Medical Care, 43(7), pp. 5-14.

Sculpher, M.J., Claxton, K., Drummond, M., McCabe, C. (2006). Whither trial-based economic evaluation for health care decision making? Health Economics, 15(7), pp.677–687.

Sculpher, M. (2015). Clinical Trials Provide Essential Evidence, but Rarely Offer a Vehicle for Cost-Effectiveness Analysis. Value in health, 18(2), pp. 141–142.

Sen, A. (1986). Commodities and capabilities. Amsterdam: North Holland.

Sen, A. (1992). Inequality reexamined. New York: Russell Sage Foundation.

Sendi, P.P., Craig, B.A., Pfluger, D., Gafni, A. and Bucher, H.C. (1999). Systematic validation of disease models for pharmacoeconomic evaluations: Swiss HIV Cohort Study. Journal of Evaluation in Clinical Practice, 5(3), pp. 283-95.

Seixas, B. V. (2017). Welfarism and extra-welfarism: a critical overview. Cadernos de Saúde Pública, 33( 8 ), e00014317. Sheldon, T.A (1996) Problems of using modelling in the economic evaluation of health care. Health Economics, 5(1), pp. 1-11.

Sherman, K. J., Cherkin, D. C., Erro, J., Miglioretti, D. L. and Deyo, R. A. (2005). Comparing Yoga, Exercise, and a Self-Care Book for Chronic Low Back Pain, A Randomized, Controlled Trial. Annals of Internal Medicine, 143(12), pp. 849-856.

Shiri, R., Karppinen, J., Leino-Arjas, P., Solovieva, S., Varonen, H., Kalso, E., Ukkola, O. and Viikari-Juntura, E. (2007). Cardiovascular and lifestyle risk factors in lumbar radicular pain or clinically defined sciatica: a systematic review. European Spine Journal, 16(12), pp. 2043-2054.

Shiri, R., Karppinen, J., Leino-Arjas, P., Solovieva, S. and Viikari-Juntura, E. (2010a). The association between obesity and low back pain: a meta-analysis. American Journal of Epidemiology, 171(2), pp. 135-154.

Shiri, R., Karppinen, J., Leino-Arjas, P., Solovieva, S. and Viikari-Juntura, E. (2010b). The association between smoking and low back pain: a meta-analysis. The American Journal of Medicine, 123(1), 87.e7.e35.

Siebert, U., Alagoz, O., Bayoumi, A.M., Jahn, B., Owens, D.K., Cohen, D.J. and Kuntz, K.m. (2012). State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. Value in Health, 15(6), pp. 812-820.

Simonetti, R., Himmelweit, S., Mackintosh, M. and Stone, H. (2010). Industrial capitalism, markets and economics in Simonetti, R et al. [eds] (2010) Doing Economics: People, Markets and Policy. Milton Keynes: Open University.

Skidmore, G., Ackerman, S.J., Bergin, C., Ross, D., Butler, J., Suthar, M. and Rittenberg, J. (2011). Cost-effectiveness of the X-STOP (R) Interspinous Spacer for Lumbar Spinal Stenosis A Comparison with Conservative Care and Laminectomy. Spine, 36(5): E345-E356.

Soto, J. (2002) Health economic evaluations using decision analytic modeling: principles and practices. Utilization of a checklist to their development and appraisal. International Journal Technology Assessment, 18(1), pp. 94-111.

Spiegelhalter, D.J. and Best, N.G. (2003) Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Statistics in Medicine, 22(23), pp. 3687–3709.

Stanton, T. R., Henschke, N., Maher, C. G., Refshauge, K. M., Latimer, J. and Mcauley, J. H. (2008). After an episode of acute low back pain, recurrence is unpredictable and not as common as previously thought. Spine, 33(26), pp. 2923-2928.

Stevens, C. D., Dubois, R. W., Larequi-Lauber, T. and Vader, J. P. (1997). Efficacy of lumbar discectomy and percutaneous treatments for lumbar disc herniation. Sozial- und Praventivmedizin, 42(6), pp. 367-379.

Stevenson, M.D., Oakley, J. and Chilcott, J.B (2004). Gaussian process modelling in conjunction with individual patient simulation modelling. A case study describing the calculation of cost-effectiveness ratios for the treatment for established osteoporosis. Medical Decision Making, 24(1), pp. 89–100.

Stevenson, M.D., Brazier, J.E., Calvert, N.W., Lloyd-Jones, M., Oakley, J. and Kanis, J.A. (2005). Description of an individual patient methodology for calculating the cost-effectiveness of treatments for osteoporosis in women. Journal of Operational Research Society, 56(2), pp. 214–221.

Stevenson, M., Davis, S., Lloyd-Jones, M. and Beverley, C. (2007). The clinical effectiveness and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women, *Health Technology Assessment,* 11(4), pp. 1-80.

Stromqvist, B., Fritzell, P., Hagg, O. and Jonsson, B. (2008). National Swedish Register for Lumbar Spine Surgery report 2008. Available from: www.4s.nu/pdf/Ryggregisterrapport_2008_eng_version.pdf. [Authors Accessed July 15, 2014].

Strong, M., Oakley, J.E. and Brennan, A. (2014) Estimating multi-parameter partial Expected Value of Perfect Information from a probabilistic sensitivity analysis sample: a non-parametric regression approach. *Medical Decision Making*, 34(3), PP. 311-26.

Sutton, A.J. and Abram, K.R. (2001). Bayesian methods in meta-analysis and evidence synthesis. Statistical Methods in Medical Research, 10(4), pp. 277-303.

Svensson, H. O. and Andersson, G. B. (1982). Low back pain in forty to forty-seven year old men. I. Frequency of occurrence and impact on medical services. Scandinavian journal of rehabilitation medicine, 14(2), pp. 47-53.

Taguchi, T., Igarashi, A., Watt, S., Parsons, B., Sadosky, A., Nozawa, K., Hayakawa, K., Yoshiyama, T., Ebata, N. and Fujii, K. (2015). Effectiveness of pregabalin compared with usual care for the treatment of chronic low back pain with a neuropathic component in Japan; the patient's perspective. Journal of Pain Research, 8, pp, 487–497.

Tamcan, O., Mannion, A. F., Eisenring, C., Horisberger, B., Elfering, A. and Muller, U. (2010). The course of chronic and recurrent low back pain in the general population. Pain, 150(3), pp. 451-457.

Tapp, S.J., Martin, B.I., Tosteson, T.D., Lurie, J.D., Weinstein, M.C., Deyo, R.A., Mirza, S.K. and Tosteson, A.N.A. (2018). Understanding the value of minimally invasive procedures for the treatment of lumbar spinal stenosis: the case of interspinous spacer devices. Spine Journal, 18(4), pp. 584-592.

Taylor, R., and Taylor, R. (2005). Spinal cord stimulation for failed back surgery syndrome: A decision-analytic model and cost-effectiveness analysis, International Journal of Technology Assessment in Health, 21(3) pp. 351 – 358.

Thomas, K.J., MacPherson, H., Thorpe, L., Brazier, J., Fitter, M., Campbell, M.J., Roman, M., Walters, S.J. and Nicholl, J. (2006). Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. BMJ, 2006; 623.

Tosteson, A.N., Tosteson, T.D., Lurie, J.D., Abdu, W., Herkowitz, H., Andersson, G., Albert, T., Bridwell, K., Zhao, W., Grove, M.rR., Weinstein, M.C. and Weinstein, J.N. (2011). Comparative effectiveness evidence from the spine patient outcomes research trial surgical versus nonoperative care for spinal stenosis, degenerative spondylolisthesis, and intervertebral disc herniation. Spine, 36(24), pp. 2061–2068.

Tosteson, A.N., Lurie, J.D., Tosteson, T.D., Skinner, J.S., Herkowitz, H., Albert, T., Boden, S.D., Bridwell, K., Longley, M., Andersson, G.B., Blood, E.A., Grove, M.R., Weinstein, J.N. (2008). Surgical treatment of spinal stenosis with and without degenerative spondylolisthesis: cost-effectiveness after 2 years. Annals of Internal Medicine, 149(12), pp. 845–853.

Tosteson, A.N., Skinner, J.S., Tosteson, T.D., Lurie, J.D., Andersson, G.B., Berven, S., Grove, M.R., Hanscom, B., Blood, E.A., Weinstein, J.N. (2008). The cost effectiveness of surgical versus nonoperative treatment for lumbar disc herniation over two years: evidence from the Spine Patient Outcomes Research Trial (SPORT). Spine, 33(19), pp. 2108-2115.

Tubach, F., Beauté, J. and Leclerc, A. (2004). Natural history and prognostic indicators of sciatica. Journal of clinical epidemiology, 57(2), pp. 174-179.

Udeh, B.L., Costandi, S., Dalton, J.E., Ghosh, R., Yousef, H. and Mekhail, N. (2015). The 2-year cost-effectiveness of 3 options to treat lumbar spinal stenosis patients. Pain Practice, 15(2):107-116.

Valat, J. P., Genevay, S., Marty, M., Rozenberg, S. and Koes, B. (2010). Sciatica. Best practice and research. Clinical rheumatology, 24(2), pp. 241-252.

van den Hout, W.B., Peul, W.C., Koes, B.W., Brand, R., Kievit, J. and Thomeer, R.T. (2008). The Hague Spine Intervention Prognostic Study Group. Prolonged conservative care versus early surgery in patients with sciatica from lumbar disc herniation: cost-utility analysis alongside a randomised controlled trial. BMJ, 336(7657), pp. 1351-1354.

van Hout, B., Janssen, M.F., Feng, Y.S. Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P. and Pickard, A.S. (2012). Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health, 15(5), pp. 708-15.

van der Windt, D., Hay, E.M., Jellema, P. and Main, C.J. (2008). Psychosocial interventions for low back pain in primary care: lessons learned from recent trials. Spine, 33(1), pp. 81–89

Van Oostrom, S. H., Monique Verschuren, W. M., De Vet, H. C. W. and Picavet, H. S. J. (2011). Ten year course of low back pain in an adult population-based cohort - The Doetinchem Cohort Study. European Journal of Pain, 15(9), pp. 993-998.

Van Tulder, M.W., Koes, B.W. and Bouter, L.M. (1995). A cost-of-illness study of back pain in The Netherlands. Pain, 62(2), pp. 233-240.

Van Tulder, M. W. and Waddell, G. (2005) Evidence-based medicine for non-specific low back pain. Best Practice and Research Clinical Rheumatology, 19, vii-ix.

Van Tulder, M., Peul, W. and Koes, B. (2010). Sciatica: what the rheumatologist needs to know. Nature Reviews Rheumatology, 6(3), pp. 139-145.

Van Tulder, M. W., Koes, B., Seitsalo, S. and Malmivaara, A. (2006). Outcome of invasive treatment modalities on back pain and sciatica: an evidence-based review. European Spine Journal, 15(1), S82-S92.

Vertuani, S., Nilsson, J., Borgman, B., Buseghin, G., Leonard, C., Assietti, R. and Quaraishi, N.A. (2015). A Cost-Effectiveness Analysis of Minimally Invasive versus Open Surgery Techniques for Lumbar Spinal Fusion in Italy and the United Kingdom. Value in Health, 18(6), pp. 810-816.

Verwoed, A., Luijsterburg, P., Lin, C., Jacobs, W., Koes, B. and Verhagen, A. (2013). Systematic review of prognostic factors predicting outcome in non-surgically treated patients with sciatica. European Journal of Pain, 17(8), pp. 1126–37.

Vingård, E., Mortimer, M., Wiktorin, C., Pernold, G., Fredriksson, K., Németh, G. and Alfredsson, L. (2002). Seeking care for low back pain in the general population. A two-year follow-up study: Results from the MUSIC-Norrtälje study. Spine, 27(19), pp. 2159-2165.

Von Korff, M., Crane, P., Lane, M., Miglioretti, D. L., Simon, G., Saunders, K., Stang, P., Brandenburg, N. and Kessler, R. (2005). Chronic spinal pain and physical-mental comorbidity in the United States: results from the national comorbidity survey replication. Pain, 113(3), pp. 331-339.

Vroomen, P. C. A. J., De Krom, M. C., Slofstra, P. D. and Knottnerus, J. A. (2000). Conservative treatment of sciatica: a systematic review. Clinical Spine Surgery, 13(6), pp. 463-469.

Waddell, G., Mcculloch, J. A., Kummel, E. D. and Venner, R. M. (1980). Nonorganic physical signs in low-back pain. Spine, 5(2), pp. 117-125.

Walker, B. F. (2000). The prevalence of low back pain: A systematic review of the literature from 1966 to 1998. Journal of Spinal Disorders, 13(3), pp. 205-217.

Weinstein, M.C., Toy, E.L., Sandberg, E.A., Neumann, P.J., Evans, J.S., Kuntz, K.M., Graham, J.D. and Hammitt, J.K. (2001) Modelling for health care and other policy decisions: uses, roles, and validity. Value in Health, 4(5), pp. 348-61

Weinstein, M.C. (2006) Recent Developments in Decision-Analytic Modelling for Economic Evaluation. Pharmacoeconomics 2006; 24 (11): 1043-1053

Weinstein JN, Lurie JD, Tosteson TD, et al., (2006). Surgical versus nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT) observational cohort. JAMA, 296(20), pp. 2451–2459.

Weinstein, J.N., Lurie, J.D., Tosteson, T.D., Hanscom, B., Tosteson, A.N., Blood, E.A., Birkmeyer, N.J., Hilibrand, A.S., Herkowitz, H., Cammisa, F.P., Albert, T.J., Emery, S.E., Lenke, L.G., Abdu, W.A., Longley, M., Errico, T.J. and Hu, S.S. (2007). Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. New England Journal of Medicine, 356(22), pp. 2257–2270.

Weinstein, J.N., Lurie, J.D., Tosteson, T.D., Tosteson, A.N., Blood, E.A., Abdu, W.A., Herkowitz, H., Hilibrand, A., Albert, T., Fischgrund, J. (2008) Surgical versus nonoperative treatment for lumbar disc herniation: four-year results for the Spine Patient Outcomes Research Trial (SPORT). Spine, 33(25), pp. 2789–2800.

Weinstein, J.N., Lurie, J.D., Tosteson, T.D., Zhao, W., Blood, E.A., Tosteson, A.N., Birkmeyer, N., Herkowitz, H., Longley, M., Lenke, L., Emery, S. and Hu, S.S. (2009). Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis.

Four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. The Journal of Bone and Joint Surgery, 91(6), pp. 1295-1304.

Weiner, S. S. and Nordin, M. (2010) Prevention and management of chronic back pain. Back Pain and Non-Inflammatory Spinal Disorders, 24(2), pp. 267-279.

Whitehurst, D.G., Bryan, S., Lewis, M., Hill, J., Hay, E.M. (2012). Exploring the cost–utility of stratified primary care management for low back pain compared with current best practice within risk-defined subgroups. Annals of the rheumatic diseases. 71(11), pp. 1796-802.

Whitehurst, D.G., Bryan, S., Lewis, M., Hay, E.M., Mullis, R. and Foster, N.E. (2015). Implementing Stratified Primary Care Management for Low Back Pain: Cost-Utility Analysis Alongside a Prospective, Population-Based, Sequential Comparison Study. Spine, 40(6), pp. 405-14.

WHO (2003) The burden of musculoskeletal conditions at the start of the new millennium, Geneva: World Health Organisation Technical Report Series, 919, pp. 1–218.

Whynes, D.K., McCahon, R.A., Ravenscroft, A., Hardman, J. (2012). Cost effectiveness of epidural steroid injections to manage chronic lower back pain. BMC Anesthesiology, 12:26

Wielage, R., Bansal, M., Andrews, J.S., Wohlreich, M.M., Klein, R.W. and Happich, M. (2013) The Cost-Effectiveness of Duloxetine in Chronic Low Back Pain: A US Private Payer Perspective. Value in Health, 16(2), pp. 334-344.

Wielage, R., Bansal, M., Wilson, K., Klein, R. and Happich, M. (2013). Cost-effectiveness of duloxetine in chronic low back pain: a Quebec societal perspective. Spine, 38(11), pp. 936-946.

Wieser, S., Horisberger, B., Schmidhauser, S., Eisenring, C., Brugger, U., Ruckstuhl, A., Dietrich, J., Mannion, A. F., Elfering, A., Tamcan, O. and Muller, U. (2011). Cost of low back pain in Switzerland in 2005. The European journal of health economics, 12(5), pp. 455-467.

Wilson, E.C.F. (2015). A Practical Guide to Value of Information Analysis, Pharmacoeconomics, 33(2), pp. 105.

Wynne-Jones, G, Artus, M, Bishop, A et al. (2018). Effectiveness and costs of a vocational advice service to improve work outcomes in patients with musculoskeletal pain in primary care. Pain, 159(1), pp. 128-138

Yaghoubi, M., Moradi-Lakeh, M., Moradi-Joo, M., Rahimi-Movaghar, V., Zamani, N. and Naghibzadeh-Tahami, A. (2016). The cost-effectiveness of dynamic and static interspinous spacer for lumbar spinal stenosis compared with laminectomy. Medical Journal of the Islamic Republic of Iran, 30, pp. 339-352.

Yano, S., Hida, K., Seki, T., Aoyama, T., Akino, M. and Iwasaki, Y. (2008). A new ceramic interspinous process spacer for lumbar spinal canal stenosis. Neurosurgery, 61(1); ONS108-ONS113

Yildirim, K., Şışecıoğlu, M., Karatay, S., Erdal, A., Levent, A., Uğur, M. and Şenel, K. (2003). The effectiveness of gabapentin in patients with chronic radiculopathy. The Pain Clinic, (15), pp. 213-218.

Zucherman, J.F., Hsu, K.Y., Hartjen, C.A., Mehalic, T.F., Implicito, D.A., Martin, M.J., Johnson, D.R., Skidmore, G.A., Vessa, P.P., Dwyer, J.W., Puccio, S.T., Cauthen, J.C. and Ozuna, R.M. (2005) A multicenter, prospective, randomized trial evaluating the X-STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication. Spine, 30(12), pp. 1351–8.

# Appendices

## Appendix 1 Quality assessment in decision-analytic models: a suggested checklist (Philips et al., 2004)

The following table provides a checklist for the critical appraisal of decision-analytic models developed for health technology assessment. The format is taken from the work by Sculpher and colleagues.[7]

| Dimension of quality | | Attributes of good practice | Questions for critical appraisal |
|---|---|---|---|
| ***Structure*** | | | |
| S1 | Statement of decision problem/objective | There should be a clear statement of the decision problem prompting the analysis | Is there a clear statement of the decision problem? |
| | | The objective of the evaluation and of the model should be defined | Is the objective of the evaluation and model specified and consistent with the stated decision problem? |
| | | The primary decision-maker should be stated clearly | Is the primary decision-maker specified? |
| S2 | Statement of scope/perspective | The perspective of the model (relevant costs and consequences) should be stated clearly, and the model inputs should be consistent with the stated perspective and overall objective of the model | Is the perspective of the model stated clearly? |
| | | | Are the model inputs consistent with the stated perspective? |
| | | The scope of the decision model should be specified and justified | Has the scope of the model been stated and justified? |
| | | The outcomes of the model should reflect the perspective and scope of the model and should be consistent with the objective of the evaluation | Are the outcomes of the model consistent with the perspective, scope and overall objective of the model? |
| S3 | Rationale for structure | The structure of the model should be consistent with a coherent theory of the health condition under evaluation, and the treatment pathways (disease states or branches) should be chosen to reflect the underlying biological process of the disease in question and the impact of the intervention. The structure should not be dictated by current patterns of service provision | Is the structure of the model consistent with a coherent theory of the health condition under evaluation? |
| | | | Are the sources of data used to develop the structure of the model specified? |
| | | All sources of evidence used to develop and inform the structure of the model (i.e. the theory of disease) should be described. The structure should be consistent with this evidence | Are the causal relationships described by the model structure justified appropriately? |
| S4 | Structural assumptions | All structural assumptions should be transparent and justified. They should be reasonable in the light of the needs and purposes of the decision-maker. | Are the structural assumptions transparent and justified? |
| | | | Are the structural assumptions reasonable given the overall objective, perspective and scope of the model? |
| S5 | Strategies/comparators | There should be a clear definition of the options under evaluation | Is there a clear definition of the options under evaluation? |
| | | All feasible and practical options relating to the stated decision problem should be evaluated. Options should not be constrained by the immediate concerns of the decision-maker or data availability, or limited to current clinical practice | Have all feasible and practical options been evaluated? |
| | | | Is there justification for the exclusion of feasible options? |
| S6 | Model type | The appropriate model type will be dictated by the stated decision problem and the choices made regarding the causal relationships within the model | Is the chosen model type appropriate given the decision problem and specified causal relationships within the model? |

continued

| Dimension of quality | | Attributes of good practice | Questions for critical appraisal |
|---|---|---|---|
| S7 | Time horizon | A model's time horizon should extend far enough into the future for it to reflect important differences between options | Is the time horizon of the model sufficient to reflect all important differences between options? |
| | | It is important to distinguish between the time horizon of the model, the duration of treatment and the duration of treatment effect | Are the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified? |
| S8 | Disease states/pathways | Disease states/pathways should reflect the underlying biological process of the disease in question and the impact of interventions | Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions? |
| S9 | Cycle length | For discrete time models, the cycle length should be dictated by the natural history of disease. It should be the minimum interval over which the pathology or symptoms are expected to alter | Is the cycle length defined and justified in terms of the natural history of disease? |
| *Data* | | | |
| D1 | Data identification | Methods for identifying data should be transparent and it should be clear that the data identified are appropriate given the objectives of the model | Are the data identification methods transparent and appropriate given the objectives of the model? |
| | | There should be justification of any choices that have been made about which specific data inputs are included in a model | Where choices have been made between data sources, are these justified appropriately? |
| | | It should be clear that particular attention has been paid to identifying data for those parameters to which the results of the model are particularly sensitive | Has particular attention been paid to identifying data for the important parameters in the model? |
| | | Where expert opinion has been used to estimate particular parameters, sources and methods of elicitation should be described | Has the quality of the data been assessed appropriately? |
| | | | Where expert opinion has been used, are the methods described and justified? |
| D2 | Data modelling | All data modelling methodology should be described and based on justifiable statistical and epidemiological methods. Specific issues to consider include those listed under D2a–d, below | Is the data modelling methodology based on justifiable statistical and epidemiological techniques? |
| D2a | Baseline data | Baseline probabilities may be based on natural history data derived from epidemiological/observational studies or relate to the control group of an experimental study | Is the choice of baseline data described and justified?<br>Are transition probabilities calculated appropriately? |
| | | Rates and interval probabilities should be transformed into transition probabilities appropriately. If there is evidence that time is an important factor in the calculation of transition probabilities in state transition models, this should be incorporated | Has a half-cycle correction been applied to both cost and outcome? |
| | | If a half-cycle correction has not been used on all transitions in state transition model (costs and outcomes), this should be justified | If not, has this omission been justified? |

*continued*

| Dimension of quality | Attributes of good practice | Questions for critical appraisal |
|---|---|---|
| D2b Treatment effects | Relative treatment effects derived from trial data should be synthesised using recognised meta-analytic techniques | If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques? |
| | The methods and assumptions that are used to extrapolate short-term results to final outcomes should be documented and justified. This should include justification of the choice of survival function (e.g. exponential or Weibull forms). Alternative assumptions should be explored through sensitivity analysis | Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified? Have alternative assumptions been explored through sensitivity analysis? |
| | Assumptions regarding the continuing effect of treatment once treatment is complete should be documented and justified. If evidence regarding the long-term effect of treatment is lacking, alternative assumptions should be explored through sensitivity analysis | Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified? Have alternative assumptions been explored through sensitivity analysis? |
| D2c Costs | Costing and discounting methods should accord with standard guidelines for economic evaluation | Are the costs incorporated into the model justified? |
| | | Has the source for all costs been described? |
| | | Have discount rates been described and justified given the target decision-maker? |
| D2d Quality of life weights (utilities) | Utilities incorporated into the model should be appropriate for the specified decision problem | Are the utilities incorporated into the model appropriate? |
| | | Is the source for the utility weights referenced? |
| | | Are the methods of derivation for the utility weights justified? |
| D3 Data incorporation | All data incorporated into the model should be described and the sources of all data should be given and reported in sufficient detail to allow the reader to be aware of the type of data that have been incorporated | Have all data incorporated into the model been described and referenced in sufficient detail? |
| | Where data are not mutually consistent in the model, the choices and assumptions that have been made should be explicit and justified | Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)? |
| | The process of data incorporation should be transparent. It should be clear whether data are incorporated as a point estimate or as a distribution. If data have been incorporated as distributions as part of probabilistic analysis, the choice of distribution and its parameters should be described and justified | Is the process of data incorporation transparent? |
| | | If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified? |
| | | If data have been incorporated as distributions, is it clear that second order uncertainty is reflected? |
| D4 Assessment of uncertainty | In assessing uncertainty, modellers should distinguish between the four principal types of uncertainty | Have the four principal types of uncertainty been addressed? |
| | | If not, has the omission of particular forms of uncertainty been justified? |
| D4a Methodological | Methodological uncertainty relates to whether particular analytical steps taken in the analysis are the most appropriate | Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions? |

*continued*

| Dimension of quality | Attributes of good practice | Questions for critical appraisal |
|---|---|---|
| D4b Structural | There should be evidence that structural uncertainties have been evaluated using sensitivity analysis | Is there evidence that structural uncertainties have been addressed via sensitivity analysis? |
| D4c Heterogeneity | It is important to distinguish between uncertainty resulting from the process of sampling from a population and variability due to heterogeneity (i.e. systematic differences between patient subgroups) | Has heterogeneity been dealt with by running the model separately for different subgroups? |
| D4d Parameter | Where data have been incorporated into the model as point estimates, the ranges used for sensitivity analysis should be stated and justified | Are the methods of assessment of parameter uncertainty appropriate? |
| | Probabilistic analysis is the most appropriate method of handling parameter uncertainty because it facilitates assessment of the joint effect of uncertainty over all parameters (see 'Data incorporation', p. 92) | If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified? |
| **Consistency** | | |
| C1 Internal consistency | There should be evidence that the internal consistency of the model has been evaluated in terms of its mathematical logic | Is there evidence that the mathematical logic of the model has been tested thoroughly before use? |
| C2 External consistency | The results of a model should be explicable. Either results should make intuitive sense or counterintuitive results should be fully explained | Are any counterintuitive results from the model explained and justified? |
| | All relevant available data should be incorporated into a model. Data should not be withheld for purposes of assessing external consistency | If the model has been calibrated against independent data, have any differences been explained and justified? |
| | The results of a model should be compared with those of previous models and any differences should be explained | Have the results of the model been compared with those of previous models and any differences in results explained? |

## Appendix 2: Search strategies for Chapter 4

### Search Strategy for Medline, hosted by OVID.

backache.ti,ab.

backache/

(spin* adj5 (disease or stenosis)).ti,ab.

exp Spinal Stenosis/

spinal diseases/

(ischi* or sciatic*).ti,ab.

sciatica/

radicul*.ti,ab.

Radiculopathy/

polyradiculopath*.ti,ab.

Polyradiculopathy/

(nerve adj5 (pain or syndrome*)).ti,ab.

Nerve Compression Syndromes/

spondylosis.ti,ab.

spondylosis/

spondylitis.ti,ab.

spondylitis/

exp Intervertebral Disc Displacement/ or exp Intervertebral Disc Degeneration/

(disc* adj5 (displacement* or protru* or avulsion or degeneration*)).ti,ab.

herniat*.ti,ab.

back injur*.ti,ab.

exp Back Injuries/

(leg adj5 pain).ti,ab.

(refer* adj5 pain).ti,ab.

lumbago.ti,ab.

lumbago/

1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26

(spinal or spine).ti,ab.

spine/

spinal canal/

(lumba$ or lumbo$).ti,ab.

lumbar vertebrae/

exp Intervertebral Disc/

Neuropathic.ti,ab.

exp Spinal Nerve Roots/

((disc or nerve) adj5 sacral).ti,ab.

exp back/

back.ti,ab.

28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38

exp pain/

(pain or painful or ach*).ti,ab.

40 or 41

39 and 42

27 or 43

expenditure$.ti,ab.

econom*.ti,ab.

health care rationing/

(cost or costs).ti,ab.

exp economics/

"Quality of Life"/

models, economic/

value of information analys$.ti,ab.

45 or 46 or 47 or 48 or 49 or 50 or 51 or 52

monte carlo.mp.

state-transition.mp.

markov.mp.

(decision* adj5 (analytic* or analys#s or tree*)).mp.

exp decision theory/

individual sampling.mp.

individual patient level.mp.

system dynamic*.mp.

discrete event simulation.mp.

model*.ti,ab.

54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63

44 and 53 and 64

limit 65 to humans

limit 66 to english language

**Search strategy for PsychINFO, hosted by OVID.**

backache.ti,ab.

(spin* adj5 (disease or stenosis)).ti,ab.

(ischi* or sciatic*).ti,ab.

radicul*.ti,ab.

polyradiculopath*.ti,ab.

(nerve adj5 (pain or syndrome*)).ti,ab.

spondylosis.ti,ab.

spondylitis.ti,ab.

(disc* adj5 (displacement* or protru* or avulsion or degeneration*)).ti,ab.

herniat*.ti,ab.

back injur*.ti,ab.

(leg adj5 pain).ti,ab.

(refer* adj5 pain).ti,ab.

lumbago.ti,ab.

exp Back Pain/

exp neuropathic pain/

1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16

(spinal or spine).ti,ab.

exp Spinal Nerves/

(lumba$ or lumbo$).ti,ab.

exp Lumbar Spinal Cord/

Neuropathic.ti,ab.

exp neuropathy/

((disc or nerve) adj5 sacral).ti,ab.

exp "back anatomy"/

back.ti,ab.

18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26

(pain or painful or ach*).ti,ab.

exp chronic pain/ or exp pain/

28 or 29

30 and 27

31 or 17

expenditure$.ti,ab.

econom*.ti,ab.

exp Economics/

(cost or costs).ti,ab.

exp "Costs and cost analysis"/

quality of life.ti,ab.

exp "Quality of life"/

value of information.ti,ab.

33 or 34 or 35 or 36 or 37 or 38 or 39 or 40

monte carlo.mp.

state transition.mp.

markov.mp.

exp Markov chains/

(decision* adj5 (analytic* or analys#s or tree*)).mp.

decision theory.mp.

individual sampling.mp.

individual patient level.mp.

system dynamic*.mp.

discrete event simulation.mp.

model*.ti,ab.

42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52

32 and 41 and 53

limit 54 to human

**Search strategy for NHS EED, and HTA databases, hosted by Cochrane**

ID Search Hits

#1 MeSH descriptor: [Back Pain] explode all trees

#2 back pain:ti,ab,kw

#3 backache*:ti,ab,kw

#4 MeSH descriptor: [Spinal Diseases] explode all trees

#5 spin* near/5 (disease or stenosis):ti,ab,kw

#6 spondyl*:ti,ab,kw

#7 "Intervertebral disc":ti,ab,kw

#8 disc* near/5 (displacement* or protu* or prolapse* or avulsion or degeneration* or hernia*):ti,ab,kw

#9 MeSH descriptor: [Sciatica] explode all trees

#10 sciatic*:ti,ab,kw

#11 MeSH descriptor: [Radiculopathy] explode all trees

#12 radicul*:ti,ab,kw

#13 polyradicul*:ti,ab,kw

#14 MeSH descriptor: [Nerve Compression Syndromes] explode all trees

#15 nerve near/5 (pain or syndrome* or compression* or inflammation*):ti,ab,kw

#16 MeSH descriptor: [Back Injuries] explode all trees

#17 back injur*:ti,ab,kw

#18 MeSH descriptor: [Pain, Referred] explode all trees

#19 refer* near/5 pain:ti,ab,kw

#20 MeSH descriptor: [Neuralgia] explode all trees

#21 MeSH descriptor: [Sacrum] explode all trees

#22 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21

#23 MeSH descriptor: [Leg] explode all trees

#24 leg:ti,ab,kw

#25 MeSH descriptor: [Spine] explode all trees

#26 spin*:ti,ab,kw

#27 lumbar*:ti,ab,kw

#28 lumbo*:ti,ab,kw

#29 MeSH descriptor: [Intervertebral Disc] explode all trees

#30 intervertebral:ti,ab,kw

#31 MeSH descriptor: [Spinal Nerve Roots] explode all trees

#32 neuropathic:ti,ab,kw

#33 MeSH descriptor: [Back] explode all trees

#34 back:ti,ab,kw

#35 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34

#36 MeSH descriptor: [Pain] explode all trees

#37 pain:ti,ab,kw

#38 #36 or #37

#39 #35 and #38

#40 #22 or #39

#41 MeSH descriptor: [Monte Carlo Method] explode all trees

#42 "Monte Carlo"

#43 MeSH descriptor: [Markov Chains] explode all trees

#44 Markov

#45 MeSH descriptor: [Decision Support Techniques] explode all trees

#46 MeSH descriptor: [Decision Theory] explode all trees

#47 decision near/5 (analytic* or analy?s or tree* or theor*)

#48 Individual near/5 (sampling or patient level)

#49 MeSH descriptor: [Systems Analysis] explode all trees

#50 system near/2 dynamic*

#51 Dynamic near/2 transition*

#52 "Discrete Event Simulation"

#53 Model:ti,ab,kw

#54 #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53

#55 #40 and #54 in Technology Assessments and Economic Evaluations

## Appendix 3: Search strategies for Chapter 5

### Search strategy for Embase

1. (econom* adj3 (evaluation* or analy* or stud* or health)).ti,ab.
2. health economics/ or device economics/
3. (cost* adj3 (effect* or benefit* or utilit* or consequence* or minimi* or analy*)).mp.
4. exp cost effectiveness analysis/
5. 1 or 2 or 3 or 4
6. "state transition".mp.
7. exp monte carlo method/
8. markov.mp.
9. "decision tree".mp.
10. (decision* adj3 (analytic* or analys#s)).mp.
11. decision theory/
12. "monte carlo".mp.
13. (individual* adj3 (patient* or sampling)).mp.
14. (dynamic adj3 (system* or transition*)).mp.
15. (simulation adj3 (individual* or patient* or discrete)).mp.
16. (model* adj3 (based or economic* or decision* or simulat* or analy*)).mp.
17. 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
18. exp personalized medicine/
19. ("system medicine" or "systems medicine").mp.
20. (stratifi* adj3 (care or medicine* or treatment* or therap* or intervention*)).mp.
21. (personal* adj3 (care or medicine* or treatment* or therap* or intervention*)).mp.
22. (individual* adj3 (care or medicine* or treatment* or therap* or intervention*)).mp.
23. (precisi* adj3 (care or medicine* or treatment* or therap* or intervention*)).mp.
24. (target* adj3 (care or medicine* or treatment* or therap* or intervention*)).mp.
25. (guide* adj3 (care or medicine* or treatment* or therap* or intervention*)).mp.
26. 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25

27. (risk adj3 (classifi* or stratifi* or assess*)).mp.

28. (care or medicine* or treatment* or therap* or intervention*).mp.

29. 27 and 28

30. 26 or 29

31. 5 and 17 and 30

32. animal/ not human/

33. 31 not 32

34. limit 33 to embase

**Search strategy for DARE, CDSR, HTA, NHS EED**

#1 MeSH descriptor: [Monte Carlo Method] this term only

#2 "Monte Carlo"

#3 MeSH descriptor: [Markov Chains] this term only

#4 Markov

#5 MeSH descriptor: [Decision Support Techniques] explode all trees

#6 MeSH descriptor: [Decision Theory] explode all trees

#7 decision near/3 (analytic* or analys?s or theor*)

#8 "decision tree"

#9 MeSH descriptor: [Systems Analysis] explode all trees

#10 Individual* near/3 (sampling or patient*)

#11 dynamic near/3 (system* or transition*)

#12 simulation* near/3 (individual* or patient* or discrete)

#13 model* near/3 (based or decision* or analy* or economic* or simulat*)

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

#15 Econom* near/3 (evaluation* or analy* or stud* or health*):ti,ab,kw

#16 MeSH descriptor: [Health Expenditures] explode all trees

#17 MeSH descriptor: [Health Care Rationing] explode all trees

#18 MeSH descriptor: [Costs and Cost Analysis] explode all trees

#19 cost* near/3 (effect* or benefit* or utilit* or consequence* or minimi* or analy*)

#20 #15 or #16 or #17 or #18 or #19

#21 stratifi* near/3 (care or treatment* or medicine* or therap* or intervention*)

#22 individual* near/3 (care or treatment* or medicine* or therap* or intervention*)

#23 personal* near/3 (care or treatment* or medicine* or therap* or intervention*)

#24 target* near/3 (care or treatment* or medicine* or therap* or intervention*)

#25 precisi* near/3 (care or treatment* or medicine* or therap* or intervention*)

#26 guide* near/3 (care or treatment* or medicine* or therap* or intervention*)

#27 "systems medicine" or "system medicine"

#28 #21 or #22 or #23 or #24 or #25 or #26 or #27

#29 risk near/3 (classif* or strati* or assess*)

#30 (care or treatment* or medicine* or therap* or intervention*)

#31 #29 and #30

**Appendix 4: Expert consultations and the model building process**

| | | |
|---|---|---|
| April 2018 | Keele Stratified Care Research Group (SCRG) KK, SJ, RaO, and ML also present | Discussion of health states, RMDQ cut-offs, model assumptions, model population, costs, and sensitivity analyses |
| August 2018 | Dr John Bedson (GP), Keele | Discussion of health states, RMDQ cut-offs, model assumptions, and sensitivity analyses |
| August 2018 | Dr Elizabeth Cottrell (GP), Keele | Discussion of health states, RMDQ cut-offs, model |

| | | |
|---|---|---|
| | | assumptions, and sensitivity analyses |
| March 2019 | Dr Katrina Humphreys (physiotherapist), and Dr Adrian Chudyk (GP), Keele | Discussion of health states, RMDQ cut-offs, model assumptions, and sensitivity analyses |
| April 2018 | Dr. Pelham Barton (Health Economist), Birmingham | Initial model conceptualisation ideas |
| June 2018 | Dr. Pelham Barton (Health Economist), Birmingham | Feedback upon model structure, help with transition probabilities |
| February 2019 | Dr. Pelham Barton (Health Economist), Birmingham | Value of Information Analyses |

**Appendix 5: Function at 7-years, dependent upon both function at base and function at twelve months.**

7YR Base function if base function =0 AND m12 function equals 2

| | n | % |
|---|---|---|
| Low | 0 | 0 |
| Medium | 2 | 25 |
| High | 6 | 75 |
| Totals | 8 | |

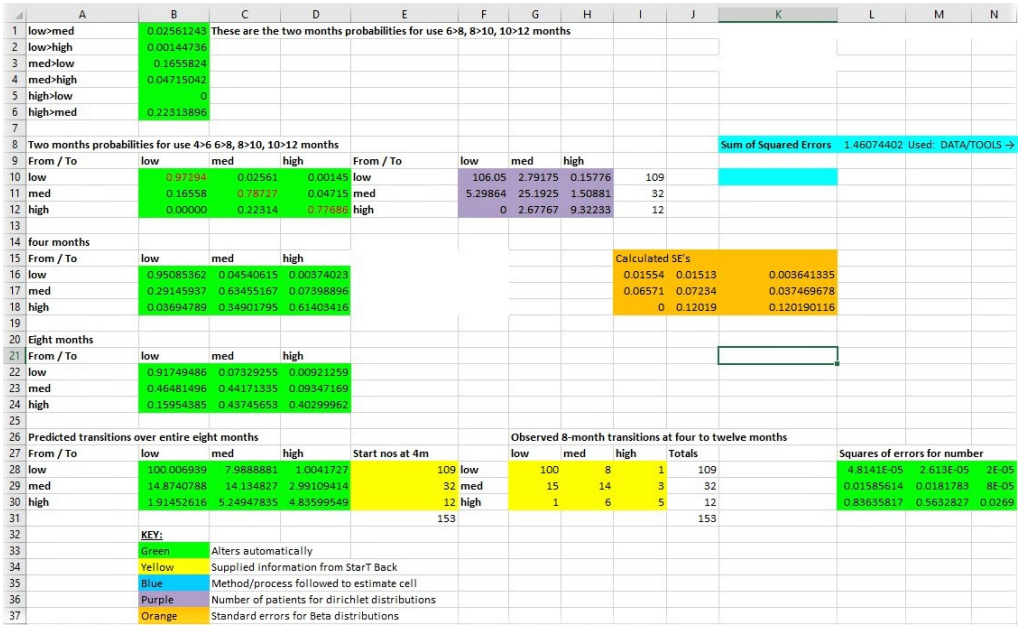
7YR Base function if base function =1 AND m12 function equals 0

| | n | % |
|---|---|---|
| Low | 4 | 66.67 |
| Medium | 2 | 33.33 |
| High | 0 | 0 |
| Totals | 6 | |

7YR Base function if base function =2 AND m12 function equals 0

| | n | % |
|---|---|---|
| Low | 0 | 0 |
| Medium | 0 | 0 |
| High | 0 | 0 |
| Totals | | |

## Appendix 6: Details of Matrix algebraic methods used to transform transition probabilities



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | low>med | 0.02561243 | These are the two months probabilities for use 6>8, 8>10, 10>12 months | | | | | | | | | | | |
| 2 | low>high | 0.00144736 | | | | | | | | | | | | |
| 3 | med>low | 0.1655824 | | | | | | | | | | | | |
| 4 | med>high | 0.04715042 | | | | | | | | | | | | |
| 5 | high>low | 0 | | | | | | | | | | | | |
| 6 | high>med | 0.22313896 | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | | |
| 8 | Two months probabilities for use 4>6 6>8, 8>10, 10>12 months | | | | | | | | | | Sum of Squared Errors | 1.46074402 | Used: DATA/TOOLS → | |
| 9 | From / To | low | med | high | From / To | low | med | high | | | | | | |
| 10 | low | 0.97294 | 0.02561 | 0.00145 | low | 106.05 | 2.79175 | 0.15776 | 109 | | | | | |
| 11 | med | 0.16558 | 0.78727 | 0.04715 | med | 5.29864 | 25.1925 | 1.50881 | 32 | | | | | |
| 12 | high | 0.00000 | 0.22314 | 0.77686 | high | 0 | 2.67767 | 9.32233 | 12 | | | | | |
| 13 | | | | | | | | | | | | | | |
| 14 | four months | | | | | | | | | | | | | |
| 15 | From / To | low | med | high | | | | | Calculated SE's | | | | | |
| 16 | low | 0.95085362 | 0.04540615 | 0.00374023 | | | | | 0.01554 | 0.01513 | 0.003641335 | | | |
| 17 | med | 0.29145937 | 0.63455167 | 0.07398896 | | | | | 0.06571 | 0.07234 | 0.037469678 | | | |
| 18 | high | 0.03694789 | 0.34901795 | 0.61403416 | | | | | 0 | 0.12019 | 0.120190116 | | | |
| 19 | | | | | | | | | | | | | | |
| 20 | Eight months | | | | | | | | | | | | | |
| 21 | From / To | low | med | high | | | | | | | | | | |
| 22 | low | 0.91749486 | 0.07329255 | 0.00921259 | | | | | | | | | | |
| 23 | med | 0.46481496 | 0.44171335 | 0.09347169 | | | | | | | | | | |
| 24 | high | 0.15954385 | 0.43745653 | 0.40299962 | | | | | | | | | | |
| 25 | | | | | | | | | | | | | | |
| 26 | Predicted transitions over entire eight months | | | | | | Observed 8-month transitions at four to twelve months | | | | | | | |
| 27 | From / To | low | med | high | Start nos at 4m | | low | med | high | Totals | | Squares of errors for number | | |
| 28 | low | 100.006939 | 7.9888881 | 1.0041727 | 109 | low | 100 | 8 | 1 | 109 | | 4.8141E-05 | 2.613E-05 | 2E-05 |
| 29 | med | 14.8740788 | 14.134827 | 2.99109414 | 32 | med | 15 | 14 | 3 | 32 | | 0.01585614 | 0.0181783 | 8E-05 |
| 30 | high | 1.91452616 | 5.24947835 | 4.83599549 | 12 | high | 1 | 6 | 5 | 12 | | 0.83635817 | 0.5632827 | 0.0269 |
| 31 | | | | | 153 | | | | | 153 | | | | |
| 32 | | KEY: | | | | | | | | | | | | |
| 33 | | Green | Alters automatically | | | | | | | | | | | |
| 34 | | Yellow | Supplied information from StarT Back | | | | | | | | | | | |
| 35 | | Blue | Method/process followed to estimate cell | | | | | | | | | | | |
| 36 | | Purple | Number of patients for dirichlet distributions | | | | | | | | | | | |
| 37 | | Orange | Standard errors for Beta distributions | | | | | | | | | | | |

**Figure A6-1 Transition matrices in Excel**

Before entering formulae into the matrix, six separate rows were created to represent each of the movements *between* states (Cells A1:B6), e.g. low to med, med to high, high to low, and initially entered a value half of the actually observed proportions in the data. The aim of this is to create an approximation from which to allow the Microsoft excel solver to create their actual value.

Each of the movements represented in the six rows (A1:B6) were linked to a transition matrix, (A9: D12), such that, for example, the low to medium movement in the matrix (C10) was set as equal to the value in the row "low to medium" (B1). As the probability of remaining in the same state was not allocated a row, in the transition matrix the probability was set as one minus the other two transitions on that matrix row, e.g. cell B10 set equal to (1-C10-D10).

Seven other matrices were set up on the page in order to complete this process. First, Matrix algebra was used to transform eight-month transitions (A21:D24) into a two-

monthly transition matrix (A9:D12). Next, an observed eight-month transition matrix was created (G27:J31) using STaRT Back data, to represent the actual number of patients moving between states over eight months, as well as the number of patients in each of starting risk states at 4-months (J28:J30). Next, a prediction transition matrix was created (A27:E31) where start numbers at four months (E28:E31) were set equal to observed totals (J28:J31). The cells representing actual transitions in this matrix (B28:D30) were set equal to the estimated eight-month transitions from the matrix algebra (B22:D24) multiplied by the start numbers at four months (E28:D30).

A matrix of errors (L28:N30) was set up to compare the differences between the observed numbers (G28:I30) and the predicted numbers (B28:D30). For example Cell L28 was set as equal to the square of G28-B28. A cell was set up to represent the sum of squared errors (L8), and the set equal to the sum of all the squared errors (L28:L30). Finally, using the solver function in Excel, the sum of these errors were minimised by varying the six arbitrary probabilities assigned to the rows reflecting the six movements between the states (B1:B6). The matrix, therefore, will yield the best possible fit to the data. The additional orange (I15:K18) and purple matrix (E9:H12) were used to create distributions for the PSA.

The same method was employed to estimate two-monthly transition probabilities for STaRT Back at four to twelve months, as well as two-monthly transition probabilities for the first four months.

**Appendix 7: RMDQ scores 12-months and 7-years in BeBack data**

| Risk Group | RMDQ 12-months | RMDQ 5 years |
|---|---|---|
| Poor Function | 15.30 | 15.42 |
| Moderate Function | 7.00 | 7.024 |
| Good Function | 1.369 | 1.297 |

**Appendix 8: Expert consultations and the model building process**

| | | |
|---|---|---|
| January 2019 | Keele Stratified Care Research Group (SCRG) KK, RaO, and ML also present | Discussion of health states, model assumptions, model population, costs, and sensitivity analyses |
| February 2019 | Dr. Pelham Barton (Health Economist), Birmingham | Initial model conceptualisation ideas, feedback upon model structure, help with transition probabilities |
| March 2019 | Dr Katrina Humphreys (physiotherapist), and Dr Adrian Chudyk (GP), Keele | Discussion of health states, model assumptions, sensitivity analyses, and treatment protocols. |