Serious Pathologies in the Lumbar Spine:

Prevalence and Diagnostic Accuracy of Red Flag Questions

Kathryn J. Street

A thesis submitted to Auckland University of Technology in fulfilment of the requirements for the degree of Masters of Health Science (MHSc)

2016

School of Physiotherapy

Supervisors

Steven G. White

Associate Professor Alain C. Vandal

Contents

List of Figures

List of Tables

List of Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
CT	Computed Tomography
CES	Cauda Equina Syndrome
DEXA	Dual-energy X-ray absorptiometry
HIV	Human Immunodeficiency Virus
LBP	Low Back Pain
LR+	Positive Likelihood Ratio
LR-	Negative Likelihood Ratio
MRI	Magnetic Resonance Imaging
NZD	New Zealand Dollar
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
SPECT	Single Photon Emission Computed Tomography
STARD	Standards for Reporting of Diagnostic Accuracy Studies
Sn	Sensitivity
Sp	Specificity
95% CI	95% Confidence Interval

Glossary

Diagnostic Accuracy – Expresses the test's ability to discriminate between patients with and without the target condition

Diagnostic utility – Refers to the intended use of the test, i.e. whether the test is used for diagnosis, screening, staging, monitoring, surveillance, prediction, prognosis, or other reasons

Incidence – The number of new cases of a condition during a specific time period in a given population

Index test – The test under evaluation

Likelihood Ratio – The likelihood that a given test result would be predicted in a patient with the target condition compared to the likelihood of the same result in a patient without the target condition

Prevalence – The baseline risk of a condition within the population of interest

Point Prevalence – The proportion of a population that has the condition at a specific point in time

Period Prevalence – The proportion of a population that has the condition at some time during a given period (e.g. 12 months), and includes people who have the condition at the start of the study period as well as those who acquire it during that period

Sensitivity – Proportion of those with the target condition who test positive with the index test

Specificity – Proportion of those without the target condition who test negative with the index test

Target condition – The disease or condition that the index test is expected to detect

Reference standard – The best available method for establishing the presence or absence of the target condition; a gold standard would be an error-free reference standard

Receiver Operating Characteristic (ROC) Curve – A ROC curve plots the sensitivity in function of the false positive rate (1- specificity)

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signed: 

Dated: 10.05.2016

Research Outputs Resulting from this Thesis

National conference presentations

Street, K., Mistry, D., White, S., & Vandal, A. (2014). Red flags in the spine. Invited workshop presentation: General Practice Conference and Medical Exhibition, Rotorua, NZ.

Street, K., Mistry, D., White, S., & Vandal, A. (2016). Red flags in the spine - updated. Invited workshop presentation: General Practice Conference and Medical Exhibition, Rotorua, NZ.

Local presentations

Street, K., White, S., & Vandal, A. (2015). Screening for serious pathologies in the lumbar spine. Invited oral presentation: Middlemore Hospital, Auckland, NZ.

Street, K., White, S., & Vandal, A. (2015). Red flags in the lumbar spine. Invited oral presentation: Auckland Physiotherapy, Auckland, NZ.

Planned poster presentations

Middlemore Science Fair (2016).

Other contributions

Advisory board for the development of a clinical pathway for acute low back pain (2015). This pathway was developed in conjunction with local experts for use in general practice, district health boards, and by St John Ambulance Services. Contributions included a lead role in the pathway development, preparation of proposals, and presentation to district health board managers.

Acknowledgements

Thank you firstly to my supervisors, Steve White and Alain Vandal. As my primary supervisor, Steve has provided endless support, endured lengthy meetings, and offered outstanding expertise. Alain Vandal, my secondary supervisor, has provided guidance and support with his exceptional statistical knowledge. Thank you also to Mr Brendan Coleman for his assistance in instigating this research at Middlemore Hospital.

Thank you to my husband and my parents for their love, encouragement and generosity, which kept me going. Thank you also to my colleagues for their support and for giving me the time to complete this project.

I would like to acknowledge all the radiology and nursing staff who made this project possible. Without their generous support it would not have been conceivable to undertake a project of this magnitude. Also thank you to Tariq Al-Shatanawi and Andrew Fegan who provided some assistance with data entry, which was kindly funded by the biostatistics and epidemiology department at Auckland University of Technology and by Auckland Physiotherapy.

This research was gratefully supported by sponsorship from the Maataatupu fund at Counties Manukau District Health Board. Funding towards university fees was also received from Middlemore Hospital.

Ethics

Ethical approval for this study was granted on the 6th of June 2013 by the Auckland University of Technology Ethics Committee (reference 13/120). Copies of ethical approval, locality approval, participant information sheets and consent forms are included in Appendices A.1-A.5 and B.1-B.2.

Abstract

Low back pain is a common problem that can be seriously or chronically disabling. It is one of the most common reasons for people to visit their general practitioner and is associated with high healthcare costs. Low back pain is frequently benign, but in rare cases may be due to underlying serious pathology. However, the actual likelihood of a patient presenting with a serious pathology to primary, secondary, or tertiary care is largely unknown. Without knowledge of prevalence it is not possible for clinicians to estimate the probability of a patient having a serious pathology. Additionally, knowledge of incidence is required to understand disease aetiology, including which age or ethnic groups may be more at risk.

Clinical identification of serious pathologies can be difficult, with evidence that the diagnosis of these cases is often delayed and not uncommonly missed. Therefore, screening questions, known as ‘red flag’ questions, have been widely recommended to assist with early recognition of serious pathologies in the lumbar spine and are included in a number of international guidelines. However, the diagnostic accuracy and utility of these questions is poorly understood, and several authors have expressed concerns regarding their use, given the lack of supportive evidence.

The aims of this thesis were to determine the prevalence of serious pathologies that commonly affect the lumbar spine (vertebral fracture, malignancy, cauda equina syndrome, and infection) in secondary and tertiary care settings, and to determine the incidence of serious pathologies in the geographic region of Counties Manukau, Auckland, New Zealand. This thesis also aimed to determine the diagnostic accuracy of red flag questions commonly used to screen for serious pathologies in patients presenting with low back pain.

The primary component of this research was a retrospective audit of 2,383 lumbar MRI scans. Adult participants who were referred for a lumbar MRI scan for the investigation of low back pain were consecutively recruited over a 10-month period. Target condition prevalence was calculated as a percentage of the study population and the prevalence specific to secondary care and tertiary care was also calculated. The incidence of serious pathologies was determined by comparing the 2013 census results from the Counties Manukau region to the data collected from Middlemore Hospital

(tertiary care). Data was subcategorised into age groups, gender, and ethnicity, to allow investigation of which groups may be most at risk.

The prevalence of serious pathologies varied from 0.12% for spinal infection in secondary care to 6.7% for vertebral fractures in tertiary care. The prevalence was significantly higher in the tertiary care setting than in the secondary care setting for all serious pathologies. The total incidence varied from 2.5 per 100,000 person-years for cauda equina syndrome to 12.9 per 100,000 person-years for vertebral fracture. Overall there was no significant difference between genders. However, the risk of developing a serious pathology increased significantly with age and peaked at 249 per 100,000 person-years in the 85 years and over group.

To determine the diagnostic accuracy of the commonly recommended red flag questions, 552 consecutive participants referred for an MRI scan for LBP were consecutively and prospectively recruited. All participants were required to complete a questionnaire that contained 37 questions related to specific spinal pathologies and a body chart to provide further detail of their symptoms. This cohort was a subgroup of the full cohort included in the study designed to determine prevalence and incidence. Data collection for these two studies occurred concurrently.

This study demonstrated that a number of red flag questions or index tests had negative likelihood ratios less than 0.1, indicating that the absence of these findings results in a conclusive shift in probability that the target pathology will be absent. These index tests were: age greater than 35 years for vertebral fracture, age greater than 42 years or ‘worsening pain’ for malignancy, and age greater than 55 years, insidious onset of pain or ‘night pain that wakes you from sleep’ for spinal infection. Hence, these index tests have sufficient diagnostic accuracy to suggest that they can be used as screening tests. For cauda equina syndrome, no index test had a negative likelihood ratio less than 0.1.

This study also demonstrated that only two red flag questions had positive likelihood ratios >10, indicating that the presence of these findings results in a conclusive shift in the probability of that pathology being present. With respect to vertebral fracture, the only index test that met this criteria (LR+ >10) was a history of concomitant HIV or AIDS. For spinal infection, a history of immunosuppressant use was the only red flag question with a positive likelihood ratio greater than 10.. Hence,

these tests have potential diagnostic utility as risk factors. No red flag questions for cauda equina syndrome or malignancy displayed a conclusive increase in probability.

This thesis has provided important new information related to the prevalence and incidence of serious pathologies within a population of low back pain patients presenting to secondary or tertiary care. In addition, this thesis has determined the diagnostic accuracy of all commonly recommended red flag questions to screen for serious pathologies. Hence, this study has provided information that can be used to determine pre and post-test probability to assist with clinical decision-making, and facilitate early diagnosis and treatment to improve patient outcomes.

Chapter 1 **Introduction**

1.1 The Problem

Low back pain is a highly prevalent problem, affecting around two thirds of people at some stage during adulthood (Andersson, 1999). It is one of the most common reasons to visit a general physician, second only to upper respiratory problems (Casazza, 2012; Gary Hart, Deyo, & Cherkin, 1995). Low back pain imposes a substantial economic and social burden on communities and can be seriously and/or chronically disabling (Deyo & Weinstein, 2001; Henschke et al., 2009). In the 2010 Global Burden of Diseases study (Murray et al., 2012), low back pain was identified as the leading cause of disability-adjusted life years in Australasia. In New Zealand, the Accident Compensation Corporation (ACC) covers the costs associated with accident-related low back pain, and each year they spend in excess of $130 million on accident-related low back pain alone (Physiotherapy New Zealand, n.d.).

In the majority of cases, low back pain is benign and self-limiting (Henschke et al., 2009). However, in rare cases, it may be due to underlying serious pathology that requires urgent medical management. The most common serious pathologies that affect the lumbar spine are fracture, malignancy, cauda equina syndrome, or spinal infection (Deyo, Rainville, & Kent, 1992). Inflammatory back pain is often considered a serious pathology, but does not require urgent management, and although other serious conditions such as abdominal aortic aneurysm and inflammatory bowel disease can masquerade as low back pain, they do not arise from the lumbar spine.

Serious pathologies are reported to account for around 1-5% of all patients presenting with low back pain (Chou et al., 2007; Henschke et al., 2009; Wilk, 2004). However, due to the rarity of these pathologies, there is limited research regarding prevalence in clinical settings, and the incidence of serious pathologies within the population is largely unknown (Henschke et al., 2013; Williams et al., 2013). Without knowledge of the risk of developing a serious pathology within a population, or full understanding of the disease aetiology, it is challenging for clinicians to recognise the rare cases of serious pathology amongst large numbers of patients presenting with low back pain.

To assist with clinical decision-making and recognition of serious pathologies, a series of questions have been recommended as useful questions to ask patients with low

back pain during subjective history taking (Koes et al., 2010; Lurie, 2005; Sizer Jr, Brismee, & Cook, 2007). These questions are commonly termed ‘red flag’ questions, as they are thought to raise the suspicion of potential underlying serious pathology. The term ‘red flag’ was introduced by the clinical standards advisory group in 1994 (Higginson, 1994), although similar screening questions were recommended by earlier authors including Mennell in 1952 and Cyriax in 1982 (Greenhalgh & Selfe, 2006). The presence of red flags should alert the clinician to perform further questioning or testing to obtain more information that determines the likely presence or absence of a serious cause of lower back pain. Alternatively, if the level of suspicion is high enough, referral for medical assessment is indicated (Accident Compensation Corporation [ACC], 1999).

Whilst red flag screening questions appear to be widely employed and are recommended in a number of low back pain guidelines internationally (Bach & Holten, 2009; Chou et al., 2007; Koes et al., 2010; Van Tulder et al., 2006), few studies have provided high quality evidence of the diagnostic accuracy of such questions. Some authors have expressed concern over the use of red flags given this lack of evidence (Harding, Davies, Buchanan, & Fairbank, 2005; Henschke & Maher, 2006; Underwood, 2009). In particular, high false positive rates leading to inappropriate referral for diagnostic work up or specialist review are not only costly but place additional strain on busy secondary and tertiary care services, and can expose patients to unnecessary radiation or interventional procedures and inappropriate diagnostic labelling (Deyo, Mirza, Turner, & Martin, 2009; Flynn, Smith, & Chou, 2011; Williams et al., 2010).

Most low back pain guidelines regarding the indications for diagnostic imaging, including plain radiography and advanced imaging such as magnetic resonance imaging (MRI) or computed tomography (CT), recommend that diagnostic imaging should be reserved for patients where there is suspicion of serious pathology or who may require surgical management or other interventional procedures (Chou et al., 2007; Koes et al., 2010; Rubinstein & van Tulder, 2008). Although routine use of medical imaging has been widely discouraged (ACC, 1999; Koes et al., 2010; National Institute for Health and Care Excellence [NICE], 2009; Van Tulder et al., 2006) many clinicians are either not aware of or do not adhere to these low back pain guidelines (Williams et al., 2010). There is evidence that indicates that the use of expensive advanced imaging such as MRI is growing at an unsustainable rate (Dagenais, Galloway, & Roffey, 2014; Oikarinen, Karttunen, Pääkkö, & Tervonen, 2013; Perez & Jarvik, 2012). While some

authors recommend reduced access to advanced imaging to reduce cost and potential harm associated with diagnostic imaging (Chou, Deyo, & Jarvik, 2012; Flynn et al., 2011; Webster, Choi, Bauer, Cifuentes, & Pransky, 2014), others argue that it should be directly accessible to primary care clinicians, such as general practitioners, to improve patient waiting times and management (Algra, 2010). However, allowing access to a significantly larger group of medical professionals would invariably lead to increased use of a service that is currently being overwhelmed (Chou, Deyo, & Jarvik, 2012).

Despite the high use of diagnostic imaging, several studies have reported that serious pathologies are often missed. For example, one study (Grigoryan, Guermazi, Roemer, Delmas, & Genant, 2003) reported that up to 70% of vertebral fractures were misdiagnosed on clinical assessment and another found that up to 75% of patients with spinal infection are initially misdiagnosed (Darouiche, 2006; Patel et al., 2014).

Early diagnosis is crucial to prevent potential adverse outcomes, such as neurological compromise, systemic illness, spread of disease, pathological fracture, spinal deformity, spinal cord compression, and ultimately mortality (Cook & Hegedus, 2012; Edmond, Kiel, Samelson, Kelly-Hayes, & Felson, 2005). Complications such as neurological compromise and progression of disease can significantly impact the patient's livelihood, quality of life, and function. These outcomes can be dire for patients, but can also come at great expense for health funders such as ACC, district health boards, and Work and Income New Zealand, as patients may require lifelong management, and may not be able to return to work (Davis et al., 2004; Todd, 2011). Hence, it is essential that clinicians screen for these pathologies and understand the diagnostic accuracy and utility of screening questions.

1.1.1 Summary of the problem

Low back pain is a common complaint worldwide. Whilst the prevalence of serious pathologies is relatively low, the consequences of missed or delayed diagnosis are high. The challenge for clinicians is to decide whether low back pain is benign in nature, or whether underlying serious pathology is suspected and warrants onward referral, follow-up, or diagnostic imaging. Although the use of red flag questions has been widely recommended to screen for serious pathologies, the diagnostic accuracy of such questions is largely unknown, and there is currently no convincing evidence to support or refute their use in clinical practice (Henschke & Maher, 2006; Henschke et al., 2009; Van den Hoogen et al., 1995). Similarly, there is a dearth of evidence

regarding either the clinical prevalence or population incidence of serious pathologies in the lumbar spine (Deyo, Jarvik, & Chou, 2014; Henschke et al., 2013; Williams et al., 2013). Without this information it is difficult for clinicians to adequately assess a patient's risk of serious pathology. Hence, the current study was undertaken to address this gap in the literature with respect to four of the most common serious pathologies that affect the lumbar spine.

1.2 Thesis Aims

The primary aims of this thesis are:

1. To determine the clinical prevalence of vertebral fracture, malignancy, cauda equina syndrome and spinal infection amongst patients with low back pain presenting for an MRI scan in secondary (private) or tertiary (university teaching hospital) care settings.

2. To determine the diagnostic accuracy of red flag questions for the identification of vertebral fracture, malignancy, cauda equina syndrome, and spinal infection amongst patients with low back pain.

The secondary aim was:

1. To determine the population incidence of vertebral fracture, malignancy, cauda equina syndrome, and spinal infection in the geographic region of Counties Manukau in Auckland, New Zealand.

1.3 Overview of the Thesis

The first chapter of this thesis has provided an overview of the importance of recognising serious pathologies amongst patients presenting with low back pain. This chapter has also discussed the proposed use of red flag questions and the apparent lack of evidence regarding prevalence and incidence of serious pathologies in the lumbar spine.

Chapter 2 consists of an in-depth systematic review of the literature related to the prevalence, incidence, and use of red flags to screen for serious pathologies in the lumbar spine. This chapter is comprised of a short introduction followed by four systematic reviews. Each systematic review is presented following the Cochrane Collaboration review format. The first review focuses on vertebral fractures in patients

presenting with low back pain. Similarly, the subsequent reviews focus on the remaining target conditions of interest, i.e. malignancy, cauda equina syndrome, and spinal infection.

Chapter 3 consists of an observational study investigating the clinical prevalence and population incidence of fractures, malignancy, cauda equina syndrome and infection in the lumbar spine. This chapter begins with a brief introduction to prevalence and incidence, followed by a description of the study methodology with regard to the data collection and analysis. The results are also reported and discussed in this chapter.

Chapter 4 consists of a diagnostic accuracy study that investigates the ability of red flag questions (index tests) to discriminate between patients with and without the aforementioned target conditions in a low back pain population. This chapter begins with a short introduction, followed by the study methodology related to the development of the study questionnaire that was used to examine the red flag questions, collection of data, and statistical analysis. The study results are also reported, summarised and discussed in this chapter.

This thesis concludes with Chapter 5, which provides a summary of the key findings alongside discussion of the clinical implications of this research and recommendations for future research.

1.4 Significance of the research

To the author's knowledge this is the first study to investigate to the incidence of serious pathologies in the lumbar spine. It is also the first study to investigate the prevalence of cauda equina syndrome or spinal infection in patients presenting to secondary or tertiary care with low back pain.

Establishing prevalence of serious pathologies is vastly important as it allows calculation of the pre-test probability of a patient presenting with a disease. Knowledge of incidence permits estimation of the number of *new* cases of serious pathologies that can be expected each year within a specific population and improves understanding of disease aetiology. Awareness of prevalence and incidence also allows healthcare funders to more accurately plan the provision of services to appropriately manage these patients.

This study is the first to investigate the diagnostic accuracy of red flag questions and to discriminate between patients with and without cauda equina syndrome or spinal infection in patients with low back pain population presenting to secondary or tertiary care. Currently, although red flag questions are widely endorsed, there is a lack of good quality evidence to support their use in screening for any serious pathology in the lumbar spine. Determining the diagnostic accuracy of red flag questions enables clinicians to calculate an individual patient's risk of having a serious pathology given the findings obtained from utilisation of these questions. This in turn enables the clinician to make an informed decision as to whether a patient requires further investigation, urgent hospital admission, specialist review, or perhaps whether conservative management is appropriate.

Finally, this research may enhance the likelihood of the early detection of the common serious pathologies in the lumbar spine, and is therefore of great benefit to patients with low back pain, and to all parties involved in the management of low back pain.

Chapter 2 Systematic Review of the Literature

2.1 Introduction

This chapter reviews the literature relevant to the prevalence and incidence of serious pathologies and the diagnostic accuracy of red flag questions amongst patients with low back pain. The most common serious pathologies (target conditions) to affect the lumbar spine are vertebral fracture, malignancy, infection, inflammatory disease, and cauda equina syndrome (Henschke et al., 2009). For this review, only serious pathologies that require urgent investigation, acute management, or hospital admission have been included. Hence, literature relevant to inflammatory back pain was excluded. Each of the remaining conditions is considered separately. To reduce repetition for the reader, a full explanation of search methodology is provided in association with the first systematic review (vertebral fracture). This same methodology was employed for each of the subsequent reviews, and differences in key words have been detailed in the relevant sections. Full details of the individual search strategy tables and results are provided in Appendix C (Tables C.1-C.3 and Figures C.1-C.3).

Prior to reporting the findings of these literature reviews, clarification is warranted regarding how the following terms have been employed and should be interpreted throughout this thesis.

2.1.1 Prevalence and incidence

Although serious pathologies account for less than 5% of all cases of low back pain (Deyo et al., 1992; Williams et al., 2013) it is important for clinicians to be aware of the prevalence of a given pathology within the population they are working in. Prevalence in a specific clinical population (in this case lower back pain) is more correctly described as clinical prevalence. Hereafter, clinical prevalence will be referred to as prevalence. Other considerations for prevalence are whether it has been recorded over a period of time (period prevalence), or at a specific point in time (point prevalence). Knowledge of the prevalence provides the clinician with an indication of the probability of the condition of interest being present in the individual patient presenting for assessment. In populations where there is an increased clinical prevalence (e.g. secondary versus primary healthcare), the pre-test probability of the presence of the condition is higher. It is also important to understand the population incidence, as an increased in incidence reflects an increased risk of developing the target condition.

Population incidence relates to the number of *new* cases of a certain condition that arise over a set period of time in a specific population (e.g. patients with low back pain). Population incidence will hereafter be referred to as incidence. The following reviews consider any research that has reported prevalence or incidence of the selected target conditions in a population of people with low back pain in primary, secondary, or tertiary care.

2.1.2 Index tests

For the purposes of this thesis, an index 'test' finding is considered to be any information obtained from questions employed during the patient history-taking component of a clinical examination. This thesis focuses specifically on subjective red flag questions (index tests) that may indicate the presence of or raise the suspicion of a serious pathology. Hence, findings obtained from the physical examination have not been included. A finding that raises the suspicion of a serious pathology is commonly referred to as a red flag and the terms red flag or index test may be used interchangeably (Sizer Jr et al., 2007).

2.1.3 Reference standard

A reference standard should be the best available method for establishing the presence or absence of the target condition (Bossuyt et al., 2015). When investigating prevalence, incidence or diagnostic test accuracy, the quality of the reference standard is of great importance. Ideally a "gold standard" or error-free reference standard should be used to ensure that the presence or absence of a target condition can be determined with a high level of precision (Eusebi, 2013). If a reference standard has low sensitivity there may be a high number of false negative results and the prevalence of a condition may be under-reported (Leeflang, Deeks, Takwoingi, & Macaskill, 2013). Conversely, if a reference standard has low specificity, there may be a higher number of false positives leading to prevalence being over-reported (Leeflang et al., 2013). The most appropriate reference standard for one pathology may not be the same as that for another. For example, CT has good precision for the diagnosis of bony pathologies such as vertebral fractures (Jarvik & Deyo, 2002), but does not clearly visualise cauda equina compression (Coscia, Leipzig, & Cooper, 1994). Conversely, MRI can clearly visualise all serious pathologies in the lumbar spine with a high level of precision (Coscia et al., 1994; Gold, 2016). A variety of reference standards were employed in the studies

identified by the current literature search. The appropriateness of the reference standard employed in each study is discussed within the relevant review.

2.1.4 Objectives

This review of the literature has gathered results from studies based on the target population of patients presenting with low back pain, and has synthesised the results from studies with lower risk of bias.

The primary aims of the literature reviews were to answer the following questions:

1. What is the prevalence and incidence of serious pathologies amongst patients with low back pain?

2. What is the diagnostic accuracy of subjective red flag questions to screen for serious pathologies in patients with low back pain?

Guidance for the methods of this review was taken from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (2013) and from the Cochrane handbook for systematic reviews of diagnostic test accuracy (Deeks, Wisniewski, & Davenport, 2013). In accordance to the Cochrane Collaboration format, previous systematic reviews are first discussed within the introduction of each systematic review to provide a summary of the current knowledge base and the rationale for undertaking a further systematic review. Prior to consideration of these previous reviews, detail regarding the known prevalence and incidence of each target condition is presented along with information regarding the relevant index tests and reference standard for the condition of interest.

2.2 Target Condition: Vertebral Fracture

2.2.1 Background

Vertebral fractures can be traumatic or may occur due to insufficiency of the bone secondary to osteoporosis or other pathologies such as infection or bony metastases (Bouxsein et al., 2006). Several fracture classification systems have been proposed in the literature, although there is a lack of consensus regarding the ‘best’ system. To address this problem, the international AO (Arbeitsgemeinschaft für Osteosynthesefragen) Spine Foundation recently assigned a group of specialists to establish a uniform classification system. The AO spine classification group proposed

that vertebral body fractures should be defined by the presence or absence of posterior wall and endplate involvement. The group proposed that the following classifications are utilised: wedge/impaction, split/pincer, incomplete burst, and complete burst fracture (Reinhold et al., 2013). Unstable fractures are rare, generally associated with high impact trauma, and require immediate medical attention as they may result in severe neurological injury (Zhang et al., 2015). The most common type of vertebral fracture is a wedge or compression fracture (Cooper, O'Neill, & Silman, 1993; Hu, Mustard, & Burns, 1996). Such fractures may result from trauma, but more commonly occur in association with bone insufficiency (Bouxsein et al., 2006). Vertebral fractures may be the first sign of osteoporosis and are independent indicators of future hip or vertebral fractures (Black, Arden, Palermo, Pearson, & Cummings, 1999; Melton, 2003), with research demonstrating that there is a 4-5 fold increase in future risk of vertebral fracture following initial fracture (Melton, 2003). Early diagnosis and management of vertebral fractures is crucial, as they can lead to spinal deformity, significant pain, depression, functional impairments, and increased mortality risk (Cooper et al., 1993; Edmond et al., 2005; Melton, 2003). Delayed diagnosis of insufficiency fractures results in a consequent delay in appropriate pharmaceutical treatment. A missed fracture could result in application of contra-indicated treatments such as spinal manipulation (Waddell, 2004).

Prevalence/incidence. Vertebral fracture is the most common serious pathology to affect the lumbar spine (Henschke et al., 2009). In the year 2000, there were 1.4 million recorded vertebral fractures diagnosed worldwide (Johnell & Kanis, 2006). However, incidence rates are likely to be much higher, as it is estimated that less than one third of vertebral fractures are correctly diagnosed (Grigoryan et al., 2003; Papaioannou et al., 2002). The incidence of vertebral fracture in New Zealand is unknown, although in 2007 an average of 300 hospital beds were occupied every day with people recovering from osteoporotic fractures (Brown, McNeill, Radwan, & Willingale, 2007). This occupancy costs the health sector around $325,000 New Zealand dollars each day (Brown et al., 2007). Osteoporosis New Zealand has predicted a significant rise in the incidence of fractures and associated cost as New Zealand’s baby boomers age, due to the increased fracture risk in older adults (Brown et al., 2007).

The prevalence of vertebral fractures in patients presenting at primary healthcare with low back pain has been reported to be between 0.7 and 4.5% (Downie et al., 2013; Williams et al., 2013). In contrast, the prevalence in patients presenting at accident and

emergency departments may be as high as 29% (Henschke, Maher, & Refshauge, 2008). This higher prevalence most likely reflects the inclusion of fractures secondary to trauma.

Index tests. Red flag questions have been widely recommended to assist with clinical screening for vertebral fractures (Chou et al., 2007; Van Tulder et al., 2006). A review (Koes et al., 2010) of international clinical guidelines for low back pain reported that the most commonly employed red flag questions for fracture were related to a history of trauma, osteoporosis, or prolonged corticosteroid use, as well as older age and female gender. The New Zealand ACC acute low back pain guidelines (ACC, 1999) include questions relating to these same risk factors (significant trauma, steroid use, and age greater than 50 years).

Reference standards for identification of vertebral fracture. Currently there is debate in the literature regarding the most appropriate method for diagnosis and therefore the best reference standard for vertebral fractures. The American College of Physicians Guidelines (Chou, Qaseem, Owens, & Shekelle, 2011) recommend plain radiography for suspected fracture if pain does not improve after a trial of therapy. In contrast the United Kingdom National Institute for Health and Care Excellence (NICE) guidelines (NICE, 2009) recommend *against* the use of plain radiography in the lumbar spine, and recommend consideration of MRI for suspected fracture.

Another tool that has been recommended for "vertebral fracture assessment" is dual-energy X-ray absorptiometry (DEXA) scanning. However, there is a relatively high chance of error when reporting bone density in the lumbar spine, as this can be altered by patient positioning, anatomical variations, the presence of bony sclerosis, or osteophyte formation associated with degenerative changes which may overestimate bone density (Deleskog, Laursen, Nielsen, & Schwarz, in press). This study by Deleskog et al. reported that 18.5% of DEXA scans for vertebral fracture were considered unreadable compared to only 2% of plain radiographs.

A review by Della-Giustina and colleagues recommended that MRI is the best method of diagnostic imaging for low back pain, but that plain radiographs are suitable for cases of suspected fracture (Della-Giustina, 2015). This recommendation is supported by the findings of another review (Gold, 2016), which concluded that MRI is

the best reference standard for accurately imaging the lumbar spine for all serious pathologies.

Previous systematic reviews. A preliminary search of the literature in regard to screening for vertebral fractures in the lumbar spine uncovered three relevant systematic reviews (Downie et al., 2013; Henschke et al., 2008; Williams et al., 2013). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used to assess the quality of the published systematic reviews. The PRISMA guidelines were primarily developed to provide direction for authors conducting systematic reviews or meta-analyses to ensure clear reporting and adequate conduct. However, these guidelines can also be used to evaluate the quality of systematic reviews (Liberati et al., 2009; Moher, Liberati, Tetzlaff, & Altman, 2009).

Henschke and colleagues (2008) reviewed the literature that related to the use of red flag questions to screen for fractures. This review included 12 studies and investigated 51 clinical features, 23 of which could be considered as red flag questions. Whilst the wording of these questions varied widely across studies included in this review, many were related to similar themes, i.e. age, gender, history of trauma, previous corticosteroid use, and altered consciousness or other distracting injury. Henschke and colleagues provided clear objectives and detail of the search strategy and the characteristics of the identified studies. The majority of studies included in this review were conducted in populations of patients with lower back pain. However, they also included studies with patients presenting with blunt or multi-trauma. All studies were assessed for bias using the validated QUADAS tool that is recommended for assessment of methodological quality of diagnostic accuracy studies (Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003). Henschke et al. (2008) reported that all of the studies in their review were at risk of bias as a result inadequate reporting and/or poor methodological quality. However, despite this risk of bias, they concluded that five red flags were ‘useful’ for screening for vertebral fractures i.e age greater than 50 years, female gender, history of major trauma, pain and tenderness, and ‘distracting painful injury’. However, none of these red flags had negative likelihood ratios (LR-) that suggest that the absence of the factor significantly decreases the probability of a fracture being present. Only ‘history of major trauma’ was reported to have a positive likelihood ratio that suggests that its presence conclusively increases the probability of a fracture being present (Scavone, Latshaw, & Rohrer, 1981).

A Cochrane review by Williams, Henschke and colleagues in 2013 identified and evaluated the studies that had investigated screening for vertebral fracture amongst patients presenting with lower back pain (Williams et al., 2013). This review included eight studies and used the QUADAS tool to assess the methodological quality of studies included in the review. Again, although significant risk of bias was identified by the reviewers for some of the included studies, diagnostic accuracy was reported for all studies, without consideration of bias and methodological flaws. These authors concluded that most individual red flags had high false positive rates, did not have meaningful likelihood ratios and were not clinically useful. Whilst they considered that a history of significant trauma (LR+ 3.42 – 12.85), age >74 years (LR+ 3.69 – 9.39), and history of corticosteroid use (LR+ 48.50) had modest to good positive likelihood ratios, the confidence intervals around the point estimates were very wide, suggesting that the estimates were imprecise. Williams and colleagues also combined the results of red flag questions such as female gender and older age using the data published by original studies in an attempt to achieve higher positive likelihood ratios. However, this yielded very high false negative rates, for example, combining age >74 years and female gender increased the positive likelihood ratio to 16 and the false negative rate to 75%.

Another systematic review investigating red flags to screen for malignancy and fracture in patients with low back pain was published by Downie, Williams, Henschke and colleagues (2013). This review seems to be largely a combination of the findings from two previous Cochrane reviews performed by the same group of authors (Henschke et al., 2013; Williams et al., 2013). Downie and colleagues (2013) drew similar conclusions that a history of prolonged corticosteroid use, severe trauma or older age increased the likelihood of spinal fracture. However, ‘severe’ trauma is incorrect and should be ‘significant’ trauma (major in young people and minor in older people), as this conclusion is based on the findings of three studies that all investigated significant trauma (Deyo & Diehl, 1986; Henschke et al., 2009; Scavone, Latshaw, & Rohrer, 1981). The conclusion that a history of prolonged corticosteroid use is a useful finding is not supported by their data analysis. As previously mentioned this finding has a high positive likelihood ratio of 48.5 but is imprecise (95% CI 11.48, 204.99), due to the fact that only two of the 1,172 participants had a true positive finding, and the false positive rate was 75%.

Another concern arising from the recommendations from all of the above three reviews is that they focus on red flag questions with high positive likelihood ratios. This is worrying, as high positive likelihood ratios express a high increase in the likelihood of a disease being present. Whilst this might be useful, the potential consequences of a missed diagnosis of a serious pathology makes it is important to focus on questions with a very low false negative rate identifying. Hence, screening questions should exhibit high sensitivity and low negative likelihood ratios.

Despite the limitations of these systematic reviews, there is consensus across all reviews that individual red flag questions hold limited diagnostic value and that they should not be relied upon to identify or rule out the presence of a vertebral fracture. No investigations of heterogeneity were performed by any systematic review, and all reviews were mainly descriptive without attempting to combine any results through meta-analysis.

Rationale for undertaking the current review. Although three systematic reviews on the use of red flags to screen for vertebral fractures already exist in the literature, there are some inconsistencies in their analysis. Due to these previously discussed inconsistencies, a new, independent systematic review with particular attention to research with low risk of bias and low concern for applicability is appropriate. Hence, the following review was conducted. The findings of this review will be used to inform the design and conduct of the studies presented in following chapters of this thesis.

2.2.2 Objectives

The aims of this review reflect the aims of this chapter (as stated in 2.1.4 Objectives). Specifically, this review aims to determine the prevalence and incidence of vertebral fracture and the diagnostic accuracy of red flag questions in screening for such fractures in patients with low back pain.

Investigation of sources of heterogeneity. Potential sources of bias and heterogeneity will be assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS 2). Factors such as differences in reference standards, patient selection and study setting, that can contribute to heterogeneity, will be investigated.

2.2.3 Methods

Criteria for consideration of studies for this review

Types of studies. Primary diagnostic studies that have compared risk factors identified from the patient history or demographic information to an appropriate reference standard for the identification of vertebral fracture were considered for inclusion in this review. Cohort or cross-sectional studies that were either prospective or retrospective were considered appropriate, provided they presented sufficient data to enable calculation of the diagnostic accuracy of index tests. Studies published in English and in full text were eligible for this review.

Participants. Studies investigating patients presenting with the primary complaint of low back pain or patients referred for lumbar spine examination were considered eligible. Studies investigating adult populations were considered for inclusion. Any studies investigating patients with a primary complaint of high-energy trauma or known serious pathologies were excluded from this review. Eligible study settings included primary, secondary or tertiary care.

Index tests. Any information gathered and recorded during patient history taking was considered as an index test. This included patient demographics such as age and sex and subjective questions such as a history of trauma, pain or corticosteroid use. "Clinical diagnosis" or "clinical suspicion" of vertebral fracture that was not further defined was excluded due to poor reproducibility.

Target condition prevalence/incidence. Studies investigating the prevalence or incidence of vertebral fracture amongst patients presenting with low back pain or for lumbar spine examination were considered for inclusion in the review.

Reference standards. Studies that employed MRI, CT, DEXA, and plain radiography were considered for this review. Although 'long term follow-up' may under-report the prevalence of fractures, studies utilising this reference standard were also considered provided the fractures were confirmed at follow-up with one of the above diagnostic tools.

Search methods for identification of studies

Electronic searches. A literature search was completed on 20th of December 2013 using the following electronic databases: Scopus, EMBASE, MEDLINE,

CINAHL, SPORTDiscus (via EBSCO). Studies published in “all years” to 2013 were considered for inclusion if they were available in full text and in English. Only studies on adult humans were considered for eligibility.

Searching other resources. Reference lists of all included publications, relevant systematic reviews and narrative literature reviews were hand searched to ensure no eligible articles were missed. A broad search of the literature using the key words “red flag*” OR “serious pathology*” AND “low* back pain” was also completed to ensure no articles were missed and to pick up any reviews on red flags that may have been missed by searching specific pathologies.

Data collection and analysis

Selection of studies. Title and abstract screening was completed by the author of this thesis. Duplicates were removed and full publications were retrieved for any citation that potentially met the inclusion criteria. Reference lists were searched and final selection was based on review of the full text of identified publications.

Data extraction and management. A single author extracted all data including study design and methods, participant characteristics, and index and reference tests. The study setting (primary, secondary or tertiary care), reason for presentation e.g. low back pain or referral for lumbar X-ray, and whether enrolment was consecutive or non-consecutive was also determined. A record was made of total number of recruited participants, number enrolled in study, number receiving index tests, number receiving the reference standard, and any withdrawals. Where possible, data (true positive, false positive, true negative and false negative) was extracted to allow calculation of the sensitivity, specificity and likelihood ratios for each risk factor investigated. Finally, the reported prevalence of vertebral fractures was recorded.

Assessment of methodological quality. The QUADAS-2 tool was used to assess for potential bias or applicability concerns of the individual papers selected. The original QUADAS (Quality Assessment of Diagnostic Accuracy Studies) tool was revised by the original authors (Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003) following feedback from users and the Cochrane collaboration that some items either overlapped or were difficult to score (Whiting et al., 2011). The QUADAS-2 is therefore the redesigned, improved tool for rating diagnostic accuracy studies (see Appendix D). The QUADAS-2 is composed of a 4-stage process suggested by Moher

and colleagues (Moher, Schulz, Simera, & Altman, 2010). The process for implementing of the QUADAS-2 consisted of tailoring the tool content, development of rater guidelines, piloting the tool, and then applying it to all studies (Whiting et al., 2011). This design assesses the value of each study with respect to the specific review question. The QUADAS-2 tool assesses risk of bias across four domains: patient selection, index test, reference standard, and flow and timing. It then assesses applicability concerns across the first three domains. The tool includes recommended questions for each domain such as "is the reference standard likely to correctly classify the condition?" These questions assist the rater in deciding whether there is 'low,' 'high,' or 'unclear' risk of bias or concerns for applicability. These question are answered 'yes,' 'no,' or 'unclear' and can be adjusted to suit the specific review during the tailoring and piloting stages. For this review, the standard questions were applicable and did not require tailoring, as two independent raters were consistent with scoring during the pilot phase.

Whiting and colleagues (2011) emphasised that the QUADAS 2 tool should not be used to generate an overall quality score. Rather, studies should be judged as 'low risk of bias' or 'at risk of bias', and 'low concern regarding applicability' or as having 'concerns regarding applicability'. They recommended that if any study is judged 'high' or 'unclear' in 1 or more domains related to bias or applicability, the study should be considered 'at risk of bias', or as having 'concerns regarding applicability' for the latter. Any study judged as having 'high risk of bias' in four or more of the seven domains was excluded from subsequent statistical analysis.

Any studies investigating prevalence or incidence alone were assessed for quality using the STROBE (Strengthening The Reporting of Observational studies in Epidemiology) guidelines (von Elm et al., 2008).

Statistical analysis and data synthesis. Individual study prevalence, demographic and patient history data (index tests) related to diagnostic accuracy were extrapolated from original studies. In cases where diagnostic accuracy calculations were not provided in the original publication, but sufficient raw data was published, the reviewer calculated diagnostic accuracy. True positive, false positive, false negative, and true negative data from original publications was entered into Statistical Package for the Social Sciences (SPSS) software, Version 22 (IBM© Corporation, 2013) to construct 2 x 2 contingency tables. Sensitivity, specificity and likelihood ratios with

95% confidence intervals were calculated using the Clinical Calculator 1 available at http://vassarstats.net/clin1.html.

Likelihood ratios were used to assess whether an index test had diagnostic utility for screening or diagnosis of vertebral fracture (Eusebi, 2013; Sackett, 1992). Likelihood ratios were interpreted based on the guidelines from Jaeschke and colleagues (1994). Positive likelihood ratios (LR+) between 5 and 10 were considered to indicate a moderate increase in the probability the target condition is present, and likelihood ratios greater than 10 indicate a large and often conclusive increase in this probability. Negative likelihood ratios (LR-) from 0.1 to 0.2 were considered to indicate a moderate decrease in the probability of the target condition being present, whilst those less than 0.1 indicate a large reduction in this probability. Therefore, an index test with a negative likelihood ratio of ≤ 0.2 (with the upper bound of the 95% confidence interval <0.5) may have utility for screening, and a positive likelihood ratio ≥ 5 (with the lower bound of the 95% confidence interval >1) may have utility for diagnosis or as a risk factor (Deeks, 2001; Jaeschke et al., 1994).

Investigation of heterogeneity. Preliminary analysis of clinical and methodological heterogeneity determined that it was inappropriate to perform any simple meta-analysis. This was determined by the low number of studies included in this review, and the differences in patient recruitment, setting, index tests, and reference standards between studies. Meta-regression analysis was considered beyond the scope of this thesis.

2.2.4 Results

Results of the search. Table 2.1 provides detail of key words used in this search. Figure 2.1 displays the flow of study screening and selection. The combined electronic searches for the prevalence or incidence of vertebral fracture amongst adult patients presenting with low back pain resulted in 1,141 titles. A second combined search for the use of red flags amongst patients with vertebral fractures presenting with low back pain resulted in 1,933 titles. Following removal of duplicates and exclusion of studies that clearly did not meet the inclusion criteria, 482 abstracts were screened. A further 420 studies were removed as they did not meet inclusion criteria. All reference lists of relevant reviews were searched and 3 additional studies were identified and included. Of the 65 full texts considered, 57 were excluded. The primarily reasons for exclusion

were inappropriate study population or design. Hence, eight studies were included for qualitative synthesis.

Table 2.1 *Search terms for prevalence of and screening for vertebral fracture in the lumbar spine*

Search	Subject headings and search terms	Results
1	((lumbar OR lumbo* OR "low* back" OR spin*) AND pain)	130,041
2	fracture*	284,234
3	prevalence OR incidence OR epidemiology	2,517,780
4	red flag* OR screening OR finding* OR "patient history" OR evaluation OR "medical history" OR "history taking" OR "clinical decision" OR (clinical* N8 sign) OR (clinical* N8 symptom*) OR (clinical* N8 presentation)	3,561,687
Combine searches	1 AND 2 AND 3	1,141
	1 AND 2 AND 4	1,933

Note. * = truncation, N = proximity search for EMBASE, MEDLINE, CINAHL, SPORTDiscus (W/ used for proximity search in Scopus).

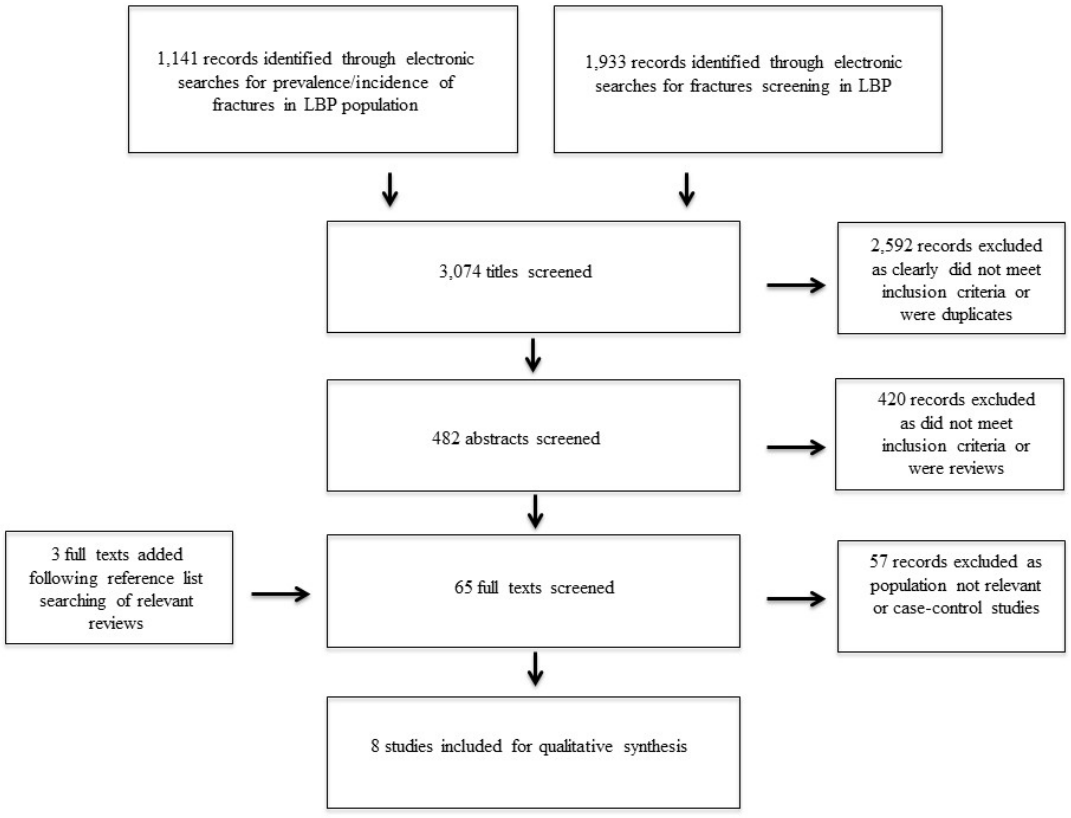


Figure 2.1 Search flow chart

Methodological quality of included studies. Table 2.2 below provides detail in regard to the assessment of study bias using the QUADAS-2 tool. Figure 2.2 and Figure 2.3 provide an illustration of the overall risk of bias and applicability of the included studies. The overall quality of the studies included in this review was relatively poor with all studies being 'at risk of bias'. In six of the eight studies, two domains were rated as 'low risk of bias' (commonly patient selection and index test domains). The remaining two studies (Reinus, Strome, & Zwemer Jr, 1998; Roman et al., 2010) had one domain (patient selection) that was considered low risk. The study by Roman and colleagues was unclear due to inadequate reporting of whether or not the index tests had been interpreted independently of the reference standards and visa versa. However, the study by Reinus and colleagues had high risk of bias. All studies had potential risk of bias with respect to the choice of reference standard, and seven of the studies had concerns for risk of bias in regard to flow and timing.

Half of the studies were judged as having 'low concern regarding applicability.' Two studies (Reinus et al., 1998; Scavone et al., 1981) included patients referred for lumbar X-ray and had concerns regarding patient selection. The study by Scavone et al. did not clearly define whether they included patients with a primary complaint of low back pain. Also, in this study, 68% of patients with a history of trauma and 36% of patients without trauma received the reference standard. Eight percent of the participants in the study did not have a primary complaint of low back pain.

Concerns regarding applicability of index tests were identified in two studies (Reinus et al., 1998; van den Bosch, Hollingworth, Kinmonth, & Dixon, 2004). Reinus and colleagues (1998) used chart review to collate index tests such as 'history of trauma.' However, only participants with positive X-ray findings had their charts reviewed. This resulted in only 196 of the 482 participants undergoing chart review, which questions the applicability of these findings as only a select group received the index test. Van den Bosch (2004) did not report the number of participants without fracture who had a history of trauma. Whilst they reported the number of 'traumatic fractures,' it was unclear as to how they established and defined a history of trauma.

The majority of studies used X-ray as their reference standard and had 'low concern regarding applicability' in this domain. One study (Henschke et al., 2009) was rated as high risk given the use of long term follow-up as their reference standard, a method known to lead to under-reporting of fractures (Grigoryan et al., 2003). Fifty

precent of studies (Deyo & Diehl, 1986; Gibson & Zoltie, 1992; Patrick, Doris, Mills, Friedman, & Johnston, 1983; Scavone et al., 1981) did not provide a clear explanation in respect to participant withdrawals.

Ideally a systematic review would include findings from those studies that are applicable and have low risk of bias. None of the studies identified by the current search met this criterion. Consequently, findings of the studies included in the review need to be interpreted cautiously. The study by Reinus and colleagues (1998) was particularly concerning as it showed high risk of bias in three of the four domains and applicability concerns in two of the three domains. Also, as this study only included chart review of patients with positive X-rays, it was not be possible to construct 2x2 contingency tables to establish diagnostic accuracy. This study was therefore excluded from index test evaluation.

Table 2.2 *Assessment of study quality for fracture using the QUADAS-2*

Author	Risk of bias				Applicability concerns		
	Patient selection	Index tests	Reference Standard	Flow & Timing	Patient selection	Index tests	Reference Standard
Deyo (1986)	L	L	H	H	L	L	L
Gibson (1992)	L	L	H	H	L	L	L
Henschke (2009)	L	L	H	H	L	L	H
Patrick (1983)	L	?	?	L	L	L	L
Reinus (1998)	L	H	H	H	?	H	L
Roman (2010)	L	?	?	H	L	L	L
Scavone (1981)	L	L	?	?	?	L	L
van den Bosch (2004)	L	L	?	?	L	?	L

Note. L = Low risk, ? = Unclear, H = High risk

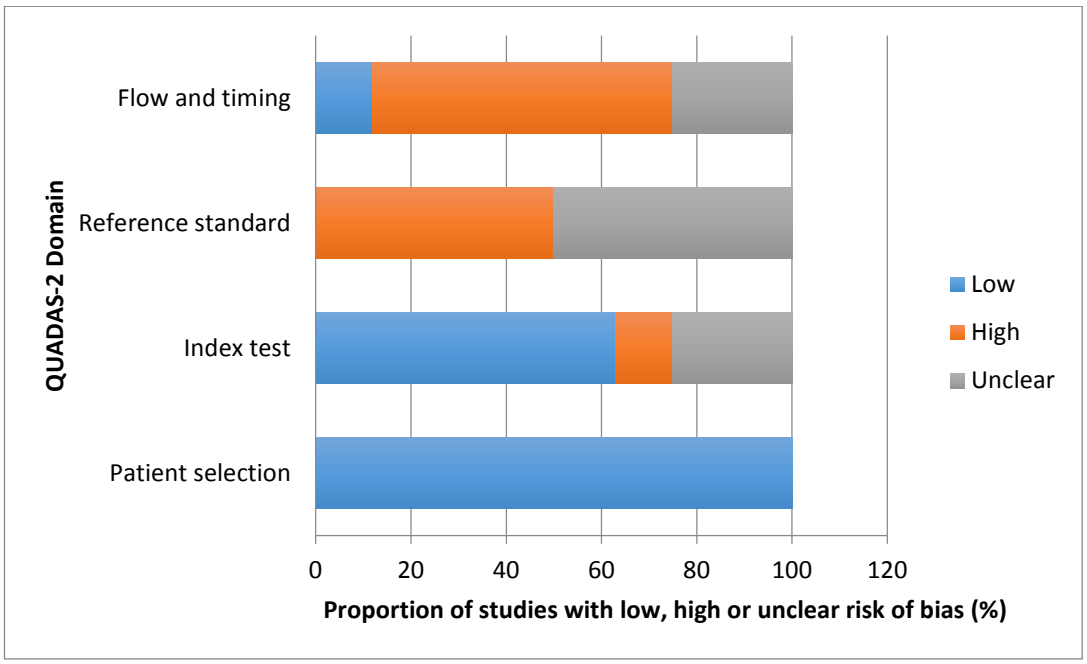


Figure 2.2 QUADAS-2 combined study results to illustrate overall risk of bias

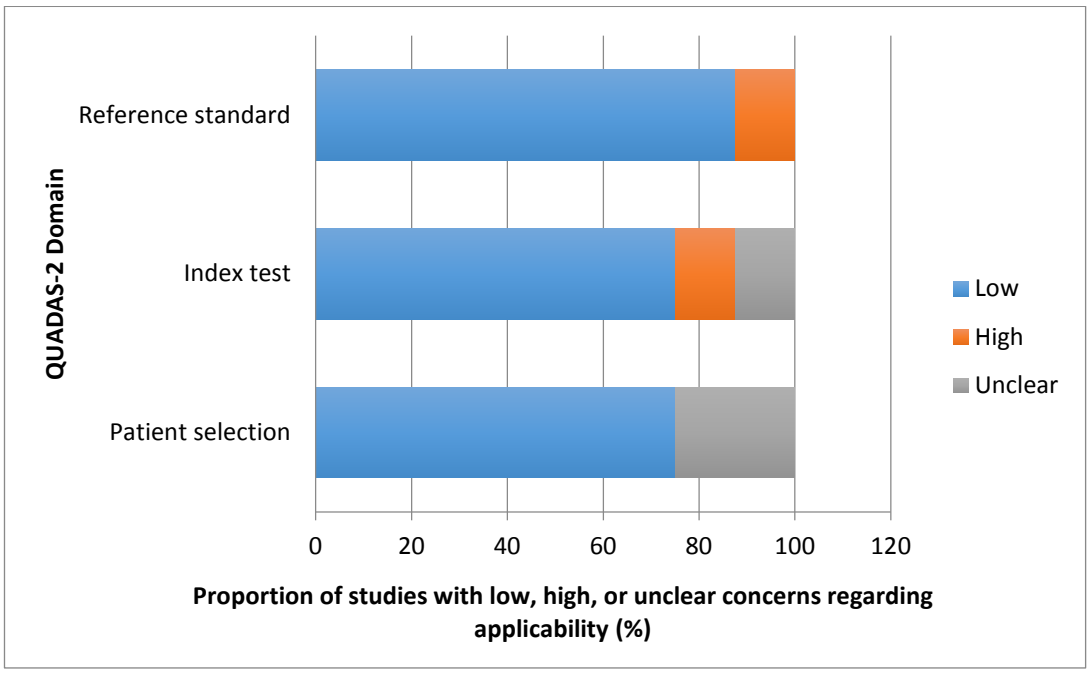


Figure 2.3 QUADAS-2 combined study results to illustrate overall applicability

Description of studies. Study characteristics of the eight studies that met all inclusion criteria are detailed in Table 2.3. Half of the studies were conducted in a primary care setting (Deyo & Diehl, 1986; Henschke et al., 2009; Scavone et al., 1981; van den Bosch et al., 2004), one study was conducted in a secondary care setting (Roman et al., 2010), and the remaining three studies were conducted in tertiary care settings (Gibson & Zoltie, 1992; Patrick et al., 1983; Reinus et al., 1998). The total

number of participants across all studies from primary care was 4671. In secondary care settings there were a total of 1448 participants, and a total of 1259 participants were included from the tertiary care setting. Four studies (Deyo & Diehl, 1986; Gibson & Zoltie, 1992; Henschke et al., 2009; Reinus et al., 1998) were prospective cohort studies. All of these but one (Deyo & Diehl, 1986) reported consecutive recruitment of participants. Four studies (Patrick et al., 1983; Roman et al., 2010; Scavone et al., 1981; van den Bosch et al., 2004) were conducted as retrospective chart reviews.

Plain radiology was the primary reference standard in seven of the eight studies. Roman and colleagues (2010) used either plain radiology or CT as their reference standard. No study reported a specific assessment method or criterion used to define vertebral fracture. One study (Henschke et al., 2009) used 12-month follow-up as their reference standard and confirmed any 'suspected' fractures with diagnostic imaging.

Study findings. The prevalence of vertebral fracture across these studies varied from 0.7% in primary care to 7.2% in tertiary care (see Table 2.3 for detail). Table 2.4 provides detail of the diagnostic accuracy of variables used to screen for vertebral fractures. All studies except one (Reinus et al., 1998) have been included in Table 2.4, although the study by Reinus et al. (1998) met initial inclusion criteria it could not be used for diagnostic accuracy calculations due to missing data and high risk of bias. Overall, 21 different index tests related to patient demographics and history findings were investigated. Three studies (Henschke et al., 2009; Roman et al., 2010; van den Bosch et al., 2004) investigated 'older age' as an index test. However, a different cutoff point was used with each study (i.e. age >52, >54, and >70 years). Gender was investigated as an index test by two studies (Roman et al., 2010; van den Bosch et al., 2004). A history of trauma was investigated by five studies, with three studies (Deyo & Diehl, 1986; Henschke et al., 2009; Scavone et al., 1981) using the term 'significant trauma,' while the other two (Gibson & Zoltie, 1992; Patrick et al., 1983) did not provide a definition of trauma. History of corticosteroid use was investigated by two studies and the presence of absence of leg pain by two studies. Otherwise, there was limited consensus on index tests between studies, with nine index tests investigated by just one study. Two studies (Henschke et al., 2009; Roman et al., 2010) developed diagnostic rules using combinations of four of five index test results.

Due to heterogeneity between studies, no meta-analysis could be performed and results have not been pooled.

Table 2.3 *Study characteristics for fracture*

Authors	Study Design	Setting	Patients	Prevalence of Fracture	Reference Standard	Withdrawals
Deyo (1986)	Prospective cohort (Sampling unclear)	Primary care walk-in clinic	621 patients referred for LBP treatment, 311 received lumbar spine X-rays	4.5% (n=14)	X-ray	Unclear
Gibson (1992)	Prospective cohort (Consecutive enrolment)	Tertiary care Accident & Emergency department	225 patients with acute LBP, 108 receiving lumbar spine X-rays	6.5% (n=7)	X-ray (68% with history of trauma & 36% without trauma received reference standard)	Unclear
Henschke (2009)	Prospective cohort (Consecutive enrolment)	Primary care General Practitioner, Physiotherapy & Chiropractor clinics	1172 patients receiving primary care for acute LBP	0.7 % (n=8)	12 month follow up	12 cases healthcare provider followed up rather than participant, no withdrawals
Patrick (1983)	Retrospective chart review	Tertiary care Accident & Emergency department	552 consecutive patients referred for lumbar spine X-ray (99% with LBP)	7.2% (n=40)	X-ray	Unclear
Reinus (1998)	Prospective cohort (Consecutive enrolment)	Tertiary care Accident & Emergency department	482 patients referred for lumbar spine X-ray 92% with back pain	11 %, 2.1% acute (n=10 acute, n= 24 indeterminate age, n=21 chronic)	X-ray	No withdrawals, chart review conducted in only 196 participants
Roman (2010)	Retrospective chart review	Secondary care spine clinic	1448 consecutive patients seen at a hospital spine surgery centre	2.6% (n=38)	X-ray or CT	No withdrawals
Scavone (1981)	Retrospective chart review	Radiology department	871 patients referred from primary care for lumbar spine X-rays	3% (n=26)	X-ray	History unable to be obtained in 57 participants
Van den Bosch (2004)	Retrospective chart review	Radiology department	2,007 patients referred from primary care for X-ray for LBP	4.1% (n=83)	X-ray	93 participants excludes as no records available, no withdrawals

Note. LBP = Low back pain, n = sample size

Primary care. Four studies were conducted in a primary healthcare setting (Deyo & Diehl, 1986; Henschke et al., 2009; Scavone et al., 1981; van den Bosch et al., 2004). The study sample size of these 4 studies ranged from 311 to 2,007, with a total sample of 4,361. Three of the four studies (Deyo & Diehl, 1986; Scavone et al., 1981; van den Bosch et al., 2004) used X-ray as their reference standard, whilst the remaining study (Henschke et al., 2009) used phone follow-up over a 12-month period. In this study follow-up calls were made at 6 weeks, 3 months and 12 months. During these calls, participants were asked if they had been diagnosed with a spinal fracture, infection, arthritis or cancer. Any participant who answered 'yes' was subsequently reviewed by a rheumatologist and underwent diagnostic imaging or further investigations to confirm the diagnosis. Seven different index tests were investigated across the four studies.

The prevalence of vertebral fractures in primary care reported by these authors ranged from 0.7% (Henschke et al., 2009) to 4.5% (Deyo & Diehl, 1986), with most reporting a prevalence around 3-4%. With respect to the diagnostic accuracy of the index tests, Henschke and colleagues (2009) reported positive likelihood ratios greater than 10 for three tests, i.e age >70 years, history of significant trauma and prolonged corticosteroid use. Whilst a point estimate for a likelihood ratio greater than 10 suggests a conclusive shift in probability of the presence of a vertebral fracture, the 95% confidence intervals indicate that this estimate needs to be considered with caution. In particular, the 95% confidence intervals (CI) of the positive likelihood ratio for prolonged corticosteroid use were very wide (11.62 – 165.22), as two of the eight participants with vertebral fractures reported a history of steroid use. Also, eight participants of the entire sample of 1,172 had used steroids. The study by Deyo et al. (1986) investigated significant trauma and steroid use but the likelihood ratios they reported indicated a slight shift in probability (LR+ 3.42 [95% CI 1.57,7.45]). Scavone and colleagues (1981) also found that a history of significant trauma conclusively shifted the probability of a positive diagnosis (LR+ 12.85 [95% CI 8.58,19.24]). All of the findings with useful positive likelihood ratios had poor negative likelihood ratios and low sensitivities.

Henschke and colleagues (2009) also presented a diagnostic rule that they considered might be useful for considering combinations of variables that improve the precision of identification of vertebral fractures. They included four variables with the highest positive likelihood ratios that had a statistically significant association ($p<0.1$)

with vertebral fracture. Three index tests were selected: a history of corticosteroid use, a history of significant trauma, and age >70 years. Although the diagnostic accuracy of female gender was not reported along with the other index tests, these authors subsequently included it in this diagnostic rule without any explanation for this decision. Henschke and colleagues reported estimates of the sensitivity and specificity (without confidence intervals) and the positive likelihood ratios for this diagnostic rule (see Table 2.4 for detail). Unfortunately, they did not provide sufficient data to allow reconstruction of 2 x 2 contingency tables and confirmation of their findings. Henschke and colleagues (2009) stated that for three or more positive findings, the specificity was 100%, sensitivity was 38%, and the positive likelihood ratio was 218.3 (45.6-953.8). Although the likelihood ratio is very high, the very wide confidence interval reflects the small number of participants with three positive findings (n=3). No participant in their study met all four criteria within the diagnostic rule.

Secondary care. One of the identified studies (Roman et al., 2010) was performed in secondary care. The sample size in this study was 1448 and the prevalence of vertebral fracture was 2.6% (n=38). Roman et al. (2010) used either X-ray or CT as their reference standard and investigated the diagnostic accuracy of eight index tests as individual measures of osteoporotic or wedge fractures. Roman et al. considered a positive likelihood ratio of >1.5, and a negative likelihood ratio of < 0.5 'useful'. The six index tests with useful positive likelihood ratios were: age >52 years, body mass index ≤ 22, female gender, does not exercise regularly, sitting decreases pain, concomitant osteoarthritis and no leg or buttock pain. The index tests with useful negative likelihood ratios were: age, body mass index, and does not exercise regularly. However, age >52 years was the only index test with a negative likelihood ratio that indicated a moderate shift in probability (LR- 0.14 [95% CI 0.03,0.45]).

Roman and colleagues also presented a diagnostic rule developed through the use of backward stepwise logistic regression analysis. These authors first identified conditionally independent variables with 'useful' likelihood ratios. They then excluded variables with *p*-values greater than 0.1 (concomitant osteoarthritis). The remaining five index tests were entered into the diagnostic rule. Roman et al. reported specificity of 96% and 99% and positive likelihood ratios of 9.6 and 9.3 for the presence of four out of five or five out of five positive tests respectively. These likelihood ratios indicated a moderate shift in probability of vertebral fracture being present in the event of a patient meeting these criteria. However, this ability to identify vertebral fractures was not

matched by an ability to rule out the presence of such a fracture given the low sensitivities (37% and 3%) and poor negative likelihood ratios (0.65 and 0.97) respectively.

Tertiary care. Three studies (Gibson & Zoltie, 1992; Patrick et al., 1983; Reinus et al., 1998) were conducted in tertiary care settings (accident and emergency departments). The study by Reinus et al. (1998) had a sample size of 482, and reported a prevalence of 2.1% for acute fractures and 11% including all vertebral fractures. This study was removed from any further analysis due to poor methodology, and the prevalence is not considered to be reliable. The remaining two studies (Gibson & Zoltie,1992; Patrick et al., 1983) had sample sizes of 108 and 552, and reported prevalence of 6.5% and 7.2% respectively. Both studies used X-ray as the reference standard. However, in the study by Gibson & Zoltie (1992), 68% of participants with a history of trauma and 36% of patients without trauma received the reference standard. These studies (Gibson & Zoltie, 1992; Patrick et al., 1983) investigated the relationship between a history of trauma, and found Gibson et al. reported a sensitivity of 100% and a negative likelihood ratio of 2.06 for this finding. These results were not dissimilar to those of Patrick et al. who reported sensitivity of 80% and negative likelihood ratio of 1.77 respectively. Both studies reported poor specificity and the positive likelihood ratios less than five. These studies suggest that a history of trauma does not have sufficient diagnostic accuracy to enable the identification or exclusion of a vertebral fracture.

Table 2.4 *Clinical signs and diagnostic accuracy data extracted from eligible studies for fracture*

Index test	LR+ (95% CI)	LR- (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Demographic				
Age > 70 years (Henschke et al., 2009)	11.19 (4.65, 19.48)	0.52 (0.23, 0.82)	0.50 (0.16, 0.84)	0.96 (0.94, 0.97)
Age >52 years (Roman et al., 2010)	1.50 (1.3, 1.5)	0.14 (0.03, 0.45)	0.95 (0.83, 0.88)	0.39 (0.38, 0.40)
Age >54 years (van den Bosch et al., 2004)	1.72 (1.54, 1.91)	0.33 (0.20, 0.53)	0.83 (0.73, 0.90)	0.52 (0.49, 0.54)
Gender (female) (Roman et al., 2010)	1.50 (1.3, 1.6)	0.26 (0.10, 0.60)	0.90 (0.76, 0.96)	0.41 (0.41, 0.42)
Gender (female) (van den Bosch et al., 2004)	1.26 (1.1, 1.45)	0.65 (0.46, 0.92)	0.72 (0.61, 0.82)	0.43 (0.41, 0.45)
History				
Significant trauma (major in young, minor in elderly) (Henschke et al., 2009)	10.03 (2.87, 35.13)	0.77 (0.52, 1.15)	0.25 (0.03, 0.65)	0.98 (0.96, 0.98)
Significant trauma (Deyo & Diehl, 1986)	3.42 (1.57, 7.45)	0.72 (0.48, 1.06)	0.36 (0.13, 0.65)	0.90 (0.86, 0.93)
Significant trauma (Scavone et al., 1981)	12.85 (8.58, 19.24)	0.36 (0.21, 0.62)	0.65 (0.44, 0.83)	0.95 (0.93, 0.96)
Trauma (not defined) (Gibson & Zoltie, 1992)	2.06 (1.68, 2.52)	0.00 (0.00, NaN)	1.00 (0.56, 1.00)	0.51 (0.42, 0.61)
Trauma (not defined) (Patrick et al., 1983)	1.77 (1.48, 2.13)	0.36 (0.20, 0.68)	0.8 (0.50, 0.59)	0.55 (0.50, 0.59)
BMI ≤ 22 (Roman et al., 2010)	2.30 (1.4, 3.4)	0.74 (0.54, 0.91)	0.38 (0.24, 0.55)	0.83 (0.82, 0.84)
No gait abnormality (Roman et al., 2010)	0.86 (0.65, 1.02)	1.5 (0.91, 2.20)	0.66 (0.50, 0.79)	0.23 (0.22, 0.23)
Does not exercise regularly (Roman et al., 2010)	1.50 (1.20, 1.60)	0.43 (0.20, 0.80)	0.81 (0.65, 0.91)	0.44 (0.43, 0.45)
Sitting decreases pain (Roman et al., 2010)	1.60 (1.20, 1.90)	0.87 (0.82, 0.92)	0.29 (0.27, 0.32)	0.81 (0.79, 0.83)

Index test	LR+ (95% CI)	LR- (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Concomitant Osteoarthritis (Roman et al., 2010)	1.10 (0.70, 1.40)	0.97 (0.67, 1.30)	0.50 (0.35, 0.65)	0.52 (0.51, 0.52)
No leg or buttock pain (Roman et al., 2010)	2.20 (1.2, 3.60)	0.81 (0.58, 0.97)	0.31 (0.16, 0.49)	0.86 (0.85, 0.87)
Hip/leg pain (Scavone et al., 1981)	0.21 (0.01, 3.35)	1.07 (1.02, 1.14)	0.00 (0.00, 0.13)	0.91 (0.89, 0.93)
Sciatica (Scavone et al., 1981)	0.42 (0.06, 2.92)	1.06 (0.98, 1.15)	0.04 (0.00, 0.20)	0.91 (0.89, 0.93)
Steroid use (Deyo & Diehl, 1986)	3.97 (0.20, 79.15)	0.97 (0.89, 1.07)	0.00 (0.00, 0.23)	0.99 (0.98, 1.00)
Prolonged corticosteroid use (Henschke et al., 2009)	48.50 (11.62, 165.22)	0.75 (0.41, 0.93)	0.25 (0.03, 0.65)	0.99 (0.99, 1.00)
Altered sensation from trunk down (Henschke et al., 2009)	0.00 (0.00, 21.01)	1.02 (0.22, 0.79)	0.00 (0.00, 0.37)	0.98 (0.97, 0.99)
Diagnostic rules				
Diagnostic rule[1] 1 positive (Henschke et al., 2009)	1.80 (1.10, 2.00)	?	0.88 (?)	0.50 (?)
Diagnostic rule[1] ≥2 positive (Henschke et al., 2009)	15.50 (7.20, 24.60)	?	0.63 (?)	0.96 (?)
Diagnostic rule[1] ≥ 3 positive (Henschke et al., 2009)	218.30 (45.60, 953.80)	?	0.38 (?)	1 (?)
1/5 positive tests [2] (Roman et al., 2010)	1.04 (0.92, 1.10)	0.39 (0.07, 2.10)	0.97 (0.89, 0.99)	0.06 (0.06, 0.07)
2/5 positive tests [2] (Roman et al., 2010)	1.40 (1.30, 1.80)	0.16 (0.04, 0.51)	0.95 (0.83, 0.99)	0.34 (0.33, 0.34)
3/5 positive tests [2] (Roman et al., 2010)	2.50 (1.90, 2.80)	0.34 (0.19, 0.46)	0.76 (0.61, 0.87)	0.68 (0.68, 0.69)
4/5 positive tests [2] (Roman et al., 2010)	9.60 (3.70, 14.90)	0.65 (0.50, 0.79)	0.37 (0.24, 0.51)	0.96 (0.95, 0.97)
5/5 positive tests [2] (Roman et al., 2010)	9.30 (1.40, 60.20)	0.97 (0.92, 0.99)	0.03 (0.01, 0.08)	0.99 (0.98, 0.99)

Note. LR = likelihood ratio, CI=Confidence Interval, BMI ,= Body Mass Index, NaN, = calculation cannot be performed as 2 x 2 tables include one or more zeros, ? = not reported and insufficient raw data provided to calculate.

[1] Henschke diagnostic rule (includes: female gender, age > 70 years, significant trauma (major in young, minor in elderly) and prolonged corticosteroid use).

[2] Roman diagnostic rule (includes age >52 years, no buttock or leg pain, BMI ≤ 22, does not exercise regularly and female gender).

2.2.5 Discussion

The aim of this review was to provide information on the clinical prevalence of vertebral fractures and to determine the diagnostic accuracy of red flag questions for screening for fracture in patients with low back pain. A total of eight studies involving patients with a primary complaint of low back pain were included in this review. One study (Reinus et al., 1998) was removed from index test analysis as poor methodology and reporting of data meant that further statistical analysis could not be performed. Amongst the remaining seven studies, 21 different index tests were investigated. The following discussion has been divided into specific subheadings as the recommended by the Cochrane guidelines, this format will be repeated for each of the reviews reported in this Chapter.

Factors affecting interpretation

Prevalence and population. Studies conducted in primary, secondary, and tertiary healthcare were included in this review so that the results could be considered across each of these settings. Whilst all of the included studies calculated clinical prevalence, no study investigated incidence, which limits the generalisability of the study findings to wider populations. In most studies it was unclear whether they only included new fractures or also included all existing fractures. Hence, these studies were considered as having reported point prevalence. One study (Henschke et al., 2009) followed participants over a 12-month period, enabling them to report period prevalence.

Variation in study setting and characteristics of the included population commonly influences the prevalence of a condition of interest (Schmidt & Factor, 2013). The current review of the literature has demonstrated that the highest prevalence of vertebral fractures (6.5-7.2%) was observed in studies conducted in the tertiary care setting, i.e accident and emergency clinics (Gibson & Zoltie, 1992; Patrick et al., 1983). In contrast, the prevalence in a population of patients referred from primary care for lumbar X-ray ranged from 3% to 4.5% (Deyo & Diehl, 1986; Scavone et al., 1981; van den Bosch et al., 2004). Interestingly, a lower prevalence (2.6%) was reported by the one study (Roman et al., 2010) set in a secondary care spine clinic. A higher prevalence of fractures in the accident and emergency environment would be expected, as that is the most likely place that patients with acute fractures would present for assessment and treatment. The lowest prevalence (0.7%) reported across all of the included studies was

that by Henschke and colleagues (set in primary care). However, these authors considered diagnoses made over a 12 month follow-up period to determine prevalence. This method has been shown to misdiagnose fractures, as up to 70% of fractures may be missed on clinical examination (Grigoryan et al., 2003). Also, the study by Henschke and colleagues may have also been affected by differential verification bias, as only a subset of patients with suspected serious pathologies underwent diagnostic imaging. If all participants had undergone similar assessment, a number of additional fractures may have been identified (Schmidt & Factor, 2013).

Index tests. Although three systematic reviews (Downie et al., 2013; Henschke et al., 2008; Williams et al., 2013) have previously been conducted to investigate the use of red flags to 'screen for' fractures, the authors have focused on the utility of such questions to identify fractures. Hence, these reviews have focused on positive likelihood ratios and have suggested red flags that may be useful for diagnosis rather than screening. It would also be useful to know which negative findings can be used to lower the suspicion of fracture, so that clinicians can confidently refer these patients for a trial of conservative management without undergoing unnecessary investigations such as plain radiography.

When tests are used to screen for serious pathologies, a false negative test can have significant adverse consequences. Hence, such tests need to have high sensitivity. To be of significant benefit clinically, the test should also have a low negative likelihood ratio reflecting that a negative test result leads to a conclusive reduction in the probability of the condition of interest being present. The current review has revealed that one red flag question that met these criteria was age greater than 52 years, and this was based on a single study (Roman et al., 2010). Conversely, three index tests (age> 70 years, a history of significant trauma, and prolonged corticosteroid use) demonstrated high specificity (95-99%) and positive likelihood ratios greater than 10. However, a history of corticosteroid use was based on two positive findings in a cohort of 1,172 participants and is therefore uninformative (Henschke et al., 2009). Whilst these specific red flag questions might be useful for the identification of vertebral fractures, the absence of a positive test cannot be relied upon to rule out such pathology due to poor sensitivity.

Several authors (Deyo & Diehl, 1988; Henschke et al., 2009; Roman et al., 2010) have discussed the benefits of combining red flag questions to improve diagnostic

accuracy and utility. The Cochrane review by Williams et al. (2013) provided detail of the enhanced diagnostic utility of combinations of red flag questions over single questions used as standalone tests. For example, they reported that age > 74 years combined with female gender results in a positive likelihood ratio of 16.17 (compared to LR+ 3.69-11.5 and LR+1.26-1.5 respectively for these as standalone tests). Whilst it is more likely that females older than 74 years will have a vertebral fracture, the downside of this combination is that it also results in a high false negative rate (sensitivity 25-45%) and misdiagnosis of up to 75% of the fractures (Henschke et al., 2009; van den Bosch et al., 2004).

Other authors (Henschke et al., 2009; Roman et al., 2010) have attempted to develop diagnostic rules using a combination of index tests to enhance the process of identification of vertebral fractures. However, these diagnostic rules need to be considered cautiously. Neither of the proposed diagnostic rules has subsequently been validated and therefore cannot yet be recommended for use in clinical practice. Additionally, both diagnostic rules prioritise specificity over sensitivity, and implementation without sound clinical reasoning could lead to large numbers of missed cases.

The diagnostic rule developed by Henschke and colleagues (2009) included four red flag questions, and reported high specificity and conclusive positive likelihood ratios for two out of four (LR+ 15.5 [7.2-24.6]) and three out of four positive findings (LR+ 218.3 [45.6-953.8]). The latter likelihood ratio appears unusually high, and it is even more unusual that the two systematic reviews (Downie et al., 2013; Williams et al., 2013) conducted by the same group of authors, managed to recalculate the likelihood ratio and found different likelihood ratios of 916 (95% CI 50-16,300), and 906.11 (95% CI 50 -16,299) respectively. Due to the trade of high specificity for poor sensitivity in this rule, up to 63% of vertebral fractures in their sample would be missed.

The diagnostic rule proposed by Roman and colleagues (2010) was well designed using backward logistic regression analysis to elect the five red flags included in the rule. They reported that the specificity of four out of five positive findings was 96%, and although the positive likelihood ratio suggested a moderate to conclusive change in probability (LR+9.6), it lacked precision, with a confidence interval of 3.7-14.9. Also, the sensitivity was 37%, suggesting high false negative rates.

Reference standard. Plain radiography was the most commonly used reference standard in this review. While plain radiography is useful to correctly classify vertebral fractures, it is currently overused, and many clinical guidelines (ACC, 1999; Koes et al., 2010) recommend against routine use. Concerns have arisen regarding the implications of radiation exposure, unnecessary cost, and issues associated with diagnostic labelling (Chou et al., 2007; Flynn et al., 2011). Some clinical guidelines have proposed the use of MRI rather than plain radiography in patients where there is suspicion of fracture (National Institute for Health and Care Excellence, 2009). Other authors (Chou et al., 2012; Deyo et al., 2014) have suggested the use of red flag questions to guide selective ordering of plain radiographs or advanced diagnostic imaging to reduce the current overuse. However, the current review has revealed that this proposal is unrealistic at the present point in time, given that there is little evidence to support the diagnostic utility of red flag questions.

Conclusion. This review has demonstrated that the prevalence of vertebral fracture is relatively low and varies depending on the clinical setting. There is moderate evidence that the prevalence is likely to be higher in tertiary care accident and emergency settings. There is suspicion that clinical assessment is likely to misdiagnose a number of cases, as evidenced by the low prevalence in the study by Henschke et al. (2009). This is also supported by previous research suggesting that up to 70% of fractures may be missed due to inadequate assessment (Grigoryan et al., 2003). The prevalence between primary and secondary care settings was similar, with a moderately small range from 2.6-4.5% between studies that used diagnostic imaging as the reference standard. It is also apparent from this review that the incidence of vertebral fractures within a population of patients presenting with low back pain has not been investigated and is therefore unknown.

The current evidence suggests that few red flag questions can be used as stand-alone tests to either rule in or rule out vertebral fractures. Whilst there is some consensus that red flag questions may be useful in raising the suspicion of the presence of a vertebral fracture (e.g. age > 70 years, or a history of significant trauma), these features currently lack precision and generalisability. It is clear that further questioning and examination is required before a clinician can confidently determine whether or not vertebral fracture is likely to be present and when referral for diagnostic imaging is appropriate. Uncritical acceptance of information obtained from red flags questions

could lead to over-investigation and overuse of diagnostic tests such as plain radiography, or alternatively, misdiagnosis.

With respect to screening for vertebral fractures, the only red flag question that demonstrated that a negative outcome could result in a conclusive reduction in the probability of a fracture being present was age greater than 52 years. However, this finding is not intuitive, in that it would seem inappropriate to assume that all patients under 52 years of age are unlikely to have a vertebral fracture. Therefore, this should be applied cautiously with further questioning to exclude any risk factors such as a history of significant trauma or prolonged corticosteroid use that could raise the suspicion of a fracture.

Further research is required to investigate the diagnostic accuracy of combinations of red flags to identify or rule out vertebral fracture. At present there is insufficient evidence to support the use of red flag questions to obtain a confident estimate of the likelihood of the presence or absence of a fracture.

Weaknesses of this review. The literature search, data extraction and analysis were performed by a single author, opening the possibility of study selection or interpretation bias. Whilst this review included only studies published in English, the systematic review by Williams et al. (2013), which did not exclude such studies, did not identify any further studies.

The overall methodological quality of original publications was poor, with potential risk of bias in all studies. Most studies included in this review were not conducted primarily as diagnostic accuracy studies. Therefore, there was unclear or high risk of bias in several domains due to poor reporting. Due to risk of bias and heterogeneity between studies, statistical pooling of diagnostic accuracy findings could not be performed, and a pooled prevalence could not be established.

Strengths of this review. This systematic review followed the Cochrane Collaboration and PRISMA guidelines. A comprehensive search of the databases and review of all relevant systematic reviews ensured that there was low risk of missing any potentially eligible studies. Critical analysis of all identified studies was performed using a validated tool to assess the methodological quality of selected studies. Studies that were at high risk of bias were excluded from further investigation to provide a more

accurate overview of the current literature. Hence, this systematic review provided an independent summary of the existing relevant literature.

Applicability of findings to this thesis. This review has demonstrated that there is currently insufficient high quality evidence regarding the prevalence or incidence of vertebral fractures amongst patients presenting to secondary or tertiary care complaining of low back pain. Additionally, in the current literature there is a lack of good quality evidence regarding the diagnostic accuracy of red flag questions and their use cannot be supported or refuted based on this evidence. Hence, this thesis explored the prevalence of vertebral fractures in secondary and tertiary care, and the incidence of fractures in tertiary care. This thesis also investigated the diagnostic accuracy of red flag questions in secondary and tertiary care settings. Based on previous research and clinical guidelines, the following red flag questions were included as index tests in this study: 'history of significant trauma', concomitant osteoarthritis or osteoporosis, older age, gender, and 'history of prolonged corticosteroid use'.

This review highlighted the need to follow the STARD guidelines (Bossuyt et al., 2015) for reporting of diagnostic accuracy studies to ensure transparency and study reproducibility. This review also emphasised the need for consecutive, prospective recruitment of participants, and correct conduct and blinding of index tests and the reference standard. Several studies were found to be at high or unclear risk of bias due to the reference standard or flow and timing of the study. Therefore, the red flag questions (index tests) in this thesis were answered directly before the reference standard, and the reference standard with the highest precision for classification of vertebral fractures (MRI) was selected.

2.3 Target Condition: Spinal Malignancy

2.3.1 Background

Malignancy is reported to be the second most common serious pathology to affect the lumbar spine (Deyo et al., 1992; van den Bosch et al., 2004). For the purposes of this review, the term 'spinal malignancy' was considered to include any primary malignant or secondary metastatic tumour affecting the spine. Early diagnosis of spinal malignancy is essential to reduce the risk of pathological fracture, further haematogenous or lymphatic spread of the disease, spinal cord compression and mortality (Loblaw, Perry, Chambers, & Laperriere, 2005). Around 10% of all malignancies will initially present in the spine, with patients generally presenting with

back pain, which can be local, mechanical or radicular in nature (Sciubba & Gokaslan, 2006). The most common malignancies to affect the spine are metastases from prostate, breast, or lung carcinoma, followed by multiple myeloma and non-Hodgkin's lymphoma (Motamedi, Ilaslan, & Seeger, 2004; Sciubba & Gokaslan, 2006). Primary tumours are rare, but occasionally malignant tumours such as chordomas, chondrosarcomas, or malignant peripheral nerve sheath tumours affect the lumbar spine (Dang et al., 2015).

Prevalence/incidence. Spinal malignancy is rare and affects less than 1% of patients presenting to primary care clinicians with low back pain (Henschke et al., 2013). However, incidence is likely to rise over time, with evidence that rates of cancer in New Zealand have soared over the past decade (2000-2010). Cancer registrations during this period have risen 19% (Ministry of Health, 2013a). An aging population and advances in cancer treatment have led to increased life expectancy and an increased number of people living with illness. As survival rates increase, it is likely that the prevalence of spinal metastases will continue to increase. Ten to 30% of cancer patients have metastases during the course of their illness, and the spine is the most common site for metastases, with evidence that around 90% of cancer patients have spinal metastases on post-mortem examination (Sciubba & Gokaslan, 2006; Walsh et al., 1997).

In New Zealand, all newly diagnosed primary tumours are reported to the national cancer registry. However, there is no registry for metastatic or secondary cancers. With respect to spinal cancer, the cancer registry classifies primary malignant tumours depending on site, either in the vertebral column (not specific to cervical, thoracic or lumbar) or in the pelvic bones, sacrum and coccyx (Ministry of Health, 2013). The combined annual incidence of such tumours varied from 0.12-0.37 per 100,000 per year between 1995 and 2012 (Dwyer, 2015) (New Zealand Cancer Registry data requested from the national office). In contrast, one study (Dreghorn, Newman, Hardy, & Dickson, 1990), based in England, reported a much higher incidence of primary tumours affecting the spine, i.e. 2.5-8.5 per 100,000 per year. Another study by Dang et al. (2015) performed a retrospective audit investigating the clinical features of consecutive 438 patients with primary spinal tumours who were admitted to a university teaching hospital in China over an 8 year period. Dang and colleagues reported that these tumours occur predominately between the ages of 18 and 59 years and that the risk increases significantly after 40 years of age. They also reported that primary tumours in the lumbar spine were less common than in the cervical or thoracic spine.

Index tests. Red flag questions specific to cancer that have commonly been recommended in clinical guidelines are: insidious onset of pain, age >50 years, previous history of cancer, no improvement after one month of conservative management, no relief with bed rest, unexplained weight loss, and being systemically unwell (Koes et al., 2010). In New Zealand, the ACC acute low back pain guidelines (Accident Compensation Corporation, 1999) include these same questions but also an additional question related to the presence of 'severe worsening pain, especially at night or when lying down'.

Reference standards for identification of malignancy. The single best non-invasive reference standard to screen for malignancy in the lumbar spine is MRI (Kosuda et al., 1996). MRI has high sensitivity and specificity, reported to be between 83-98% and 90-98% (respectively) when compared to autopsy or surgery as the gold standard for the diagnosis of spinal malignancy (Joines, McNutt, Carey, Deyo, & Rouhani, 2001; Kosuda et al., 1996). The positive and negative likelihood ratios (LR+ 8-31; LR- 0.07-0.19) for MRI indicate that it has utility for both ruling in or out spinal malignancy (Jarvik & Deyo, 2002). Although bone scanning (scintigraphy) or Single Emission Photon Computed Tomography (SPECT) can identify metastatic lesions earlier than plain radiography, these imaging techniques have low image resolution and poor specificity (Morris et al., 2002; Uchida et al., 2013).

The use of plain radiographs in the lumbar spine to screen for tumours is known to be problematic. Although, the majority of spinal metastatic lesions are osteolytic (destroy bone), up to 50% of bone must be eroded before a lytic lesion can be detected on plain radiography (Sciubba & Gokaslan, 2006). Hence, plain radiography has poor sensitivity for malignancy (60%) and negative likelihood ratios that indicate that the absence of findings on an X-ray only slightly reduces probability of the presence of this pathology (LR- 0.4-0.42) (Jarvik & Deyo, 2002).

Previous systematic reviews. A preliminary search of the literature identified three systematic reviews (Downie et al., 2013; Henschke et al., 2013; Henschke, Maher, & Refshauge, 2007) that have investigated the use of red flag questions to screen for malignancy in patients with low back pain. All three systematic reviews were assessed for quality using the PRISMA guidelines (Liberati et al., 2009). The Cochrane review by Henschke and colleagues (2013) is an update of the authors' previous review published in 2007. The authors published the updated review following the publication

of their own study investigating the prevalence and screening for serious spinal pathologies in primary care (Henschke et al., 2009). The 2013 review considered 20 different index tests across 8 primary studies. Seven of these tests were investigated by more than one study. Sixteen of the 20 index tests could be obtained from the patient history and were therefore considered relevant to the current review. Henschke et al. (2013) considered that there was some evidence to suggest that a 'previous history of cancer' had sufficient diagnostic accuracy to be considered useful as a stand-alone test. These authors reported that no other red flag questions had proven diagnostic utility. These authors suggested that a solution that might enhance the identification of spinal cancer would be to consider information obtained from combinations of red flags. The review by Henschke et al. (2013) followed the Cochrane Collaboration guidelines and recommended search strategy. Overall, the review was well conducted. However, although the authors highlighted the fact that several studies had significant methodological flaws, all studies were included in the analysis. Therefore, their recommendations were based on some evidence obtained from poor quality studies.

The systematic review by Downie and colleagues (2013) included the same studies that were included in the systematic review by Henschke and colleagues (2013). Downie et al. drew the same conclusion, which was that only a history of malignancy had evidence that supported that it was a risk factor that increased the likelihood of spinal malignancy being present. Downie and colleagues provided the true positive, false positive, true negative and false negative data that they had extrapolated from the original studies. Although this allowed transparency of their analysis, it also allowed the author of this thesis to identify several mistakes in their data entry. An example was that the main conclusion is based on the findings of only two studies (Deyo & Diehl, 1988; Reinus et al., 1998), and the study by Reinus and colleagues had significant methodological flaws that should not have allowed calculation of diagnostic accuracy (as mentioned previously 2.2.4, p. 34-35). Reinus and colleagues only reported a history of cancer in patients with suspected cancer or fracture, and therefore the true negative and false negative data required to calculate likelihood ratios is unknown.

The authors of each of these reviews all suggested that a combination of test findings might prove to be more useful than individual test findings. However, they reported that there was currently insufficient evidence to specify which combinations may be useful. Also, due to the lack of good quality research and differences in study

settings, index tests, reference standard, and patient selection between studies, no systematic review performed any meta-analyses.

Interestingly, although the primary aim of the reviews was to consider the utility of red flag question for screening for malignancy, all reviews focused on highly specific tests with good positive likelihood ratios, rather than highly sensitive tests. It is generally considered important that tests used for screening should be sensitive so that the number of false negatives is minimised (Grimes & Schulz, 2002).

Rationale for undertaking the current review. Although three systematic reviews have been published in this area, all of the reviews have drawn conclusions focusing on red flag questions with high specificity, and did not discuss useful screening questions. Also, although these reviews recognised the poor methodological quality of some of the selected studies, they did not consider this during index test analysis. Therefore, the reviewers all drew conclusions based on studies that were potentially at high risk of bias. Hence, it was decided that an updated, independent review of the literature focusing on studies that were at lower risk of bias would be valuable.

2.3.2 Objectives

This review reflects the aims of this chapter (as stated in 2.1.4 Objectives). Specifically, this review aims to determine the prevalence and incidence of malignancy in the lumbar spine and the diagnostic accuracy of red flag questions to screen for malignancy in patients with low back pain.

2.3.3 Methods

Criteria for consideration of studies for this review

Types of studies. Primary diagnostic studies comparing patient demographics or subjective history findings to an appropriate reference standard were considered for inclusion in this review. Prospective or retrospective cohort or cross-sectional studies were considered for inclusion in this review if they published sufficient data to construct diagnostic 2 x 2 tables to assess the diagnostic accuracy of demographic data or patient history findings.

Participants. Studies were considered eligible if they included adult human participants with the primary complaint of low back pain. Primary, secondary and tertiary care settings were all eligible.

Index tests. Information that can be gathered during patient history taking was considered an index test. Demographic data such as age and gender were also considered.

Target condition prevalence/incidence. Studies investigating the prevalence or incidence of spinal malignancy in patients presenting with low back pain were considered for inclusion in this review.

Reference standards. Plain radiography, CT, MRI, bone scintigraphy and long term follow-up were considered appropriate reference standards, with the recognition that plain radiography may under-report malignancy, as 50% bone loss is required to detect a lytic lesion on X-ray (Sciubba & Gokaslan, 2006).

Search methods for identification of studies

Electronic searches. A literature search was completed on the 20th of December 2013 using the electronic databases Scopus, EMBASE, MEDLINE, CINAHL, and SPORTDiscus (via EBSCO). Full text publications from "all years" to 2013 were considered for inclusion. The reference lists of all included studies and relevant reviews were all searched to ensure that no eligible publications were missed.

Data collection and analysis

Selection of studies. A single author completed the title and abstract screening of potentially eligible studies. Final selection was based on review of the full text of identified publications.

Data extraction and management. Data related to study design and characteristics, participants, prevalence, reference standard and index tests was extracted by a single author. Where possible, index test data included true positive, false positive, true negative and false negative numbers for use to construct diagnostic 2 x 2 tables to assess diagnostic accuracy.

Assessment of methodological quality. Potential sources of bias and applicability concerns were assessed using the QUADAS-2 tool as previously described in 2.2.3: Methods.

Statistical analysis and data synthesis. Study prevalence was calculated from data extrapolated from original studies. Diagnostic 2 x 2 tables were constructed for each index test to calculate sensitivity, specificity, and likelihood ratios with 95% confidence intervals. Heterogeneity between studies meant that prevalence and diagnostic accuracy findings could not be pooled.

Investigations of heterogeneity. Due to the limited number of studies investigating similar index tests, factors influencing heterogeneity could not be investigated.

2.3.4 Results

Results of the search. The key words specific to malignancy included: cancer, tumor or tumour, carcinoma, neoplasm, sarcoma, metastases or malignancy with truncations. See Appendix C.1 for detail of key words and flow chart of search results. The combined search for the prevalence of malignancy amongst patients with low back pain resulted in 996 titles. A combined search for the diagnostic accuracy of red flag questions for malignancy amongst patients presenting with low back pain resulted in 3,091 titles. Following removal of duplicates and titles that clearly did not fit the inclusion criteria, 503 abstracts were screened. A further 425 studies were removed as they did not meet inclusion criteria. Reference lists of relevant reviews were searched and an additional two studies were added for full text screening. A total of 80 full texts were screened. 12 studies were identified that met all inclusion criteria and were selected for qualitative synthesis. The majority of full texts were excluded as they did not investigate a low back pain population. Single case studies and case-control design studies were also excluded due to poor methodology.

Methodological quality of included studies. The results of the assessment of methodological quality are displayed in Table 2.5. Figure 2.4 illustrates the overall risk of bias in each of the domains: patient selection, index tests, reference standard, and flow and timing. Figure 2.5 illustrates the applicability of each study in the domains: patient selection, index test and reference standard. Overall, one study (Cook, Ross, Isaacs, & Hegedus, 2012) showed low risk of bias and low concern for applicability in

all domains. All other studies had unclear or high risk in at least two domains. With regard to the index tests, two thirds of studies showed low concern for applicability. However, 50% of studies (Fernbach, Langer, & Gross, 1976; Frazier, Carey, Lyles, Khayrallah, & McGaghie, 1989; Jacobson, 1997; Khoo et al., 2003; Reinus et al., 1998; Slipman et al., 2003) showed high or unclear risk of bias. This was commonly due to lack of blinding whilst interpreting index tests or poor reporting of the process. The reference standard was applicable in most cases; whilst one study (Slipman et al., 2003) had a high concern for applicability, as they screened clinical notes to identify cases of malignancy and may have missed several cases. The two domains of concern for risk of bias were reference standard and flow and timing. Overall, the most common issue with the reference standard was that it was either interpreted without blinding of index test results, or that this was not clearly reported in the study methodology. Risk of bias was usually introduced to patient flow when studies did not preform the same reference standard on all participants (Deyo & Diehl, 1986; Frazier et al., 1989; Reinus et al., 1998; Roman et al., 2010; Slipman et al., 2003) or when there was an inappropriate delay between the index test and reference standard (Henschke et al., 2009).

Although 11 of the 12 studies were found to have potential bias, three studies (Frazier et al., 1989; Reinus et al., 1998; Slipman et al., 2003) were found to have concerning risk of bias or concerns regarding applicability in over 50% of domains, and have therefore been removed from further statistical analysis of index tests.

Table 2.5 *Assessment of study quality for malignancy using the QUADAS-2*

Author	Risk of bias				Applicability concerns		
	Patient selection	Index tests	Reference Standard	Flow & Timing	Patient selection	Index tests	Reference Standard
Cook (2012)	L	L	L	L	L	L	L
Deyo (1986)	L	L	H	H	L	L	L
Deyo (1988)	L	L	?	?	L	L	L
Donner-Banzhoff (2006)	L	L	?	H	L	L	?
Fernbach (1976)	L	H	?	H	L	L	L
Frazier (1989)	L	?	H	H	L	L	?
Henschke (2009)	L	L	H	H	L	L	L
Jacobson (1997)	L	?	?	L	L	L	L
Khoo (2003)	L	H	L	L	L	H	L
Reinus (1998)	L	H	H	H	?	H	L
Slipman (2003)	L	H	L	H	L	H	H
Van den Bosch (2004)	L	L	?	?	L	?	L

Note. L = Low risk, ? = Unclear, H = High risk

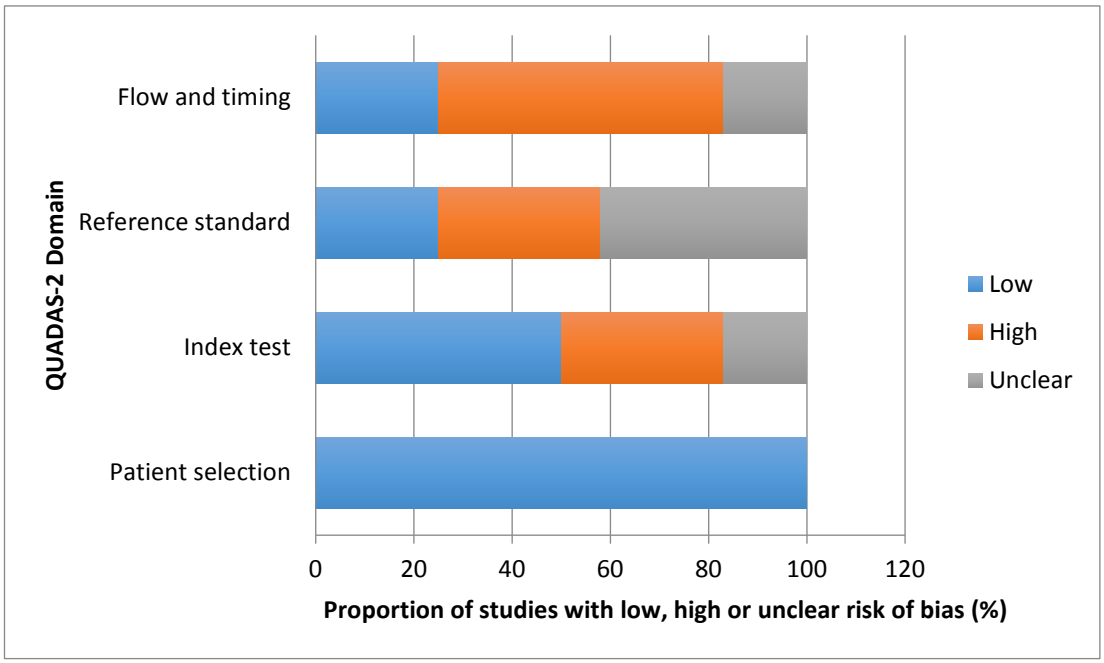


Figure 2.4 Combined QUADAS-2 results to illustrate overall risk of bias

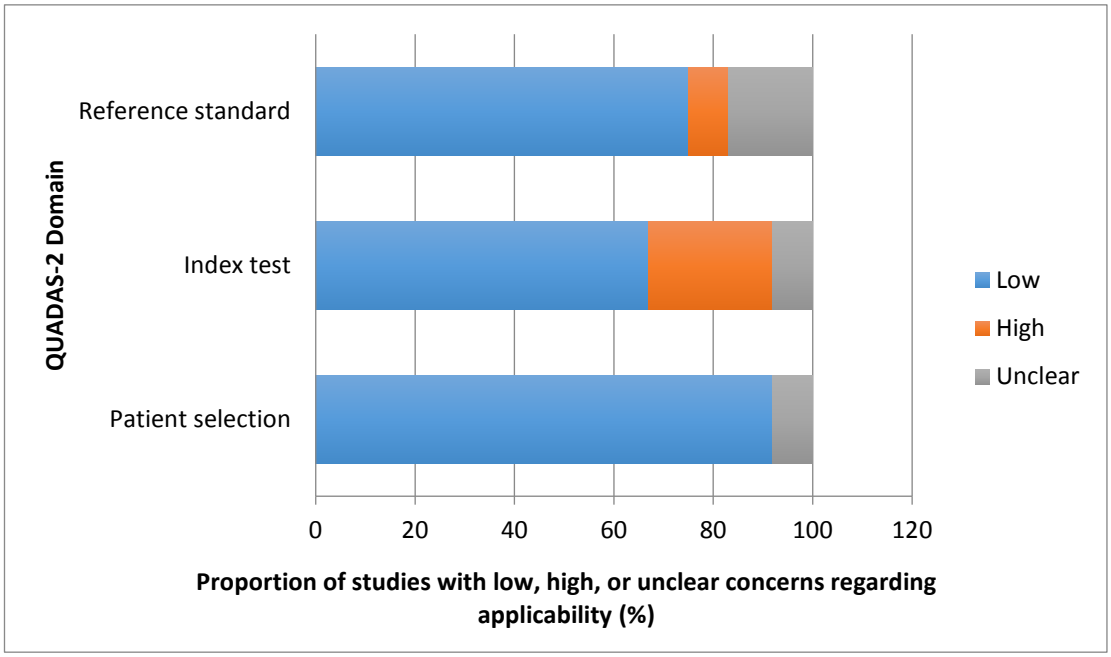


Figure 2.5 Combined QUADAS-2 results to illustrate overall applicability

Description of studies. Study characteristics of the included 12 studies are detailed in Table 2.6. Seven studies (Deyo & Diehl, 1986, 1988; Donner-Banzhoff et al., 2006; Frazier et al., 1989; Henschke et al., 2009; Khoo et al., 2003; van den Bosch et al., 2004) were performed in primary care settings. One study (Jacobson, 1997) was performed in secondary care and three studies (Cook et al., 2012; Fernbach et al., 1976;

Reinus et al., 1998) were performed in tertiary care settings. One study (Slipman et al., 2003) was performed in both secondary and tertiary settings. The total number of participants across all 12 studies was 30,144, comprised of 8,466 primary care participants with 28 cases of malignancy, 14,797 secondary care participants with 36 cases of malignancy, and 6,881 tertiary care participants with 124 cases of malignancy.

Fifty percent of the included studies (Deyo & Diehl, 1986, 1988; Donner-Banzhoff et al., 2006; Henschke et al., 2009; Khoo et al., 2003; Reinus et al., 1998) were conducted prospectively, with consecutive enrolment of participants in most studies. Five studies (Fernbach et al., 1976; Frazier et al., 1989; Jacobson, 1997; Slipman et al., 2003; van den Bosch et al., 2004) were conducted as retrospective reviews and one study (Cook et al., 2012) was a retrospective cohort design.

The most commonly utilised reference standard was diagnostic imaging, where three studies (Khoo et al., 2003; Reinus et al., 1998; van den Bosch et al., 2004) used plain radiographs, one study (Jacobson, 1997) used bone scintigraphy, and one study (Cook et al., 2012) used MRI. Laboratory tests were used by one study (Fernbach et al., 1976) and long term follow-up by three studies (Donner-Banzhoff et al., 2006; Frazier et al., 1989; Henschke et al., 2009). The two studies by Deyo and Diehl (1986, 1988) searched the institutional tumour registry to identify cases of cancer, and one study (Slipman et al., 2003) searched the medical notes to identify tumour cases.

Study findings. Across the 12 identified studies, the prevalence of malignancy ranged from 0 to 0.7% in primary care, 0.12 to 7% in secondary care, and 0.69 to 5.9% in tertiary care (see Table 2.6 for detail). Three studies (Frazier et al., 1989; Reinus et al.,1998; Slipman et al., 2003) were excluded from further statistical analysis due to a high risk of bias (see Table 2.5). One of these studies (Frazier et al., 1989) was conducted in a primary care setting (reporting a prevalence of 0.21%), one study (Slipman et al., 2003) was conducted in both secondary and tertiary care settings (reporting prevalence of 0.12% and 0.69% respectively), and the final study (Reinus et al., 1998) was based in a tertiary care setting (reporting a prevalence of 1.45%). The low prevalence in both secondary and tertiary care reported by Slipman and colleagues may have been due to under-reporting, resulting from their decision to identify cases via a retrospective review of medical records. This choice of reference standard is likely to lead to missed cases (Henschke et al., 2013). No study investigated incidence.

Table 2.6 *Study characteristics for malignancy*

Authors	Study Design	Setting	Participants	Prevalence of Malignancy	Reference Standard	Withdrawals
Cook (2012)	Retrospective cohort study	Spine clinic, tertiary care	1109 with LBP presenting to spine clinic	5.9% (n=66)	MRI	42 of the original 1161 were excluded due to incomplete data
Deyo (1986)	Prospective, consecutive	Walk-in clinic, primary care	621 patients with a primary complaint of back pain (311 received reference standard)	0.64% (n=4)	Hospital tumour registry and discharge records	Unclear
Deyo (1988)	Prospective, consecutive	Walk-in clinic, primary care	1975 patients with a primary complaint of back pain	0.66% (n=13)	Tumour registry	Not reported
Donner-Banzhoff (2006)	Prospective, consecutive	GP clinics, primary care	1353 patients with low back pain (1190 available at follow up)	0.07% (n=1)	12 month follow up	163 lost to follow up
Fernbach (1976)	Respective chart review	Orthopaedic spine clinic at university teaching hospital, tertiary care	259 patients with LBP over 50 years of age were compared to 259 patients with LBP under 50 years (total 518).	3.46% (n=18)	ESR and serum concentrations of alkaline phosphatase and calcium	No withdrawals
Frazier (1989)	Retrospective chart review	Medical walk-in clinics, primary care	471 patients with acute lumbosacral back pain	0.21% (n=1)	Clinical notes from visits up to 6 months after the initial assessment, 99 received roentgenograms	392 did not meet inclusion criteria, 174 excluded without explanation
Henschke (2009)	Prospective, consecutive	Primary care	1172 patients with acute LBP	0% (n=0)	12 month follow up	12 cases healthcare provider flowed up rather than participant

Authors	Study Design	Setting	Participants	Prevalence of Malignancy	Reference Standard	Withdrawals
Jacobson (1997)	Retrospective review	Secondary care	491 patients, 257 complaining of back pain who were referred for bone scans (without prior history of malignancy)	7% (n=18)	Bone scan	No withdrawals
Khoo (2003)	Prospective, consecutive	Primary care radiology clinic	1030 patients referred from general practice for a lumbar spine radiographs	0.1% (n=1)	X-ray	No withdrawals
Reinus (1998)	Prospective, consecutive	Tertiary care, Accident & Emergency department	482 patients referred for lumbar spine X-ray	1.45% (n=7)	X-ray	No withdrawals, chart review conducted in only 196 participants
Slipman (2003)	Retrospective chart review	3 multidisciplinary spine centres – 1 academic (tertiary care) and 2 private (secondary care)	19,312 patients referred to spine clinics (4,772 tertiary, 14,540 secondary)	0.26% (n=33 tertiary (0.69%), n=18 secondary (0.12%))	Review of the medical notes, MRI to confirm some cases	No withdrawals, detailed chart review in 51 of 19,312 participants
Van den Bosch (2004)	Retrospective review	Primary care radiology department	2,007 patients referred with LBP for lumbar radiographs (sample chosen randomly from 6,269)	0.7% (n=8)	X-ray	93 participants excluded as no records available

Note. n = number, LBP = low back pain, GP = general practitioner, ESR = erythrocyte sedimentation rate, MRI = magnetic resonance imaging

Study findings continued. A total of 25 index tests were investigated across the 12 studies (see Table 2.7 for detail). The most commonly investigated index test was age, which was investigated by 6 studies, using 4 different cut-off points (>44, >54, ≥50, or >50 years). Eight index tests were investigated by a single study only. The study by Henschke and colleagues (2009) did not find any cases of malignancy and therefore no sensitivity or likelihood ratio statistics could be calculated.

Primary care. Seven studies were conducted in primary care settings. One study (van den Bosch et al., 2004) reported high sensitivity and a negative likelihood ratio (LR) that indicated a moderate shift in probability for age greater than 44 years (Sensitivity 93%, LR- 0.20). However, age greater than 50 years and age greater than 54 years were less informative with low to modest diagnostic accuracy (LR- 0.3-0.4). Findings that had specificity for malignancy were history of cancer (Specificity 98%, LR+15.27), unexplained weight loss (Specificity 94%, LR+ 2.57), and failed conservative management (Specificity 90%, LR+ 2.61-3.08) (Deyo & Diehl, 1986; 1988). A 'history of cancer' was the only index test with a positive likelihood ratio that indicated a conclusive shift in the probability of a malignancy being present when the test is positive. Deyo and Diehl (1988) also reported that 'tried bed rest no relief' had a specificity of 100%. However, this was only asked in four of the 13 participants with cancer, and therefore lacked precision (95% CI 0.40-1).

The study by Deyo and Diehl (1988) was the only study to investigate a combination of red flags. They selected the four red flag questions with the highest likelihood ratios, and considered these as risk factors (age >50 years, history of cancer, unexplained weight loss, or failure to improve with conservative therapy). Using the four red flags, they found that every patient with cancer had at least one positive finding. Therefore, if a patient had no positive findings they were considered 'low risk', and cancer could be excluded with a sensitivity of 100% without further investigations. The red flag question with the highest positive likelihood ratio was history of cancer, and 9% of patients with a history of cancer had spinal malignancy. Therefore, Deyo and Diehl (1988) recommended that any patient with a history of cancer is considered 'high risk' and these patients should undergo expeditious plain radiographs and ESR. Deyo and Diehl (1988) cautioned that the use of history of cancer and ESR alone misclassified 6 out of 13 cancer patients as non-cancer. However, no patient was misclassified using ESR and plain radiography combined. The 'intermediate risk' included patients over 50 years of age, with no history of cancer, and patients with

unexplained weight loss or signs of systemic illness. For this group the prevalence was 1.2%, therefore the authors suggested ESR alone could be used to raise or lower the suspicion of cancer. In this study no participant with cancer had a normal ESR (<20mm/hr).

Secondary care. The study by Jacobson (1997) was the only study conducted in solely in secondary care and investigated patients referred for a bone scintigraphy. Another study by Slipman et al. (2003) was conducted in secondary and tertiary care and used chart review as the reference standard to identify malignancy. Jacobson found a high prevalence of 7%, compared to Slipmans' 0.12% in private spine clinics. The only index test investigated by Jacobson was age greater than 50 years, which had good sensitivity and a negative likelihood ratio, which indicated a moderate shift in probability (Sensitivity 94%, LR- 0.11). However, this was traded for poor specificity (41%).

Tertiary care. Three studies were conducted in tertiary care, one study was set in an accident and emergency department (Reinus et al., 1998), and two studies took place in spine clinics of university hospitals (Cook et al., 2012; Slipman et al., 2003). Although the two studies by Cook et al. and Slipman et al. were conducted in similar settings, they found significantly different prevalence of 5.9% and 0.69% respectively. The choice of reference standard is likely to have contributed to this variance, as Cook and colleagues used MRI for the reference standard, which is considered as the best available reference standard, whereas Slipman used chart review. The studies by Reinus and Slipman were excluded from the index test analysis due to methodological concerns and inadequate reporting. The remaining study by Cook and colleagues investigated "not depressed or anxious" as an index test and reported good sensitivity (94%) but very poor specificity (2%) and the likelihood ratios (LR+0.96, LR-3.16) did not significantly shift the probability of a diagnosis.

Table 2.7 *Clinical signs and diagnostic accuracy data extracted from eligible studies for malignancy*

Index test	LR+ (95% CI)	LR- (95% CI)	Sensitivity (%)	Specificity (%)
Demographic				
Age >44 years (van den Bosch et al., 2004)	1.39 (1.21-1.60)	0.20 (0.03-1.35)	0.93 (0.66-1.00)	0.33 (0.31-0.35)
Age > 54 years (van den Bosch et al., 2004)	1.60 (1.24-2.07)	0.40 (0.14-1.10)	0.80 (0.51-0.95)	0.50 (0.48-0.52)
Age ≥ 50 years (Jacobson, 1997)	1.71 (1.53-1.90)	0.00 (0.00-NaN)	1.00 (0.78-1.00)	0.41 (0.35-0.48)
Age ≥ 50 years (Fernbach et al., 1976)	1.95 (1.68-2.25)	0.11 (0.02-0.73)	0.94 (0.71-1.00)	0.52 (0.47-0.56)
Age > 50 years (Deyo & Diehl, 1986)	2.50 (1.40-4.46)	0.36 (0.07-1.95)	0.75 (0.22-0.99)	0.70 (0.66-0.74)
Age > 50 years (Deyo & Diehl, 1988)	2.65 (1.95-3.60)	0.32 (0.12-0.88)	0.77 (0.46-0.94)	0.71 (0.69-0.73)
Age at onset <20 or >55 years (Henschke et al., 2009)	NaN	NaN	NaN	0.76 (0.73-0.78)
History				
Unexplained weight loss (>4.5kg in 6 months) (Henschke et al., 2009)	NaN	NaN	NaN	1.00 (0.99-1.00)
Unexplained weight loss (Deyo & Diehl, 1988)	2.57 (0.71-9.31)	0.90 (0.71-9.31)	0.15 (0.03-0.46)	0.94 (0.93-0.95)
Previous history of cancer (Henschke et al., 2009)	NaN	NaN	NaN	0.96 (0.95-0.97)
Previous history of cancer (Deyo & Diehl, 1988)	15.27(6.38-36.55)	0.70 (0.49-1.01)	0.31 (0.10-0.61)	0.98 (0.97-0.99)
Duration > 1 month (Deyo & Diehl, 1988)	2.62 (1.48-4.67)	2.63 (1.48-4.67)	0.50 (0.22-0.78)	0.81 (0.79-0.83)
Failed conservative care after 1 month (Deyo & Diehl, 1986)	2.61 (0.47-14.52)	0.83 (0.47-1.46)	0.25 (0.0-0.78)	0.90 (0.88-0.93)
Failed conservative care after 1 month (Deyo & Diehl, 1988)	3.08 (1.35-7.04)	0.77 (0.53-1.1)	0.31 (0.10-0.91)	0.90 (0.88-0.91)
Tried bed rest, but no relief (Henschke et al., 2009)	NaN	NaN	NaN	0.84 (0.81-0.86)

Index test	LR+ (95% CI)	LR- (95% CI)	Sensitivity (%)	Specificity (%)
History continued				
Tried bed rest, but no relief (Deyo & Diehl, 1988)	1.85 (1.75-1.96)	0.00 (0.00-NaN)	1.00 (0.40-1.00)	0.46 (0.43-0.49)
Recent back injury (Deyo & Diehl, 1988)	0.00 (0.00-NaN)	1.22 (1.22-1.22)	0.00 (0.00-0.28)	0.82 (0.80-0.84)
Thoracic pain (Deyo & Diehl, 1988)	1.04 (0.29-3.7)	0.99 (0.77-1.28)	0.17 (0.03-0.49)	0.84 (0.82-0.86)
Insidious onset (Henschke et al., 2009)	NaN	NaN	NaN	0.83 (0.80-0.85)
Insidious onset (Deyo & Diehl, 1988)	1.06 (0.69-1.63)	0.92 (0.46-1.82)	0.62 (0.32-0.85)	0.42 (0.40-0.44)
Systemically unwell (Henschke et al., 2009)	NaN	NaN	NaN	0.98 (0.97-0.98)
Constant, progressive, non-mechanical pain (Henschke et al., 2009)	NaN	NaN	NaN	0.97 (0.96-0.98)
Altered sensation from trunk down (Henschke et al., 2009)	NaN	NaN	NaN	0.98 (0.97-0.99)
Not depressed or anxious (Cook et al., 2012)	0.96 (0.90-1.02)	3.16 (1.11-8.99)	0.94 (0.84-0.98)	0.02 (0.01-0.03)
Is the low back pain familiar? (Donner-Banzhoff et al., 2006)	0.00 (0.00-NaN)	1.21 (1.21-1.21)	0.00 (0.00-0.95)	0.83 (0.81-0.85)
Combined results				
Age < 50 years, no history of cancer, no weight loss or other sign of systemic illness, no history of failed conservative management (Deyo & Diehl, 1988)	2.48 (2.35-2.61)	0.00 (0.00-NaN)	1.00 (0.72-1.00)	0.60 (0.57-0.62)

Note. NaN = unable to be calculated as values entered include one or more zeros, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, CI = confidence interval

2.3.5 Discussion

The primary aim of this review was to provide information on the prevalence or incidence of spinal malignancy, and the diagnostic accuracy of red flags to screen for malignancy in patients presenting with low back pain. A total of 12 original publications were included in this review. Following exclusion of three studies with methodological concerns (Frazier et al., 1989; Reinus et al., 1998; Slipman et al., 2003).

Factors affecting interpretation

Prevalence and setting. Study setting appeared to influence prevalence. Studies conducted in primary care settings found low prevalence of malignancy (0-0.7%), whereas studies based in secondary and tertiary care reported relatively higher prevalence of 3.5-7% when diagnostic imaging was used as the reference standard. Variation in prevalence between clinical settings is important for clinicians to recognise. The likelihood of assessing a patient with underlying malignancy may be more than 10 times higher if a clinician is working in an orthopaedic spine clinic, compared to a general practice or private physiotherapy clinic (Donner-Banzhoff et al., 2006; Henschke et al., 2009; Jacobson, 1997). Awareness of increased prevalence of malignancy can allow clinicians to calculate pre-test probability and assess the likelihood of a patient presenting with malignancy within the clinical setting they are working in. Hence, it is a useful indicator that more thorough investigation may be required within secondary and tertiary settings.

Index tests. When used in isolation, most red flags were uninformative. The most commonly investigated red flag question (index test) was older age. Negative likelihood ratios varied from 0 – 0.4, indicating a slight to conclusive reduction in the probability of the disease being present with a negative test result. However, specificity for older age was poor and there were high false positive rates. The only red flag question considered to be independently informative by previous reviews (Downie et al., 2013; Henschke et al., 2013) was a history of cancer. This conclusion has been based on the high positive likelihood ratios of 15.27 and 31.67 in two studies (Deyo & Diehl, 1988; Reinus et al., 1998). However, these likelihood ratios lacked precision, with wide confidence intervals, and the study by Reinus et al. was at high risk of bias. Deyo and Diehl's (1988) study was conducted on an indigenous population with the majority of participants of low socioeconomic groups, poor education and mostly Mexican-American decent, which also limits the generalisability of the study. Their

study was at risk of bias, as not all participants received the same reference standard or completed all of the index tests. Deyo and Diehl's (1988) study was also completed before MRI was widely available, and the use of a highly sensitive reference standard would improve the precision of their results. Hence, there is minimal evidence to suggest that a history of cancer is an independently informative red flag question.

Deyo and Diehl (1988) investigated a combination of red flags and suggested a classification system to assist with patient management. They used red flag questions to categorise patients into high, intermediate, or low risk of malignancy, to guide clinicians through the decision making process. From their results this appears to be a potentially useful system. However, clinical application of this categorisation system would require validation and assessment of reliability in different settings and populations.

Reference standard. Most low back pain guidelines recommended that advanced imaging should be reserved for patients who have clinical suspicion of serious pathology, or who may be surgical candidates (Koes et al., 2010). Therefore, patients undergoing advanced diagnostic imaging may be more likely to have serious pathologies. Studies using advanced imaging as their reference standard may therefore be affected by patient selection bias.

Prevalence did appear to be influenced by the reference standard. This was evidenced by the similarities in prevalence between studies that utilised similar reference standards. The two studies (Donner-Banzhoff et al., 2006; Henschke et al., 2009) used long-term follow-up as their reference standard and both found low prevalence (0-0.07%). Another two studies (Frazier et al., 1989; Slipman et al., 2003) based in different clinical settings used chart review as their reference standard and found similar prevalence of 0.21 and 0.26%. Two studies (Deyo & Diehl, 1986, 1988) used the tumour registry as the reference standard and again found similar prevalence (0.64 and 0.66%), and the two studies with more sensitive reference tests (MRI and bone scintigraphy) found significantly higher prevalences (5.9 and 7%). These results would suggest that reference standards with lower diagnostic accuracy may have misclassified some cases of malignancy and may have introduced spectrum bias. Sensitive tests such as MRI will detect early stage malignancy. However, plain radiography, chart review, and clinical follow-up are all unlikely to detect early changes (Schmidt & Factor, 2013; Sciubba & Gokaslan, 2006).

Conclusion. Although spinal malignancy is rare in primary care, clinicians working in secondary or tertiary care may be more likely to assess patients with malignancy and need to ensure they adequately screen all patients. The incidence of malignancy was not reported by any study and is therefore unknown. If there is suspicion of malignancy, advanced diagnostic imaging such as MRI is recommended.

Red flag questions used in isolation are uninformative, and although combinations of questions may be useful, further research is required to establish their diagnostic accuracy and utility. The most useful findings were that if a patient is under 45 years of age they are less likely to have spinal malignancy, and there is limited evidence that the greatest risk factor for spinal malignancy is a history of cancer.

Weaknesses of this review. All study selection, data extraction, and analysis was performed by a single author, which may have introduced study selection and review bias. Also, due to the low number of studies, heterogeneity between studies, and a lack of consistency between reported index tests, no meta-analysis could be performed.

Strengths of this review. This review followed the Cochrane Collaboration and PRISMA guidelines. This study included four additional references (Cook et al., 2012; Fernbach et al., 1976; Slipman et al., 2003; van den Bosch et al., 2004) that were not included in other published systematic reviews (Downie et al., 2013; Henschke et al., 2013; Henschke et al., 2007). The addition of these references allowed for a wider overview of the prevalence of malignancy in different settings, as well as supplementary index test information.

Applicability of findings to this thesis. This review has illustrated that further research is required to investigate the diagnostic accuracy and utility of red flag questions for screening or identification of spinal malignancy.

This review revealed that little is known about the prevalence of spinal malignancy in patients presenting with low back pain, and incidence of malignancy in a low back pain population is unknown. Hence, more research is required to establish the prevalence and incidence, particularly in secondary and tertiary care.

This thesis therefore investigated the prevalence of spinal malignancy in secondary and tertiary care, and the incidence in tertiary care. MRI was utilised as the reference standard to ensure accurate diagnosis of malignancy. The red flag questions

that required further investigation based on previous studies and clinical guidelines were: older age, insidious onset of pain, worsening pain, constant pain, night pain, unexplained weight loss, history of cancer, systemically unwell, and no relief with bed rest. Therefore, these questions were all included in the diagnostic accuracy study in this thesis.

2.4 Target Condition: Cauda Equina Syndrome

2.4.1 Background

Cauda equina syndrome is the term used to classify severe neurological injury that can result from compression of the cauda equina (Dinning & Schaeffer, 1993). It was first described by Dandy in 1929 as loose cartilage from an intervertebral disc that simulated a tumour of the spinal cord (Dandy, 1989). Cauda equina syndrome was first recognised as requiring emergency surgical decompression by Mixter (1934). The most common cause of cauda equina syndrome is disc prolapse, followed by tumour, infection, stenosis and haematoma (Fraser, Roberts, & Murphy, 2009; Kostuik, Harrington, & Alexander, 1986).

Cauda equina syndrome can be classified as partial or complete, depending on the extent of compression of the descending sacral nerves. Distinction between partial or complete compression is very important for the prognosis. Studies have shown that patients with unilateral loss of sacral nerves following surgical resection can still have near-normal pelvic autonomic function (Stener & Gunterberg, 1978). Complete lesions have a poor prognosis and usually occur secondary to a massive central or paracentral disc prolapse, most commonly at the L4/5 or L5/S1 level (Gleave & Macfarlane, 2002). A massive disc prolapse can cause compression of the bilateral descending nerves, which results in loss of parasympathetic supply to the pelvic viscera, and loss of sensory supply to the perineum (Gleave & Macfarlane, 2002). The small unmyelinated and myelinated nerve fibres that supply parasympathetic function and pain sensation are significantly less resilient to mechanical compression than larger fibres that supply motor power, light touch and proprioception. Reversibility of altered parasympathetic function is dependent on timely relief of mechanical pressure, ischaemia and venous congestion (Gleave & Macfarlane, 2002). Hence, cauda equina syndrome is considered a surgical emergency and acute surgical decompression is required (Accident Compensation Corporation, 1999; Shapiro, 2000).

Ahn and colleagues (2000) performed a meta-analysis of surgical outcomes and found significant advantages of preforming surgery within 48 hours of the onset of symptoms. Delayed diagnosis can result in ongoing neurological compromise, including paraplegia, neurogenic bladder and/or bowel abnormalities, altered sexual function (including erectile dysfunction) and perineal or saddle anaesthesia (Ahn et al., 2000; Nater & Fehlings, 2015; Shapiro, 2000). These outcomes can be dire for both patients and healthcare funders. A study by Todd (2011) investigated cauda equina syndrome in medico-legal practice and found that only 11% of people with complete cauda equina compression were able to return to normal work, and 11% returned to modified work.

Prevalence/incidence. Cauda equina syndrome has been reported in the literature to account for 1-10% of all surgical discectomies (Choudhury & Taylor, 1980; Jennett, 1956; Kostuik, Harrington, Alexander, Rand, & Evans, 1986; Robinson, 1965; Shephard, 1959). Most studies have been based on surgical populations, and the prevalence and incidence of cauda equina syndrome in people with low back pain is unknown.

Index tests. Cauda equina syndrome is variable in presentation, and consequently a review of 105 cauda equina syndrome articles found 14 different descriptions of bladder involvement, 10 descriptions of bowel involvement, six of pain and five of sexual dysfunction (Fraser et al., 2009). Common clinical features of cauda equina syndrome include severe low back pain, unilateral or bilateral radicular pain, perianal or saddle anaesthesia, lower limb motor weakness, sensory deficits, urinary retention or overflow, and bowel incontinence (Aho, Auranen, & Pesonen, 1969; Mahadevappa, Persi, & Nesathurai, 2015; Malloch, 1965; Shapiro, 1993, 2000). Back pain is the most common complaint and is reported in 94-100% of cases (Gooding, Higgins, & Calthorpe, 2013; Jalloh & Minhas, 2007) followed by sciatica which is reported in 83-100% of cases (Buchner & Schiltenwolf, 2002; Korse, Jacobs, Elzevier, & Vleggeert-Lankamp, 2013; O'Laoire, Crockard, & Thomas, 1981).

A recent systematic review by Korse and colleagues (2013) investigated complaints of altered micturition, defecation and sexual function in cauda equina syndrome due to lumbar disc herniation. Their review included 15 original studies published from 1956-2011 and included 464 patients. All included studies were retrospective, had small sample sizes (14-54) and were all based in secondary or tertiary surgical settings. They found that cauda equina syndrome was most common in middle-

aged patients with an overall mean age of 43.5 years between studies. There was no clinically significant difference between genders. On average 89% of patients complained of dysfunctional micturition at presentation. Dysfunction of defecation was recorded in 8 of the 15 studies, and on average 47% of patients complained of changes in defecation on presentation.

Another recent study by Gooding and colleagues (2013) investigated the diagnostic accuracy of clinical features in a population on 57 cases of suspected cauda equina syndrome who were undergoing MRI. They investigated subjective and objective findings, including: back pain, sciatica, lower limb sensory change, lower limb weakness, abnormal lower limb reflexes, urinary symptoms, urinary retention with post void volume >500mls on bladder scan, bowel incontinence or constipation, and digital rectal examination. Cauda equina syndrome was confirmed on MRI in only 23% of clinically suspected cases. Back pain had a sensitivity of 100% but was non-specific, and all other clinical findings had poor sensitivity, ranging from 8% for urinary frequency and bowel incontinence to 54% for bilateral sciatica. The authors concluded that due to the poor diagnostic accuracy of clinical findings, no discreet clinical protocol could be utilised to rule out or raise the suspicion of cauda equina syndrome. Other studies (Bell, Collie, & Statham, 2007; Domen, Hofman, Van Santbrink, & Weber, 2009) conducted on populations of patients with suspected cauda equina syndrome undergoing MRI have also reinforced that signs and symptoms are variable and of little diagnostic value.

In New Zealand the Accident Compensation Corporation (ACC) Acute Low Back Pain Guidelines recommend the use of red flag questions to identify potential cases of cauda equina syndrome (ACC, 1999). ACC states that if some or all of the following red flags are present the patient should be referred to hospital urgently: urinary retention, faecal incontinence, widespread neurological signs and symptoms in the lower limbs, gait abnormality, saddle area numbness or a lax anal sphincter (ACC, 1999). However, there is no evidence to support the use of these red flag questions for cauda equina syndrome and their diagnostic value is unknown.

The variable presentation of cauda equina syndrome makes clinical diagnosis difficult. Bell and colleagues (2007) investigated the correlation between clinical assessment and MRI findings. Their research was conducted in a tertiary care neurosurgical centre in London, and they included a total of 23 patients with suspected

cauda equina syndrome over a 4-month period. Within their sample only five were found to have cauda equina compression on MRI. Bell and colleagues (2003) investigated the ability of middle grade medical staff to clinically diagnose cauda equina syndrome and found that the diagnostic accuracy was 0.56, which is similar to chance. Some explanations of why cauda equina syndrome is so difficult to diagnose clinically have been proposed. Clinical guidelines usually recommend that if a patient presents with urinary retention or incontinence it may be indicative of cauda equina syndrome. However, difficulty passing urine is also associated with pain and may occur due to increased sympathetic tone (Bell et al., 2007). This means that while urinary changes cannot be ignored, they often falsely indicate neurogenic bladder symptoms.

Sun and colleagues (2014) performed a meta-analysis investigating the progression pattern of cauda equina syndrome in an attempt to improve early recognition and diagnosis. They suggested that, due to the expense and legal implications of a missed or late diagnosis, cauda equina syndrome should be diagnosed early with the onset of progression of lower limb sensory-motor deficits. They advised that waiting for the onset of sphincter dysfunction may increase the likelihood of irreversible changes. Sun et al. (2014) found that cauda equina syndrome commonly progresses through three stages: early, incomplete, and cauda equina syndrome in retention. The early stage is characterised by progression from unilateral to bilateral lower limb symptoms, the incomplete stage is characterised by reduction in sphincter function, and the final stage by complete sphincter dysfunction. They reported that 99.4% of patients had experienced early symptoms without being diagnosed. However, this study by Sun et al. did not consider methodological quality or standards of reporting of the primary studies. They also did not include any subjects without cauda equina syndrome and therefore could not investigate false positive rates or establish likelihood ratios.

Unfortunately, none of the original studies or reviews mentioned above included patients without confirmed or suspected cauda equina syndrome, therefore diagnostic accuracy could not be extrapolated for use in a low back pain population. Whilst an understanding of common symptoms patients with cauda equina syndrome may present with is useful to raise the suspicion of a possible diagnosis, this provides little diagnostic value.

Reference standard for the identification of cauda equina syndrome. The best available reference standard for diagnosis of cauda equina syndrome is MRI (Coscia, Leipzig, & Cooper, 1994; Domen et al., 2009). MRI is the diagnostic imaging modality of choice, as it is currently the only non-invasive modality that can clearly visualise the nerve roots and accurately assess compression (Coscia et al., 1994).

Rationale. Although several other authors have published reviews, systematic reviews, and meta-analyses (Buchner & Schiltenwolf, 2002; Fraser et al., 2009; Korse et al., 2013; Sun et al., 2014) on common symptoms and post-operative outcomes associated with cauda equina syndrome, no systematic review investigating patients presented with low back pain has been published. At present the prevalence and incidence of a cauda equina syndrome in a low back pain population is largely unknown.

2.4.2 Objectives

This review addresses the aims of this chapter, as stated in 2.1.4 Objectives. Specifically, this review aims to determine the prevalence or incidence of cauda equina syndrome and the diagnostic accuracy of index tests to screen for cauda equina syndrome in patients presenting with low back pain.

2.4.3 Methods

Criteria for consideration of studies for this review

Types of studies. This review considered all primary studies that compared index tests, such as patient demographics or history findings, to an appropriate reference standard in order to identify cauda equina syndrome. Prospective or retrospective cohort or cross-sectional studies were considered for inclusion if they presented adequate data to calculate the diagnostic accuracy of index tests. Full text articles published in English were considered for eligibility.

Participants. Only studies investigating adult patients presenting with a primary complaint of low back pain were considered for inclusion in this review. Studies based in primary, secondary or tertiary care settings were all considered for inclusion.

Index tests. An index test was considered as any finding from the patient history or patient demographics that could raise the suspicion of cauda equina syndrome.

Target condition prevalence/incidence. Studies investigating the prevalence or incidence of cauda equina syndrome within a population of patients with low back pain were considered for eligibility.

Reference standards. MRI or surgery were considered appropriate reference standards for the diagnosis of cauda equina syndrome.

Search methods for identification of studies

Electronic searches. On the 20th of December 2013 an electronic literature search of the following databases was completed: Scopus, EMBASE, MEDLINE, CINAHL, and SPORTDiscus (via EBSCO). Primary publications from "all years" to 2013 were considered for eligibility. Reference lists of all relevant reviews and primary studies were searched to ensure no eligible studies were missed.

Data collection and analysis

Selection of studies. Title and abstract screening and final selection of studies was completed by a single author. Final study selection was completed following review of all full text publications that were potentially eligible.

Data extraction and management. Data concerning study characteristics, design, participants, reference standard, prevalence and index tests was extracted by a single author. True positive, false positive, true negative and false negative data related to index tests was extracted and used to construct diagnostic 2 x 2 tables for diagnostic accuracy calculations.

Assessment of methodological quality. Sources of potential bias and any concerns regarding applicability were assessed using the QUADAS-2 tool (see Methods 2.3.3).

Statistical analysis and data synthesis. Study prevalence was extrapolated from original studies. Sensitivity, specificity and likelihood ratios with 95% confidence intervals were calculated using diagnostic 2 x 2 tables.

Investigations of heterogeneity. No investigations into heterogeneity could be performed, as only one study met the inclusion criteria.

2.4.4 Results

Results of the search. The key words specific to cauda equina syndrome were "cauda equina" and "spinal cord compression." The search terms and search strategy flow chart are detailed in Appendix C.2. The electronic search for the prevalence of cauda equina syndrome amongst patients with low back pain resulted in 352 titles. The electronic search for the diagnostic accuracy of red flags for cauda equina syndrome in patients with low back pain resulted in 1,050 titles. A total of 1,402 titles were screened, and 1,239 were excluded as they were duplicates, irrelevant, or conducted on animals. 163 abstracts were screened and a further 134 were excluded as they were reviews, editorials, letters, or did not meet inclusion criteria. 30 full texts were screened and 28 were excluded, due to study design (case study, case-series, or case-control) or population. The majority of studies were excluded as they were based on populations of participants with suspected cauda equina, known cauda equina, or surgical populations, and did not involve patients presenting with a primary complaint of low back pain which our study focused on. One study met the inclusion criteria for this review.

Methodological quality of the included study. Results of the methodological quality assessment are shown in Table 2.8. As only one study was included in this review, no graphs to illustrate overall risk of bias and applicability could be constructed. The study by Henschke et al. (2009) had low concerns for applicability in all domains. However, it did show high risk of bias for flow and timing as not all patients received MRI, which was the reference standard used to diagnose cauda equina syndrome. There may have also been an inappropriate delay between the index test and reference standard.

Table 2.8 *Assessment of study quality for cauda equina syndrome using the QUADAS-2*

Author	Risk of bias				Applicability concerns		
	Patient selection	Index tests	Reference Standard	Flow & Timing	Patient selection	Index tests	Reference Standard
Henschke (2009)	L	L	L	H	L	L	L

Note. L = Low risk, ? = Unclear, H = High risk

Table 2.9 *Study characteristics for cauda equina syndrome (CES)*

Authors	Study Design	Setting	Patients	Prevalence	Reference Standard	Withdrawals
Henschke (2009)	Prospective cohort (Consecutive enrolment)	Primary care General Practitioner, Physiotherapy and Chiropractor clinics	1172 patients receiving primary care for acute LBP	0.1 % (n=1)	12 month follow up (confirmation of CES on MRI)	12 cases healthcare provider followed up rather than participant

Note. CES = cauda equina syndrome, MRI = magnetic resonance imaging

Table 2.10 *Clinical signs and diagnostic accuracy data extracted from eligible studies for CES*

Index tests	LR+ (95% CI)	LR- (95% CI)	Sensitivity (%)	Specificity (%)
Acute onset of urinary retention or overflow incontinence (Henschke et al., 2009)	0.00 (0.00, NaN)	1.00 (1.00, 1.00)	0.00 (0.00, 0.95)	1.00 (0.99, 1.00)
Loss of anal sphincter tone or faecal incontinence (Henschke et al., 2009)	0.00 (0.00, NaN)	1.00 (1.00, 1.00)	0.00 (0.00, 0.95)	1.00 (0.99, 1.00)
Saddle anaesthesia about the anus, perineum or genitals (Henschke et al., 2009)	0.00 (0.00, NaN)	1.00 (1.00, 1.00)	0.00 (0.00, 0.95)	1.00 (0.99, 1.00)
Widespread (greater than 1 nerve root) or progressive motor weakness in the legs or gait disturbances (Henschke et al., 2009)	0.00 (0.00, NaN)	1.00 (1.00, 1.00)	0.00 (0.00, 0.95)	1.00 (0.99, 1.00)

Note. NaN = unable to be calculated as values entered include one or more zero

Description of study. One study by Henschke and colleagues (2009) was included for review. Their study was a prospective cohort study based in primary care physiotherapy, chiropractic and general practice clinics. Consecutive patients presenting with a primary complaint of low back pain were considered for recruitment. The total number of participants was 1,172 and they found one case of cauda equina syndrome. Their reference standard was long-term follow-up. However, diagnosis was confirmed on MRI.

Study findings. The prevalence of cauda equina syndrome in primary care was 0.1%. Henschke and colleagues (2009) investigated the diagnostic accuracy of four index tests in relation to cauda equina syndrome. Unfortunately, their prevalence was very low, and the single patient with cauda equina syndrome did not complain of any of the typical symptoms that were considered as index tests. Therefore, the analysis was very limited and all tests were calculated as 0% sensitivity and 100% specificity, as reported in Table 2.10. No specific symptoms or demographic data from the single case of cauda equina syndrome were published, so no further tests could be investigated.

2.4.5 Discussion

The primary aim of this review was to establish the prevalence and incidence of cauda equina syndrome and the diagnostic accuracy of red flags to screen for cauda equina syndrome in patients presenting with low back pain. Unfortunately, only one study met the inclusion criteria, and reported only a single case of cauda equine syndrome. The lack of research in this area may be partially due to the fact that cauda equina syndrome is so variable in presentation and can only be diagnosed on MRI or surgically.

Factors affecting interpretation

Prevalence and setting. There is a lack of research investigating the prevalence amongst a population of patients presenting with low back pain. One study, by Henschke et al. (2009), reported a low prevalence of 0.1% in primary care settings. The majority of published studies were conducted in tertiary settings on surgical populations and were therefore excluded from this review. More research is required within different populations to establish generalisable results.

In the wider literature, the prevalence amongst patients undergoing lumbar discectomy varies from 1-10% (Choudhury & Taylor, 1980; Jennett, 1956; Kostuik,

Harrington, Alexander, et al., 1986; Robinson, 1965; Shephard, 1959). It is not possible to estimate prevalence in a low back pain population from a surgical population. However, to give an idea of the rarity of cauda equina syndrome, we can consider the incidence of spine surgery. In the United States, spine surgery rates were 4.5 per 1,000 Medicare enrolees (Deyo & Mirza, 2006), and in New Zealand spine surgery rates were 60% lower than in the United States (Deyo & Mirza, 2009). Therefore, a prevalence of 1-10% within a surgical population would equate to a very low prevalence amongst patients presenting with low back pain.

Index tests. Unfortunately, this review was unable to support or refute the use of any red flag questions for cauda equina syndrome. More research is required to investigate the use of red flag questions in this area, as well as combinations of red flag questions. Development of a screening questionnaire would be particularly beneficial in this area, due to the personal nature of common symptoms such as bladder, bowel and sexual dysfunction. People with these symptoms may not immediately report them to their doctor or clinician as they may be too embarrassed to discuss them. However, they may be more likely to disclose these issues in a questionnaire (Palmieri & Stern, 2009).

Reference standard. To correctly classify cauda equina syndrome, a reference standard with high precision and clear visualisation of cauda equina nerve roots is required, namely MRI or surgery (Coscia et al., 1994).

Conclusion. Very little is known about the prevalence of cauda equina syndrome in populations of patients presenting with low back pain. The incidence within a low back pain population is unknown. Clinical history findings may be used to increase the suspicion of cauda equina syndrome but there is no research to support the use of red flag questions. There is some evidence that red flag questions may have poor sensitivity ranging from 8% for urinary frequency and bowel incontinence, to 54% for bilateral sciatica in populations with suspected cauda equina syndrome (Gooding et al., 2013). Therefore, cauda equina syndrome cannot be ruled out in the absence of these features. Further research is required to assess the diagnostic accuracy and utility of these index tests within populations of patients complaining of low back pain.

Weaknesses of the review. A single author conducted this literature search and review. Unfortunately, although there is a wealth of papers describing the signs and symptoms and surgical outcomes associated with cauda equina, there is very limited

research investigating cauda equina syndrome in a population with back pain. Therefore, this review was unable to establish prevalence, incidence, or provide any useful recommendations to screen for cauda equina syndrome in clinical practice.

Strengths of this review. To the author's knowledge this is the first review of the prevalence of cauda equina syndrome. This is also the first review to investigate the use of the red flag questions to screen for cauda equine syndrome in patients with low back pain.

Applicability of findings to this thesis. This systematic review highlights the need for further research to establish the prevalence and incidence of cauda equina syndrome. It also highlights the need for additional research to investigate the diagnostic accuracy of the red flag questions that are commonly recommended to screen for cauda equina syndrome. The red flag questions that arose from the background review of the literature and require further investigation were: urinary retention or overflow, urinary or faecal incontinence, sexual dysfunction, perineal or saddle anaesthesia, back pain, sciatica, bilateral sensorimotor symptoms, and gait abnormality. Therefore, these questions were included for further investigation in the diagnostic accuracy study of this thesis.

2.5 Target Condition: Spinal Infection

2.5.1 Background

For the purposes of this review, 'spinal infection' is considered an umbrella term to include infective spondylitis, discitis (or diskitis), vertebral osteomyelitis, and paraspinal muscle abscess (commonly psoas abscess). Spinal infections are rare but may be becoming more prevalent due to increasing rates of spinal surgery and interventional procedures, increasing chronic and antibiotic resistant diseases, and an aging population (World Health Organisation, 2014). Immunocompromised patients have an increased risk of infection, and with advances in organ transplants, and treatment for illnesses such as Human Immunodeficiency Virus (HIV), Acquired Immune Deficiency Syndrome (AIDS), *Myobacterium tuberculosis* (TB) and cancer, the number of immunocompromised hosts is rising. Inappropriate use of antibiotics worldwide has also led to increased diversity of micro-organisms and an increase in antibiotic resistant micro-organisms such as methicillin-resistant *Staphylococcus aureus* (MRSA) (Nagashima, Yamane, Nishi, Nanjo, & Teshima, 2010).

Common pathogens that cause spinal infections are pyogenic micro-organisms such as gram negative bacteria, commonly *Staphylococcus aureus* which has been documented to account for 40-49% of cases (Carragee, 1997; Nagashima et al., 2010). However, this percentage has been declining in recent decades due to an increasing variety of organisms (Nagashima et al., 2010). Hadjipavlou et al. (2000) found that MRSA accounted for 10-61% of all *Staphylococcus aureus* isolates in patients with spinal infections (Hadjipavlou et al., 2000; Nagashima et al., 2010). Infections may also be non-pyogenic granulomatous infections from TB, brucellosis, fungi or parasites. The spine is the most common extra-pulmonary site for TB, and TB spondylitis accounts for 10-40% of all cases of TB (Denis-Delpierre et al., 1998; Dolberg, Schlaeffer, Greene, & Alkan, 1991; Weir & Thornton, 1985).

Routes of contamination can be via the blood stream (haematogenous spread), post-surgical, direct implantation during procedures such as lumbar puncture or the infection may spread from an adjacent tissue focus such as a psoas abscess (Nagashima et al., 2010). Early diagnosis of spinal infection is crucial to reduce the risk of systemic illness, neurological compromise, epidural abscess, osteolysis, or ultimately, mortality (Rees, 2013). Epidural abscess affects 4-38% of patients presenting with non-postoperative spondylodiscitis, and is a serious complication, due to the increased risk of significant neurological compromise from a combination of mechanical compression and vascular compromise to the spinal cord (Cottle & Riordan, 2008; Hadjipavlou et al., 2000).

Prevalence/incidence. Spinal infection is rare, and little is recorded in the literature with regard to the prevalence of spinal infection amongst patients presenting with low back pain. Spinal infection is known to account for 2-4% of all bone and joint infections (Jevtic, 2004; Khan, Vaccaro, & Zlotolow, 1999; Tali, 2004) and the lumbar spine is the most commonly affected site in the spine. Hadjipavlou et al. (2000) found that 56.1% of all cases of spinal infection affected the lumbar spine. Nagashima et al. (2010) found that over time the lumbar spine has become the predominant site of infection. They analysed spinal infections over 50 years, and found that between 1956-1965, 58% of cases were found in the lumbar region, and by 1996 – 2005 81% of cases were in the lumbar spine. Males are predominantly affected, with the percentage of males ranging from 58-91% between studies (Beronius, Bergman, & Andersson, 2001; Friedman, Maher, Quast, McClelland, & Ebersold, 2002; Kemp, Jackson, Jeremiah, & Hall, 1973; Nagashima et al., 2010).

The majority of published studies investigating spinal infection report case studies or small case series studies amongst disease populations such as tuberculosis (D'Agostino et al., 2010; Desai, 1994; Wibaux et al., 2013). Some epidemiological studies calculated the incidence in specific countries, and in Denmark between 1978-82 they found a total incidence of 0.053 per 100,000 person-years of vertebral osteomyelitis (Krogsgaard, Wagn, & Bengtsson, 1998). Another study based in France found a higher annual incidence of vertebral osteomyelitis: 2.4/100,000 person-years in 2002/2003 (Grammatico et al., 2008).

In Denmark, a 14-year population-based study investigating pyogenic spondylodiscitis found an increasing overall incidence from 2.2 to 5.8 per 100,000 person-years between 1995-2008 (Kehrer, Pedersen, Jensen, & Lassen, 2014). Despite advancements in antibiotic therapy, several authors have reported an increasing incidence of spinal infections over time (Beronius et al., 2001; Collert, 1977; Espersen, Frimodt-Moller, Rosdahl, Skinhoj, & Bentzon, 1991). Increased incidence may be due to high-risk behaviours such as drug and alcohol abuse, an increased number of people living with immunosuppressive diseases, and an increase in interventional spine procedures (Guglielmi, De Serio, Leone, Agrosì, & Cammisa, 2000). Incidence of spinal infection in western societies has been reported between 0.047 and 5.8 per 100,000 person-years (Grammatico et al., 2008; Jiménez-Mejías et al., 1999; Krogsgaard et al., 1998; Lam & Webb, 2004).

Pyogenic vertebral osteomyelitis is also known as spondylodiscitis, as it is a bacterial infection of the vertebral bodies with extension into the adjacent intervertebral disc spaces. Lora-Tamayo and colleagues (2011) retrospectively investigated the trend of pyogenic vertebral osteomyelitis in a university teaching hospital (tertiary care) in Spain. They found that between 1991 and 2009 the incidence increased from 0.047 – 0.059 per 100,000 person-years. They also found that 60% of patients were male and the mean age was 66 years. Lora-Tamayo colleagues also investigated granulomatous vertebral osteomyelitis, which is commonly associated with TB and is otherwise known as Pott’s disease. They found that these patients had an increased risk of morbidity and mortality.

Several authors (Beronius et al., 2001; Espersen et al., 1991; Kehrer et al., 2014; Lora-Tamayo et al., 2011) have reported trends of increasing incidence of spinal infection over the past decade. The prevalence and trends of infectious diseases also

varies between countries. In 2014 there were a reported 1.5 million deaths worldwide secondary to TB. There were 6 million new cases of TB recorded including 297 reported cases in New Zealand. In New Zealand the incidence of TB was 7.4 per 100,000 in 2014 and was higher in the surrounding Pacific Islands (World Health Organisation, 2015). Fiji had a prevalence rate of 67 per 100,000, which is problematic for New Zealand as the Pacific Islands are included in our catchment area for the treatment of spinal cord injuries.

Index tests. Diagnosis of spinal infections is challenging for medical practitioners working on orthopaedics, emergency departments, and even more so in primary care. Recognition of this pathology is challenging due to its low prevalence and variable clinical presentation. The rates of missed or delayed diagnosis are reported to vary from 11-75% (Darouiche, 2006; Patel et al., 2014; World Health Organisation, 2014). Davis and colleagues (2004) found that 51% of patients present to the emergency department two or more times before they are diagnosed, and delayed diagnosis leads to a nearly 4-fold increase in the likelihood of ongoing residual motor weakness.

The “classic triad” of symptoms for spinal epidural abscess is commonly considered as fever (≥38°C), back pain and neurological deficit (Davis et al., 2004). Davis and colleagues conducted a retrospective case-control study in a tertiary care hospital over a 10-year period. They found that this triad of symptoms had high specificity of 99%, but poor sensitivity of 8% amongst the 74 cases of spinal epidural abscess. Davis et al. compared this to the presence of one or more of the following risk factors: intravenous (IV) drug use, immunocompromised, alcohol abuse, recent spine procedure, distant site of infection, diabetes, indwelling catheter, recent spine fracture, chronic renal failure or cancer. Davis et al. reported that the sensitivity with one or more risk factor findings was 98% with a negative likelihood ratio of 0.02, indicating a moderate shift in the probability of spinal epidural abscess being present. The specificity was also reasonably high at 79% with a positive likelihood ratio of 4.6, indicating a slight shift in probability. Other authors have also highlighted the increased risk of infection among HIV positive patients and IV drug users (Bigos et al., 1995; Della-Giustina, 2015; Toloba et al., 2011). Toloba and colleagues retrospectively investigated 178 cases of spinal TB and found that 5.2% of patients were HIV positive. Della-Giustina even suggested that if an IV drug user presents with insidious onset low back pain this should be assumed to be spinal infection until it has been ruled out via diagnostic imaging (Della-Giustina, 2015).

Constitutional symptoms such as unexplained weight loss, fever, chills, night sweats, or malaise are often associated with an increased risk of infection (Bigos et al., 1995). These findings are particularly concerning when they are found in patients with underlying concomitant illness, such as recent bacterial infection (including urinary tract infections), being immunocompromised, or when associated with IV drug abuse. The Accident Compensation Corporation acute low back pain guidelines (ACC, 1999) recommend fever, intravenous drug use and severe, unremitting night-time pain as red flags that may raise the suspicion of spinal infection. However, these recommendations are not based on any strong evidence.

The presence or absence of fever on clinical presentation has been shown to vary depending on the site of the infection. Deyo and colleagues reviewed the literature and found that fever had low sensitivity ranging from 27-50% between studies investigating osteomyelitis (Deyo et al., 1992). Studies investigating discitis found higher sensitivities of 60-70%, and studies investigating epidural abscess found the highest sensitivities of 66-83% (Cottle & Riordan, 2008; Darouiche, 2006; Reihsaus, Waldbaur, & Seeling, 2000). Fever is even less common in granulomatous vertebral osteomyelitis compared to pyogenic osteomyelitis, with fevers present in 17% and 48% of patients respectively (Kim et al., 2010). Deyo et al. (1992) also found that around 2% of patients presenting with low back pain will have fever due to unrelated viral infections; they recommend that this group require further questioning, examination, and consideration of imaging and blood work to rule out infection. Due to the variable presence of fever, Çigdem Ataman and colleagues recommended that spinal infection be considered as a possible diagnosis in any patient with back pain who has a raised inflammatory markers (C-reactive protein (CRP) or Erythrocyte Sedimentation Rate (ESR)). They found that CRP was raised in 64% and ESR in 77% of patients with spondylodiscitis (Çigdem Ataman et al., 2013).

Mylona and colleagues (2009) completed a review of the literature on pyogenic vertebral osteomyelitis, and included 14 studies with a total of 1,008 patients, all with confirmed pyogenic vertebral osteomyelitis. The mean age ranged from 46-72 years (median 59 years); the lower mean age of 46 years was found in two studies that included a higher proportion of IV drug users. Most patients had concomitant illnesses; most commonly diabetes mellitus (24%), followed by IV drug users (11%). 7% were immunocompromised, 6% had malignancy, and 5% abused alcohol.

Lora-Tamayo and colleagues found that 33% of patients with vertebral osteomyelitis present with neurological deficit within the first 4 weeks, 40% from 1-3 months, and greater than 3 months in 27% of patients. Neurological changes can present as radiculopathy, myelopathy or with cauda equina syndrome symptoms (Lora-Tamayo et al., 2011). Paraplegia can also result secondary to compression of the spinal canal. Back pain was the most common complaint in 86%, fever in 60%, and neurological changes such as sensory or motor deficits or urine retention in 34% (Lora-Tamayo et al., 2011; Mylona et al., 2009).

Reference standards for the identification of spinal infection. MRI is the modality of choice for spinal infections (Gold, 2016; Wilmink, 1999). It is more sensitive than plain radiology and more specific than bone scintigraphy for the diagnosis of spinal TB (Desai, 1994). Lury and colleagues found that plain radiographs can appear normal for the first two to three weeks following the onset of spinal infection (Lury, Smith, & Castillo, 2006). CT is more sensitive than plain radiography, and it has higher resolution and can identify bony destruction early (Fernandez-Ulloa, Vasavada, Hanslits, Volarich, & Elgazzar, 1985). However, MRI is more accurate in diagnosing spinal infection or epidural abscess and it allows visualisation of any spinal cord compression and epidural abscess (Diehn, 2012; Sans et al., 2012). This was also supported by a review of diagnostic imaging by Jarvik and Deyo (2002) who found that MRI was the most sensitive (96%) and specific (92%) test for infection, with positive and negative likelihood ratios suggesting a conclusive shift in probability (LR+12, LR-0.04). Plain radiography had poor specificity (57%) and poor likelihood ratios (LR+1.9, LR-0.32).

Rationale. Although several reviews and meta-anaylses have investigated red flags and spinal infection, no review has investigated red flag questions within a low back pain population. Also, the prevalence and incidence has not been documented within a low back pain population by any review. Therefore, systematic review of the literature investigating spinal infection in a low back pain population is warranted.

2.5.2 Objectives

The aim of this review reflects the aims of this chapter (see Objectives 2.1.4). This review specifically aims to determine the prevalence or incidence of spinal infections, and to investigate the diagnostic accuracy of red flag questions to screen for spinal infection in patients presenting with low back pain.

2.5.3 Methods

Criteria for consideration of studies for this review.

Types of studies. Primary diagnostic studies comparing index tests such as patient demographics or subjective history questions to an appropriate reference standard were considered for inclusion in this review. Studies investigating the prevalence or incidence of spinal infection amongst patients with low back pain were also considered for inclusion. Prospective or retrospective cross-sectional or cohort studies were considered eligible if they reported prevalence data or published sufficient data to construct diagnostic 2 x 2 tables.

Participants. Studies including adult human participants with low back pain were considered eligible for inclusion. Studies conducted in primary, secondary or tertiary care settings were all considered for inclusion.

Index tests. An index test was considered to be any question or demographic findings that could raise or lower the suspicion of spinal infection. A requirement was that all index tests could be asked or collected during patient history taking.

Target condition prevalence/incidence. Primary studies investigating the prevalence or incidence of spinal infection amongst patients with low back pain were considered eligible for inclusion in this review.

Reference standards. Plain radiography, advanced imaging (such as MRI, CT, SPECT or bone scintigraphy), or long-term follow-up were considered eligible reference standards.

Search methods for identification of studies

Electronic searches. An electronic literature search was completed on the 20th of December 2013 via the following databases: Scopus, EMBASE, MEDLINE, CINAHL, and SPORTDiscus (via EBSCO). Full text publications from “all years” to 2013 were considered for eligibility. Reference lists of all potentially eligible full texts or relevant reviews were searched to ensure no publications were missed.

Data collection and analysis

Selection of studies. Title and abstract screening was completed by a single author. Final selection was based on review of all potentially eligible full texts.

Data extraction and management. A single author extracted data relevant to study design, study characteristic, participants, target condition prevalence, reference standard and index tests. True positive, false positive, true negative and false negative data for each index test was extracted and used to construct diagnostic 2 x 2 tables to assess diagnostic accuracy.

Assessment of methodological quality. Possible sources of bias or concerns regarding study applicability were assessed used the QUADAS-2, as previously described in 2.2: Methods.

Statistical analysis and data synthesis. The prevalence of spinal infection was extrapolated from original studies. Index test data was used to construct diagnostic 2 x 2 tables, and sensitivity, specificity and likelihood ratios were calculated. Diagnostic accuracy findings could not be pooled as no index test was investigated by more than one study.

Investigations of heterogeneity. Secondary to the limited number of studies and lack of homogeneity between patient selection, study setting, index tests and reference standards, factors influencing heterogeneity could not be investigated and no meta-analysis could be performed.

2.5.4 Results

Results of the search. The electronic searches used the following key words specific to spinal infection: “discitis” or “diskitis” or “spondylodiscitis” or “osteomyelitis” or “abscess” or “spondylodiscitis” or “infection” or “infective spondylitis.” The search results and key words are reported in Appendix C.3. The combined search for the prevalence of spinal infection amongst patients presenting with low back pain resulted in 884 titles. The combined search for the diagnostic accuracy of red flags for spinal infection in patients with low back pain resulted in 1,498 titles. Following title screening of the 2,382 titles, 2,061 were removed, as they were duplicates, irrelevant, or animal studies. The remaining 321 abstracts were screened and an additional 265 articles were removed as they were reviews, editorials, letters, or did

not meet the inclusion criteria. Following hand searching of reference lists of any relevant reviews, 59 articles were selected for full text screening. 53 articles were removed primarily due to inappropriate study population or study design (case study/series or case-control). Three studies met the inclusion criteria and were selected for qualitative synthesis.

Methodological quality of included studies. Results of methodological quality assessment are reported in Table 2.11. Due to the small number of studies eligible for inclusion in the review, the combined results were not collated into graphs as seen in the first two systematic reviews in this chapter.

All studies showed risk of bias in half of the domains investigated. All studies showed unclear risk for the reference standard domain due to poor reporting of reference standard blinding, and for one study (Henschke et al., 2009) it was unclear if the reference standard could correctly classify the target condition. The study by Khoo and colleagues (2003) recorded the clinical indication on the referral, which could have been considered an index test. However, they did not provide any true positive or false negative data to allow diagnostic evaluation. Their study was therefore rated as having high applicability concerns for their index tests. It was also unclear if the index tests were interpreted without knowledge of the reference standard, which may have introduced bias. The flow and timing domain described bias that may be introduced via an inappropriate interval between the index test and reference standard, or if patients were given different reference tests, did not all receive the reference test, or were not all included in the analysis. In the study by Henschke et al. it was unclear if the interval between index test and reference standard was appropriate, and as the study by van den Bosch (2004) was conducted retrospectively, 93 participants were excluded due to missing details. All studies showed low risk of bias for patient selection and low concern regarding applicability for patient selection.

Table 2.11 *Assessment of study quality for spinal infection using the QUADAS-2*

Author	Risk of bias				Applicability concerns		
	Patient selection	Index tests	Reference Standard	Flow & Timing	Patient selection	Index tests	Reference Standard
Henschke (2009)	L	L	?	?	L	L	L
Khoo (2003)	L	?	?	L	L	H	L
van den Bosch (2004)	L	L	?	?	L	?	L

Note. L = Low risk, ? = Unclear, H = High risk

Description of studies. Table 2.12 provides detail of the study characteristics of the three eligible studies. All three studies were conducted in primary care settings. One study (Henschke et al., 2009) prospectively and consecutively recruited participants from physiotherapy, chiropractic and general practice clinics. The other studies were conducted by consecutively reviewing patients with low back pain who were referred from primary care for a lumbar X-ray. One study (van den Bosch et al., 2004) was completed as a retrospective chart review, and the other (Khoo et al., 2003) was a prospective analysis. The total number of participants across the three studies was 4,209, with only 3 cases of infection diagnosed.

Two studies (Khoo et al., 2003; van den Bosch et al., 2004) used plain radiography as the reference standard, and one study (Henschke et al., 2009) used 12-month follow-up. Khoo and colleagues reported prevalence findings but did not report any index test data. Henschke and colleagues investigated six index tests which are considered relevant to spinal infection but did not find any cases of infection, and van den Bosch et al. investigated age and gender in relation to infection (see Table 2.13).

Study findings. All included studies were conducted in primary care settings. The prevalence varied from 0-0.2% between studies. The two studies (Khoo et al., 2003; van den Bosch et al., 2004) that found cases of infection were conducted in radiology departments and found prevalences of 0.2 and 0.05% respectively. Eight index tests were investigated across the studies and no study investigated the same index test. 75% of the index tests were investigated by Henschke and colleagues (2009) who did not find any cases of infection. Van den Bosch and colleagues found that the single case of confirmed spinal infection was female and over 54 years of age.

Table 2.12 *Study characteristics for spinal infection*

Authors	Study Design	Setting	Patients	Prevalence of Spinal Infection	Reference Standard	Withdrawals
Henschke (2009)	Prospective, consecutive	Primary care	1172 patients receiving primary care for acute LBP	0 % (n=0)	12 month follow up	
Khoo (2003)	Prospective, consecutive	Radiology department	1030 patients referred from general practice for a lumbar spine radiographs	0.2% (n=2)	X-ray	No withdrawals
Van den Bosch (2004)	Retrospective chart review	Radiology department	2,007 patients referred from primary care for X-ray for LBP	0.05% (n=1)	X-ray	93 excluded due to missing radiographic or demographic details

Note. n = number, LBP = low back pain

Table 2.13 *Clinical signs and diagnostic accuracy data extracted from eligible studies for infection*

Index tests	LR+ (95% CI)	LR- (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)
Systemically unwell (Henschke et al., 2009)	NaN	NaN	NaN	0.98 (0.97, 0.98)
Constant, progressive, non-mechanical pain (Henschke et al., 2009)	NaN	NaN	NaN	0.97 (0.96, 0.98)
Recent bacterial infection eg. urinary tract or skin infection (Henschke et al., 2009)	NaN	NaN	NaN	0.98 (0.97, 0.98)
Intravenous drug abuse (Henschke et al., 2009)	NaN	NaN	NaN	1.00 (1.00, 1.00)
Immune suppression from steroids, transplant or HIV (Henschke et al., 2009)	NaN	NaN	NaN	1.00 (0.99, 1.00)
Altered sensation from the trunk down (Henschke et al., 2009)	NaN	NaN	NaN	0.98 (0.97, 0.99)
Age <54 years (van den Bosch et al., 2004)	1.78 (1.69, 1.87)	0.00 (0.00, NaN)	1.00 (0.05, 1.00)	0.44 (0.41, 0.47)
Female gender (van den Bosch et al., 2004)	1.73 (1.67, 1.80)	0.00 (0.00, NaN)	1.00 (0.05, 1.00)	0.42 (0.40, 0.44)

Note. NaN = calculation could not be performed as the values entered contain zero

2.5.5 Discussion

The aim of this review was to provide information on the prevalence or incidence of spinal infection, and to investigate the use of red flags to screen for spinal infection. Following study selection, three studies were considered eligible for inclusion in this review, with the majority excluded, as participants had known or suspected infection rather than a primary complaint of low back pain.

Prevalence and setting. Spinal infection is very rare, with only 3 cases seen within the total sample of 4,209 participants. All studies were conducted in primary care, and as patients with spinal infection may be acutely unwell, they may be more likely to present to secondary or tertiary services such as emergency departments. Therefore, more research is required to investigate prevalence in secondary and tertiary care.

Index tests. The only useful index tests from this systematic review were age greater than 54 years and female gender, which both had sensitivities of 100%. However, this was based on a single patient with discitis and is therefore not informative. The findings from the study by van den Bosch et al. (2004) also contradict the findings from previous research that suggest infection is more common in males (Beronius et al., 2001; Friedman et al., 2002; Kemp et al., 1973; Nagashima et al., 2010). However, other studies (Lora-Tamayo et al., 2011) provide some support that risk of infection increases with older age. Another review by Mylona et al. (2009) also supported this finding and reported that the median age for spinal infection was 59 years.

Reference standard. The choice of reference standard is likely to have affected prevalence. The study with the highest prevalence (Khoo et al., 2003) used X-ray as their reference standard, which was shown by van den Bosch and colleagues to over-report infection. Conversely, Henschke and colleagues (2009) used 12 month follow-up as their reference standard which could potentially miss patients with spinal infection as screening usually involves review of inflammatory markers (CRP and ESR), white blood cell count, blood cultures, and MRI. The study by van den Bosch et al. (2004) discussed the fact that patients with mild spinal infections may settle without treatment and therefore could be missed if screening was not adequate. Hence, MRI is the recommended reference standard of choice if spinal infection is suspected (Nagashima et al., 2010; Tyrrell, Cassar-Pullicino, & McCall, 1999).

Conclusions. Although there is limited research investigating prevalence of spinal infection in primary care, there is a dearth of research investigating prevalence in secondary and tertiary care settings. No study investigated incidence in a low back pain population and therefore incidence is unknown. This review was unable to recommend any red flag questions to screen for spinal infection. Further research is required to investigate the diagnostic accuracy of commonly recommended red flag questions such as fever, IV drug use, and severe unremitting night pain, or other risk factors such as immunocompromisation, or recent invasive spine procedure (ACC, 1999; Davis et al., 2004).

Weaknesses of the review. Study selection, data extraction and analysis were all performed by a single author. Due to the low number of studies included in this review, no investigations into heterogeneity or meta-analysis could be performed.

Strengths of the review. To the author's knowledge this is the first review to investigate prevalence and diagnostic accuracy of red flag questions to screen for spinal infection in a population of patients presenting with low back pain. This review has uncovered a significant gap in the evidence base, and has demonstrated clear rationale that further research in this area is required.

Applicability of findings to this thesis. At present very little is known about the prevalence of spinal infection in patients presenting with a primary complaint of low back pain. The prevalence in secondary and tertiary care, and the incidence is unknown. Therefore, this thesis will further investigate the prevalence amongst patients presenting to secondary and tertiary care with low back pain. This study will also be the first to investigate the incidence of spinal infection within a low back pain population.

The findings of this systematic review highlight the need for further research to support or refute the use of red flag questions to raise or lower the suspicion of spinal infection in clinical practice. The findings from this review were used to inform the choice of red flag questions for the diagnostic accuracy study in Chapter 4. Red flag questions that were considered to require further investigation included older age, gender, risk taking behaviours such as recreational IV drug use, constitutional symptoms such as fevers, unexplained weight loss, night sweats, night pain, malaise, and systemically unwell, and other risk factors such as HIV, AIDS and recent infections.

Chapter 3 Clinical Prevalence and Population Incidence of Serious Pathologies Amongst Patients Undergoing Magnetic Resonance Imaging for Low Back Pain

3.1 Introduction

Low back pain is a common problem, and in rare cases it may be due to underlying serious pathology (de Schepper et al., 2016; Henschke et al., 2009). The most common serious pathologies to affect the lumbar spine are fracture and malignancy. The reported prevalence of fracture varies from 0.7 to 7%, whilst that for malignancy lies between 0.07 and 7% (Henschke et al., 2013; Williams et al., 2013). The variation in reported prevalence most likely reflects the different patient cohorts, settings, and reference standards of the studies that have investigated prevalence. Cauda equina syndrome and spinal infection are rare, with reported prevalences of 0.1% and 0.05-0.2% respectively in primary care (Henschke et al., 2009; Khoo et al., 2003; van den Bosch et al., 2004). The incidence of serious pathologies in the lumbar spine is largely unknown.

Establishing disease prevalence in clinical populations is important to gain an understanding of likelihood of a patient presenting with that disease (pre-test probability). Prevalence can then be used to estimate the probability of a patient having a positive diagnosis based on the result of an index test (Fletcher, 2014). The post-test odds are calculated by multiplying the pre-test odds by the likelihood ratio for that test. If the prevalence of a condition is very low, even a test with a high level of diagnostic accuracy may have poor predictive value (Fletcher, 2014). For example, the prevalence of malignancy in tertiary care spine clinic reported by Jacobson (1997) is 7%. Deyo and Diehl (1988) have reported that the positive likelihood ratio of the index test ‘previous history of cancer’ is 15. Given these parameters, the probability of malignancy in a patient with a history of cancer would be 53%. However, the prevalence of malignancy in the primary care population is much lower i.e. 0.7% (Donner-Banzhoff et al., 2006). A positive result for this same index test in this population alters the post-test probability to just 9.6%. Hence, it is important for clinicians to consider prevalence when they are assessing risk of underlying serious pathology and deciding whether a patient may or may not require further investigation.

Knowledge of incidence is useful for understanding both the risk of developing the target condition, and disease aetiology (Fletcher, 2014). Incidence can indicate whether certain age groups or ethnicities may be more at risk of *developing* a serious pathology. Awareness of incidence and prevalence provides useful information for healthcare services to allow provision of funding and services to treat patients suffering from serious pathologies, and to plan for the number of new cases that are likely to arise over a period of time.

There is evidence that the prevalence of serious pathologies such as cancer and infection have been rising over recent decades (Kehrer et al., 2014; Ministry of Health, 2013a). The cause of the rise in prevalence of these pathologies is probably multifactorial. The increasing number of people living with co-morbidities such as obesity and diabetes, which are associated with increased risk of both infection and cancer (Mylona et al., 2009) is likely to be a key factor. Another factor associated with increasing prevalence is the aging population. Life expectancy has increased as a result of improvements in diagnosis and medical management of disease. However, as a result of this longer life span, more people are *living with* disease (Bossuyt, Reitsma, Linnet, & Moons, 2012).

Rising rates of serious pathologies may also be due to increased recognition of serious pathologies with advances in medical imaging and improved accuracy of diagnosis (Joines et al., 2001). Due to its high resolution and accuracy, MRI is a valuable tool for definitive diagnosis or exclusion of many spinal pathologies. However, MRI is an expensive test and is not widely accessible to primary care clinicians. Despite the costs and accessibility issues, many authors (Chou et al., 2012; Deyo, 1994; Pham, Landon, Reschovsky, Wu, & Schrag, 2009) have expressed concerns regarding the unsustainable rise in the number of referrals for MRI. In New Zealand, $115 million dollars was spent on diagnostic imaging (not limited to the lumbar spine) in 2015 (ACC, 2015). Some authors have suggested that patients are often referred for MRI in the absence of clear clinical rationale (Boden & Swanson, 1998; Pham et al., 2009). Others (Deyo et al., 2009) have warned that MRI usage may be higher in populations where imaging is funded by private insurers.

On the basis of the evidence considered in the preceding chapter, it is clear that additional high quality research is required to establish the prevalence and incidence of

serious pathologies in the lumbar spine that may present clinically as low back pain. Hence, the following study was conducted to address this gap in the literature.

3.1.1 Study aims

The aims of this study were to:

1. Investigate the prevalence of vertebral fracture, malignancy, cauda equina syndrome and spinal infection in patients with low back pain referred for a lumbar MRI scan in both a secondary care (private) and tertiary care (university teaching hospital) settings.

2. Determine the incidence of vertebral fracture, malignancy, cauda equina syndrome, and spinal infection in the geographic region of Counties Manukau in Auckland, New Zealand.

3.1.2 Methodological considerations

Several methodological considerations for conducting an observational epidemiology study were identified by the literature review reported in Chapter 2. On the basis of this information, the following decisions were made. One important consideration was whether data should be collected prospectively or retrospectively. Lijmer et al. (1999) investigated associations between study characteristics and diagnostic accuracy in 184 diagnostic accuracy studies using a regression model. They established that there was no difference between results from data that was captured and analysed prospectively, compared to retrospectively. it was not pragmatic to collect data prospectively within the limited timeframe allowed for this thesis. Therefore, given the findings of Lijmer et al. (1999), data for this study was collected retrospectively.

Another important methodological consideration is appropriate choice of reference standard. For this study MRI was chosen as the reference standard, given that it has been shown to be the single best non-invasive test available for the diagnosis of serious pathologies in the lumbar spine (Jarvik & Deyo, 2002; Kosuda et al., 1996). Although it would be ideal to have all MRI scans double-read by blinded, experienced radiologists (using a standardised classification system), it was not possible to obtain either the funding or additional personnel necessary for this to occur. However, MRI results were reported by experienced radiologists who were blinded to the index test results. Guidance for this study was taken from the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (von Elm et al., 2008).

This study also adhered to the layout of headings and subheadings recommended by the STROBE guidelines.

3.2 Study Methods

3.2.1 Ethics

Auckland University of Technology Ethics Committee (AUTEC) ethical approval to perform a retrospective audit of all lumbar MRI scans during the study period from 1st of October 2013 – 31st of July 2014 was approved on 30th June 2014 (see Appendix A.2). Locality approval for the retrospective audit was received via email correspondence (see Appendix A.5).

3.2.2 Study design

Retrospective audit.

3.2.3 Participants

Consecutive patients referred for lumbar MRI over a 10-month period were considered for inclusion in this study. This study was conducted as a retrospective observational study, and all reported data was de-identified to maintain the privacy and confidentiality of all participants. Data collection took place across two settings: a secondary care private musculoskeletal radiology practice (Specialist Radiology Group) and a tertiary care teaching hospital (Middlemore Hospital).

Eligibility criteria. All patients who were referred to Specialist Radiology Group or Middlemore Hospital for a lumbar MRI between the 1st of October 2013 and the 31st of July 2014 were considered eligible participants if they met the criteria below.

Inclusion criteria. All patients who received an MRI scan for low back pain and were 16 years of age or over were considered for inclusion in this study.

Exclusion criteria. Patients under 16 years of age, patients with known serious pathologies or patients undergoing lumbar MRI for reasons other than back pain (e.g. for structural or congenital abnormalities not associated with back pain) were excluded.

3.2.4 Test methods

Target conditions. This study investigated four serious pathologies that affect the lumbar spine: vertebral fracture, spinal malignancy, cauda equina syndrome and spinal infection. Vertebral fractures were defined as ‘any fracture affecting the vertebral body’.

Pars interarticularis or pedicle fractures were excluded. Spinal malignancy included any metastatic or malignant tumour that could potentially be causing the participant's complaint of low back pain. Cauda equina syndrome was classified as compression of the cauda equina nerves. Spinal infection included vertebral osteomyelitis, spondylodiscitis, discitis, epidural abscess and paravertebral muscle abscess.

Reference standard. MRI was utilised as the reference standard. The MRI scanner at SRG was a 3 Tesla Philips Achieva and the MRI scanner at Middlemore Hospital was a 1.5 Tesla Siemens Avanto. The MRI protocol included T1 and T2 weighted sagittal and coronal images, plus STIR or fat-suppressed images if indicated. Gadolinium contrast was given in limited cases at the radiologist's discretion, if indicated.

Sample size. The primary aim of this study was to determine the prevalence of serious spinal pathologies; therefore sample size was evaluated on the basis of the projected accuracy of the prevalence estimate. We expressed this accuracy as the expected confidence interval width for the prevalence. Estimation of sample size requires some knowledge of the prevalence of serious pathologies. As this detail was unknown in New Zealand, it was not possible to calculate an accurate sample size. Studies based overseas have shown that prevalence across various settings ranged from 1 - 5% for serious pathologies in the lumbar spine (Chou et al., 2007; Henschke et al., 2009; Wilk, 2004). Due to the low prevalence of serious pathologies it was considered advantageous to recruit as many participants as possible over a 10-month period. A minimum sample size of 1250 was required to allow an expected maximum confidence interval width ranging between 1.2 percentage points (for a prevalence of 1%) and 2.5 percentage points (for a prevalence of 5%).

Data collection. The Decision Support team at Middlemore Hospital and the information technology support person at SRG exported the full list of national health index (NHI) numbers, age, gender and ethnicity data for all patients who had received a lumbar MRI scan during the 10-month study period. The primary researcher, with some assistance from a research assistant, then retrieved the radiology report via the online pacs system. The information was then copied into an Excel spreadsheet for analysis. A separate spreadsheet was set up for secondary care and tertiary care to allow comparison between groups. All spreadsheets were stored securely as password-protected files.

Resources and budget. The primary resource required for this study was the researcher's time. However, due to the volume of work some assistance was required from research assistants. Tariq Al-Shatanawi assisted with data input and was kindly funded by the Department of Biostatistics and Epidemiology at Auckland University of Technology, and Andrew Fegan assisted with final data coding and received funding from Auckland Physiotherapy (the primary researcher's private practice).

3.2.5 Data analysis

The primary researcher screened all participants and excluded any prospective participants that did not meet the eligibility criteria. The primary researcher then analysed data from individual MRI reports that had been entered into an Excel spreadsheet. The data was coded to convert the information from MRI reports into numeric data, to define the type of serious pathology and the lumbar vertebral affected (see Appendix E.1 for the data coding manual). The data coding forms were developed specifically for this study in consultation with experienced musculoskeletal radiologists. A random selection of participants were double-read (5%) by a research assistant to assess the degree of error in interpretation. The primary researcher and research assistant were blinded to the index test results.

Following the completion of data entry, the total number of participants with each serious pathology was determined. Overall prevalence was calculated as a percentage of the study population and the prevalence specific to secondary care and tertiary care. 95% confidence intervals were calculated.

The Mid-P Exact method was employed to determine prevalence rate ratios to allow comparisons between secondary and tertiary prevalence (Martin & Austin, 1996). A freely available web-based calculator was used to perform these calculations, this is available at www.openepi.com/Menu/OE_Menu.htm. Findings were reported as point estimates, along with 95% confidence intervals. If a confidence interval (CI) for a prevalence ratio or rate ratio includes 1, there is no significant difference between groups at the 5% level. Therefore, we elected not to present *p*-values in this chapter, as statistically significant difference can be determined using these ratios.

To determine incidence in the geographic region of Counties Manukau, data collected from participants recruited from Middlemore Hospital (Counties Manukau District Health Board) was examined. The population was subcategorised based on age,

gender, and ethnicity to allow comparison between the 2013 Census data (Statistics New Zealand., 2014) from the Counties Manukau region. The census data considered age in 10-year bands starting from 15-24 years, up to 85 years and over. Under 16 year-olds were excluded from this study, therefore we estimated the 16-24 year-old age group population to be 90% of the 15-24 year-old age group population reported in the census data. Ethnicity-specific prevalence is based on prioritised ethnicity where respondents are allocated to a single ethnic group using the prioritised system (Māori > Pacific > Asian > European/Other). The denominator is population-based on total response where each respondent is counted in each ethnic group they selected. Therefore, the sum of the total response group will exceed the total population, as respondents may have been counted more than once (Borman, n.d.). In 2013, at most 13.8% of the population aged 15 and over declared multiple ethnicities. Hence, ethnicity-specific responses are underestimated. We note that our incidence estimates are also approximate as the 2013 population was used and this study was completed on the 31st of July 2014. Incidence was calculated by determining the number of new cases of serious pathology diagnosed over the study period. Population incidence was based on population estimates with respect to age, gender and ethnicity for each target condition. Incidence rate ratios were computed to compare groups. Confidence intervals were computed using the Clopper-Pearson method (Clopper & Pearson, 1934) for rates and the Mid-P Exact method (Martin & Austin, 1996) for rate ratios.

3.3 Results

3.3.1 Participants

A total of 2,383 participants referred for lumbar MRI scans were included in this study. MRI reports were obtained for all participants. The secondary care practice Specialist Radiology Group contributed 71% (1,681) of these scans and the remaining 29% (702) of the scans were from the tertiary care institute Middlemore Hospital.

Baseline characteristics of participants. The median age across all participants was 52 years, with an interquartile range of 39-64 years. Participants presenting to secondary care were significantly younger ($p < 0.05$) than tertiary care participants, with mean ages of 49 and 57 years respectively (see Table 3.1). The female to male ratio was even in secondary care. However, there were significantly more females (57%) in the tertiary care group ($p < 0.05$).

Table 3.1 *Baseline characteristics*

Participant demographics		Combined	Secondary care	Tertiary care
Age (years)	Mean age (SD)	52 (17)	49 (16)	57 (18)
	Median (Interquartile range)	52 (39-64)	49 (37-61)	59 (45-71)
Gender	Female (%)	1,235 (52%)	835 (50%)	400 (57%)
	Male (%)	1,148 (48%)	846 (50%)	302 (43%)
	Total	2,383	1,681	702

Note. SD = Standard deviation.

3.3.2 Prevalence

Table 3.2 provides detail regarding the prevalence of identified serious pathologies. Prevalence was significantly higher in the tertiary care group than in the secondary care group for all of the pathologies investigated in this study. The presence of malignancy was rare in the secondary care group with a prevalence of 0.29%, and was 15 times higher in tertiary care group with a prevalence of 4.42%. The prevalence of vertebral fracture was significantly higher in tertiary care with a prevalence rate ratio (PRR) of 3 and the confidence interval did not include one. Both groups showed higher prevalence of compression or insufficiency fractures, compared to traumatic fractures. Cauda equina compression was rare in secondary care, with a prevalence of 0.59%, and was significantly higher in tertiary care (PRR 3.8 [95% CI 1.7,8.4]). The largest difference between groups was noted for infection, which was rare in secondary care with a prevalence of 0.12%, and was nearly 30 times more common in tertiary care (PRR 28.7 [95% CI 6.8,121]).

Five percent of the study results were double-read to ensure there were no diagnostic coding errors. This process established 100% agreement between researchers and therefore there was low risk of error in data coding for prevalence.

Table 3.3 *Prevalence of serious pathology in secondary and tertiary care*

Target condition	SP 2° & 3° care (n)	SP 2° care (n)	SP 3° care (n)	Prevalence % 2° + 3° (95% CI)	Prevalence % 2° care (95% CI)	Prevalence % 3° care (95% CI)	Prevalence rate ratios: 3°/2° (95% CI)
Malignancy	36	5	31	1.51 (1.06, 2.09)	0.30 (0.097, 0.69)	4.42 (3.02, 6.21)	14.9 (6.12, 43.0)
Vertebral Fracture Total	84	37	47	3.52 (2.82, 4.35)	2.20 (1.55, 3.02)	6.70 (4.96, 8.80)	3.04 (1.98, 4.71)
Compression Fracture[1]	59	28	31	2.48 (1.89, 3.18)	1.67 (1.11, 2.40)	4.42 (3.02, 6.21)	2.65 (1.59, 4.45)
Traumatic Fracture[1]	25	9	16	1.05 (0.68, 1.54)	0.54 (0.25, 1.01)	2.28 (1.31, 3.67)	4.28 (1.89, 10.1)
Cauda Equina Syndrome	26	10	16	1.09 (0.71, 1.59)	0.59 (0.29, 1.09)	2.28 (1.31, 3.67)	3.81 (1.74, 8.77)
Spinal infection	26	2	24	1.09 (0.71, 1.59)	0.12 (0.014, 0.43)	3.42 (2.20, 5.04)	28.7 (7.94, 180)
Multiple Serious Pathologies	15	1	14	0.63 (0.35, 1.04)	0.059 (0.0015,0.033)	1.99 (1.09, 3.32)	33.5(5.94, 716)
Total	157	53	104	6.59 (5.63, 7.66)	3.15 (2.37, 4.10)	14.8 (12.2-17.7)	4.70 (3.39-6.58)

Note. SP = serious pathology, 2° = secondary care, 3° = tertiary care, n= frequency, 95% CI = 95% confidence interval, [1] = compression and traumatic fractures are subcategories of vertebral fractures.

3.3.3 Incidence

Table 3.3 provides detail regarding the incidence of serious pathologies across various ethnic, gender, and age groups. According to the 2013 New Zealand census, 211,038 Europeans, 104,673 Pacific Islanders, 101,520 Asians, and 67,944 Māori live in the Counties Manukau area. The total incidence of serious pathologies amongst patients referred for lumbar MRI for low back pain between the 1st of October 2013 and 31st of July 2014 was 25.8 per 100,000 person-years (p-y). Incidence increased with age, peaking at 249 per 100,000 p-y in the 85 years and over age group. Serious pathologies were slightly but not significantly more common in males with a rate ratio (RR) of 1.20 (95% CI [0.80,1.81]). Europeans had the highest risk of developing a serious pathology, followed by Pacific Islanders (rate ratio with respect to Europeans [RR] 0.791, 95% CI [0.471,1.33]), Māori (RR 0.548, 95% CI [0.270,1.113]), and Asians (RR 0.285, 95% CI [0.130,0.623]). In Europeans, incidence peaked at 85 years and over, and at 74-84 years in all other ethnicities.

Overall the serious pathology with the highest incidence was vertebral fracture (see Table 3.4). With further regard to vertebral fractures, the ethnicity-specific highest total incidence was found amongst Europeans, at 20 per 100,000 p-y, and incidence increased with older age. There was also a slight peak in incidence in males aged 25-34 years in for both European and Pacific populations. Overall incidence was higher for males in European and Pacific populations, but higher for females in Māori and Asian populations. There was no statistically significant difference between total incidence for males and females (RR 1.02, 95% CI [0.57,1.80]).

Detail regarding the incidence of malignancy is provided in Table 3.5. The peak incidence occurred in the 74-84 years age group at 78.4 per 100,000 p-y. The risk of malignancy was greatest for Māori with a total incidence of 25.6 per 100,000 p-y, and a peak of 495.3 per 100,000 p-y in the 74-84 years age group. The incidence amongst Asians was significantly lower with an overall incidence of 1.2 per 100,000 p-y. In European and Pacific populations incidence was similar between males and females (RR 1.06, 95% CI [0.42,2.7] in Europeans, RR 1.07, 95% CI [0.14,7.6] in Pacific populations), and incidence was slightly but not significantly lower for males in Māori (RR 0.85, 95% CI [0.19,3.8]).

Table 3.6 provides detail of the incidence with respect to cauda equina syndrome. The overall incidence was the lowest of all serious pathologies at 4.4 per

100,000 p-y. The highest ethnicity specific total incidence was found in Europeans, followed by Māori (RR with respect to Europeans 0.69, 95% CI [0.15,3.2]), Asians (RR 0.23, 95% CI [0.03,1.8]) and Pacific Islanders (RR 0.22, 95% CI [0.03, 1.8]). The peak incidence by age group varied between ethnicities. Cauda equina syndrome generally affected older age groups, with incidence peaking in the 74-84 year old age group for Europeans, and 55-64 year old age group for Māori and Asians. However, Pacific Islanders were younger, with a peak in the 16-24 year old age group. Incidence was higher in males, but not significantly so, in European and Māori populations (RR 1.32, 95% CI [0.35,4.9] for Europeans; 1.13, 95% CI [0.07,18.1] for Māori).

The incidence of spinal infections has been reported in Table 3.7. Incidence was highest in Pacific Islanders and was significantly lower for Europeans (RR with respect to Pacific 0.35, 95% CI [0.13,0.91]), but not for Asian populations (RR 0.41, 95% CI [0.13,1.3]). Māori had the lowest incidence (RR with respect to Pacific 0.15, 95% CI [0.02,1.2]). Incidence increased with age for Europeans, Pacific Islanders, and Asians. However, in Māori the incidence peaked at the younger age group of 25-34 years at 14.9 per 100,000 p-y. Pacific Islanders also had a peak of 27.5 per 100,000 p-y in the 25-34 year age group. Overall incidence was significantly higher in males than females (RR 5.3, 95% CI [1.8,15.5]), with the largest gender difference in Pacific Islanders, and only minimal difference in Asians (RR male to female 1.04, 95% CI [0.15,7.4]).

Table 3.3 *Incidence table (per 100,000 person-years) for serious pathologies in Counties Manukau*

Age group	European Estimate[1] (95% CI)			Māori Estimate[1] (95% CI)			Pacific Estimate[1] (95% CI)			Asian Estimate[1] (95% CI)			Grand Total Estimate (95% CI)		
16-24 All	5.1	(0.1,	28.2)	0.0	(0.0,	39.9)	19.9	(4.1,	58.1)	8.2	(0.2,	45.7)	10.2	(3.3,	23.7)
Female	10.1	(0.3,	56.2)	0.0	(0.0,	75.5)	12.8	(0.4,	71.6)	0.0	(0.0,	64.6)	8.2	(1.0,	29.6)
Male	0.0	(0.0,	37.4)	0.0	(0.0,	84.7)	27.4	(3.4,	98.9)	15.4	(0.4,	85.9)	12.2	(2.5,	35.6)
25-34 All	5.9	(0.1,	32.6)	29.9	(3.7,	107.8)	36.7	(10.0,	93.9)	6.6	(0.2,	37.0)	17.5	(7.6,	34.5)
Female	0.0	(0.0,	40.5)	25.4	(0.6,	141.5)	16.9	(0.4,	93.9)	0.0	(0.0,	46.6)	8.2	(1.0,	29.6)
Male	12.5	(0.3,	69.9)	36.2	(0.9,	201.7)	60.5	(12.5,	176.6)	14.0	(0.4,	77.9)	28.3	(10.3,	61.5)
35-44 All	12.9	(2.7,	37.6)	14.7	(0.4,	81.9)	20.0	(2.5,	72.3)	0.0	(0.0,	29.9)	12.1	(4.4,	26.4)
Female	24.0	(4.9,	70.2)	25.9	(0.7,	144.3)	0.0	(0.0,	68.1)	0.0	(0.0,	54.6)	15.0	(4.0,	38.5)
Male	0.0	(0.0,	33.9)	0.0	(0.0,	125.2)	43.7	(5.3,	157.8)	0.0	(0.0,	66.0)	8.7	(1.1,	31.6)
45-54 All	15.5	(4.2,	39.6)	33.2	(4.1,	119.9)	11.9	(0.3,	66.5)	8.8	(0.2,	48.8)	18.0	(8.3,	34.2)
Female	22.7	(4.7,	66.3)	0.0	(0.0,	110.8)	22.2	(0.6,	123.9)	0.0	(0.0,	61.5)	19.3	(6.3,	44.9)
Male	7.9	(0.2,	44.0)	74.3	(9.0,	268.0)	0.0	(0.0,	95.0)	18.5	(0.5,	103.0)	16.7	(4.5,	42.7)
55-64 All	38.5	(16.7,	75.9)	26.7	(0.7,	148.8)	97.9	(31.8,	228.3)	25.5	(3.1,	92.0)	48.7	(28.9,	77.0)
Female	28.4	(5.8,	83.0)	47.6	(1.2,	264.8)	75.6	(9.2,	272.8)	47.8	(5.8,	172.5)	47.2	(21.6,	89.5)
Male	48.9	(15.9,	114.1)	0.0	(0.0,	225.3)	121.7	(25.1,	355.4)	0.0	(0.0,	100.6)	50.3	(23.0,	95.5)
65-74 All	77.2	(39.9,	134.9)	0.0	(0.0,	204.5)	70.6	(8.6,	254.8)	26.3	(0.7,	146.6)	67.0	(38.3,	108.8)
Female	62.0	(20.1,	144.7)	0.0	(0.0,	378.5)	0.0	(0.0,	245.4)	51.3	(1.3,	285.7)	48.5	(17.8,	105.6)
Male	93.6	(37.6,	192.7)	0.0	(0.0,	445.4)	150.6	(18.3,	543.0)	0.0	(0.0,	199.3)	87.0	(41.8,	159.9)
74-84 All	142.0	(73.4,	248.0)	495.3	(102.2,	1441.7)	186.5	(22.6,	672.3)	69.8	(1.7,	388.5)	182.8	(113.2,	279.3)
Female	151.1	(60.8,	311.2)	263.3	(6.7,	1459.7)	0.0	(0.0,	572.4)	0.0	(0.0,	489.4)	157.7	(75.7,	289.9)
Male	131.0	(42.6,	305.4)	866.3	(105.1,	3100.3)	465.7	(56.4,	1673.9)	146.5	(3.7,	814.0)	213.7	(106.7,	382.0)
85 & over All	286.9	(137.7,	527.2)	0.0	(0.0,	4420.1)	0.0	(0.0,	1696.0)	0.0	(0.0,	1587.3)	249.0	(119.5,	457.4)
Female	267.7	(98.2,	581.8)	0.0	(0.0,	6035.7)	0.0	(0.0,	2335.8)	0.0	(0.0,	2535.1)	231.5	(85.0,	503.4)
Male	321.6	(87.7,	821.8)	0.0	(0.0,	13982.5)	0.0	(0.0,	5582.3)	0.0	(0.0,	4172.0)	280.7	(76.5,	717.5)
Female total	31.2	(20.8,	45.1)	13.4	(3.6,	34.3)	11.2	(3.6,	26.1)	7.0	(1.5,	20.4)	23.5	(17.1,	31.5)
Male total	27.1	(17.2,	40.6)	19.0	(6.2,	44.3)	35.9	(20.1,	59.2)	9.7	(2.6,	24.9)	28.3	(21.1,	37.4)
Grand Total	29.2	(21.7,	38.4)	16.0	(7.3,	30.4)	23.1	(14.1,	35.7)	8.3	(3.4,	17.1)	25.8	(20.8,	31.7)

Note. All = Male and female combined, 95% CI = Confidence Interval, [1] = Ethnicity-specific prevalences are underestimated as the numerator is frequency based on prioritised ethnicity and denominator is population based on total response

Table 3.4 *Incidence table (per 100,000 person-years) for vertebral fractures in Counties Manukau*

Age group	European Estimate[1] (95% CI)			Māori Estimate[1] (95% CI)			Pacific Estimate[1] (95% CI)			Asian Estimate[1] (95% CI)			Grand Total Estimate (95% CI)		
16-24 All	0.0	(0.0,	18.6)	0.0	(0.0,	39.9)	0.0	(0.0,	24.5)	0.0	(0.0,	30.3)	0.0	(0.0,	7.5)
Female	0.0	(0.0,	37.2)	0.0	(0.0,	75.5)	0.0	(0.0,	47.3)	0.0	(0.0,	64.6)	0.0	(0.0,	15.2)
Male	0.0	(0.0,	37.4)	0.0	(0.0,	84.7)	0.0	(0.0,	50.5)	0.0	(0.0,	56.9)	0.0	(0.0,	14.9)
25-34 All	5.9	(0.1,	32.6)	0.0	(0.0,	55.0)	9.2	(0.3,	51.1)	0.0	(0.0,	24.5)	4.4	(0.5,	15.8)
Female	0.0	(0.0,	40.5)	0.0	(0.0,	93.6)	0.0	(0.0,	62.2)	0.0	(0.0,	46.6)	0.0	(0.0,	15.0)
Male	12.5	(0.3,	69.9)	0.0	(0.0,	133.5)	20.2	(0.5,	112.3)	0.0	(0.0,	51.6)	9.4	(1.1,	34.0)
35-44 All	4.3	(0.1,	23.8)	0.0	(0.0,	54.2)	0.0	(0.0,	36.9)	0.0	(0.0,	29.9)	2.0	(0.0,	11.3)
Female	8.0	(0.2,	44.6)	0.0	(0.0,	95.6)	0.0	(0.0,	68.1)	0.0	(0.0,	54.6)	3.8	(0.1,	20.9)
Male	0.0	(0.0,	33.9)	0.0	(0.0,	125.2)	0.0	(0.0,	80.5)	0.0	(0.0,	66.0)	0.0	(0.0,	16.1)
45-54 All	7.7	(0.9,	28.0)	16.6	(0.4,	92.5)	0.0	(0.0,	44.1)	0.0	(0.0,	32.3)	6.0	(1.3,	17.6)
Female	7.6	(0.2,	42.1)	0.0	(0.0,	110.8)	0.0	(0.0,	82.0)	0.0	(0.0,	61.5)	3.9	(0.1,	21.4)
Male	7.9	(0.2,	44.0)	37.1	(0.9,	206.7)	0.0	(0.0,	95.0)	0.0	(0.0,	68.2)	8.3	(1.0,	30.1)
55-64 All	19.2	(5.2,	49.3)	0.0	(0.0,	98.6)	0.0	(0.0,	72.2)	12.7	(0.3,	71.0)	13.5	(4.4,	31.6)
Female	18.9	(2.3,	68.4)	0.0	(0.0,	175.4)	0.0	(0.0,	139.3)	0.0	(0.0,	88.1)	10.5	(1.3,	37.8)
Male	19.6	(2.4,	70.7)	0.0	(0.0,	225.3)	0.0	(0.0,	149.5)	27.3	(0.7,	151.8)	16.8	(3.4,	49.0)
65-74 All	83.7	(44.6,	143.0)	55.5	(1.4,	308.9)	70.6	(8.6,	254.8)	52.7	(6.4,	190.1)	75.4	(44.7,	119.1)
Female	74.4	(27.3,	161.9)	102.8	(2.6,	571.5)	66.6	(1.7,	370.5)	102.6	(12.5,	370.3)	80.8	(38.8,	148.6)
Male	93.6	(37.6,	192.7)	0.0	(0.0,	445.4)	75.3	(1.9,	418.9)	0.0	(0.0,	199.3)	69.6	(30.1,	137.0)
74-84 All	82.8	(33.3,	170.7)	165.1	(4.2,	916.9)	0.0	(0.0,	343.5)	69.8	(1.7,	388.5)	87.1	(41.8,	160.0)
Female	64.8	(13.4,	189.2)	263.3	(6.7,	1459.7)	0.0	(0.0,	572.4)	0.0	(0.0,	489.4)	78.9	(25.6,	184.0)
Male	104.8	(28.6,	268.1)	0.0	(0.0,	1587.3)	0.0	(0.0,	855.9)	146.5	(3.7,	814.0)	97.1	(31.5,	226.5)
85 & over All	200.8	(80.8,	413.4)	0.0	(0.0,	4420.1)	0.0	(0.0,	1696.0)	433.1	(10.9,	2393.7)	199.2	(86.0,	392.2)
Female	178.4	(48.7,	456.3)	0.0	(0.0,	6035.7)	0.0	(0.0,	2335.8)	694.5	(17.6,	3819.3)	192.9	(62.7,	449.7)
Male	241.2	(49.8,	703.5)	0.0	(0.0,	13982.5)	0.0	(0.0,	5582.3)	0.0	(0.0,	4172.0)	210.5	(43.4,	614.3)
Female total	18.9	(11.1,	30.3)	6.7	(0.8,	24.2)	2.2	(0.0,	12.4)	7.0	(1.5,	20.4)	12.8	(8.3,	19.1)
Male total	21.2	(12.5,	33.5)	3.8	(0.1,	21.1)	4.8	(0.6,	17.3)	4.9	(0.6,	17.6)	13.0	(8.3,	19.5)
Grand Total	20.0	(14.0,	27.8)	5.3	(1.1,	15.7)	3.5	(0.7,	10.1)	6.0	(1.9,	13.8)	12.9	(9.5,	17.1)

Note. All = Male and female combined, 95% CI = Confidence Interval, [1] = Ethnicity-specific prevalences are underestimated as the numerator is frequency based on prioritised ethnicity and denominator is population based on total response

Table 3.5 *Incidence table (per 100,000 person-years) for malignancy in Counties Manukau*

Age group	European Estimate[1] (95% CI)			Māori Estimate[1] (95% CI)			Pacific Estimate[1] (95% CI)			Asian Estimate[1] (95% CI)			Grand Total Estimate (95% CI)		
16-24 All	0.0	(0.0,	18.6)	0.0	(0.0,	39.9)	0.0	(0.0,	24.5)	0.0	(0.0,	30.3)	0.0	(0.0,	7.5)
Female	0.0	(0.0,	37.2)	0.0	(0.0,	75.5)	0.0	(0.0,	47.3)	0.0	(0.0,	64.6)	0.0	(0.0,	15.2)
Male	0.0	(0.0,	37.4)	0.0	(0.0,	84.7)	0.0	(0.0,	50.5)	0.0	(0.0,	56.9)	0.0	(0.0,	14.9)
25-34 All	0.0	(0.0,	21.6)	14.9	(0.4,	83.1)	0.0	(0.0,	33.8)	6.6	(0.2,	37.0)	4.4	(0.5,	15.8)
Female	0.0	(0.0,	40.5)	25.4	(0.6,	141.5)	0.0	(0.0,	62.2)	0.0	(0.0,	46.6)	4.1	(0.1,	22.8)
Male	0.0	(0.0,	46.3)	0.0	(0.0,	133.5)	0.0	(0.0,	74.3)	14.0	(0.4,	77.9)	4.7	(0.1,	26.2)
35-44 All	4.3	(0.1,	23.8)	14.7	(0.4,	81.9)	10.0	(0.3,	55.7)	0.0	(0.0,	29.9)	6.1	(1.3,	17.8)
Female	8.0	(0.2,	44.6)	25.9	(0.7,	144.3)	0.0	(0.0,	68.1)	0.0	(0.0,	54.6)	7.5	(0.9,	27.2)
Male	0.0	(0.0,	33.9)	0.0	(0.0,	125.2)	21.8	(0.6,	121.7)	0.0	(0.0,	66.0)	4.4	(0.1,	24.4)
45-54 All	11.6	(2.4,	33.9)	16.6	(0.4,	92.5)	11.9	(0.3,	66.5)	0.0	(0.0,	32.3)	10.0	(3.2,	23.3)
Female	15.1	(1.8,	54.6)	0.0	(0.0,	110.8)	22.2	(0.6,	123.9)	0.0	(0.0,	61.5)	11.6	(2.4,	33.7)
Male	7.9	(0.2,	44.0)	37.1	(0.9,	206.7)	0.0	(0.0,	95.0)	0.0	(0.0,	68.2)	8.3	(1.0,	30.1)
55-64 All	9.6	(1.2,	34.8)	26.7	(0.7,	148.8)	19.6	(0.5,	109.1)	0.0	(0.0,	47.0)	10.8	(2.9,	27.7)
Female	9.5	(0.3,	52.7)	47.6	(1.2,	264.8)	37.8	(0.9,	210.4)	0.0	(0.0,	88.1)	15.7	(3.2,	46.0)
Male	9.8	(0.3,	54.4)	0.0	(0.0,	225.3)	0.0	(0.0,	149.5)	0.0	(0.0,	100.6)	5.6	(0.1,	31.1)
65-74 All	45.1	(18.1,	92.8)	0.0	(0.0,	204.5)	35.3	(0.9,	196.6)	0.0	(0.0,	97.1)	33.5	(14.5,	66.1)
Female	37.2	(7.7,	108.7)	0.0	(0.0,	378.5)	0.0	(0.0,	245.4)	0.0	(0.0,	189.2)	24.3	(5.0,	70.9)
Male	53.5	(14.6,	136.8)	0.0	(0.0,	445.4)	75.3	(1.9,	418.9)	0.0	(0.0,	199.3)	43.5	(14.1,	101.5)
74-84 All	59.2	(19.2,	138.0)	495.3	(102.2,	1441.7)	0.0	(0.0,	343.5)	0.0	(0.0,	257.3)	78.4	(35.8,	148.7)
Female	43.2	(5.3,	155.9)	263.3	(6.7,	1459.7)	0.0	(0.0,	572.4)	0.0	(0.0,	489.4)	63.1	(17.2,	161.5)
Male	78.6	(16.2,	229.5)	866.3	(105.1,	3100.3)	0.0	(0.0,	855.9)	0.0	(0.0,	539.2)	97.1	(31.5,	226.5)
85 & over All	0.0	(0.0,	105.7)	0.0	(0.0,	4420.1)	0.0	(0.0,	1696.0)	0.0	(0.0,	1587.3)	0.0	(0.0,	91.8)
Female	0.0	(0.0,	164.4)	0.0	(0.0,	6035.7)	0.0	(0.0,	2335.8)	0.0	(0.0,	2535.1)	0.0	(0.0,	142.3)
Male	0.0	(0.0,	296.3)	0.0	(0.0,	13982.5)	0.0	(0.0,	5582.3)	0.0	(0.0,	4172.0)	0.0	(0.0,	258.6)
Female total	10.0	(4.6,	19.1)	13.4	(3.6,	34.3)	4.5	(0.5,	16.2)	0.0	(0.0,	8.6)	8.6	(4.9,	13.9)
Male total	10.6	(4.8,	20.1)	11.4	(2.4,	33.2)	4.8	(0.6,	17.3)	2.4	(0.0,	13.6)	8.5	(4.7,	14.0)
Grand Total	10.3	(6.1,	16.2)	12.5	(5.0,	25.6)	4.6	(1.3,	11.8)	1.2	(0.0,	6.7)	8.5	(5.8,	12.2)

Note. All = Male and female combined, 95% CI = Confidence Interval, [1] = Ethnicity-specific prevalences are underestimated as the numerator is frequency based on prioritised ethnicity and denominator is population based on total response

Table 3.6 *Incidence table (per 100,000 person-years) for cauda equina syndrome in Counties Manukau*

Age group	European Estimate[1] (95% CI)			Māori Estimate[1] (95% CI)			Pacific Estimate[1] (95% CI)			Asian Estimate[1] (95% CI)			Grand Total Estimate (95% CI)		
16-24 All	5.1	(0.1,	28.2)	0.0	(0.0,	39.9)	6.6	(0.2,	36.9)	0.0	(0.0,	30.3)	4.1	(0.5,	14.7)
Female	10.1	(0.3,	56.2)	0.0	(0.0,	75.5)	12.8	(0.4,	71.6)	0.0	(0.0,	64.6)	8.2	(1.0,	29.6)
Male	0.0	(0.0,	37.4)	0.0	(0.0,	84.7)	0.0	(0.0,	50.5)	0.0	(0.0,	56.9)	0.0	(0.0,	14.9)
25-34 All	0.0	(0.0,	21.6)	0.0	(0.0,	55.0)	0.0	(0.0,	33.8)	0.0	(0.0,	24.5)	0.0	(0.0,	8.1)
Female	0.0	(0.0,	40.5)	0.0	(0.0,	93.6)	0.0	(0.0,	62.2)	0.0	(0.0,	46.6)	0.0	(0.0,	15.0)
Male	0.0	(0.0,	46.3)	0.0	(0.0,	133.5)	0.0	(0.0,	74.3)	0.0	(0.0,	51.6	0.0	(0.0,	17.3)
35-44 All	0.0	(0.0,	15.8)	0.0	(0.0,	54.2)	0.0	(0.0,	36.9)	0.0	(0.0,	29.9)	0.0	(0.0,	7.4)
Female	0.0	(0.0,	29.6)	0.0	(0.0,	95.6)	0.0	(0.0,	68.1)	0.0	(0.0,	54.6)	0.0	(0.0,	13.9)
Male	0.0	(0.0,	33.9)	0.0	(0.0,	125.2)	0.0	(0.0,	80.5)	0.0	(0.0,	66.0)	0.0	(0.0,	16.1)
45-54 All	0.0	(0.0,	14.2)	16.6	(0.4,	92.5)	0.0	(0.0,	44.1)	0.0	(0.0,	32.3)	2.0	(0.0,	11.2)
Female	0.0	(0.0,	27.8)	0.0	(0.0,	110.8)	0.0	(0.0,	82.0)	0.0	(0.0,	61.5)	0.0	(0.0,	14.2)
Male	0.0	(0.0,	29.2)	37.1	(0.9,	206.7)	0.0	(0.0,	95.0)	0.0	(0.0,	68.2)	4.2	(0.1,	23.2)
55-64 All	4.8	(0.1,	26.8)	26.7	(0.7,	148.8)	0.0	(0.0,	72.2)	12.7	(0.3,	71.0)	10.8	(2.9,	27.7)
Female	0.0	(0.0,	34.9)	47.6	(1.2,	264.8)	0.0	(0.0,	139.3)	23.9	(0.6,	133.1)	15.7	(3.2,	46.0)
Male	9.8	(0.3,	54.4)	0.0	(0.0,	225.3)	0.0	(0.0,	149.5)	0.0	(0.0,	100.6)	5.6	(0.1,	31.1)
65-74 All	19.3	(4.0,	56.5)	0.0	(0.0,	204.5)	0.0	(0.0,	130.2)	0.0	(0.0,	97.1)	16.8	(4.6,	42.9)
Female	12.4	(0.3,	69.2)	0.0	(0.0,	378.5)	0.0	(0.0,	245.4)	0.0	(0.0,	189.2)	8.1	(0.2,	45.0)
Male	26.7	(3.3,	96.5)	0.0	(0.0,	445.4)	0.0	(0.0,	277.5)	0.0	(0.0,	199.3)	26.1	(5.4,	76.3)
74-84 All	35.5	(7.3,	103.8)	0.0	(0.0,	607.4)	0.0	(0.0,	343.5)	0.0	(0.0,	257.3)	34.8	(9.5,	89.1)
Female	43.2	(5.3,	155.9)	0.0	(0.0,	967.4)	0.0	(0.0,	572.4)	0.0	(0.0,	489.4)	47.3	(9.7,	138.3)
Male	26.2	(0.7,	145.9)	0.0	(0.0,	1587.3)	0.0	(0.0,	855.9)	0.0	(0.0,	539.2)	19.4	(0.5,	108.2)
85 & over All	28.7	(0.7,	159.8)	0.0	(0.0,	4420.1)	0.0	(0.0,	1696.0)	0.0	(0.0,	1587.3)	24.9	(0.6,	138.7)
Female	0.0	(0.0,	164.4)	0.0	(0.0,	6035.7)	0.0	(0.0,	2335.8)	0.0	(0.0,	2535.1)	0.0	(0.0,	142.3)
Male	80.4	(2.0,	447.3)	0.0	(0.0,	13982.5)	0.0	(0.0,	5582.3)	0.0	(0.0,	4172.0)	70.2	(1.7,	390.4)
Female total	4.5	(1.2,	11.4)	3.3	(0.1,	18.6)	2.2	(0.0,	12.4)	2.3	(0.0,	13.0)	4.8	(2.2,	9.1)
Male total	5.9	(1.9,	13.8)	3.8	(0.1,	21.1)	0.0	(0.0,	8.8)	0.0	(0.0,	8.9)	4.0	(1.6,	8.2)
Grand Total	5.2	(2.4,	9.8)	3.6	(0.4,	12.8)	1.2	(0.0,	6.4)	1.2	(0.0,	6.7)	4.4	(2.5,	7.1)

Note. All = Male and female combined, 95% CI = Confidence Interval, [1] = Ethnicity-specific prevalences are underestimated as the numerator is frequency based on prioritised ethnicity and denominator is population based on total response

Table 3.7 *Incidence table (per 100,000 person-years) for spinal infection in Counties Manukau*

Age group	European Estimate[1] (95% CI)			Māori Estimate[1] (95% CI)			Pacific Estimate[1] (95% CI)			Asian Estimate[1] (95% CI)			Grand Total Estimate (95% CI)		
16-24 All	0.0	(0.0,	18.6)	0.0	(0.0,	39.9)	0.0	(0.0,	24.5)	8.2	(0.2,	45.7)	2.0	(0.0,	11.3)
Female	0.0	(0.0,	37.2)	0.0	(0.0,	75.5)	0.0	(0.0,	47.3)	0.0	(0.0,	64.6)	0.0	(0.0,	15.2)
Male	0.0	(0.0,	37.4)	0.0	(0.0,	84.7)	0.0	(0.0,	50.5)	15.4	(0.4,	85.9)	4.1	(0.1,	22.6)
25-34 All	0.0	(0.0,	21.6)	14.9	(0.4,	83.1)	27.5	(5.6,	80.5)	0.0	(0.0,	24.5)	8.8	(2.4,	22.4)
Female	0.0	(0.0,	40.5)	0.0	(0.0,	93.6)	0.0	(0.0,	62.2)	0.0	(0.0,	46.6)	0.0	(0.0,	15.0)
Male	0.0	(0.0,	46.3)	36.2	(0.9,	201.7)	60.5	(12.5,	176.6)	0.0	(0.0,	51.6)	18.9	(5.1,	48.3)
35-44 All	0.0	(0.0,	15.8)	0.0	(0.0,	54.2)	10.0	(0.3,	55.7)	0.0	(0.0,	29.9)	2.0	(0.0,	11.3)
Female	0.0	(0.0,	29.6)	0.0	(0.0,	95.6)	0.0	(0.0,	68.1)	0.0	(0.0,	54.6)	0.0	(0.0,	13.9)
Male	0.0	(0.0,	33.9)	0.0	(0.0,	125.2)	21.8	(0.6,	121.7)	0.0	(0.0,	66.0)	4.4	(0.1,	24.4)
45-54 All	0.0	(0.0,	14.2)	0.0	(0.0,	61.3)	0.0	(0.0,	44.1)	8.8	(0.2,	48.8)	2.0	(0.0,	11.2)
Female	0.0	(0.0,	27.8)	0.0	(0.0,	110.8)	0.0	(0.0,	82.0)	0.0	(0.0,	61.5)	0.0	(0.0,	14.2)
Male	0.0	(0.0,	29.2)	0.0	(0.0,	136.9)	0.0	(0.0,	95.0)	18.5	(0.5,	103.0)	4.2	(0.1,	23.2)
55-64 All	9.6	(1.2,	34.8)	0.0	(0.0,	98.6)	58.7	(12.1,	171.6)	12.7	(0.3,	71.0)	18.9	(7.6,	39.0)
Female	0.0	(0.0,	34.9)	0.0	(0.0,	175.4)	0.0	(0.0,	139.3)	23.9	(0.6,	133.1)	5.2	(0.1,	29.2)
Male	19.6	(2.4,	70.7)	0.0	(0.0,	225.3)	121.7	(25.1,	355.4)	0.0	(0.0,	100.6)	33.5	(12.3,	73.0)
65-74 All	12.9	(1.6,	46.5)	0.0	(0.0,	204.5)	35.3	(0.9,	196.6)	26.3	(0.7,	146.6)	16.8	(4.6,	42.9)
Female	0.0	(0.0,	45.8)	0.0	(0.0,	378.5)	0.0	(0.0,	245.4)	51.3	(1.3,	285.7)	8.1	(0.2,	45.0)
Male	26.7	(3.3,	96.5)	0.0	(0.0,	445.4)	75.3	(1.9,	418.9)	0.0	(0.0,	199.3)	26.1	(5.4,	76.3)
74-84 All	35.5	(7.3,	103.8)	0.0	(0.0,	607.4)	186.5	(22.6,	672.3)	0.0	(0.0,	257.3)	52.2	(19.1,	113.7)
Female	43.2	(5.3,	155.9)	0.0	(0.0,	967.4)	0.0	(0.0,	572.4)	0.0	(0.0,	489.4)	31.5	(3.9,	113.9)
Male	26.2	(0.7,	145.9)	0.0	(0.0,	1587.3)	465.7	(56.4,	1673.9)	0.0	(0.0,	539.2)	77.7	(21.2,	198.8)
85 & over All	0.0	(0.0,	105.7)	0.0	(0.0,	4420.1)	0.0	(0.0,	1696.0)	0.0	(0.0,	1587.3)	0.0	(0.0,	91.8)
Female	0.0	(0.0,	164.4)	0.0	(0.0,	6035.7)	0.0	(0.0,	2335.8)	0.0	(0.0,	2535.1)	0.0	(0.0,	142.3)
Male	0.0	(0.0,	296.3)	0.0	(0.0,	13982.5)	0.0	(0.0,	5582.3)	0.0	(0.0,	4172.0)	0.0	(0.0,	258.6)
Female total	2.2	(0.2,	8.0)	0.0	(0.0,	12.4)	0.0	(0.0,	8.2)	4.7	(0.5,	16.8)	2.1	(0.6,	5.4)
Male total	5.9	(1.9,	13.8)	3.8	(0.1,	21.1)	23.9	(11.5,	44.0)	4.9	(0.6,	17.6)	11.3	(6.9,	17.5)
Grand Total	4.0	(1.6,	8.2)	1.8	(0.1,	9.9)	11.5	(5.5,	21.3)	4.8	(1.3,	12.2)	6.6	(4.3,	9.8)

Note. All = Male and female combined, 95% CI = Confidence Interval, [1] = Ethnicity-specific prevalences are underestimated as the numerator is frequency based on prioritised ethnicity and denominator is population based on total response

3.3.4 Discussion

This study has provided important new information regarding the prevalence and incidence of serious pathologies in the lumbar spine. The prevalence of all target conditions has been investigated in both secondary and tertiary care settings, and incidence has been investigated in the geographic region of Counties Manukau using data from Middlemore Hospital (Counties Manukau District Health Board).

Prevalence. Establishing the prevalence in a population is important, as it allows an increased understanding of the proportion of patients within that population who will be living with a serious pathology at that point in time. Knowledge of this prevalence is useful for planning and provision of services to manage and treat patients suffering from these pathologies. It is also important for clinical decision-making as prevalence dictates the pre-test probability of a patient having a serious pathology.

With respect to prevalence, our study has demonstrated several similarities and some differences compared to previous findings reported in the literature. For vertebral fracture, our systematic review of the literature found that one previous study (Roman et al., 2010) was conducted in secondary care. Roman and colleagues reported a prevalence of 2.6%, which was similar to the prevalence of 2.2% in our secondary care group. In tertiary care our study demonstrated a prevalence of 6.7%, which was between the 6.5% and 7.2% reported by two other studies conducted in tertiary care settings (Gibson & Zoltie, 1992; Patrick et al., 1983). Interestingly, our study prevalence was comparable to that of similar previous studies, although we used MRI as the reference standard and all the aforementioned studies used X-ray as their reference standard. Another recently published study (de Schepper et al., 2016) used MRI as their reference standard and reported a prevalence of 2.5%, which was similar to our secondary care prevalence. However, their study was conducted in a primary care setting.

With regard to malignancy, previous studies based in secondary care settings have reported a fairly large range in prevalence between 0.1% (Slipman et al., 2003) and 7% (Jacobson, 1997). Our study reported a prevalence of 0.3% in secondary care, which was on the lower side of the previously reported range. This range was likely due to heterogeneity in patient selection, setting, and reference standard between published studies, which make comparison difficult. The study by Slipman and colleagues (2003) was conducted in private clinics and used review of the medical notes as their reference

standard, whereas Jacobson (1997) included only patients referred for bone scan, which is less likely to misclassify malignancy. The prevalence in our tertiary care group was 4.4%, similar to the prevalence of 5.9% reported by Cook and colleagues (2012) who also used MRI as their reference standard.

Cauda equina syndrome was rare with a prevalence of 0.6% in our secondary care group and 2.3% in our tertiary care group. No study investigated cauda equina syndrome in a similar setting, but one study (Henschke et al., 2009) found a prevalence of 0.1% for cauda equina syndrome in primary care using long term follow-up with diagnostic confirmation using MRI.

In our study the prevalence of spinal infection was low at 0.1% in the secondary care group, and significantly higher at 3.4% in the tertiary care group. The prevalence of infection ranged from 0-0.2% in the primary care studies included in the systematic review (Henschke et al., 2009; Khoo et al., 2003; van den Bosch et al., 2004). A higher primary care prevalence of 0.73% was reported by de Schepper and colleagues (2016). This may have been due to the increased accuracy of their reference standard, as they used MRI rather than X-ray or long-term follow-up. No published study investigated prevalence in secondary or tertiary care, therefore no comparisons could be made.

Study setting. Significant differences in the prevalence of serious pathologies were found between secondary (private) and tertiary care (public health) settings. The overall prevalence of serious pathologies in secondary care was 3.2%. This was similar to the prevalence of 3% found by the primary care study by de Schepper and colleagues (2015) based in the Netherlands. In contrast, the prevalence was significantly higher in our tertiary care group at 15%. The cause of the difference between settings is likely multifactorial. However, a major contributor to the higher prevalence in the tertiary care (Middlemore Hospital) population may be socioeconomic and ethnic variance. Middlemore Hospital services the geographic region of Counties Manukau, and has a very large over-representation of people living in deprivation, compared to the national average (Ministry of Health, 2013b). Conversely, patients attending the private musculoskeletal practice (Specialist Radiology Group) are likely to be in higher socioeconomic groups. Specialist Radiology Group is located in Greenlane which is one of the highest decile (least deprived) areas in Auckland (White, Gunston, Salmond, Atkinson, & Crampton, 2008).

The Counties Manukau region has a larger population of Māori and Pacific Islanders than any other region in New Zealand (Health Partners Consulting Group, 2012; Ministry of Health, 2015b). Around 70% of Māori and 75% of Pacific Islanders live in high deprivation areas. Māori have the highest rates of avoidable mortality, followed by Pacific Islanders (Powell & Wolfgramm, 2013). Avoidable mortality arises secondary to poor management and delayed diagnosis of diseases such as diabetes, heart disease and cancer (Borman, n.d.). Increased numbers of people living with these concomitant diseases is likely to influence the prevalence of serious pathologies, as diseases such as lung and breast cancer may metastasise to the spine, and diabetes increases risk of infection and malignancy (Coussens & Werb, 2002; Sciubba & Gokaslan, 2006).

Differences between secondary and tertiary care prevalence may be partially due to the clinicians' threshold for referral. Clinicians working in tertiary care settings may be more likely to adhere to guidelines that recommend that only patients who have suspected serious pathology or neurological deficits and are potential candidates for invasive interventions should be referred for MRI (Chou et al., 2012). Conversely, despite warnings regarding the potential adverse effects of referring patients for imaging unnecessarily, clinicians working in private spine clinics have been shown to be less adherent to such guidelines and continue to refer patients who do not strictly meet these criteria (Carey & Garrett, 1996; Lurie, Birkmeyer, & Weinstein, 2003). One study (Boden & Swanson, 1998) found that 25% of patients referred for advanced imaging met lenient referral criteria. Over-referral for diagnostic imaging may also be influenced by patient expectation, as patients attending private clinics often expect to undergo imaging as part of their management (Verbeek, Sengers, Riemens, & Haafkens, 2004).

Incidence. Knowledge of the incidence of serious pathologies as the underlying cause of low back pain is useful for healthcare planning to enable more accurate forecasting of healthcare expenditure for district health boards. Population incidence also improves understanding of disease aetiology.

The results of this study clearly demonstrate that the incidence of fracture increased with older age, with peak incidence for both the European and Asian populations occurring at 85 years and over. Interestingly, for Māori and Pacific populations the peak ages were lower (74-84 years and 65-74 years respectively). This variance may reflect differences in life expectancy. Recent research has shown that

Māori and Pacific Islanders have a reduced life expectancy of 73 and 73.5 years for males and 77.1, and 77.9 years for females, respectively. This is nearly 7 years younger than the life expectancy for non-Māori and non-Pacific (80.3 years for males and 83.9 years for females) (Health Partners Consulting Group, 2012; Ministry of Health, 2015a). Aside from the increased risk with older age, there was also a peak in incidence for males in the 25-34 year age group for European and Pacific Islanders. This may be associated with increased risk-taking behaviour and involvement in adventure sport in this age group. A study investigating ACC injury claims related to adventure tourism and sports from 2004-2005 found that the majority of claims were from males aged 20-50 years (Bentley, Macky, & Edwards, 2006). Other authors (Henschke et al., 2009; Roman et al., 2010) have found higher prevalence of vertebral fracture in women. However, this study did not find any significant difference between genders.

Our study provides some evidence that risk of spinal malignancy increases with age. Māori had a significantly higher risk of spinal malignancy than European or Pacific populations. Asians had a significantly lower risk. The finding that spinal malignancy was more common in Māori is supported by the findings of another study that reported significantly higher rates of lung cancer mortality and cancer registrations in Māori compared to non-Māori (Borman, n.d.).

Cauda equina syndrome is rare and does not appear to be gender specific. In our study, risk generally increased with older age, although this was not significant. Within Maori and Asian populations the peak incidence occurred in the 55-64 year age group. The peak age was older for Europeans and younger for Pacific Islanders. The younger age at onset in Pacific Islanders may be related to the pathophysiology of cauda equina syndrome, as it most commonly occurs following a massive disc prolapse, and disc prolapse usually occurs between 30 and 50 years of age (Dunsmuir, 2004).

The risk of spinal infection increases with age and there is a significantly higher incidence in males. Previous studies have also reported predominance in males, with the percentage of males affected ranging from 58-91% (Beronius et al., 2001; Friedman et al., 2002; Kemp et al., 1973; Nagashima et al., 2010). Although risk of infection increased with age, there was a small peak at 25-34 years in Māori and Pacific Islanders. This could be associated with increased risk-taking behaviour such as IV drug use in this age group (Guglielmi et al., 2000). Spinal infection was significantly more common in Pacific Islanders compared to Europeans. The overall incidence of spinal

infection in this study was 6.6 per 100,000 p-y. This incidence rate is higher than reported rates for vertebral osteomyelitis in Denmark (0.053/100,000 p-y), or France (2.4/100,000 p-y) (Grammatico et al., 2008; Krogsgaard et al., 1998). Our population incidence is also higher than the peak incidence of pyogenic spondylodiscitis determined over a 14-year period in Denmark (Kehrer et al., 2014). However, in our study, cases of vertebral osteomyelitis and spondylodiscitis were all included under the umbrella of spinal infection, which would increase incidence.

3.3.5 Limitations

To ensure accurate classification of serious pathologies and to give a true representation of prevalence we chose to use MRI as the reference standard for this study. MRI has the highest sensitivity and specificity of any non-invasive diagnostic test for serious pathologies in the lumbar spine, and therefore has the highest precision for diagnosis. Ideally, all MRI scans would be double-read by blinded, experienced radiologists using a standardised reporting system. However, due to the limited budget and time frame of a Master's thesis, this was not possible. Despite the limitations of MRI scans being reported by a single radiologist, this does reflect standard clinical practice.

Another potential limitation is that the incidence may be under-estimated for fractures. Fractures are commonly identified via plain radiographs or CT. Hence, some patients with previously observed spinal fractures may not have been referred for MRI and would therefore not have been included in our study. However, our study prevalence was similar to other studies that used plain radiographs as their reference standard. The majority of patients with a new diagnosis of malignancy, cauda equina syndrome or infection should have been included in this study as it is standard practice that they would be referred for MRI. Nevertheless, there may be a small number of patients who were unable to undergo MRI due to pacemakers, presence of metal fragments, implantable cardiac defibrillator, or weight exceeding the limit for the MRI scanner.

3.3.6 Conclusion

This study has established the prevalence of serious pathologies in secondary and tertiary care settings in Central and South Auckland, New Zealand. It has also established the population incidence in South Auckland (Counties Manukau). The prevalence of serious pathologies was significantly higher in tertiary care than in

secondary care settings. The incidence of serious pathologies increased with older age. Overall, there was no significant gender difference. However, infection affected significantly more males than females. Māori had an increased risk of malignancy, Europeans had the highest incidence for fractures and cauda equina syndrome, and Pacific Islanders had the highest incidence for infection. Asians had the lowest overall risk of developing a serious pathology.

To the author's knowledge this is the first study to investigate the prevalence and incidence of serious pathologies in the lumbar spine, in New Zealand. This is also the first study to report both prevalence and incidence of serious pathologies within a low back pain population. This study has not only investigated prevalence within a low back pain population, but has allowed comparison between secondary and tertiary care settings. Knowledge of variation in prevalence based on setting is useful for all professionals working in these settings and also increases the generalisability of this study. Findings from this study may be useful for healthcare professionals, policy makers and healthcare funders for provision of services and to inform clinical guidelines.

Conflicts of interest. There were no conflicts of interest.

Chapter 4 Diagnostic Accuracy of Red Flag Questions to Screen for Serious Pathologies Amongst Patients Referred for Magnetic Resonance Imaging for Low Back Pain

4.1 Introduction

Although serious pathologies are a rare cause of lower back pain, they cannot be ignored or disregarded by clinicians. Delayed diagnosis of a serious pathology can have dire consequences and could ultimately lead to chronic disability, morbidity, or mortality. Missed or delayed diagnosis therefore may come at great expense for the patient in terms of their livelihood, health and wellbeing, and for the funder in terms of health dollars spent on additional care and management.

Red flag questions have been recommended for use by numerous clinical guidelines to assist with early recognition and management of serious pathologies (ACC, 1999; Chou et al., 2007; Koes et al., 2010; Van Tulder et al., 2006). However, there is very little evidence to support or refute the use of red flag questions in clinical practice. There is also a lack of consistency regarding which red flag questions are most useful. This uncertainty regarding diagnostic utility is likely to contribute to a lack of adherence to recommendations in primary care. One study by Bishop and Wing (2006) found that fewer than 5% of primary care physicians routinely use red flag screening.

Without the use of adequate screening questions, clinicians are reliant on their own experience and clinical reasoning to identify potential serious pathologies. However, if a clinician is not aware of the patterns of signs and symptoms related to a specific pathology, they may not recognise it and will miss the diagnosis. Primary care clinicians therefore tend to have increased reliance on laboratory tests and diagnostic imaging to assist with clinical diagnosis and screening, which has led to overuse (Pham et al., 2009). To reduce current overuse of diagnostic imaging, several authors (Chou et al., 2012; Deyo, 1994) have recommended selective ordering based on positive red flag findings. However, others (Henschke & Maher, 2006; Underwood, 2009) have argued that red flags need more evaluation before they can be used for these purposes.

4.1.1 Diagnostic accuracy

Consideration of whether red flag questions are useful to aid in diagnosis or to screen for serious pathology is dependent on the diagnostic accuracy of these questions.

Diagnostic accuracy refers to the ability of a test to differentiate between the presence and absence of the target condition (Eusebi, 2013). Diagnostic accuracy is regarded as the level of agreement between index test results (such as a red flag answer) and a reference standard. The reference standard is defined as the best available method of confirming or excluding the target condition, which in this case of serious pathologies is MRI (Eusebi, 2013; Gold, 2016). Comparison of the results of the index test to those of the reference standard allows diagnostic accuracy to be calculated using measures of accuracy such as specificity, sensitivity, and likelihood ratios. Specificity is the proportion of patients without the target condition who have a negative test (Bossuyt et al., 2015). Sensitivity is the proportion of patients with the target condition who have a positive result (Bossuyt et al., 2015). A test that has 100% sensitivity will always be positive in patients with the target condition. Hence, a negative test result suggested that the target condition can be ruled out. Conversely, a test with 100% specificity will always be negative in people without the target condition. Consequently, a patient with a positive result is likely to have the target condition (Bossuyt et al., 2003).

Knowledge of sensitivity and specificity is necessary to understand the diagnostic utility of a question or test. Diagnostic utility describes the relevance or usefulness of a test (Eusebi, 2013). If a test has high specificity it can be used to rule in the target condition and has diagnostic utility for diagnosis. A test with high sensitivity can be used to rule out a target condition when the test is negative, and therefore its diagnostic utility will be for screening (Grimes & Schulz, 2002). The ideal index test would perfectly discriminate between the presence and absence of a serious pathology without false positive and false negative results. However, in reality this is not the case, for example, if an index test with perfect sensitivity was chosen, the trade off would be specificity, as sensitivity and specificity are often inversely related. Therefore, the cost of misdiagnosing a patient with the target condition needs to be considered when choosing cut-off points for sensitivity or specificity (Grimes & Schulz, 2002). Receiver operator characteristic (ROC) curves are also useful to assist with this choice as they display the relationship between sensitivity and specificity (Sackett, 1992). When screening for a pathology that could potentially be life threatening, or lead to permanent disability, it is more important to ensure no cases are missed than to ensure there are no false positives. Therefore, a screening test for a serious pathology must have high sensitivity (Eusebi, 2013; Grimes & Schulz, 2002).

Likelihood ratios are more useful for assessing the diagnostic utility of a test, as they consider both sensitivity and specificity to determine the magnitude of the shift in the probability of the target condition being present or absent with a certain index test result (Grimes & Schulz, 2002). A positive likelihood ratio expresses the likelihood of a positive test result occurring in a patient with the target condition. Conversely, a negative likelihood ratio expresses the likelihood of a negative test in a patient who has the target condition (Eusebi, 2013). A likelihood ratio of 1 indicates no change in the probability of a target condition being present or absent (Eusebi, 2013; Sackett, 1992). Likelihood ratios between 2 and 5 indicate a slight increase in the probability of the target condition being present. Likelihood ratios between 5 and 10 indicate a moderate increase in probability and likelihood ratios greater than 10 indicate a conclusive or large increase in the probability of the target condition being present (Jaeschke et al., 1994). Likelihood ratios between 0.2 and 0.5 indicate a slight reduction in the probability of the target condition being present given a negative test result. Negative likelihood ratios between 0.1 and 0.2 moderately reduce the probability, and likelihood ratios less than 0.1 indicate a large and often conclusive reduction in the probability of the target condition being present (Jaeschke et al., 1994). Index tests with negative likelihood ratios that indicate a conclusive reduction in probability of a target condition being present with a negative test result can be considered to have diagnostic utility as screening tests. Conversely, index tests with positive likelihood ratios that conclusively increase the probability of a diagnosis when the test is positive can be considered as risk factors for the target condition (Bossuyt et al., 2012).

4.1.2 Study Aims

The aim of this study was to determine the diagnostic accuracy of subjective ‘red flag’ questions for the identification of vertebral fracture, malignancy, cauda equina syndrome, and spinal infection.

4.1.3 Methodological considerations

The systematic reviews of the literature reported in Chapter 2 identified a number of methodological factors that need to be considered in the design and conduct of diagnostic accuracy studies. Hence the following decisions were made regarding the current study. Firstly, a prospective, cross-sectional study design was chosen, with consecutive recruitment of participants. To reduce the risk of bias, the index tests were completed prior to the reference standard and interpreted independently from the

reference standard to ensure results were blinded. Pre-specification of cut-off points for index tests such as age or weight loss can introduce bias. Therefore, exact ages were collected (ratio data) and ranges were used for index tests such as length of time since onset, or amount of weight loss (ordinal data). Patient selection bias can occur in situations where patients who may be more at risk of serious pathologies do not take part, due to factors such as lower education, or if the patient is in too much pain. To ensure results were not affected by patient selection bias, the prevalence in the questionnaire group was compared to the clinical prevalence reported in Chapter 3. This also ensured that the study sample was representative of the low back pain population.

In diagnostic accuracy studies, bias can often be introduced by an inappropriate choice of reference standard that can either lead to over or under-estimation of the prevalence of the target condition. Therefore, the best available reference standard should be employed. Review of the literature highlighted that MRI is the single best reference standard to screen for or diagnose serious pathologies in the lumbar spine (Chou, Fu, Carrino, & Deyo, 2009; Gold, 2016; Jarvik & Deyo, 2002). While it is acknowledged that MRI is an expensive test that is not directly accessible to primary care clinicians, it was utilised as the reference standard for this study to ensure the diagnosis was accurate. All participants received the same reference standard and all participants were included in the analysis. Any withdrawals were explained. The index tests were completed in the form of a standardised questionnaire whilst participants were waiting for their MRI. Hence, there was no significant delay between the index tests and the reference standard.

Specialised musculoskeletal radiologists with a minimum of five years of experience read and reported MRI scans. The radiology report was then used to determine the diagnosis. MRI reports were analysed by the researcher, independent of the index test results, to reduce bias and ensure blinding.

To ensure adequate reporting, the STARD guidelines for reporting diagnostic accuracy studies were used in the development and conduct of this study (Bossuyt et al., 2015).

4.2 Study Methods

For completeness study methodology has been described in full below. For clarity this study was conducted concurrently with the previous study and the participants in

this group are a subgroup of the prevalence study group. The key differences in study methodology related to the recruitment of participants, and the collection and analysis of data, as participants involved in this study were required to complete a questionnaire prior to their MRI scan.

4.2.1 Ethics

Auckland University of Technology Ethics Committee (AUTEC) ethical approval was gained on the 6th of June 2013. Counties Manukau District Health Board (CMDHB) locality approval was gained on the 10th of October 2013 (see Appendix A.1 & A.3).

4.2.2 Study design

Prospective, cross-sectional study.

4.2.3 Participants

The key principles of partnership, participation and protection were considered in the research planning process (Hudson & Russell, 2009). All patients satisfying the inclusion criteria (see below) were given the opportunity to take part. The research was not targeted at a specific ethnic group, but was culturally sensitive and respectful of cultural values and rights. All participants were advised that they could withdraw from the study at any time prior to data analysis without change to their management.

Data for this study was collected concurrently, with that obtained for the prevalence study reported in Chapter Three. Concurrent collection of data with the previous study allowed for comparison between participants involved in this study and those included in the study in Chapter Three, to ensure the study sample was representative of patients receiving lumbar MRI scans during this timeframe.

Eligibility criteria. All patients referred for a lumbar MRI scan to any of the radiology departments involved in this study were invited to participate in the study. Enrolment was consecutive and took place over a ten-month period from the 10th of October 2013 to the 31st of July 2014.

Inclusion criteria. All patients with a primary complaint of low back pain requiring an MRI scan were invited to participate. People of all ethnicities were included.

Exclusion criteria. Anyone under 16 years of age or anyone unable to complete the questionnaire (secondary to language barriers or cognitive issues), or people who were unable to give informed consent were excluded. People undergoing MRI for the investigation of *known* serious pathologies (such as cancer) were also excluded.

Study sites. This study involved radiology departments in both secondary and tertiary care settings. The secondary care site was the private musculoskeletal radiology practice, Specialist Radiology Group (SRG) at Ascot Office Park, Greenlane, Central Auckland. The tertiary care department was based at Middlemore Hospital, a university teaching hospital. Radiology staff, nursing and administration staff agreed to assist with data collection and radiology department managers approved this study prior to commencement.

4.2.4 Test methods

Target conditions. This study investigated the four most common serious pathologies that affect the lumbar spine, i.e. vertebral fracture, spinal malignancy, cauda equina syndrome, and spinal infection. These pathologies were defined in detail in the previous chapter (see 3.2.4 Test methods).

Index tests. Following a review of the literature, a series of red flag questions (index tests) were compiled from clinical guidelines (ACC, 1999; Chou et al., 2007; Koes et al., 2010; Van Tulder et al., 2006) and from previous studies and reviews. One study by Henschke and colleagues (2009) proposed a diagnostic rule for vertebral fracture that included the following index tests: female gender, age greater than 70 years, history of prolonged corticosteroid use and significant trauma. Therefore, these index tests were included in our study to allow evaluation of this rule in our study population. The choice of index tests was therefore based on the best available evidence and the most commonly used red flag questions.

No pre-specified cut-off points were used for questions related to age, amount of weight loss, or time frames. These variables were either entered as ratio data (e.g. age in years) or ordinal data (e.g. time in weeks). For questions with a large number of possible answers (such as type of cancer), a blank space was left for patients to enter their specific details. See Appendix B.3 for a copy of the questionnaire.

Additional index tests and a body chart were included in the questionnaire for use in future research. These supplementary questions were related to spinal stenosis,

inflammatory back pain, and radiculopathy. Investigation into the additional pathologies and index tests was not possible within the scope of this Master's thesis, and will therefore not be discussed in any further detail.

Questionnaire development. The questionnaire was developed in conjunction with supervisors, peers, radiographers and medical specialists prior to a trial with patients to ensure there was expert consensus on the choice of index tests.

The questionnaire was initially trailed on 50 consecutive patients referred for lumbar MRI at Middlemore Hospital. The researcher then interviewed all patients to ensure that the questions were interpreted correctly. All trial participants were informed that their questionnaires would not be used for the main study and were aware that they were involved only in a pilot study. All patients gave informed consent prior to completing the questionnaire. Following the pilot study, several minor changes were made to clarify any questions that were commonly misinterpreted. Following these changes the questionnaire was finalised. A template was then set up by a technician using Orthoscope software (Scope Solutions, NZ), which has been specifically designed to allow medical questionnaires to be scanned directly into Microsoft Access to reduce error associated with data entry.

Reference standard. MRI was chosen as the best available reference standard for identification of serious pathologies in the lumbar spine is MRI. MRI scans were read and reported by experienced musculoskeletal radiologists who were blinded to the index test results. Due to resource constraints it was not possible for all MRI scans to be read by a second radiologist. Therefore, the original MRI report was used to establish the diagnosis. MRI reports were analysed by the researcher independently of index test results. If there was any uncertainty regarding the diagnosis in the radiologist's report, it was double-read by the research assistant (a physiotherapy student) to ensure agreement on the final diagnosis and coding. Examination of the reliability of MRI reporting was considered to be outside the scope of this research.

Participant recruitment. The radiology staff and nurses working within radiology departments, orthopaedic wards at Middlemore Hospital (MMH) and outpatients orthopaedic department at Manukau Super Clinic (a secondary MMH site) agreed to take part in this research and assisted with patient recruitment. Any potential participant who was referred for a lumbar MRI scan for low back pain and who met the inclusion

criteria was informed of the research by radiology or nursing staff, and was provided with a detailed participant information sheet (see Appendix B.2). This sheet included contact details of the researcher so that prospective participants could ask questions or withdraw from the study at any time prior to data analysis. If they agreed to participate, they informed nursing or radiology staff and then completed a consent form (see Appendix B.1). They were given the final version of the questionnaire to complete prior to their scheduled MRI, on the same day. Questionnaires and consent forms were then placed into a secure box behind the radiology reception area or within the charge nurses' office to ensure patient confidentiality.

Data collection

Index tests. Questionnaire and consent forms were collected from radiology departments and Orthopaedic wards each week by the researcher. The charge nurses on the Orthopaedic wards also asked inpatients who were scheduled for a lumbar MRI if they would like to take part in the research. These patients were often heavily medicated or sedated when they presented for the MRI scan, therefore it was deemed to be more suitable to collect data from the wards. Inpatients on non-orthopaedic wards who were scheduled for a lumbar MRI were asked if they would like to take part in the study by the researcher. Where possible, participants completed the forms independently. However, some inpatients were unable to write due to the severity of their condition. In these cases the researcher or nurse asked the questions and recorded the participant's responses on their behalf with their informed consent.

Following completion of data collection, questionnaire results were scanned and imported into Microsoft Access using Orthoscope software. This software recognised the questionnaires using a unique barcode printed on each page of the questionnaire. Orthoscope is a scanning system that allowed questionnaires to be scanned directly into a digital template. This reduced the time taken for data entry and ensured that data entry errors were eliminated. Data was also double-checked by the primary researcher to ensure no items were missed. Any questions that were not answered were coded as 'BLANK'. Where a participant selected both 'yes' and 'no', this was recorded as multiple answers and coded 'MULT'. Orthoscope software also allowed all data to be converted directly from Microsoft Access to a Microsoft Excel spreadsheet for analysis. Index test data entry was completed independently to reference standard data entry to ensure blinding. Results were then stored in a password-protected file using Microsoft

Excel. Original copies of the consent forms and questionnaires were stored in a locked box in the researchers office.

Reference standards. MRI reports were obtained from all participants and were analysed following completion of participant recruitment and index test data entry. MRI findings (obtained from the radiology report) were transferred to an Excel spreadsheet along with patient demographic data. The researcher then coded each scan depending on the diagnosed pathology at each anatomical level in the lumbar spine (such as L1/2). Coding focused on the presence or absence of serious pathology but also included extra information such as the presence of disc pathology, spinal stenosis, neural compromise and inflammatory changes for use in future research (see Appendix E.1 for the coding manual). The researcher was blinded to the index tests results during reference standard data entry and diagnostic coding. Results were stored in separate secure password-protected spreadsheets depending on study setting (secondary or tertiary care). Finally, following completion of data entry and coding, the index test and reference test findings were combined in Microsoft Excel for data analysis.

Sample size. Estimation of sample size requires knowledge of the prevalence of serious pathologies. As this detail was unknown in New Zealand, it was not possible to calculate sample size. Studies based overseas have shown that prevalence across various settings ranged from 1-5% for serious pathologies (Chou et al., 2007; Henschke et al., 2009; Wilk, 2004). Due to the low prevalence of serious pathologies, it was considered advantageous to recruit as many participants as possible over a 10-month period. Based on information provided by radiologists working at the sites involved in this study, it was estimated that around 60 patients would have lumbar MRI scans each week across the two sites (i.e. 250 potential participants per month). Therefore, if between 20 and 40% of patients receiving a scan were eligible and agreed to take part in the study, a sample size between 500 and 1,000 would be anticipated, and the study aimed to recruit a minimum of 500 participants. At the expected prevalences this figure allowed us to expect a confidence interval width of 0.63 around sensitivity, and 0.055 for specificity, under an assumption of specificity and sensitivity at 90% and prevalence of 1%. At 5% prevalence the confidence interval width narrowed to 0.27 for sensitivity and 0.056 for specificity (Newcombe, 1998). Hence, a minimum sample of 500 participants was considered acceptable and feasible for this study.

4.2.5 Data analysis

Following data entry and diagnostic coding, any ‘yes’ or ‘no’ answers were converted to 1 or 0 so that dichotomous data could be exported directly from Microsoft Excel to Statistical Package for the Social Sciences (SPSS) software, Version 22 (IBM© Corporation, 2013) for statistical analysis. The prevalence of each target condition was calculated and recorded. Categorical index tests were assessed for their association with each target condition using the Fishers Exact test. The null hypothesis was considered as independence between probability of a condition and probability of a criterion being positive. We applied simple logistic regression of the condition on age to determine whether participant age was significantly associated with each target condition. *P*-values less than 0.05 were considered statistically significant.

ROC curves were calculated for ratio variables that demonstrated a statistically significant relationship with the presence of a target condition. To identify the cut-off point that optimises sensitivity and specificity, Youden’s index was calculated (Youden, 1950). Due to the potentially dire consequences associated with missing a serious pathology, a second cut-off point was determined at the point where the highest specificity associated with a sensitivity of 100% occurred.

The diagnostic accuracy of each index test was evaluated with respect to the relevant target condition. True positive, true negative, false positive and false negative data was used to construct diagnostic 2 x 2 tables to calculate sensitivity, specificity and likelihood ratios for each index test using SPSS. 95% confidence intervals were constructed for all point estimates using the online clinical calculator available at http://vassarstats.net/clin1.html.

Indeterminate or missing data. Any questions (index tests) that were not completed were coded as ‘BLANK’ and in cases where participants selected multiple answers this was coded as MULT. No assumptions were made regarding participants’ answers. BLANK and MULT were considered as indeterminate data and treated as missing data during index test analysis.

4.3 Results

4.3.1 Participants

A total of 2,664 patients underwent a lumbar MRI scan during the period of data collection study. This comprised of 1708 participants in secondary care and 956

participants in tertiary care. Of these, 281 patients were excluded as they were under 16 years of age, had a *known* serious pathology or did not have low back pain. Therefore, 2,383 patients were considered eligible participants. Questionnaires were completed by 564 patients, resulting in an inclusion rate of 24%. The most common reason that the questionnaire was not completed was that staff became too busy and forgot to ask eligible patients if they would like to take part. Some patients refused or could not complete the questionnaire as they were sedated, anaesthetised or did not speak English. Following completion of the questionnaires, one participant was excluded from the tertiary care group as they did not sign the consent form. Five participants in the secondary care group were excluded as they had undergone a scan of their thoracic spine or sacrum and coccyx but not the lumbar spine (n=4), or they could not have an MRI as they exceeded the maximum weight (n=1). There were no withdrawals. For further clarification, the flow of participants through secondary and tertiary care settings are shown in Figure 4.1 and 4.2 respectively.

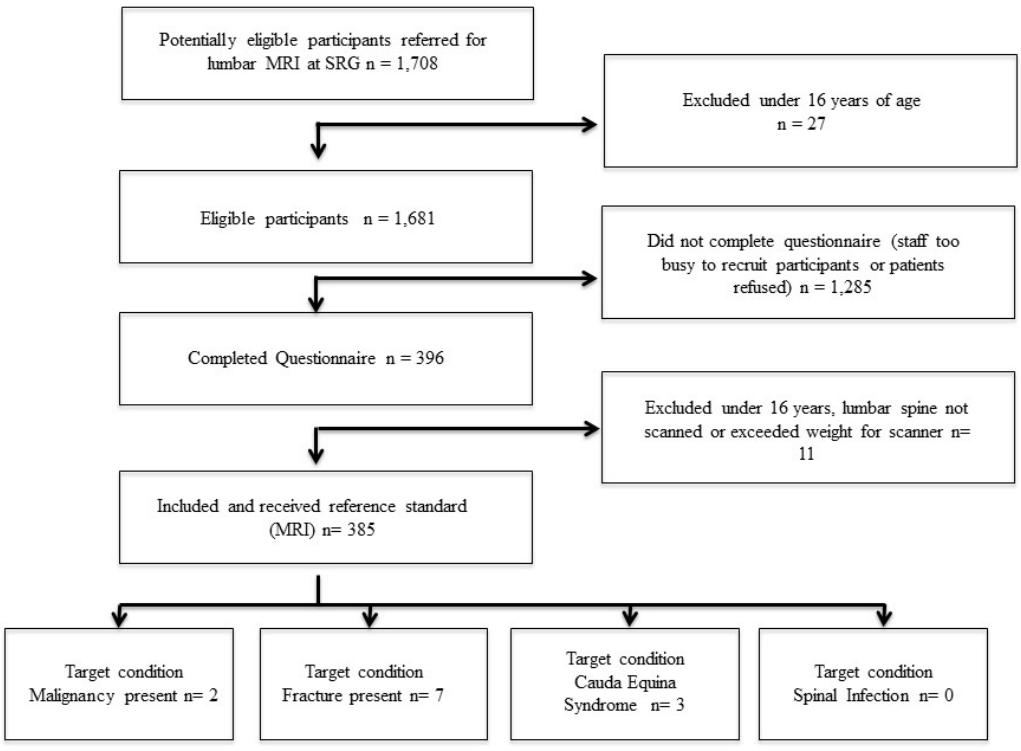


Figure 4.1 Flow of participants through study in secondary care (SRG)

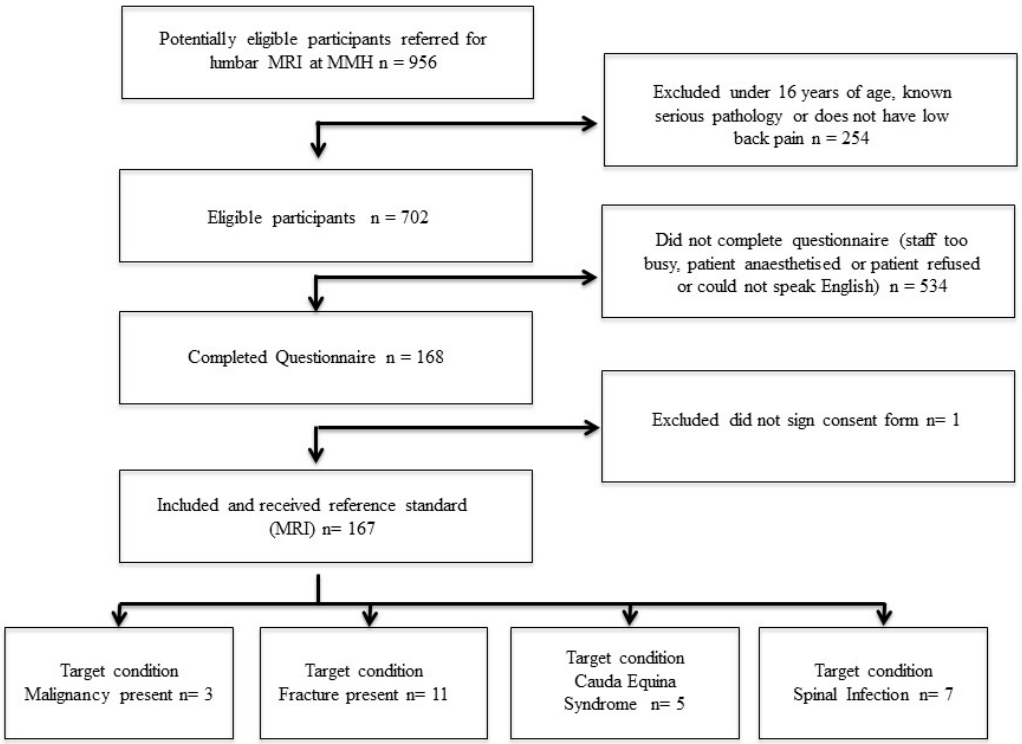


Figure 4.2 Flow of participants through study in tertiary care (MMH)

Baseline characteristics of participants. The average age of participants was 50 years, with a range from 16 to 94 years of age (see Table 4.1). The secondary care group were younger with an average age of 46 years, compared to an average age of 59 years in the tertiary care group. The proportion of male to female participants was similar in the combined group. In the secondary care group there were slightly more males (53%), and slightly more females (56%) in the tertiary care group. However, there was no clinically significant difference. The most common ethnicity was New Zealand European, with a slightly higher proportion in secondary care of 74% compared to 67% in tertiary care. Overall, there was no clinically significant difference in ethnicity between secondary and tertiary care.

Table 4.1 *Baseline characteristic*s

Participant demographics		All cases (2°+3° care combined)	Secondary care	Tertiary care
Mean age years (range)		50 years (16-94)	46 years (16-94)	59 years (16-90)
Gender n (%)	Female	273 (49%)	180 (47%)	93 (56%)
	Male	279 (51%)	205 (53%)	74 (44%)
Ethnicity n (%)	NZ European	396 (72%)	283 (74%)	113 (67%)
	Māori	31 (6%)	16 (4%)	15 (9%)
	Pacific Islanders	22 (4%)	9 (2%)	13 (8%)
	Indian	41 (7%)	34 (9%)	7 (4%)
	Chinese	8 (1%)	5 (1%)	3 (2%)
	Other	54 (10%)	37 (10%)	17 (10%)
	Total	552 (100%)	385 (100%)	167 (100%)

Note. n= number, %= proportion, 2° = secondary, 3° = tertiary

To ensure the questionnaire group study sample was representative of the population with low back pain presenting for MRI, baseline characteristics from the current study population were compared with those obtained from the previous study (reported in Chapter 3). Table 4.2 below demonstrates that there was no clinically significant difference in age or gender between the participants in the current study (the 'questionnaire' group) and those in the reported in the previous study ('prevalence study' group).

Prevalence. Comparison of the prevalence of serious pathologies between the questionnaire group and the prevalence study group revealed that there was no clinically significant difference between groups (see Table 4.2). With regard to the prevalence of individual target conditions in the questionnaire group, vertebral fracture was the most common condition with a prevalence of 3.3%, followed by cauda equina syndrome with a prevalence of 1.45%, infection with a prevalence of 1.27%, and malignancy, which was rare with a prevalence of 0.9%. The prevalence of these target conditions was similar in the prevalence study group (3.52% for vertebral fracture, 1.09% for cauda equina syndrome, 1.13% for infection, and 1.51% for malignancy).

Table 4.2 *Baseline characteristics and prevalence of serious pathologies between groups*

	Q group 2° care	PS group 2° care	Q group 3° care	PS 3° care	Q group (2°+3°)	PS group (2°+3°)
Mean age in years (range)	46 (16-94)	49 (16-100)	59 (16-90)	57 (16-100)	50 (16-94)	52 (16-100)
% Female	47	50	56	57	49	52
% Male	53	50	44	43	51	48
Prevalence of Serious Pathologies % (95% CI)	3.1(1.4,4.9)	3.2(2.4,4.1)	16(10,21)	14(12,17)	6.8(4.8,9.0)	7.3(6.2,8.3)

Note. Q = questionnaire, PS = prevalence study, 2° = secondary, 3° = tertiary, 95% CI = 95% confidence interval

4.3.2 Index test results

All of the index tests investigated were dichotomous, with the exception of age. Age was recorded as a continuous variable and therefore ROC curves were generated to determine a cutoff point if there was a significant association between age and the target condition (see Figure 4.3 – Figure 4.5). Statistically significant associations were demonstrated between age and fracture (*p*-value <0.0001) and infection (*p*-value 0.0013). Borderline association was also demonstrated between age and malignancy (*p*-value 0.060), which signalled that further investigation was warranted. There was no significant association between age and cauda equina syndrome (*p*-value 0.32). Youden's index (Youden, 1950) was used to establish the cutoff point where sensitivity and specificity were optimal. A second cutoff point was also recorded where sensitivity reached 100% (see Table 4.3, Table 4.4 and Table 4.5).

With regard to fracture, Youden's Index point occurred at 58 years of age, where sensitivity was 83% and specificity was 71%. 100% sensitivity was achieved at 35 years of age (Specificity 24%). The area under the curve was 0.835.

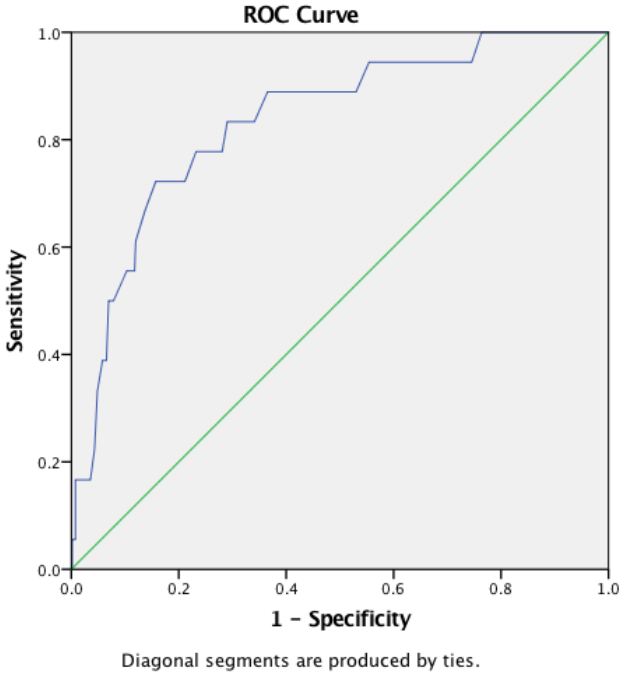


Figure 4.3 ROC Curve for age and with respect to fracture

With respect to malignancy, Youden's Index occurred at 59 years of age with a sensitivity of 80% and a specificity of 71%. 100% sensitivity was achieved at 42 years of age, where specificity was 46%. The area under the curve was 0.747.

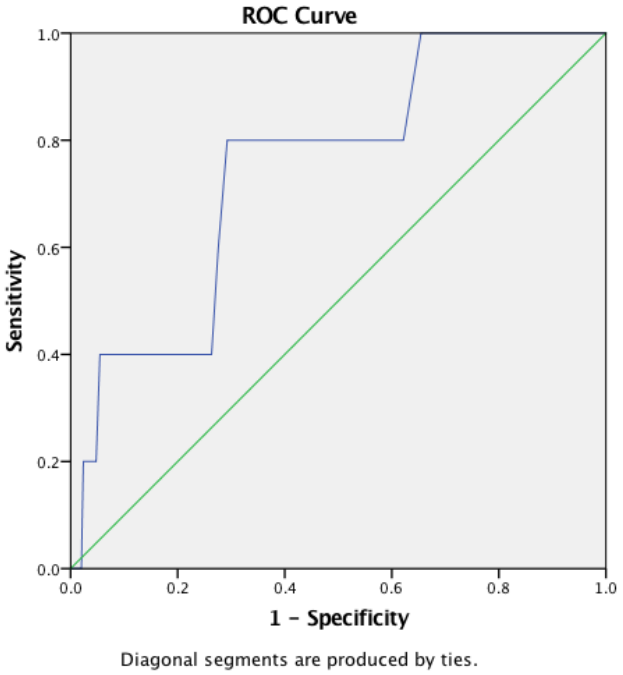


Figure 4.4 ROC Curve for age and with respect to malignancy

For spinal infection, Youden's Index occurred at 60 years of age, where sensitivity was 86% and specificity was 73%. At 55 years of age sensitivity was 100% and specificity was 65%. The area under the curve for spinal infection was 0.884.

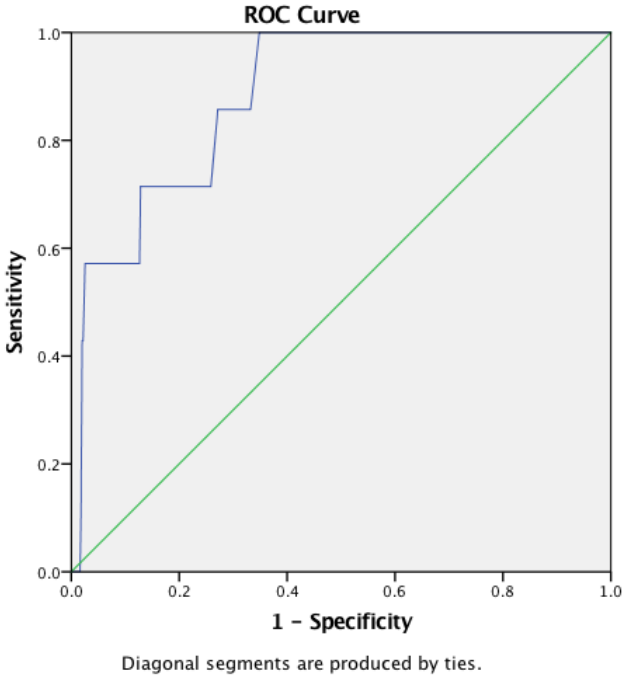


Figure 4.5 ROC Curve evaluating age and infection

Diagnostic accuracy. Tables 4.3-4.6 provide detail of the diagnostic accuracy of the index tests investigated, which showed significant association with the target condition or likelihood ratios that indicated a slight to conclusive shift in probability.

Table 4.3 provides detail of diagnostic accuracy of index tests for vertebral fracture. Several index tests (older age, insidious onset of symptoms, concomitant osteoporosis or HIV/AIDS, and a history of IV drug use) were significantly associated with a fracture (p-values ≤ 0.05). Positive likelihood ratios for these variables ranged from indicating a slight shift in probability (LR+1.3 for age >35 years) to a large shift in probability (LR+29.5 for HIV/AIDS) (Jaeschke et al., 1994). All participants with a fracture were over 35 years of age (sensitivity 100%). Only two out of the 18 participants with a fracture were under 50 years, and both under-50 year-olds had a history of significant trauma.

Diagnostic accuracy calculations for malignancy are reported in Table 4.4. Prolonged corticosteroid use was the only index test with a positive likelihood ratio that suggested a moderate increase in the probability of malignancy (LR+ 7.23 [95% CI 2.34-22.3]). Prolonged corticosteroid use was also significantly associated with malignancy. There was a slight increase in the probability of malignancy being present for the following index tests: unexplained weight loss, leg pain worse than back pain, history of cancer, age greater than 59 years, and not worse with standing and walking

(LR+ 2.35-4.61). The latter index test was significantly associated with malignancy (*p*-value 0.05). Age greater than 42 years and worsening pain both had conclusive negative likelihood ratios of zero, and 100% sensitivity. Negative likelihood ratios indicating a slight reduction in probability were also found for age >59 years, constant pain, unsteady gait, family history of cancer, not worse standing and walking, and leg pain worse than back pain (LR- 0.28-0.49).

Table 4.5 reports the diagnostic accuracy of index tests for cauda equina syndrome. Positive likelihood ratios that indicated a slight increase in probability were found for urinary overflow, urinary incontinence, faecal incontinence, change in sexual function, and perineal anaesthesia (LR+ 2.28-4.94). Absence of trauma was the only index test with a conclusive negative likelihood ratio (LR-0). Slight negative shifts in probability were found for worsening pain, unsteady gait, and worse standing and walking (LR- 0.35-0.36). The diagnostic accuracy findings of the four ACC red flags specific to cauda equina syndrome are also reported in Table 4.5. The three ACC red flags that showed a slight change in probability were: urinary retention (LR+ 2.15), faecal incontinence (LR+ 4.82), and gait abnormality (LR- 0.35). Saddle anaesthesia had poor diagnostic accuracy and no red flag question was significantly associated with cauda equina syndrome.

The diagnostic accuracy of index tests for spinal infection are reported in Table 4.6. The only index test with a positive likelihood ratio showing a large shift in probability was a history of immunosuppressant use (LR+11.7). However, this index test lacked precision, with the 95% confidence intervals ranging from 3.23 to 42.5. Age greater than 55, insidious onset, recent infection, and fevers all demonstrated slight shifts in the probability of spinal infection being present (LR+2.62-3.93) and were all significantly associated with infection. Index tests displaying slight shifts in probability, but no significant association were: systemically unwell, night sweats, and relieved with sitting. The index test age greater than 55 showed a conclusive reduction in probability (LR-0), and age greater than 60 years showed a moderate reduction in probability (LR-0.2). Negative likelihood ratios for insidious onset and night pain also showed a conclusive shift in probability (LR-0) and were significantly associated with infection. Negative likelihood ratios that slightly changed the probability were found for worsening pain, worse standing and walking, fatigue, and current or ex-smoker.

The diagnostic accuracy of the recommended ACC red flag questions to screen for any serious pathology in the lumbar spine has been reported in Table 4.7. Of the nine ACC red flags questions recommended for the identification of serious pathologies, the only index test that displayed a significant association with any serious pathology was age greater than 50 years. Likelihood ratios for IV drug use and steroid use demonstrate a slight shift in probability (LR+ 2.36 and 2.48 respectively) of a serious pathology being present when these factors are present.

The red flag questions included in the Henschke diagnostic tool for vertebral fractures (Henschke et al., 2009) were included in the current study. Table 4.8 provides detail of the combined results. Results of individual red flag questions are included in Table 4.3. Of the four individual index tests used in the diagnostic rule, only age greater than 70 years suggested a moderate change (LR+ 5.1) in the likelihood of a fracture being present. Specificity of 96% and a positive likelihood ratio (LR+ 6.25 [95% CI 2.37-16.5]) suggesting a moderate shift in the probability of a fracture being present was found for a combination of three or more positive red flags. However, this likelihood ratio lacked precision, and the high specificity was traded for low sensitivity of 22%. Combinations of one or more, or two or more positive red flags had poor diagnostic utility. Application of this rule to our study population would have missed one case of vertebral fracture. Also eight of the remaining 17 cases only had one positive finding using this diagnostic rule, which does not appear to be useful for diagnosis as the specificity was only 6% in this group.

Table 4.3 *Diagnostic accuracy of red flag questions for vertebral fracture*

Index test	Sample size (n)	Disease	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR-	P-value
Age >35 years	552	18	18	408	0	126	1.00 (0.78, 1.00)	0.24 (0.20, 0.27)	1.30 (1.25, 1.37)	0.00 (0.00, NaN)	0.002*
Age >58 years	552	18	15	155	3	379	0.83 (0.58, 0.96)	0.71 (0.67, 0.75)	2.88 (2.25, 3.67)	0.23 (0.08, 0.66)	<0.0001*
Age >70 years[1]	552	18	11	64	7	470	0.61 (0.36, 0.82)	0.88 (0.85, 0.91)	5.10 (3.30, 7.87)	0.44 (0.25, 0.78)	<0.0001*
Female gender[1]	552	18	9	264	9	270	0.50 (0.27, 0.73)	0.51 (0.46, 0.55)	1.01 (0.63, 1.62)	0.99 (0.62, 1.58)	1.00
Insidious onset of symptoms	506	18	12	185	6	303	0.67 (0.41, 0.86)	0.65 (0.58, 0.66)	1.76 (1.24, 2.49)	0.54 (0.29, 1.03)	0.01*
Concomitant osteoporosis	543	18	4	24	14	501	0.22 (0.07, 0.48)	0.95 (0.93, 0.97)	4.86 (1.88, 12.6)	0.82 (0.64, 1.04)	0.01*
Corticosteroid use[1,2]	547	18	2	30	16	499	0.11 (0.02, 0.36)	0.94 (0.92, 0.96)	1.96 (0.51, 7.57)	0.94 (0.80, 1.11)	0.28
Symptoms relieved by sitting	543	18	2	61	16	464	0.11 (0.02, 0.36)	0.89 (0.85, 0.91)	0.97 (0.25, 3.61)	1.00 (0.85, 1.18)	1.00
Not relieved by sitting	543	18	16	464	2	61	0.89 (0.64, 0.98)	0.12 (0.09, 0.15)	1.01 (0.85, 1.19)	0.97 (0.25, 3.63)	1.00
Significant trauma[1,2]	541	18	8	183	10	340	0.44 (0.22, 0.69)	0.65 (0.61, 0.69)	1.27 (0.75, 2.16)	0.85 (0.56, 1.29)	0.46
Concomitant HIV or AIDS	549	18	2	2	16	529	0.11 (0.02, 0.36)	1.00 (0.99, 1.00)	29.5 (4.40, 197)	0.89 (0.76, 1.05)	0.01*
History of IV drug use[2]	547	18	3	17	15	512	0.17 (0.04, 0.42)	0.97 (0.95, 0.98)	5.19 (1.69, 16.1)	0.86 (0.70, 1.06)	0.02*

Note. n = number, TP = true positive, FP = false positive, FN = false negative, TN = true negative, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, *p*-values calculated using Fishers exact test, * = statistically significant, NaN = cannot be calculated as a value entered contains 0, [1] = included in Henschke diagnostic rule, [2] = ACC red flags

Table 4.4 *Diagnostic accuracy of red flag questions for malignancy*

Index test	Sample size (n)	Disease	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR-	P-value
Age >42 years	552	5	5	358	0	189	1.00 (0.46, 1.00)	0.35(0.31, 0.39)	1.53 (1.43, 1.63)	0 (0.00, NaN)	0.24
Age >59 years	552	5	4	160	1	387	0.80 (0.30, 0.99)	0.70 (0.67, 0.74)	2.74 (1.73, 4.32)	0.28 (0.05, 1.63)	0.06
Unexplained weight loss[2]	489	5	2	42	3	442	0.40 (0.07, 0.83)	0.91 (0.88, 0.94)	4.61 (1.52, 14.0)	0.66 (0.32, 1.34)	0.07
Constant pain	507	5	4	269	1	233	0.80 (0.30, 0.99)	0.46 (0.42, 0.51)	1.49 (0.96, 2.33)	0.43 (0.07, 2.50)	0.38
Worsening pain	544	5	5	352	0	187	1.00 (0.46, 1.00)	0.35 (0.31, 0.39)	1.53 (1.44, 1.63)	0.00 (0.00, NaN)	0.17
Night pain[2]	543	5	4	335	1	203	0.80 (0.30, 0.99)	0.38 (0.34, 0.42)	1.28 (0.82, 2.00)	0.53 (0.9, 3.08)	0.65
Leg pain > back pain	544	5	4	190	1	349	0.80 (0.30, 0.99)	0.65 (0.61, 0.69)	2.7 (1.44, 3.62)	0.31 (0.05, 1.79)	0.06
Unsteady gait	544	5	4	318	1	221	0.80 (0.30, 0.99)	0.41 (0.37, 0.45)	1.36 (0.87, 2.11)	0.49 (0.08, 2.83)	0.65
Fatigue	543	5	3	265	2	273	0.60 (0.17, 0.93)	0.51 (0.46, 0.55)	1.22 (0.59, 2.50)	0.79 (0.27, 2.31)	0.68
History of cancer[2]	545	5	2	62	3	478	0.40 (0.07, 0.83)	0.89 (0.85, 0.91)	3.49 (1.16, 10.5)	0.68 (0.33, 1.39)	0.11
Family history of cancer	540	5	4	239	1	296	0.80 (0.30, 0.99)	0.55 (0.51, 0.60)	1.79 (1.14, 2.80)	0.36 (0.06, 2.09)	0.18
Not worse standing/walking	542	5	4	183	1	354	0.80 (0.30, 0.99)	0.66 (0.62, 0.70)	2.35 (1.49, 3.70)	0.30 (0.05, 1.75)	0.05*
Corticosteroid use[2]	547	5	2	30	3	512	0.40 (0.07, 0.83)	0.95 (0.91, 0.96)	7.23 (2.34, 22.3)	0.64 (0.31, 1.30)	0.03*

Note. n = number, TP = true positive, FP = false positive, FN = false negative, TN = true negative, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, *p*-values calculated using Fishers exact test, * = statistically significant, NaN = cannot be calculated as a value entered contains 0, [2] = ACC red flags

Table 4.5 *Diagnostic accuracy of red flag questions for cauda equina compression*

Index test	Sample size (n)	Disease	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR-	P-value
Age >53 years	552	8	6	213	2	331	0.75 (0.36, 0.96)	0.61 (0.57, 0.65)	1.27 (1.74, 2.19)	0.74 (0.30, 1.81)	0.09
Worsening pain	544	8	7	350	1	186	0.88 (0.47, 0.99)	0.35 (0.31, 0.39)	1.34 (1.02, 1.75)	0.36 (0.57, 2.27)	0.27
Worse standing and walking	542	8	7	348	1	186	0.88 (0.47, 0.99)	0.35 (0.31, 0.39)	1.34 (1.03, 1.76)	0.36 (0.06, 2.26)	0.27
Urinary retention[2]	548	8	1	49	6	492	0.14(0.01, 0.56)	0.91 (0.88, 0.93)	1.58 (0.25, 9.87)	0.94 (0.70, 1.26)	0.49
Urinary overflow	543	8	2	31	5	505	0.29 (0.05, 0.70)	0.94 (0.92, 0.96)	4.94 (1.46, 16.7)	0.76 (0.47, 1.21)	0.06
Urinary incontinence	547	8	2	66	5	474	0.29 (0.05, 0.70)	0.88 (0.85, 0.90)	2.34 (0.71, 7.71)	0.81 (0.51, 1.30)	0.21
Bowel incontinence[2]	548	8	2	28	6	512	0.25 (0.04, 0.64)	0.95 (0.93, 0.96)	4.86 (1.39, 17.0)	0.79 (0.53, 1.18)	0.07
Unsteady gait[2]	544	8	6	316	1	221	0.86 (0.42, 0.99)	0.41 (0.37, 0.45)	1.47 (1.07, 1.99)	0.35 (0.06, 2.14)	0.25
Legs feel weak[2]	544	8	6	306	2	230	0.75 (0.36, 0.96)	0.43 (0.39, 0.47)	1.31 (0.87, 1.98)	0.58 (0.17, 1.94)	0.48
Pain in both legs below the knee	547	8	2	104	6	435	0.25 (0.04, 0.64)	0.81 (0.77, 0.84)	1.29 (0.39, 4.36)	0.93 (0.62, 1.39)	0.66
Saddle anaesthesia[2]	543	8	3	159	5	376	0.38 (0.10, 0.74)	0.70 (0.67, 0.74)	1.26 (0.51, 3.12)	0.89 (0.52, 1.52)	0.70
Change in sexual function	531	8	2	51	4	474	0.33 (0.06, 0.76)	0.90 (0.87, 0.93)	3.43 (1.07, 11.0)	0.74 (0.42, 1.30)	0.11
Perineal anaesthesia	542	8	2	67	5	468	0.29 (0.06, 0.70)	0.88 (0.84, 0.90)	2.28 (0.69, 7.52)	0.82 (0.51, 1.31)	0.22
Absence of trauma	541	8	8	342	0	191	1.00 (0.60, 1.00)	0.36 (0.32, 0.40)	1.56 (1.46, 1.66)	0.00 (0.00, NaN)	0.06

Note. P&Ns = Pins and needles, n = number, TP = true positive, FP = false positive, FN = false negative, TN = true negative, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, *p*-values calculated using Fishers exact test, NaN = cannot be calculated as a value entered contains 0, [2] = ACC red flags

Table 4.6 *Diagnostic accuracy of red flag questions for spinal infection*

Index test	Sample size (n)	Disease	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR-	P-value
Age >55 years	552	7	7	190	0	355	1.00 (0.56, 1.00)	0.65 (0.61, 0.68)	2.87 (2.56, 3.22)	0.00 (0.00, NaN)	0.001*
Age>60 years	552	7	6	148	1	397	0.88 (0.42, 0.99)	0.73 (0.69, 0.76)	3.16 (2.26, 4.40)	0.20 (0.03, 1.21)	0.005*
Insidious onset of symptoms	505	7	7	190	0	308	1.00 (0.56, 1.00)	0.62 (0.57, 0.66)	2.62 (2.34, 2.93)	0.00 (0.00, NaN)	0.001*
Worsening pain	544	7	6	351	1	186	0.86 (0.42, 0.99)	0.35 (0.31, 0.39)	1.31 (0.96, 1.79)	0.41 (0.07, 2.55)	0.43
Night pain[2]	543	7	7	332	0	204	1.00 (0.56, 1.00)	0.38 (0.34, 0.42)	1.61 (1.51, 1.73)	0.00 (0.00, NaN)	0.05*
Night sweats	545	7	4	122	3	416	0.57 (0.20, 0.88)	0.77 (0.74, 0.81)	2.55 (1.31, 4.87)	0.55 (0.24, 1.30)	0.05*
Systemically unwell	546	7	2	72	5	467	0.29 (0.05, 0.70)	0.87 (0.84, 0.89)	2.14 (0.65, 7.04)	0.83 (0.52, 1.32)	0.24
Recent infections	548	7	3	59	4	482	0.43 (0.12, 0.80)	0.89 (0.86, 0.92)	3.93 (1.62, 9.54)	0.64 (0.34, 1.22)	0.03*
Worse standing and walking	542	7	6	349	1	186	0.86 (0.42, 0.99)	0.35 (0.31, 0.39)	1.31 (0.96, 1.79)	0.41 (0.07, 2.54)	0.43
Fevers, chills or sweats[2]	546	7	4	96	3	443	0.57 (0.20, 0.88)	0.82 (0.79, 0.85)	3.21 (1.65, 6.25)	0.52 (0.22, 1.23)	0.02*
Fatigue	543	7	6	262	1	274	0.86 (0.42, 0.99)	0.51 (0.47, 0.55)	1.75 (1.28, 2.40)	0.28 (0.05, 1.72)	0.07
Smoking (current or ex-smoker)	528	7	5	213	2	308	0.71 (0.30, 0.95)	0.60 (0.55, 0.63)	1.75 (1.08, 2.82)	0.48 (0.15, 1.56)	1.00
History of immunosuppressants	540	7	2	13	5	520	0.29 (0.05, 0.70)	0.98 (0.96, 0.99)	11.7 (3.23, 42.5)	0.73 (0.46, 1.17)	0.014*
Symptoms relieved with sitting	543	7	3	60	4	476	0.43 (0.12, 0.80)	0.89 (0.86, 0.91)	3.83 (1.56, 9.30)	0.64 (0.34, 1.22)	0.04*

Note. n = number, TP = true positive, FP = false positive, FN = false negative, TN = true negative, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, *p*-values calculated using Fishers exact test, * = statistically significant, NaN = cannot be calculated as a value entered contains 0, [2] = ACC red flags

Table 4.7 *ACC red flags for any serious pathology*

Index test	Sample size (n)	Disease	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR-	P-value
Age > 50 years	552	38	33	222	5	292	0.87 (0.71, 0.95)	0.57 (0.52, 0.61)	2.01 (1.72, 2.36)	0.23 (0.10, 0.53)	<0.0001*
Significant trauma	541	38	12	179	26	324	0.32 (0.18, 0.49)	0.65 (0.60, 0.69)	0.89 (0.55, 1.43)	1.06 (0.85, 1.32)	0.76
Unexplained weight loss	487	38	4	38	34	411	0.11 (0.03, 0.26)	0.92 (0.88, 0.94)	1.24 (0.47, 3.30)	0.98 (0.88, 1.09)	0.84
History of cancer	544	38	5	58	33	448	0.13 (0.05, 0.29)	0.89 (0.85, 0.91)	1.15 (0.49, 2.69)	0.98 (0.87, 1.11)	0.91
Fever, chills, or sweats	546	38	9	91	29	417	0.24 (0.12, 0.41)	0.82 (0.79, 0.85)	1.32 (0.73, 2.41)	0.93 (0.78, 1.11)	0.49
History of IV drug use	547	38	3	17	35	492	0.08 (0.02, 0.22)	0.97 (0.95, 0.98)	2.36 (0.72, 7.71)	0.95 (0.87, 1.05)	0.31
Corticosteroid use	547	38	5	27	33	482	0.13 (0.05, 0.29)	0.95 (0.92, 0.96)	2.48 (1.01, 6.07)	0.92 (0.81, 1.03)	0.12
Night pain	543	38	24	314	14	191	0.63 (0.46, 0.78)	0.38 (0.34, 0.42)	1.02 (0.79, 1.31)	0.97 (0.64, 1.49)	1.00
Worse lying down	506	38	11	166	23	306	0.32 (0.18, 0.51)	0.65 (0.61, 0.69)	0.92 (0.56 , 1.52)	1.04 (0.83, 1.32)	0.90

Note. n = number, TP = true positive, FP = false positive, FN = false negative, TN = true negative, LR+ = positive likelihood ratio, LR, = negative likelihood ratio, *p*, values calculated using Fishers exact test, * = statistically significant.

Table 4.8 *Henschke diagnostic rule for vertebral fracture*

Index test	Sample size (n)	Disease	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR,	P, value
1 or more positive red flags	552	18	17	386	1	148	0.72 (0.68, 0.76)	0.06 (0.23, 0.29)	0.77 (0.68, 0.87)	4.99 (0.68, 36.4)	0.05*
2 or more positive red flags	552	18	9	134	9	400	0.50 (0.27, 0.73)	0.75 (0.71, 0.78)	1.99 (1.23, 3.24)	0.67 (0.42, 1.06)	0.04*
3 or more positive red flags	552	18	4	19	14	515	0.22 (0.07, 0.48)	0.96 (0.94, 0.97)	6.25 (2.37, 16.5)	0.81 (0.63, 1.03)	0.01*

Note. n = number, TP = true positive, FP = false positive, FN = false negative, TN = true negative, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, red flags included in the diagnostic rule = age >70 years, significant trauma, prolonged corticosteroid use, and female gender (Henschke et al., 2009), *p*-values calculated using Fishers exact test, * = statistically significant.

4.4 Discussion

This study provides new evidence that will improve our ability to recognise risk factors for serious pathologies in the lumbar spine. It will also improve our understanding of the diagnostic accuracy and utility of red flag questions to screen for or raise the suspicion of serious pathology as the underlying cause of low back pain.

The most common serious pathology in this study was fracture, with a prevalence of 3.3%. The only red flag question (index test) with independent diagnostic utility for screening for vertebral fracture was age greater than 35 years (sensitivity 100%, LR- 0). However, the positive likelihood ratio was poor, and specificity was only 24%. Hence, the cutoff point selected with Youden's Index (age >58 years) may be more useful clinically, with a high specificity of 71%, a sensitivity of 83%, and likelihood ratios that indicate a slight to moderate shift in probability (LR+ 2.9, LR- 0.2). A previous study by Roman et al. (2010) also investigated the relationship between age and fracture in a population of patients presenting to a secondary care spine clinic with low back pain. Roman et al. reported that age greater than 52 years may be useful for vertebral fracture screening, with a negative likelihood ratio of 0.14 indicating a moderate change in probability. The accuracy of this index test was similar to our diagnostic accuracy for age greater than 58 years.

Our study also demonstrated that only two index tests had positive likelihood ratios indicating that their presence was associated with a moderate to conclusive shift in the probability of the presence of a vertebral fracture. Participants with concomitant HIV or AIDS or a history of IV drug use had an increased risk of vertebral fracture, with 97-100% specificity, and positive likelihood ratios between 5.19 and 29.5. However, the only index test with a positive likelihood ratio greater than 10 that could be considered to have utility as a risk factor was a history of HIV/AIDS (LR+29.5). It appears that no published study has investigated concomitant HIV/AIDS, or IV drug use. Hence, our findings cannot be compared with previous research. Although osteoporosis has been recommended by multiple clinical guidelines as a risk factor for vertebral fracture, our study was the first to investigate the diagnostic accuracy of a history of osteoporosis in a low back pain population. We have demonstrated that the presence of osteoporosis increases the likelihood of a fracture being present by a moderate amount (LR+4.9).

Although a history of corticosteroid use has been reported as being a risk factor for fracture (Henschke et al., 2009), this finding was not supported by our own findings. Despite a high specificity (94%), the likelihood ratios associated with this risk factor did not demonstrate a significant change in probability (LR+1.9, LR-0.94). Similarly, previous studies have reported that a history of significant trauma is associated with fracture (Henschke et al., 2009; Scavone et al., 1981). Our findings do not support this conclusion, as there was no statistically significant association between significant trauma and fracture, and the diagnostic accuracy was poor with sensitivity of 44%, specificity of 65% and inconclusive likelihood ratios (LR+1.3, LR-0.9). These differences may be partially due to variances in study setting, as both studies that reported conclusive likelihood ratios for significant trauma were set in primary care. Conversely, two studies (Gibson & Zoltie, 1992; Patrick et al., 1983) set in tertiary care also reported poor diagnostic accuracy for trauma.

Our study was unable to support the use of the diagnostic rule proposed by Henschke and colleagues (Henschke et al., 2009). The diagnostic rule contained four red flag questions: age greater than 70 years, significant trauma, a history of prolonged corticosteroid use, and female gender. Five precent of our sample had no red flags and 44% of our sample had only one positive red flag. The specificity of one positive finding in our cohort was 6%. At this value, it seems inappropriate that further investigation of patients with one or more positive findings is warranted. Henschke et al. recommend that patients with three or more positive findings should be referred for further investigations. Implementation of this recommendation in our study would have missed 78% of all fractures. The positive likelihood ratio for three or more positive red flags was also much lower in our study (LR+ 6.25 [95% CI 2.37-16.5]), than in Henschke's study (LR+ 218.3 [95% CI 45.6-953.8]). Additionally, both positive likelihood ratios lacked precision, with wide confidence intervals. In our study, age greater than 70 years was the only index test, of the four included in the Henschke diagnostic rule, that had a positive likelihood ratio that indicated that this finding had any diagnostic utility as a stand-alone test. The remaining three index tests displayed poor diagnostic accuracy in our study. These differences in diagnostic accuracy between studies are likely to have been influenced by study setting (primary versus secondary/tertiary care) and choice of reference standard (long-term follow-up versus MRI). Our findings suggest that the Henschke diagnostic rule is not applicable to our population, therefore further research would be required to validate their rule before to

could be recommended for use in clinical practice.

Our results can also be compared to a recently published study (de Schepper et al., 2016) that investigated the prevalence of spinal pathologies in patients presenting for lumbar MRI referred from primary care. The study by de Schepper et al. (2016) investigated three red flag questions to screen for fracture (trauma, age over 70 years, and female gender). Although they did not report diagnostic accuracy statistics, they published sufficient data to allow such calculation. De Schepper et al. (2016) reported that use of these three red flags for fracture would have missed 4 fractures (24%), and 7 (41%) patients with fracture only had one positive red flag finding in their study. Their findings were similar to our study findings in that female gender had poor diagnostic accuracy. They reported that trauma had poor sensitivity and a positive likelihood ratio of 2.94, indicating a slight increase in probability. Our study demonstrated an even lower positive likelihood ratio of 1.27. In the study by de Schepper et al. (2016), age greater than 70 years also had poor sensitivity but a positive likelihood ratio of 5.68, which was similar to the 5.10 found in our study, both indicating a moderate increase in the probability of a positive diagnosis. From these results it could be concluded that fracture risk increases with older age and a history of trauma may slightly increases the probability of fracture.

Malignancy had a low prevalence of 0.9% in our study. We established that the index tests age (>42 years), and worsening pain had diagnostic utility for screening, with 100% sensitivity and negative likelihood ratios of zero. Similarly, other authors have found that age greater than 50 is a useful test to screen for cancer with negative likelihood ratios between zero and 0.11 (Fernbach et al., 1976; Jacobson, 1997). Another study reported that age greater than 44 had a negative likelihood ratio of 0.2, indicating a moderate shift in probability (van den Bosch et al., 2004). To the author's knowledge no other study has investigated worsening pain. The only index test that could be considered a risk factor for malignancy was corticosteroid use with 95% specificity and a positive likelihood ratio of 7.23 (95% CI 2.34-22.34). At first glance, the association between corticosteroid use and cancer may have poor face validity. However, chronic inflammation is now recognised as a critical component in tumour progression. Recent research has established that malignancy can arise from site of infection and chronic inflammation (Coussens & Werb, 2002). Therefore, a history of corticosteroid use may be useful to signal that a patient has a history of chronic

inflammation, and hence an increased risk of cancer. However, this has not been investigated by any other study, and more research would be required for validation. Other red flag questions that suggested a slight change in probability were: unexplained weight loss, history of cancer, and constant pain. These index tests may have increased diagnostic accuracy when used in combinations. However, more research is required to investigate combinations or clusters of tests. A history of cancer has previously been regarded as a significant risk factor for spinal malignancy (Deyo & Diehl, 1988). However, this study found a specificity of 89% and a positive likelihood ratio of 3.49, and was therefore unable to conclusively support this.

The study by de Schepper and colleagues (2016) also investigated five red flags for malignancy (age at onset over 50 years, continuous back pain, back pain at night, history of malignancy, and unexplained weight loss). All patients with malignancy had at least one positive red flag. They found that of the 5 cases of malignancy, 2 had one positive red flag and 3 had two positive red flags. No patient with malignancy had continuous back pain or unexplained weight loss, and only one had a history of cancer. However, due to the low number of patients with a history of cancer, the specificity was 96% and the positive likelihood ratio indicated a moderate increase in probability (LR+ 6.46). This was higher than our study results (LR+ 3.49). In the study by de Schepper et al. (2016), all patients with malignancy had an onset of pain after age 50, which therefore had a sensitivity of 100%, compared to 80% sensitivity for age greater than 50 in our study. Two of the five patients with malignancy had a history of night pain. However, the diagnostic accuracy of night pain was poor and the false positive rate was high as 53% of the whole study sample complained of night pain. Similarly, our study found high false positive rates and poor specificity with night pain (Specificity 38%). With regard to diagnostic utility, they demonstrated that age older age was useful for screening, which was similar to our study findings. They also demonstrated that a history of cancer could be considered as a moderate risk factor, with positive likelihood ratios of 6.46 (de Schepper et al., 2016). However, our study was unable to support this conclusion.

Cauda equina syndrome was also rare, with a prevalence of 1.45%. The only index test with an informative likelihood ratio, which conclusively changed the probability of a diagnosis, was absence of trauma. Absence of trauma had utility for screening with 100% sensitivity and a negative likelihood ratio of zero. Other red flags

that slightly changed the probability, but could be useful in combinations, were bladder or bowel incontinence, urinary overflow, change in sexual function, perineal anaesthesia, worsening pain, unsteady gait, and worse standing and walking. Increased pain with standing and walking could also be a useful question to assist in differentiation between complicated and uncomplicated disc prolapses, as the majority of patients with an uncomplicated disc prolapse feel better standing and walking (Otéro & Bonnet, 2014). However, further research would be required to investigate this. Due to the lack of previous research in the area, these results were unable to be compared to previous findings in the literature.

This study found a relatively high prevalence of spinal infections at 1.27%, compared to prevalence between zero and 0.2% found by previous primary care studies (Henschke et al., 2009; Khoo et al., 2003; van den Bosch et al., 2004). Red flag questions that had diagnostic utility for screening were age greater than 55 years, insidious onset of pain, and the presence of night pain that wakes you from sleep. This finding was supported by another study by van den Bosch and colleagues (2004) that also reported 100% sensitivity with age greater than 54 years. A history of immunosuppressant use had diagnostic utility as a risk factor with a specificity of 98% and a positive likelihood ratio of 11.71 (95% CI 3.23-42.49). Other red flags that slightly changed the probability but may be useful in combinations for screening were worsening pain, fatigue, worse standing and walking, and history of smoking. Red flag questions that may be useful in combinations as risk factors were fevers, sweats, or chills, recent infection, night sweats, and systemically unwell. Davis and colleagues (2004) investigated the presence of one or more of the following risk factors: IV drug use, immunocompromised, recent spine procedure, distant site of infection, diabetes, indwelling catheter, recent spine fracture, chronic renal failure, cancer or alcohol abuse. Davis et al. found that the sensitivity with one or more risk factors present was 98% with a negative likelihood ratio of 0.02. The specificity was also reasonably high at 79% with a positive likelihood ratio of 4.6. However, the diagnostic accuracy is likely to be higher than ours as the study by Davis et al. (2004) was conducted on a group of patients with suspected spinal infection. In our study no IV drug user had infection, but as this is a known risk factor (Della-Giustina, 2015), further research would be required to investigate this before refuting its use.

4.4.1 Limitations

There were some limitations of this study. Firstly, spinal diagnosis was based on MRI reports. Ideally MRI scans would have been read and reported by an independent experienced radiologist using a standardised classification system for diagnosis. Unfortunately this study did not have the resources to fund this. However, diagnosis based on MRI reports does reflect standard clinical practice as closely as possible. A random selection of 5% of the MRI reports were double-read to ensure accurate diagnostic coding. However, ideally all reports would have been double-read. Lastly, MRI referral criterion was not restricted, as this study was intended to provide a snapshot of actual clinical practice. Therefore, referral criteria may have differed between public and private practice, and more research would be required to investigate this.

4.4.2 Conclusion

The current study provides important new information that should enhance the understanding of the likelihood of a patient presenting to secondary or tertiary care with a serious pathology as the underlying cause of their low back pain. Our study has provided original evidence regarding specific age and ethnic groups that may be more at risk of developing a serious pathology in the lumbar spine. This is also the first study to investigate the prevalence and incidence within the New Zealand population. To ensure the accuracy of our results, our study employed the best available reference standard with the highest level of precision for the recognition of serious pathologies in the lumbar spine.

Our study has also provided beneficial new diagnostic information, especially for rare pathologies such as spinal infection and cauda equina syndrome. To the author's knowledge, no other study has investigated the diagnostic accuracy of red flag questions to screen for cauda equina syndrome or spinal infection in either secondary or tertiary care. Also, whilst a small number of studies have investigated vertebral fractures and malignancy in low back pain populations in other countries, very few studies have been designed in a manner that allows any conclusive recommendations to be made regarding the diagnostic accuracy of red flag questions to screen for these pathologies.

Based on the results of our study, we recommend that most red flag questions cannot be used as stand-alone tests for the identification of serious pathologies in the lumbar spine. However, we can recommend that some red flag questions have sufficient

diagnostic accuracy for use as screening tests or as risk factors. For vertebral fractures, age greater than 35 years has sufficient diagnostic accuracy for use as a screening test, and HIV/AIDS was a risk factor for vertebral fracture. For malignancy, age greater than 42 years and worsening pain displayed diagnostic utility for screening. The absence of trauma had utility as a screening test for cauda equina syndrome. For spinal infection, age greater than 55 years, insidious onset, and night pain that wakes you from sleep all displayed utility as screening questions. Lastly, a history of immunosuppressant use had utility as a risk factor for spinal infection. Further research is required to investigate combinations of findings, which may improve both diagnostic accuracy and the clinical utility of red flag questions (Bossuyt et al., 2012).

Conflicts of interest. There were no conflicts of interest. There was a low risk of harm, as this was not an interventional study. The questions were asked as part of the patients' scheduled appointments, therefore no reimbursement was required, and no coercion or deceit could occur. Radiology staff rather than specialists asked patients if they would like to participate, in order to reduce the effect of any unforeseen power imbalance. This study did not involve children and there were no adverse outcomes.

Chapter 5 Summary, Key Findings and Conclusions

5.1 Summary

The identification of serious pathologies is an important topic that is relevant to all parties involved in the management of low back pain. Serious pathologies are rare, but delayed diagnosis can lead to dire outcomes such as progression of disease, systemic illness, irreversible neurological changes including incontinence and sexual dysfunction, pathological fracture, spinal deformity, and ultimately, mortality. At present, studies (Darouiche, 2006; Grigoryan et al., 2003; Patel et al., 2014) have shown that up to 75% of serious pathologies may be missed on initial clinical assessment. Research has also shown that these pathologies are not only difficult to diagnose in primary care, but that doctors working in secondary and tertiary care settings frequently fail to make an early diagnosis (Davis et al., 2004). Consequently, despite warnings regarding the potential harms associated with diagnostic imaging, clinicians are becoming increasingly reliant on the use of imaging for differential diagnosis (Flynn et al., 2011) and both plain radiography and MRI are currently overused (Chou et al., 2012). Red flag questions were implemented in 1994 by the clinical standards advisory board (Higginson, 1994) in an attempt to increase awareness and improve screening for serious pathologies in the lumbar spine. However, there is currently a lack of research to support or refute their use.

This thesis contributes vital new evidence related to the prevalence, incidence and the diagnostic accuracy of red flag questions. In particular the systematic review of the literature undertaken within this research established that there was a dearth of evidence related to cauda equina syndrome and spinal infection, with no previous study investigating prevalence, incidence or diagnostic accuracy in either secondary or tertiary care. Overall, there was a lack of good quality evidence with regard to the prevalence, incidence or diagnostic accuracy of any red flag questions to screen for any serious pathology in the lumbar spine. Therefore, this research contributes significantly to the current evidence base.

5.1.1 Key findings

Serious pathologies. This research has established that the risk of developing a serious pathology increased with age to a peak incidence of 2 per 1,000 person-years in people 74 years and over. There was no significant difference in serious pathology risk

between genders. European and Pacific populations had the highest overall risk of serious pathology. Asians had the lowest risk of developing serious pathology. Serious pathologies were significantly more common in tertiary care than in secondary care, with 15% of patients having a serious pathology on MRI in tertiary care compared to 3% in secondary care.

Vertebral fracture. The prevalence and incidence of vertebral fractures increased with age and there was no significant difference between genders. The incidence of fractures was highest amongst Europeans. Older age was the only red flag that had individual diagnostic utility for screening. All participants with fractures were over 35 years of age and all participants between 35 and 50 years of age had a history of significant trauma. A history of HIV or AIDS was a significant risk factor with specificity of 100% and a positive likelihood ratio of 30. A history of intravenous drug use, or concomitant osteoporosis could also be considered risk factors with good specificity (95-97%), but likelihood ratios indicating a moderate increase in the probability of a positive diagnosis (LR+ 4.9-5.2). These risk factors may be more useful in combination, however further research is required to investigate the diagnostic accuracy of combinations of red flag questions. This research was unable to support the use of the combination of red flag questions proposed as a diagnostic rule by Henschke and colleagues (2009) to screen for vertebral fractures.

Malignancy. The prevalence of malignancy was 15 times higher in tertiary care, than in secondary care. Māori had the highest risk of malignancy. Age greater than 42 years and the question “has your pain been worsening over time” both had diagnostic utility for screening with 100% sensitivity and negative likelihood ratios of zero. A history of cancer, unexplained weight loss and a history of corticosteroid use could all be considered risk factors (specificity 89-95%). However, in isolation they were not signficant risk factors as the likelihood ratios shifted the probability of a positive diagnosis a slight to moderate amount (LR+3.5-7.2).

Cauda equina syndrome. Cauda equina syndrome was the rarest of all the serious pathologies investigated with a prevalence of 1%. Cauda equina syndrome is difficult to diagnose clinically due to variable signs and symptoms. The absence of trauma may be useful as a screening test with 100% sensitivity and a negative likelihood ratio of zero. Other red flag questions that had lower diagnostic accuracy but may be useful in combinations for screening were: worsening pain, worse with standing and walking, and

unsteady gait (sensitivity 86-88%, LR- 0.4). Red flag questions that raised the probability of a positive diagnosis a slight amount, but may be useful in combinations as risk factors were: bowel or bladder incontinence (specificity 95%, 88%, LR+ 4.9, 2.3, respectively), urinary overflow (specificity 94%, LR+ 4.9), change in sexual function (specificity 88%, LR+ 2.2) and perineal anaesthesia (specifity 88%, LR+ 2.3). It is important to note that all potential risk factors had low sensitivity (25-33%), therefore cauda equina syndrome cannot be ruled out in their absence.

Spinal infection. The incidence of infection increased with age and was higher in males and in Pacific Islanders. The prevalence of spinal infection was 29 times higher in tertiary care than in secondary care. Red flag questions that had diagnostic utility for screening were age greater than 55 years, insidious onset of pain, and night pain which all had 100% sensitivity and negative likelihood ratios of zero. A history of immunosuppressant use was a significant risk factor with 98% specificity and a positive likelihood ratio of 12. Other red flags with weaker diagnostic accuracy that may be useful for screening were: worsening pain, worse standing and walking, and fatigue (sensitivity 86%, LR- 0.3-0.4). Red flag questions that may be useful as risk factors in combinations were: night sweats, systemically unwell, fevers, sweats, or chills, and recent infections (specifity 77-89%, LR+ 2.1-3.9).

5.1.2 Directions for future research

In clinical practice, clinicians would rarely rely on one test finding to make a diagnostic or management decision. Instead, it is more common to consider combinations of such findings to make a clinical diagnosis. The current study has provided information on the diagnostic accuracy and utility of individual questions. However, further research that investigates the diagnostic accuracy of combinations of red flag questions is warranted. Methods such as logistic regression, support vector machine, and classification trees should be explored to determine which method provides the best result where sensitivity and specificity are optimised. Optimal index based on indicators will generally need to be formed using different coefficients for each indicator. Such research may lead to the development of pathology-specific diagnostic tools or questionnaires to improve the early identification of serious pathologies and guide clinical management. Utilisation of red flag questions in this manner may also be valuable to support selective diagnostic imaging, although further research is required before any guidelines could be proposed.

Implementation of a computer-based diagnostic tool could allow a calculation of the probability of a patient having a serious pathology. This would not only improve early diagnosis and management, but would reduce healthcare expenditure by reducing inappropriate specialist referral and unnecessary use of diagnostic imaging. For patients with a positive result the tool could then offer advice with regard to indications for further investigations such as ESR, CRP, white blood cell count, plain radiographs, or referral for consideration of MRI / specialist review. Conduct of a study of this calibre would require funding and support from district health boards and radiology departments, as it would need to be integrated into standard practice to capture the sample size required for accurate precision. The system could then be systematically re-evaluated to update diagnostic accuracy and pre-test probability as more information was added to the database. Once this system was running smoothly at one institute, it could then be trailed in another region and the probability of disease could be adjusted to the specific population. Information gathered from this diagnostic tool could also be used to track incidence and prevalence in specific populations, and could be used by healthcare funders to plan provision of services.

The author had a key part in developing an online clinical pathway for acute lower back pain, which has been implemented for use across Auckland for general practitioners, emergency departments, and St John paramedic services. This clinical pathway includes advice on red flag screening. However, there may be scope to include a diagnostic algorithm within the pathway.

5.1.3 Clinical implications and conclusions

Our study has established the clinical prevalence of serious pathologies in the lumbar spine amongst patients presenting to secondary and tertiary care settings. This has allowed a greater understanding of pre-test probability of a patient presenting in these settings with underlying serious pathology. Our study also investigated incidence, which has aided in a better understanding of the aetiology of serious pathologies, including the age groups, gender, and ethnicities that are more likely to be at risk.

Our study has established that the majority of red flag questions when used in isolation are uninformative and lack precision. However, a small selection of red flag questions may be useful for screening or as risk factors to raise the suspicion of a diagnosis. Hence, on the basis of our findings, we advise cautious interpretation of red flag findings in combination with sound clinical reasoning. Clinicians should be aware

that most risk factors for serious pathologies hold poor sensitivity and therefore serious pathology cannot necessarily be ruled out in their absence. At this stage the presence or absence of red flag findings cannot be relied on to decide whether or not diagnostic imaging or further investigation is required. Multiple red flag findings are likely to be more useful diagnostically and further investigation in this area will be of great benefit to all parties involved in the management of low back pain.

References

Accident Compensation Corporation. (1999). *New Zealand acute low back pain guide*. Wellington, New Zealand: Author.

Accident Compensation Corporation. (2015). *Annual report*. Retrieved from http://www.acc.co.nz/PRD_EXT_CSMP/groups/external_communications/documents/reports_results/2015annualreport.pdf

Ahn, U. M., Ahn, N. U., Buchowski, J. M., Garrett, E. S., Sieber, A. N., & Kostuik, J. P. (2000). Cauda equina syndrome secondary to lumbar disc herniation. *Spine, 25*(12), 1515-1522. doi:10.1097/00007632-200006150-00010

Aho, A. J., Auranen, A., & Pesonen, K. (1969). Analysis of cauda equina symptoms in patients with lumbar disc prolapse. Preoperative and follow-up clinical and cystometric studies. *Acta Chirurgica Scandinavica, 135*(5), 413-420.

Algra, P. R. (2010). [MRI diagnostics should be available to general practice] [Abstract]. *Nederlands Tijdschrift voor Geneeskunde, 154*, A2851. Retrieved from http://www.ntvg.nl

Andersson, G. B. J. (1999). Epidemiological features of chronic low-back pain. *Lancet, 354*(9178), 581-585. doi:10.1016/S0140-6736(99)01312-4

Bach, S. M., & Holten, K. B. (2009). Guideline update: What's the best approach to acute low back pain? *The Journal of Family Practice, 58*(12). Retrieved from http://www.jfponline.com

Bell, D. A., Collie, D., & Statham, P. F. (2007). Cauda equina syndrome - What is the correlation between clinical assessment and MRI scanning? *British Journal of Neurosurgery, 21*(2), 201-203. doi:10.1080/02688690701317144

Bentley, T., Macky, K., & Edwards, J. (2006). Injuries to New Zealanders participating in adventure tourism and adventure sports: An analysis of Accident Compensation Corporation (ACC) claims. *New Zealand Medical Journal, 119*(1247), U2359. Retrieved from http://www.nzma.org.nz/journal/119-1247/2359/

Beronius, M., Bergman, B., & Andersson, R. (2001). Vertebral osteomyelitis in Göteborg, Sweden: A retrospective study of patients during 1990-95. *Scandinavian Journal of Infectious Diseases, 33*(7), 527-532. doi:10.1080/00365540110026566

Bigos, S. J., Bowyer, R. O. R., Braen, G. R., Brown, K. C., Deyo, R. A., Haldeman, S., . . . Triano, J. J. (1995). Acute low back problems in adults: Assessment and

treatment. *American Family Physician, 51*(2), 469-484. Retrieved from http://www.aafp.org.ezproxy.aut.ac.nz/journals/afp.html

Black, D. M., Arden, N. K., Palermo, L., Pearson, J., & Cummings, S. R. (1999). Prevalent vertebral deformities predict hip fractures and new vertebral deformities but not wrist fractures. *Journal of Bone and Mineral Research, 14*(5), 821-828. doi:10.1359/jbmr.1999.14.5.821

Boden, S. D., & Swanson, A. L. (1998). An assessment of the early management of spine problems and appropriateness of diagnostic imaging utilization. *Physical Medicine and Rehabilitation Clinics of North America, 9*(2), 411-417. Retrieved from http://www.pmr.theclinics.com/

Borman, B. (n.d.). *Health needs assessment Counties Manukau District Health Board.* Retrieved from https://www.health.govt.nz/system/files/documents/pages/hna-counties-manukau-dhb.pdf

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., . . . Cohen, J. F. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Britsh Medical Journal, 351*, h5527. doi:10.1136/bmj.h5527

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Croatian Medical Journal, 44*(5), 639-650. Retrieved from http://www.cmj.hr

Bossuyt, P. M., Reitsma, J. B., Linnet, K., & Moons, K. G. M. (2012). Beyond diagnostic accuracy: The clinical utility of diagnostic tests. *Clinical Chemistry, 58*(12), 1636-1643. doi:10.1373/clinchem.2012.182576

Bouxsein, M. L., Melton, L. J., 3rd, Riggs, B. L., Muller, J., Atkinson, E. J., Oberg, A. L., . . . Khosla, S. (2006). Age- and sex-specific differences in the factor of risk for vertebral fracture: a population-based study using QCT. *Journal of Bone and Mineral Research, 21*(9), 1475-1482.

Brown, P., McNeill, R., Radwan, E., & Willingale, J. (2007). *The burden of osteoporosis in New Zealand: 2007-2020.* Retrieved from osteoporosis.org.nz

Buchner, M., & Schiltenwolf, M. (2002). Cauda equina syndrome caused by intervertebral lumbar disk prolapse: Mid-term results of 22 patients and literature review. *Orthopedics, 25*(7), 727-731. Retrieved from http://www.healio.com/orthopedics/journals/ortho

Carey, T. S., & Garrett, J. (1996). Patterns of ordering diagnostic tests for patients with acute low back pain. *Annals of Internal Medicine, 125*(10), 807-814. Retrieved from http://www.annals.org/

Carragee, E. J. (1997). Pyogenic vertebral osteomyelitis. *Journal of Bone and Joint Surgery, 79*(6), 874-880. Retrieved from http://www.jbjs.org

Casazza, B. A. (2012). Diagnosis and treatment of acute low back pain. *American Family Physician, 85*(4), 343-350. Retrieved from http://www.aafp.org/online/en/home/publications/journals/afp.html

Chou, R., Deyo, R. A., & Jarvik, J. G. (2012). Appropriate use of lumbar imaging for evaluation of low back pain. *Radiologic Clinics of North America, 50*(4), 569-585. doi:10.1016/j.rcl.2012.04.005

Chou, R., Fu, R., Carrino, J. A., & Deyo, R. A. (2009). Imaging strategies for low-back pain: systematic review and meta-analysis. *The Lancet, 373*(9662), 463-472. doi:10.1016/S0140-6736(09)60172-0

Chou, R., Qaseem, A., Owens, D. K., & Shekelle, P. (2011). Diagnostic imaging for low back pain: Advice for high-value health care from the American college of physicians. *Annals of Internal Medicine, 154*(3), 181-189. Retrieved from http://www.annals.org/

Chou, R., Qaseem, A., Snow, V., Casey, D., Cross, T. J., Jr., Shekelle, P., & 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), 478-491. Retrieved from http://www.annals.org/

Choudhury, A. R., & Taylor, J. C. (1980). Cauda equina syndrome in lumbar disc disease. *Acta Orthopaedica Scandinavica, 51*(3), 493-499. Retrieved from http://www.tandfonline.com.ezproxy.aut.ac.nz/toc/iort20/current

Çigdem Ataman, H., Kader, A., Cemal, B., Zeliha Koçak, T., Yasemin Kurtoglu, G., Sami, K., & Ali Pekcan, D. (2013). Evaluation of epidemiological, clinical and laboratory characteristics of patients with spondylodiscitis. *Flora: The Journal of Infectious Diseases & Clinical Microbiology, 18*(1), 28-33. Retrieved from http://www.journals.elsevier.com/flora/

Clopper, C., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the bionomial. *Biometrika, 26*(4), 404-413. doi:10.1093/biomet/26.4.404

Collert, S. (1977). Osteomyelitis of the spine. *Acta Orthopaedica, 48*(3), 283-290. doi:10.3109/17453677708988770

Cook, C., & Hegedus, E. (2012). *Orthopedic physical examination tests: An evidence-based approach* (2nd ed.). Upper Saddle River, NJ: Pearson.

Cook, C., Ross, M. D., Isaacs, R., & Hegedus, E. (2012). Investigation of nonmechanical findings during spinal movement screening for identifying and/or ruling out metastatic cancer. *Pain Practice, 12*(6), 426-433. doi:10.1111/j.1533-2500.2011.00519.x

Cooper, C., O'Neill, T., & Silman, A. (1993). The epidemiology of vertebral fractures. *Bone, 14*(Suppl. 1), 89-97. doi:10.1016/8756-3282(93)90358-H

Coscia, M., Leipzig, T., & Cooper, D. (1994). Acute cauda equina syndrome: Diagnostic advantage of MRI. *Spine, 19*(4), 475-478. Retrieved from http://journals.lww.com/spinejournal/pages/default.aspx

Cottle, L., & Riordan, T. (2008). Infectious spondylodiscitis. *Journal of Infection, 56*(6), 401-412. doi:10.1016/j.jinf.2008.02.005

Coussens, L. M., & Werb, Z. (2002). Inflammation and cancer. *Nature, 420*(6917), 860-867. doi:10.1038/nature01322

D'Agostino, C., Scorzolini, L., Massetti, A. P., Carnevalini, M., d'Ettorre, G., Venditti, M., . . . Orsi, G. B. (2010). A seven-year prospective study on spondylodiscitis: epidemiological and microbiological features. *Infection, 38*(2), 102-107. doi:10.1007/s15010-009-9340-8

Dagenais, S., Galloway, E. K., & Roffey, D. M. (2014). A systematic review of diagnostic imaging use for low back pain in the United States. *Spine Journal, 14*(6), 1036-1048. doi:10.1016/j.spinee.2013.10.031

Dandy, W. E. (1989). Loose cartilage from intervertebral disk simulating tumor of the spinal cord. *Clinical Orthopaedics and Related Research, 238*(1), 4-8. Retrieved from http://www.springer.com/medicine/orthopedics/journal/11999

Dang, L., Liu, X., Dang, G., Jiang, L., Wei, F., Yu, M., . . . Liu, Z. (2015). Primary tumors of the spine: a review of clinical features in 438 patients. *Journal of Neuro-Oncology, 121*(3), 513-520. doi:10.1007/s11060-014-1650-8

Darouiche, R. O. (2006). Spinal epidural abscess. *New England Journal of Medicine, 355*(19), 2012-2020. doi:10.1056/NEJMra055111

Davis, D. P., Wold, R. M., Patel, R. J., Tran, A. J., Tokhi, R. N., Chan, T. C., & Vilke, G. M. (2004). The clinical presentation and impact of diagnostic delays on emergency department patients with spinal epidural abscess. *The Journal Of*

Emergency Medicine, 26(3), 285-291. Retrieved from http://www.jem-journal.com/

de Schepper, E. I. T., Koes, B. W., Veldhuizen, E. F. H., Oei, E. H. G., Bierma-Zeinstra, S. M. A., & Luijsterburg, P. A. J. (2016). Prevalence of spinal pathology in patients presenting for lumbar MRI as referred from general practice. *Family Practice, 33*(1), 51-56 56p. doi:10.1093/fampra/cmv097

Deeks, J. J. (2001). Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal, 323*(7305), 157-162. Retrieved from http://www.bmj.com/

Deeks, J. J., Wisniewski, S., & Davenport, C. (2013). Guide to the contents of a Cochrane diagnostic test accuracy protocol. In J. J. Deeks, P. M. Bossuyt, & C. A. Gatsonis (Eds.), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0.* (pp. i, 3-15). Retrieved from http://methods.cochrane.org/sdt/sites/methods.cochrane.org.sdt/files/uploads/Ch04_Sep2013.pdf

Deleskog, L., Laursen, N. Ø., Nielsen, B. R., & Schwarz, P. (in press). Vertebral fracture assessment by DXA is inferior to X-ray in clinical severe osteoporosis. *Osteoporosis International*. doi:10.1007/s00198-016-3532-8

Della-Giustina, D. (2015). Evaluation and treatment of acute back pain in the emergency department. *Emergency Medicine Clinics of North America, 33*(2), 311-326. doi:10.1016/j.emc.2014.12.005

Denis-Delpierre, N., Merrien, D., Billaud, E., Besnier, J. M., Duhamel, E., Hutin, P., . . . Raffi, F. (1998). Extrapulmonary tuberculosis in western France. Retrospective analysis of 217 cases in the GERICCO study (1991-1993). *Presse Medicale, 27*(8), 341-346. Retrieved from http://www.elsevier-masson.fr/la-presse-medicale-0755-4982.html

Desai, S. S. (1994). Early diagnosis of spinal tuberculosis by MRI. *Journal of Bone and Joint Surgery, 76*(6), 863-869. Retrieved from http://www.jbjs.org

Deyo, R. A. (1994). Magnetic resonance imaging of the lumbar spine - Terrific test or tar baby? *New England Journal of Medicine, 331*(2), 115-116. doi:10.1056/NEJM199407143310209

Deyo, R. A., & Diehl, A. K. (1986). Lumbar spine films in primary care - Current use and effects of selective ordering criteria. *Journal of General Internal Medicine, 1*(1), 20-25. doi:10.1007/bf02596320

Deyo, R. A., & Diehl, A. K. (1988). Cancer as a cause of back pain: Frequency, clinical presentation, and diagnostic strategies. *Journal of General Internal Medicine, 3*(3), 230-238. Retrieved from http://www.springer.com/medicine/internal/journal/11606

Deyo, R. A., Jarvik, J. G., & Chou, R. (2014). Low back pain in primary care. *British Medical Journal, 349*, g4266. doi:10.1136/bmj.g4266

Deyo, R. A., & Mirza, S. K. (2006). Trends and variations in the use of spine surgery. *Clinical Orthopaedics and Related Research, 443*, 139-146. doi:10.1097/01.blo.0000198726.62514.75

Deyo, R. A., & Mirza, S. K. (2009). The case for restraint in spinal surgery: Does quality management have a role to play? *European Spine Journal, 18*(Suppl. 3), S331-S337. doi:10.1007/s00586-009-0908-x

Deyo, R. A., Mirza, S. K., Turner, J. A., & Martin, B. I. (2009). Overtreating chronic back pain: Time to back off? *Journal of the American Board of Family Medicine, 22*(1), 62-68. doi:10.3122/jabfm.2009.01.080102

Deyo, R. A., Rainville, J., & Kent, D. L. (1992). What can the history and physical examination tell us about low back pain? *Journal of the American Medical Association, 268*(6), 760-765. doi:10.1001/jama.268.6.760

Deyo, R. A., & Weinstein, J. N. (2001). Low back pain. *New England Journal of Medicine, 344*(5), 363-370. doi:10.1056/NEJM200102013440508

Diehn, F. E. (2012). Imaging of spine infection. *Radiologic Clinics of North America, 50*(4), 777-798. doi:10.1016/j.rcl.2012.04.001

Dinning, T. A. R., & Schaeffer, H. R. (1993). Discogenic compression of the cauda equina: A surgical emergency. *Australian and New Zealand Journal of Surgery, 63*(12), 927-934. Retrieved from http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1445-2197

Dolberg, O. T., Schlaeffer, F., Greene, V. W., & Alkan, M. L. (1991). Extrapulmonary tuberculosis in an immigrant society: Clinical and demographic aspects of 92 cases. *Reviews of Infectious Diseases, 13*(1), 177-179. Retrieved from http://cid.oxfordjournals.org/

Domen, P. M., Hofman, P. A., Van Santbrink, H., & Weber, W. E. J. (2009). Predictive value of clinical characteristics in patients with suspected cauda equina syndrome. *European Journal of Neurology, 16*(3), 416-419. doi:10.1111/j.1468-1331.2008.02510.x

Donner-Banzhoff, N., Roth, T., Sönnichsen, A. C., Luckmann, J., Leonhardt, C., Chenot, J. F., . . . Bauma, E. (2006). Evaluating the accuracy of a simple heuristic to identify serious causes of low back pain. *Family Practice, 23*(6), 682-686. doi:10.1093/fampra/cml049

Downie, A., Williams, C. M., Henschke, N., Hancock, M. J., Ostelo, R. W. J. G., De Vet, H. C. W., . . . Maher, C. G. (2013). Red flags to screen for malignancy and fracture in patients with low back pain: Systematic review. *British Medical Journal, 347,* f7095. doi:10.1136/bmj.f7095

Dreghorn, C. R., Newman, R. J., Hardy, G. J., & Dickson, R. A. (1990). Primary tumors of the axial skeleton experience of the leeds regional bone tumor registry. *Spine, 15*(2), 137-140. Retrieved from http://journals.lww.com/spinejournal/pages/default.aspx

Dunsmuir, R. (2004). (iii) Prolapsed intervertebral discs. *Current Orthopaedics, 18*(6), 434-440. doi:10.1016/j.cuor.2005.01.006

Edmond, S. L., Kiel, D. P., Samelson, E. J., Kelly-Hayes, M., & Felson, D. T. (2005). Vertebral deformity, back symptoms, and functional limitations among older women: The Framingham Study. *Osteoporosis International, 16*(9), 1086-1095. doi:10.1007/s00198-004-1815-y

Espersen, F., Frimodt-Moller, N., Rosdahl, V. T., Skinhoj, P., & Bentzon, M. W. (1991). Changing pattern of bone and joint infections due to Staphylococcus aureus: Study of cases of bacteremia in Denmark, 1959-1988. *Reviews of Infectious Diseases, 13*(3), 347-358. Retrieved from http://cid.oxfordjournals.org/

Eusebi, P. (2013). Diagnostic accuracy measures. *Cerebrovascular Diseases, 36*(4), 267-272. doi:10.1159/000353863

Fernandez-Ulloa, M., Vasavada, P. J., Hanslits, M. L., Volarich, D. T., & Elgazzar, A. H. (1985). Diagnosis of vertebral osteomyelitis: Clinical, radiological and scintigraphic features. *Orthopedics, 8*(9), 1144-1150. Retrieved from http://www.healio.com/orthopedics/journals/ortho

Fernbach, J. C., Langer, F., & Gross, A. E. (1976). The significance of low back pain in older adults. *Canadian Medical Association Journal, 115*(9), 898-900. Retrieved from http://www.cmaj.ca

Fletcher, R., Fletcher, S., Fletcher, G. (2014). *Clinical epidemiology: The essentials* (5th ed.). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Flynn, T. W., Smith, B., & Chou, R. (2011). Appropriate use of diagnostic imaging in low back pain: A reminder that unnecessary imaging may do as much harm as good. *Journal of Orthopaedic and Sports Physical Therapy*, *41*(11), 838-846. doi:10.2519/jospt.2011.3618

Fraser, S., Roberts, L., & Murphy, E. (2009). Cauda equina syndrome: A literature review of its definition and clinical presentation. *Archives of Physical Medicine and Rehabilitation, 90*(11), 1964-1968. doi:10.1016/j.apmr.2009.03.021

Frazier, L. M., Carey, T. S., Lyles, M. F., Khayrallah, M. A., & McGaghie, W. C. (1989). Selective criteria may increase lumbosacral spine roentgenogram use in acute low-back pain. *Archives of Internal Medicine, 149*(1), 47-50.

Friedman, J. A., Maher, C. O., Quast, L. M., McClelland, R. L., & Ebersold, M. J. (2002). Spontaneous disc space infections in adults. *Surgical Neurology, 57*(2), 81-86. doi:10.1016/S0090-3019(01)00681-4

Gary Hart, L., Deyo, R. A., & Cherkin, D. C. (1995). Physician office visits for low back pain: Frequency, clinical evaluation, and treatment patterns from a U.S. national survey. *Spine, 20*(1), 11-19. Retrieved from http://journals.lww.com/spinejournal/pages/default.aspx

Gibson, M., & Zoltie, N. (1992). Radiography for back pain presenting to accident and emergency departments. *Archives of Emergency Medicine, 9*(1), 28-31.

Gleave, J. R. W., & Macfarlane, R. (2002). Cauda equina syndrome: What is the relationship between timing of surgery and outcome? *British Journal of Neurosurgery, 16*(4), 325-328. doi:10.1080/0268869021000032887

Gold, M. (2016). Magnetic resonance imaging of spinal emergencies. *Topics in Magnetic Resonance Imaging, 24*(6), 325-330. doi:10.1097/RMR.0000000000000069

Gooding, B. W. T., Higgins, M. A., & Calthorpe, D. A. D. (2013). Does rectal examination have any value in the clinical diagnosis of cauda equina syndrome? *British Journal of Neurosurgery, 27*(2), 156-159. doi:10.3109/02688697.2012.732715

Grammatico, L., Baron, S., Rusch, E., Lepage, B., Surer, N., Desenclos, J. C., & Besnier, J. M. (2008). Epidemiology of vertebral osteomyelitis (VO) in France: analysis of hospital-discharge data 2002-2003. *Epidemiology and Infection, 136*(5), 653-660. Retrieved from http://journals.cambridge.org/action/displayJournal?jid=HYG

Greenhalgh, S., & Selfe, J. (2006). *Red Flags: A guide to identifying serious pathology of the spine.*: Churchill Livingstone: Elsevier.

Grigoryan, M., Guermazi, A., Roemer, F. W., Delmas, P. D., & Genant, H. K. (2003). Recognizing and reporting osteoporotic vertebral fractures. *European Spine Journal, 12*(Suppl. 2), S104-S112. doi:10.1007/s00586-003-0613-0

Grimes, D. A., & Schulz, K. F. (2002). Uses and abuses of screening tests. *Lancet, 359*(9309), 881-884. doi:10.1016/S0140-6736(02)07948-5

Guglielmi, G., De Serio, A., Leone, A., Agrosì, L., & Cammisa, M. (2000). Combined imaging in spondylodiscitis. *Rays, 25*(1), 75-88.

Hadjipavlou, A. G., Mader, J. T., Necessary, J. T., & Muffoletto, A. J. (2000). Hematogenous pyogenic spinal infections and their surgical management. *Spine, 25*(13), 1668-1679. doi:10.1097/00007632-200007010-00010

Harding, I. J., Davies, E., Buchanan, E., & Fairbank, J. T. (2005). The symptom of night pain in a back pain triage clinic. *Spine, 30*(17), 1985-1988. doi:10.1097/01.brs.0000176313.02258.5e

Health Partners Consulting Group. (2012). *Metro-Auckland Pacific population health profile: Tomorrow's health today*. Retrieved from http://www.health.govt.nz/system/files/documents/publications/metro-auckland-pacific-population-health-profile-april2103.pdf

Henschke, N., & Maher, C. (2006). Red flags need more evaluation. *Rheumatology, 45*(7), 920-921. doi:10.1093/rheumatology/kel102

Henschke, N., Maher, C. G., Ostelo, R. W., de Vet, H. C., Macaskill, P., & Irwig, L. (2013). Red flags to screen for malignancy in patients with low-back pain. *Cochrane database of systematic reviews, 2013*(2), CD008686. doi:10.1002/14651858.CD008686.pub2.

Henschke, N., Maher, C. G., & Refshauge, K. M. (2007). Screening for malignancy in low back pain patients: A systematic review. *European Spine Journal, 16*(10), 1673-1679. doi:10.1007/s00586-007-0412-0

Henschke, N., Maher, C. G., & Refshauge, K. M. (2008). A systematic review identifies five "red flags" to screen for vertebral fracture in patients with low back pain. *Journal of Clinical Epidemiology, 61*(2), 110-118. doi:10.1016/j.jclinepi.2007.04.013

Henschke, N., Maher, C. G., Refshauge, K. M., Herbert, R. D., Cumming, R. G., Bleasel, J., . . . McAuley, J. H. (2009). Prevalence of and screening for serious

spinal pathology in patients presenting to primary care settings with acute low back pain. *Arthritis and Rheumatism, 60*(10), 3072-3080. doi:10.1002/art.24853

Higginson, G. (1994). Clinical standards advisory group [Article]. *Quality and Safety in Health Care, 3*(SUPPL), 12-15. doi:10.1136/qshc.3.Suppl.12

Hu, R., Mustard, C. A., & Burns, C. (1996). Epidemiology of incident spinal fracture in a complete population. *Spine, 21*(4), 492-499 498p. Retrieved from http://journals.lww.com/spinejournal/pages/default.aspx

Jacobson, A. F. (1997). Musculoskeletal pain as an indicator of occult malignancy: Yield of bone scintigraphy. *Archives of Internal Medicine, 157*(1), 105-109.

Jaeschke, R., Guyatt, G. H., Sackett, D. L., Guyatt, G., Bass, E., Edwards, P. B., . . . Wilson, M. (1994). Users' guides to the medical literature: Iii. How to use an article about a diagnostic test b. What are the results and will they help me in caring for my patients? *JAMA: The Journal of the American Medical Association, 271*(9), 703-707. doi:10.1001/jama.1994.03510330081039

Jalloh, I., & Minhas, P. (2007). Delays in the treatment of cauda equina syndrome due to its variable clinical features in patients presenting to the emergency department [Article]. *Emergency Medicine Journal, 24*(1), 33-34. doi:10.1136/emj.2006.038182

Jarvik, J. G., & Deyo, R. A. (2002). Diagnostic evaluation of low back pain with emphasis on imaging. *Annals of Internal Medicine, 137*(7), 586-597. Retrieved from http://www.annals.org/

Jennett, W. B. (1956). A study of 25 cases of compression of the cauda equina by prolapsed intervertebral discs. *Journal of Neurology, Neurosurgery, and Psychiatry, 19*(2), 109-116. Retrieved from http://jnnp.bmj.com

Jevtic, V. (2004). Vertebral infection. *European Radiology, 14*(Suppl.), E43-E52. doi:10.1007/s00330-003-2046-x

Jiménez-Mejías, M. E., De Dios Colmenero, J., Sánchez-Lora, F. J., Palomino-Nicás, J., Reguera, J. M., De La Heras, J. C., . . . Pachón, J. (1999). Postoperative spondylodiskitis: Etiology, clinical findings, prognosis, and comparison with nonoperative pyogenic spondylodiskitis. *Clinical Infectious Diseases, 29*(2), 339-345. Retrieved from http://cid.oxfordjournals.org/

Johnell, O., & Kanis, J. A. (2006). An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporosis International, 17*(12), 1726-1733. doi:10.1007/s00198-006-0172-4

Joines, J. D., McNutt, R. A., Carey, T. S., Deyo, R. A., & Rouhani, R. (2001). Finding cancer in primary care outpatients with low back pain a comparison of diagnostic strategies. *Journal of General Internal Medicine, 16*(1), 14-23. doi:10.1046/j.1525-1497.2001.00249.x

Kehrer, M., Pedersen, C., Jensen, T. G., & Lassen, A. T. (2014). Increasing incidence of pyogenic spondylodiscitis: A 14-year population-based study. *Journal of Infection, 68*(4), 313-320. doi:10.1016/j.jinf.2013.11.011

Kemp, H. B. S., Jackson, J. W., Jeremiah, J. D., & Hall, A. J. (1973). Pyogenic infections occurring primarily in intervertebral discs. *Journal of Bone and Joint Surgery, 55*(4), 698-714. Retrieved from http://www.jbjs.org

Khan, I. A., Vaccaro, A. R., & Zlotolow, D. A. (1999). Management of vertebral diskitis and osteomyelitis. *Orthopedics, 22*(8), 758-765. Retrieved from http://www.healio.com/orthopedics/journals/ortho

Khoo, L. A. L., Heron, C., Patel, U., Given-Wilson, R., Grundy, A., Khaw, K. T., & Dundas, D. (2003). The diagnostic contribution of the frontal lumbar spine radiograph in community referred low back pain - A prospective study of 1030 patients. *Clinical Radiology, 58*(8), 606-609. doi:10.1016/s0009-9260(03)00173-9

Kim, C. J., Song, K. H., Jeon, J. H., Park, W. B., Park, S. W., Kim, H. B., . . . Kim, N. J. (2010). A comparative study of pyogenic and tuberculous spondylodiscitis. *Spine, 35*(21), E1096-E1100. doi:10.1097/BRS.0b013e3181e04dd3

Koes, B. W., van Tulder, M., Lin, C. W., Macedo, L. G., McAuley, J., & Maher, C. (2010). An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *European Spine Journal, 19*(12), 2075-2094. Retrieved from http://www.springer.com/medicine/orthopedics/journal/586

Korse, N. S., Jacobs, W. C. H., Elzevier, H. W., & Vleggeert-Lankamp, C. L. A. M. (2013). Complaints of micturition, defecation and sexual function in cauda equina syndrome due to lumbar disk herniation: A systematic review. *European Spine Journal, 22*(5), 1019-1029. doi:10.1007/s00586-012-2601-8

Kostuik, J. P., Harrington, I., & Alexander, D. (1986). Cauda equina syndrome and lumbar disc herniation. *Journal of Bone and Joint Surgery, 68*(3), 386-391. Retrieved from http://www.jbjs.org

Kostuik, J. P., Harrington, I., Alexander, D., Rand, W., & Evans, D. (1986). Cauda equina syndrome and lumbar disc herniation. *The Journal Of Bone And Joint Surgery. American Volume, 68*(3), 386-391. Retrieved from http://www.jbjs.org

Kosuda, S., Kaji, T., Yokoyama, H., Yokokawa, T., Katayama, M., Iriye, T., . . . Kusano, S. (1996). Does bone SPECT actually have lower sensitivity for detecting vertebral metastasis than MRI? *Journal of Nuclear Medicine, 37*(6), 975-978. Retrieved from http://jnm.snmjournals.org

Krogsgaard, M. R., Wagn, P., & Bengtsson, J. (1998). Epidemiology of acute vertebral osteomyelitis in Denmark. *Acta Orthopaedica Scandinavica, 69*(5), 513-513. Retrieved from http://www.tandfonline.com.ezproxy.aut.ac.nz/toc/iort20/current

Lam, K. S., & Webb, J. K. (2004). Discitis. *Hospital Medicine, 65*(5), 280-286. Retrieved from http://www.hospitalmed.com

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration [Conference Paper]. *Journal of Clinical Epidemiology, 62*(10), e1-34. doi:10.1016/j.jclinepi.2009.06.006

Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bonsel, G. J., Prins, M. H., Van Der Meulen, J. H. P., & Bossuyt, P. M. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *Journal of the American Medical Association, 282*(11), 1061-1066. doi:10.1001/jama.282.11.1061

Loblaw, D. A., Perry, J., Chambers, A., & Laperriere, N. J. (2005). Systematic review of the diagnosis and management of malignant extradural spinal cord compression: The cancer care Ontario practice guidelines initiative's neuro-oncology disease site group. *Journal of Clinical Oncology, 23*(9), 2028-2037. doi:10.1200/JCO.2005.00.067

Lora-Tamayo, J., Euba, G., Narváez, J. A., Murillo, O., Verdaguer, R., Sobrino, B., . . . Ariza, J. (2011). Changing trends in the epidemiology of pyogenic vertebral osteomyelitis: the impact of cases with no microbiologic diagnosis. *Seminars in Arthritis and Rheumatism, 41*(2), 247-255. doi:10.1016/j.semarthrit.2011.04.002

Lurie, J. D. (2005). What diagnostic tests are useful for low back pain? *Best Practice and Research: Clinical Rheumatology, 19*(4), 557-575. doi:10.1016/j.berh.2005.03.004

Lurie, J. D., Birkmeyer, N. J., & Weinstein, J. N. (2003). Rates of advanced spinal imaging and spine surgery. *Spine, 28*(6), 616-620. doi:10.1097/00007632-200303150-00018

Lury, K., Smith, J. K., & Castillo, M. (2006). Imaging of spinal infections. *Seminars in Roentgenology, 41*(4), 363-379. doi:10.1053/j.ro.2006.07.008

Mahadevappa, K., Persi, A., & Nesathurai, S. (2015). Acute cauda equina syndrome caused by a disk lesion: Is emergent surgery the correct option? *Spine*. doi:10.1097/BRS.0000000000000790

Malloch, J. D. (1965). Acute retention due to intervertebral disc prolapse. *British Journal of Urology, 37*(5), 578. Retrieved from http://www.bjui.org/

Martin, D. O., & Austin, H. (1996). Exact estimates for a rate ratio. *Epidemiology, 7*(1), 29-33. Retrieved from http://journals.lww.com/epidem/pages/default.aspx

Melton, L. J., III. (2003). Adverse Outcomes of Osteoporotic Fractures in the General Population. *Journal of Bone and Mineral Research, 18*(6), 1139-1141. doi:10.1359/jbmr.2003.18.6.1139

Ministry of Health. (2013a). *Cancer: New registrations and deaths*. Retrieved from http://www.health.govt.nz/system/files/documents/publications/cancer-new-registrations-deaths-2010-aug13-v2.pdf

Ministry of Health. (2013b). *New Zealand Health Survey: Annual update of key findings 2012/13*. Retrieved from http://www.health.govt.nz/publication/new-zealand-health-survey-annual-update-key-findings-2012-13

Ministry of Health. (2015a). *Independent life expectancy in New Zealand 2013*. Retrieved from http://www.health.govt.nz/system/files/documents/publications/independent-life-expectancy-new-zealand-2013-jul15-v2.pdf

Ministry of Health. (2015b). *Population by DHB.* Retrieved 10th April, 2016, from http://www.health.govt.nz/our-work/populations/maori-health/tatau-kahukura-maori-health-statistics/tatauranga-taupori-demographics/population-projections

Mixter, W. J. (1934). Rupture of the intervertebral disc with involvement of the spinal canal. *New England Journal of Medicine, 211*, 210-215. Retrieved from http://www.nejm.org/

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology, 62*(10), 1006-1012. doi:10.1016/j.jclinepi.2009.06.005

Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for developers of health research reporting guidelines. *PLoS Medicine, 7*(2). doi:10.1371/journal.pmed.1000217

Morris, M. J., Akhurst, T., Osman, I., Nunez, R., Macapinlac, H., Siedlecki, K., . . . Scher, H. I. (2002). Fluorinated deoxyglucose positron emission tomography imaging in progressive metastatic prostate cancer. *Urology, 59*(6), 913-918. doi:10.1016/S0090-4295(02)01509-1

Motamedi, K., Ilaslan, H., & Seeger, L. L. (2004). Imaging of the lumbar spine neoplasms. *Seminars in Ultrasound, CT and MRI, 25*(6), 474-489. doi:10.1053/j.sult.2004.09.004

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2197-2223. doi:10.1016/S0140-6736(12)61689-4

Mylona, E., Samarkos, M., Kakalou, E., Fanourgiakis, P., & Skoutelis, A. (2009). Pyogenic vertebral osteomyelitis: A systematic review of clinical characteristics. *Seminars in Arthritis and Rheumatism, 39*(1), 10-17. doi:10.1016/j.semarthrit.2008.03.002

Nagashima, H., Yamane, K., Nishi, T., Nanjo, Y., & Teshima, R. (2010). Recent trends in spinal infections: retrospective analysis of patients treated during the past 50 years. *International Orthopaedics, 34*(3), 395-399 395p. doi:10.1007/s00264-009-0741-1

Nater, A., & Fehlings, M. G. (2015). The timing of decompressive spinal surgery in cauda equina syndrome. *World Neurosurgery, 83*(1), 19-22. doi:10.1016/j.wneu.2014.03.017

National Institute for Health and Care Excellence. (2009). *Low back pain in adults: Early management* [NICE guidelines CG88]. Retrieved November 12, 2013, from http://www.nice.org.uk/guidance/CG88/chapter/introduction

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine, 17*(8), 857-872. doi:10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

O'Laoire, S. A., Crockard, H. A., & Thomas, D. G. (1981). Prognosis for sphincter recovery after operation for cauda equina compression owing to lumbar disc

prolapse. *British Medical Journal, 282*(6279), 1852-1854. Retrieved from http://www.bmj.com/

Oikarinen, H., Karttunen, A., Pääkkö, E., & Tervonen, O. (2013). Survey of inappropriate use of magnetic resonance imaging. *Insights into Imaging, 4*(5), 729-733. doi:10.1007/s13244-013-0276-2

Otéro, J., & Bonnet, F. (2014). Low back pain: Prevalence of McKenzie's syndromes and directional preferences. *Kinesitherapie, 14*(145), 36-44. doi:10.1016/j.kine.2013.10.009

Palmieri, J. J., & Stern, T. A. (2009). Lies in the doctor-patient relationship. *Primary Care Companion to the Journal of Clinical Psychiatry, 11*(4), 163-168. doi:10.4088/PCC.09r00780

Papaioannou, A., Watts, N. B., Kendler, D. L., Chui, K. Y., Adachi, J. D., & Ferko, N. (2002). Diagnosis and management of vertebral fractures in elderly adults. *American Journal of Medicine, 113*(3), 220-228. doi:10.1016/S0002-9343(02)01190-7

Patel, A. R., Alton, T. B., Bransford, R. J., Lee, M. J., Bellabarba, C. B., & Chapman, J. R. (2014). Spinal epidural abscesses: Risk factors, medical versus surgical management, a retrospective review of 128 cases. *Spine Journal, 14*(2), 326-330. doi:10.1016/j.spinee.2013.10.046

Patrick, J. D., Doris, P. E., Mills, M. L., Friedman, J., & Johnston, C. (1983). Lumbar spine X-rays: A multihospital study. *Annals of Emergency Medicine, 12*(2), 84-87. doi:10.1016/S0196-0644(83)80378-3

Perez, F. A., & Jarvik, J. G. (2012). Evidence-based imaging and effective utilization: Lessons in neuroradiology. *Neuroimaging Clinics of North America, 22*(3), 467-476. doi:10.1016/j.nic.2012.05.002

Pham, H. H., Landon, B. E., Reschovsky, J. D., Wu, B., & Schrag, D. (2009). Rapidity and modality of imaging for acute low back pain in elderly patients. *Archives of Internal Medicine, 169*(10), 972-981. doi:10.1001/archinternmed.2009.78

Physiotherapy New Zealand. (n.d.). *Physiotherapy works: The evidence*. Retrieved from http://physiotherapy.org.nz/assets/Your-health/PNZ-Back-Pain.pdf

Powell, E., & Wolfgramm, T. (2013). *Pacific Health Development: Annual Plan 2013/14*. Retrieved from http://www.countiesmanukau.health.nz/assets/About-CMH/Reports-and-planning/Maori-and-pacific-health/2013-2015-Pacific-health-development-annual-plan.pdf

Reihsaus, E., Waldbaur, H., & Seeling, W. (2000). Spinal epidural abscess: A meta-analysis of 915 patients. *Neurosurgical Review, 23*(4), 175-204. Retrieved from http://www.springer.com/medicine/surgery/journal/10143

Reinhold, M., Audigé, L., Schnake, K. J., Bellabarba, C., Dai, L. Y., & Oner, F. C. (2013). AO spine injury classification system: A revision proposal for the thoracic and lumbar spine. *European Spine Journal, 22*(10), 2184-2201. doi:10.1007/s00586-013-2738-0

Reinus, W. R., Strome, G., & Zwemer Jr, F. L. (1998). Use of lumbosacral spine radiographs in a level II emergency department. *American Journal of Roentgenology, 170*(2), 443-447. Retrieved from http://www.ajronline.org

Robinson, R. G. (1965). Massive protrusions of lumbar disks. *British Journal of Surgery, 52*(11), 858-865. Retrieved from http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1365-2168

Roman, M., Brown, C., Richardson, W., Isaacs, R., Howes, C., & Cook, C. (2010). The development of a clinical decision making algorithm for detection of osteoporotic vertebral compression fracture or wedge deformity. *Journal of Manual and Manipulative Therapy, 18*(1), 44-49. doi:10.1179/106698110x12595770849641

Rubinstein, S. M., & van Tulder, M. (2008). A best-evidence review of diagnostic procedures for neck and low-back pain. *Best Practice and Research: Clinical Rheumatology, 22*(3), 471-482. doi:10.1016/j.berh.2007.12.003

Sackett, D. L. (1992). A primer on the precision and accuracy of the clinical examination. *JAMA: The Journal of the American Medical Association, 267*(19), 2638-2644. doi:10.1001/jama.1992.03480190080037

Sans, N., Faruch, M., Lapègue, F., Ponsot, A., Chiavassa, H., & Railhac, J. J. (2012). Infections of the spinal column-Spondylodiscitis. *Diagnostic and Interventional Imaging, 93*(6), 520-529. doi:10.1016/j.diii.2012.04.003

Scavone, J. G., Latshaw, R. F., & Rohrer, G. V. (1981). Use of Lumbar Spine Films: Statistical Evaluation at a University Teaching Hospital. *JAMA: The Journal of the American Medical Association, 246*(10), 1105-1108. doi:10.1001/jama.1981.03320100041027

Schmidt, R. L., & Factor, R. E. (2013). Understanding sources of bias in diagnostic accuracy studies. *Archives of Pathology and Laboratory Medicine, 137*(4), 558-565. doi:10.5858/arpa.2012-0198-RA

Sciubba, D. M., & Gokaslan, Z. L. (2006). Diagnosis and management of metastatic spine disease. *Surgical Oncology, 15*(3), 141-151. doi:10.1016/j.suronc.2006.11.002

Shapiro, S. (1993). Cauda equina syndrome secondary to lumbar disc herniation. *Neurosurgery, 32*(5), 743-747. Retrieved from http://journals.lww.com/neurosurgery/pages/default.aspx

Shapiro, S. (2000). Medical realities of cauda equina syndrome secondary to lumbar disc herniation. *Spine, 25*(3), 348-352. doi:10.1097/00007632-200002010-00015

Shephard, R. H. (1959). Diagnosis and prognosis of cauda equina syndrome produced by protrusion of lumbar disk. *British Medical Journal, 2*(5164), 1434-1439. Retrieved from http://www.bmj.com/

Sizer Jr, P. S., Brismee, J. M., & Cook, C. (2007). Medical screening for red flags in the diagnosis and management of musculoskeletal spine pain. *Pain Practice, 7*(1), 53-71. doi:10.1111/j.1533-2500.2007.00112.x

Slipman, C. W., Patel, R. K., Botwin, K., Huston, C., Zhang, L., Lenrow, D., & Garvan, C. (2003). Epidemiology of spine tumors presenting to musculoskeletal physiatrists. *Archives of Physical Medicine and Rehabilitation, 84*(4), 492-495. doi:10.1053/apmr.2003.50125

Statistics New Zealand. (2014). 2013 Census district health board tables. Retrieved 21 March 2016, from Statistics New Zealand. http://www.stats.govt.nz/Census/2013-census/data-tables/dhb-tables.aspx

Stener, B., & Gunterberg, B. (1978). High amputation of the sacrum for extirpation of tumors: Principles and technique. *Spine, 3*(4), 351-366. Retrieved from http://journals.lww.com/spinejournal/pages/default.aspx

Sun, J. C., Xu, T., Chen, K. F., Qian, W., Liu, K., Shi, J. G., . . . Jia, L. S. (2014). Assessment of cauda equina syndrome progression pattern to improve diagnosis. *Spine, 39*(7), 596-602. doi:10.1097/BRS.0000000000000079

Tali, E. T. (2004). Spinal infections. *European Journal of Radiology, 50*(2), 120-133. doi:10.1016/j.ejrad.2003.10.022

Todd, N. V. (2011). Causes and outcomes of cauda equina syndrome in medico-legal practice: A single neurosurgical experience of 40 consecutive cases. *British Journal of Neurosurgery, 25*(4), 503-508. doi:10.3109/02688697.2010.550344

Toloba, Y., Diallo, S., Maiga, Y., Sissoko, B., Ouattara, K., Soumaré, D., & Sidibé, S. (2011). [Spinal tuberculosis (Pott's disease): epidemiological, clinical, radiological and evolutionary aspects at the University Hospital of Point G]

[abstract]. *Le Mali Medical, 26*(2), 8-11. Retrieved from http://www.malimedical.org/

Tyrrell, P. N. M., Cassar-Pullicino, V. N., & McCall, I. W. (1999). Spinal infection. *European Radiology, 9*(6), 1066-1077. doi:10.1007/s003300050793

Uchida, K., Nakajima, H., Miyazaki, T., Tsuchida, T., Hirai, T., Sugita, D., . . . Baba, H. (2013). 18 F-FDG PET/CT for Diagnosis of osteosclerotic and osteolytic vertebral metastatic lesions: Comparison with bone scintigraphy. *Asian Spine Journal, 7*(2), 96-103. doi:10.4184/asj.2013.7.2.96

Underwood, M. (2009). Diagnosing acute nonspecific low back pain: Time to lower the red flags? *Arthritis and Rheumatism, 60*(10), 2855-2857. doi:10.1002/art.24858

van den Bosch, M. A. A. J., Hollingworth, W., Kinmonth, A. L., & Dixon, A. K. (2004). Evidence against the use of lumbar spine radiography for low back pain. *Clinical Radiology, 59*(1), 69-76. doi:10.1016/j.crad.2003.08.012

Van Tulder, M., Becker, A., Bekkering, T., Breen, A., Del Real, M. T. G., Hutchinson, A., . . . Malmivaara, A. (2006). Chapter 3: European guidelines for the management of acute nonspecific low back pain in primary care. *European Spine Journal, 15*(Suppl. 2), S169-S191. doi:10.1007/s00586-006-1071-2

Verbeek, J., Sengers, M. J., Riemens, L., & Haafkens, J. (2004). Patient expectations of treatment for back pain: A systematic review of qualitative and quantitative studies. *Spine, 29*(20), 2309-2318. doi:10.1097/01.brs.0000142007.38256.7f

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2008). The strengthening the reporting of observational studies in epidemiology (strobe) statement: Guidelines for reporting observational studies. *Journal of Clinical Epidemiology, 61*(4), 344-349. doi:10.1016/j.jclinepi.2007.11.008

Waddell, G. (2004). *The back pain revolution*. Edinburgh, United Kingdom: Churchill Livingstone.

Walsh, G. L., Gokaslan, Z. L., McCutcheon, I. E., Mineo, M. T., Yasko, A. W., Swisher, S. G., . . . Roth, J. A. (1997). Anterior approaches to the thoracic spine in patients with cancer: Indications and results. *Annals of Thoracic Surgery, 64*(6), 1611-1618. doi:10.1016/S0003-4975(97)01034-5

Webster, B. S., Choi, Y., Bauer, A. Z., Cifuentes, M., & Pransky, G. (2014). The cascade of medical services and associated longitudinal costs due to nonadherent magnetic resonance imaging for low back pain. *Spine, 39*(17), 1433-1440. doi:10.1097/BRS.0000000000000408

Weir, M. R., & Thornton, G. F. (1985). Extrapulmonary tuberculosis. Experience of a community hospital and review of the literature. *The American Journal of Medicine, 79*(4), 467-478. doi:10.1016/0002-9343(85)90034-8

White, P., Gunston, J., Salmond, C., Atkinson, J., & Crampton, P. (2008). *Atlas of Socioeconomic Deprivation in New Zealand*. Retrieved from http://www.health.govt.nz/system/files/documents/publications/auckland.pdf

Whiting, P. F., Rutjes, A. W. S., Reitsma, J. B., Bossuyt, P. M. M., & Kleijnen, J. (2003). The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology, 3*, 1-13. doi:10.1186/1471-2288-3-1

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . Bossuyt, P. M. M. (2011). Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*(8), 529-536. Retrieved from http://www.annals.org/

Wibaux, C., Moafo-Tiatsop, M., Andrei, I., Biver, E., Cotten, A., Cortet, B., . . . Flipo, R.-M. (2013). Changes in the incidence and management of spinal tuberculosis in a French University Hospital Rheumatology Department from 1966 to 2010. *Joint, Bone, Spine: Revue du Rhumatisme, 80*(5), 516-519. doi:10.1016/j.jbspin.2012.10.001

Wilk, V. (2004). Acute low back pain: Assessment and management. *Australian Family Physician, 33*(6), 403-407. Retrieved from http://www.racgp.org.au/afp

Williams, C. M., Henschke, N., Maher, C. G., van Tulder, M. W., Koes, B. W., Macaskill, P., & Irwig, L. (2013). Red flags to screen for vertebral fracture in patients presenting with low-back pain. *Cochrane database of systematic reviews, 2013*(1), CD008643. doi:10.1002/14651858.CD008643.pub2.

Williams, C. M., Maher, C. G., Hancock, M. J., McAuley, J. H., McLachlan, A. J., Britt, H., . . . Latimer, J. (2010). Low back pain and best practice care a survey of general practice physicians. *Archives of Internal Medicine, 170*(3), 271-277. doi:10.1001/archinternmed.2009.507

Wilmink, J. T. (1999). MR imaging of the spine: Trauma and degenerative disease. *European Radiology, 9*(7), 1259-1266. doi:10.1007/s003300050832

World Health Organisation. (2014). *Antimicrobial resistance; Global report on surveillance*. Retrieved from http://www.who.int/drugresistance/documents/surveillancereport/en/

World Health Organisation. (2015). *Global tuberculosis report* (20th ed. ed.). Retrieved from http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35. Retrieved from http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1097-0142

Appendix A Ethical Approval

A.1 Ethics approval Auckland University of Technology Ethics Committee

AUT UNIVERSITY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

A U T E C
S E C R E T A R I A T

6 June 2013

Steve White

Faculty of Health and Environmental Sciences

Dear Steve

Re Ethics Application: **13/120 The diagnostic accuracy of red flag questions to screen for serious pathologies in patients presenting with low back pain.**

Thank you for providing evidence as requested, which satisfies the points raised by the AUT University Ethics Committee (AUTEC).

Your ethics application has been approved for three years until 6 June 2016.

As part of the ethics approval process, you are required to submit the following to AUTEC:

- A brief annual progress report using form EA2, which is available online throughhttp://www.aut.ac.nz/researchethics. When necessary this form may also be used to request an extension of the approval at least one month prior to its expiry on 6 June 2016;
- A brief report on the status of the project using form EA3, which is available online throughhttp://www.aut.ac.nz/researchethics. This report is to be submitted either when the approval expires on 6 June 2016 or on completion of the project.

It is a condition of approval that AUTEC is notified of any adverse events or if the research does not commence. AUTEC approval needs to be sought for any alteration to the research, including any alteration of or addition to any documents that are provided to participants. You are responsible for ensuring that research undertaken under this approval occurs within the parameters outlined in the approved application.

AUTEC grants ethical approval only. If you require management approval from an institution or organisation for your research, then you will need to obtain this. If your research is undertaken within a jurisdiction outside New Zealand, you will need to make the arrangements necessary to meet the legal and ethical requirements that apply there. To enable us to provide you with efficient service, please use the application number and study title in all correspondence with us. If you have any enquiries about this application, or anything else, please do contact us at ethics@aut.ac.nz.

All the very best with your research,

Madeline Banda

Acting Executive Secretary

A.2 Amendment approval Auckland University of Technology Ethics Committee

AUT UNIVERSITY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

A U T E C
S E C R E T A R I A T

30 June 2014

Steve White
Faculty of Health and Environmental Sciences

Dear Steve

Re: Ethics Application: **13/120 The prevalence of spinal pathologies amongst patients undergoing magnetic resonance imaging (MRI) for low back pain.**

Thank you for your request for approval of an amendment to your ethics application.

I have approved the minor amendment to your ethics application allowing a retrospective audit of MRI scans, with CMDHB approval.

I remind you that as part of the ethics approval process, you are required to submit the following to the Auckland University of Technology Ethics Committee (AUTEC):

- A brief annual progress report using form EA2, which is available online through http://www.aut.ac.nz/researchethics. When necessary this form may also be used to request an extension of the approval at least one month prior to its expiry on 6 June 2016;
- A brief report on the status of the project using form EA3, which is available online through http://www.aut.ac.nz/researchethics. This report is to be submitted either when the approval expires on 6 June 2016 or on completion of the project.

It is a condition of approval that AUTEC is notified of any adverse events or if the research does not commence.

AUTEC approval needs to be sought for any alteration to the research, including any alteration of or addition to any documents that are provided to participants. You are responsible for ensuring that research undertaken under this approval occurs within the parameters outlined in the approved application.

AUTEC grants ethical approval only. If you require management approval from an institution or organisation for your research, then you will need to obtain this. If your research is undertaken within a jurisdiction outside New Zealand, you will need to make the arrangements necessary to meet the legal and ethical requirements that apply there.

To enable us to provide you with efficient service, please use the application number and study title in all correspondence with us. If you have any enquiries about this application, or anything else, please do contact us at ethics@aut.ac.nz.

All the very best with your research,

Kate O'Connor

Executive Secretary **Auckland University of Technology Ethics Committee**

A.3 Provisional locality approval Counties Manukau Health

earch Committee and Director of Hospital Services

A.4 Locality approval Counties Manukau Health

COUNTIES MANUKAU HEALTH

Ko Awatea Research Office
Counties Manukau DHB
Clinical Support Building
Room 253, Level 2
Middlemore Hospital

10 October 2013

Dear Katy Street

Thank you for the information you supplied to the Ko Awatea Research Office regarding your research proposal:

Research Registration Number: **1545**

Ethics Reference Number: **AUTEC 13/120**

Research Project Title: **The diagnostic accuracy of red flag questions to screen for serious pathologies in patients presenting with low back pain**

I am pleased to inform you that the CMDHB Research Committee and Director of Hospital Services have approved this research with you as the CMDHB Co-ordinating Investigator.

Your study is approved until 6 June 2016.

Amendments:

- All amendments to your study must be submitted to the Research Office for review.
- Any substantial amendment (as defined in the *Standard Operating Procedures for HDECs*, May 2012) must also be submitted to the Ethics Committee for approval.

All external reporting requirements must be adhered to.

Please note that failure to submit amendments and external reports may result in the withdrawal of Ethical and CMDHB Organisational approval.

We wish you well in your project. Please inform the Research Office when you have completed your study (including when a study is terminated early) and provide us with a brief final report (1-2 pages) which we will disseminate locally.

Yours sincerely

S. Everitt.

Dr Samantha Everitt
Manager Research Office
Counties Manukau District Health Board
Under delegated authority from CMDHB Research Committee and Director of Hospital Services

A.5 Amendment locality approval Counties Manukau Health

Amendment approval for 'The diagnostic accuracy of red flag questions to screen for serious pathologies in patients presenting with low back pain' (study 1545)

Hi Katy

Just to confirm, both the service manager and clinical leader have confirmed approval for the amendment to collect data from patient records for the above study. So please accept this email as confirmation of Counties approval for the amendment, pending confirmation of AUTEC approval. If you could forward a copy of your AUTEC amendment approval to Erin Eydt (cc'ed above) once received that would be great.

Best wishes

Sam

Dr Samantha Everitt PhD
Manager: Research Office
Ko Awatea I Health System Innovation and Improvement

T: +64 9 276 0044 I Ext: 8280 I M: 021 913 529 I F: +64 9 259 9665

samantha.everitt@middlemore.co.nz
Level 2 I Support Building 30
Middlemore Hospital, 100 Hospital Road, Otahuhu
Private Bag 93311, Auckland 1640
cmdhb.org.nz I koawatea.co.nz

Appendix B Participant Forms

B.1 Participant consent form

Consent Form

AUT UNIVERSITY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

Project title:	***"The prevalence of spinal pathologies amongst patients undergoing magnetic resonance imaging (MRI) for low back pain."***
Project Supervisor:	***Steve White (AUT), Brendan Coleman (MMH)***
Researcher:	***Katy Street***

- O I have read and understood the information provided about this research project in the Information Sheet dated 24th April 2013.
- O I have had an opportunity to ask questions and to have them answered.
- O I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.
- O I agree to fully complete the questionnaire.
- O I agree for the researcher to view my MRI results.
- O I agree to take part in this research.
- O I wish to receive a copy of the report from the research (please tick one): YesO NoO

Participant's signature:

.....

Participant's name:

.....

Participant's Contact Details (if you would like copies of the research report):

.....

.....

.....

.....

Date: .....

B.1 Participant consent form

Participant Information Sheet

Date Information Sheet Produced:

24th April 2013

Project Title

"The prevalence of spinal pathologies amongst patients undergoing magnetic resonance imaging (MRI) for low back pain."

An Invitation

My name is Katy Street. I am a physiotherapist and I am currently completing a Master of Health Science.

I would like to invite you to participate in this study. This study aims to improve the way we manage and diagnose specific causes of low back pain. If you choose to take part there will be no change to your planned treatment, you will simply be asked to fill in one questionnaire. The results of the questionnaires will then be used to find out whether or not any specific questions are associated with the diagnosis made on the MRI scan. MRI results will also be used to calculate the prevalence of certain pathologies amongst people with low back pain in Auckland, New Zealand.

Participation is voluntary and you may withdraw from the study at anytime prior to data analysis. You will not be disadvantaged in anyway if you choose not to take part in the study.

What is the purpose of this research?

The aim of my Master's thesis is to establish the prevalence of specific causes of low back pain and to determine whether or not any of those pathologies can be predicted from the questions we commonly ask people with low back pain.

How was I identified and why am I being invited to participate in this research?

Everyone requiring an MRI scan for low back pain at Middlemore hospital, or Specialist Radiology Group (SRG) in 2013 will be invited to participate in this research.

What will happen in this research?

If you choose to take part in this study, all you need to do is complete a questionnaire, which will only take around 10 minutes of your time. You will also need to give consent for the researcher to look at the results of your MRI results.

What are the discomforts and risks?

There will be no risks involved and no discomfort. No changes will be made to your standard care.

What are the benefits?

Gaining an understanding of the prevalence of specific causes of low back pain in our local population will help us to determine how likely it is that someone complaining of low back pain may have one of these conditions. Improving our knowledge of which specific questions can be used to rule out or diagnose specific pathologies will improve the care we deliver to all patients with back pain. We hope that this could lead to shorter waiting times, as well as many other benefits for patients.

This research is being used as part of a Masters thesis and will be published. No data that could identify anyone involved in the study will be published. Publication of the data will benefit health professionals and people with back pain in the wider community,

How will my privacy be protected?

The researchers will be the only additional people who will look at your answers to the questionnaire and your MRI results. They will keep the information confidential and your questionnaire will locked in a safe place. Before any information is published the data will be analysed and grouped together, so you will not be able to be identified in any publication of the study results.

What are the costs of participating in this research?

There is no cost involved in participating in this research. It will just take you 5 minutes to complete the questionnaire.

What opportunity do I have to consider this invitation?

You can complete the questionnaire while you wait for your MRI scan. If you change your mind and decide you would not like to participate, you can withdraw at any time prior to data analysis. Please contact the researcher for further details if required.

How do I agree to participate in this research?

Please complete the consent form given to you by the clinic staff.

Will I receive feedback on the results of this research?

If you would like to be sent a copy of the research results please contact the researcher, or simply tick "yes" you would like to receive a copy of the report on you consent form.

Researcher: Katy Street, *katy@aucklandphysiotherapy.co.nz*, +64273336680.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Steve White, *steve.white@aut.ac.nz* (+649) 921 9999 ext 7073.

Concerns regarding the conduct of the research should be notified to the Acting Executive Secretary of AUTEC, Madeline Banda, *ethics@aut.ac.nz* , (+649) 921 9999 ext 8316.

Whom do I contact for further information about this research?

Researcher Contact Details:

Katy Street, *katy@aucklandphysiotherapy.co.nz*, +64273336680.

Project Supervisor Contact Details:

Steve White, *steve.white@aut.ac.nz* (+649) 921 9999 ext 7073

Approved by the Auckland University of Technology Ethics Committee on *(06.06.2013)*, AUTEC Reference number *13/120*.

B.3 Participant questionnaire

Low Back Pain Questionaire

Last Name:	_____	NHI:	_____
First Name:	_____		_____
		Gender:	Male ○ Female

00

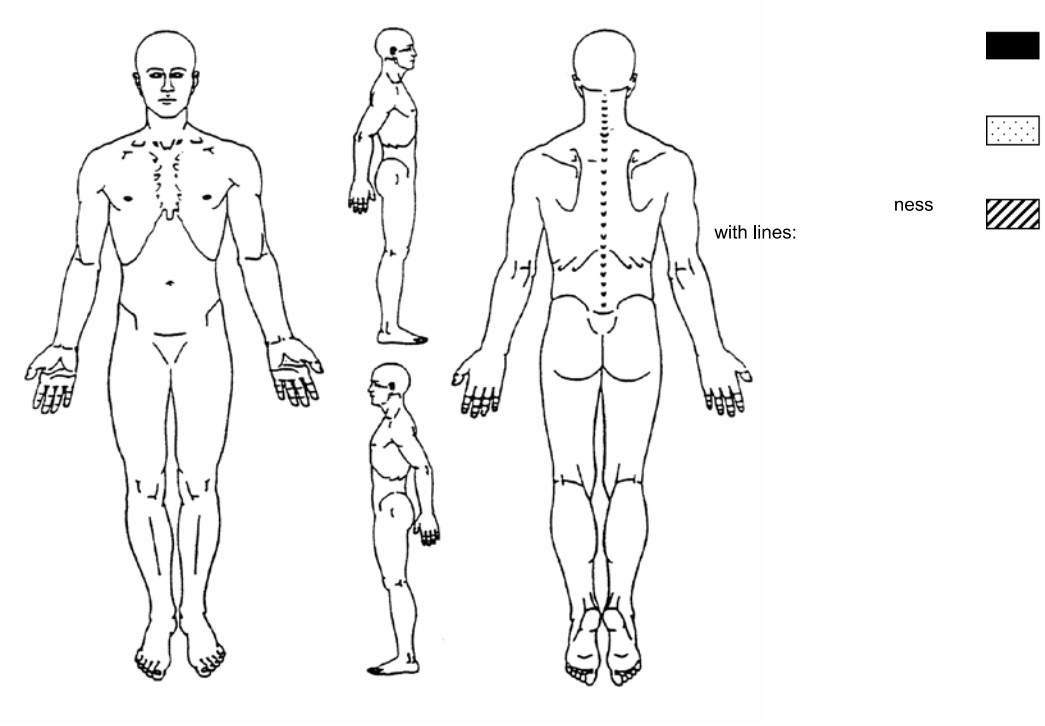
26.

Low Back Pain Questionaire (Page 2)

9. symptoms, a cold, or have you been feeling generally unwell?

14.

32.



with lines:

ness

Appendix C Search Strategy

Table C.1 *Search terms for prevalence of and screening for malignancy in the lumbar spine*

Search	Subject headings and search terms	Results
1	((lumbar OR lumbo* OR "low* back" OR spin*) AND pain)	111,014
2	cancer* OR tumor* OR tumour* OR carcinoma* OR neoplasm* OR sarcoma* OR metastas* OR malignan*	3,369,300
3	prevalence OR incidence OR epidemiology	2,166,585
4	red flag* OR screening OR finding* OR "patient history" OR evaluation OR "medical history" OR "history taking" OR (clinical* N8 sign) OR (clinical* N8 symptom*) OR (clinical* N8 presentation)	3,559,119
Combine searches	1 AND 2 AND 3	996
	1 AND 2 AND 4	3,091

Note. * = truncation, N = proximity search for EMBASE, MEDLINE, CINAHL, SPORTDiscus (W/ used for proximity search in Scopus).

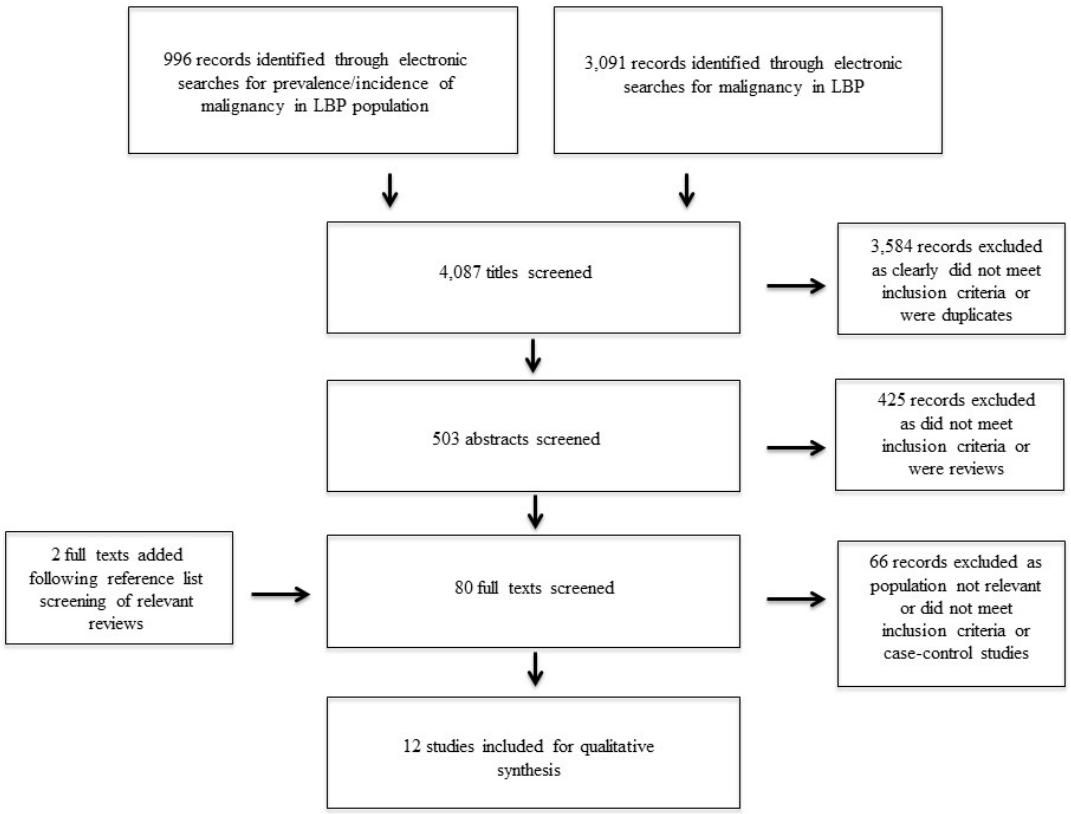


Figure C.1 Flow chart for malignancy literature search

Table C.2 *Search terms for prevalence of and screening for cauda equina syndrome in the lumbar spine*

Search	Subject headings and search terms	Results
1	((lumbar OR lumbo* OR "low* back" OR spin*) AND pain)	111,014
2	"cauda equina" OR "spinal cord compression"	17,530
3	prevalence OR incidence OR epidemiology	2,166,585
4	red flag* OR screening OR finding* OR "patient history" OR evaluation OR "medical history" OR "history taking" OR (clinical* N8 sign) OR (clinical* N8 symptom*) OR (clinical* N8 presentation)	3,559,119
Combine searches	1 AND 2 AND 3	352
	1 AND 2 AND 4	1,050

Note. * = truncation, N = proximity search for EMBASE, MEDLINE, CINAHL, SPORTDiscus (W/ used for proximity search in Scopus).

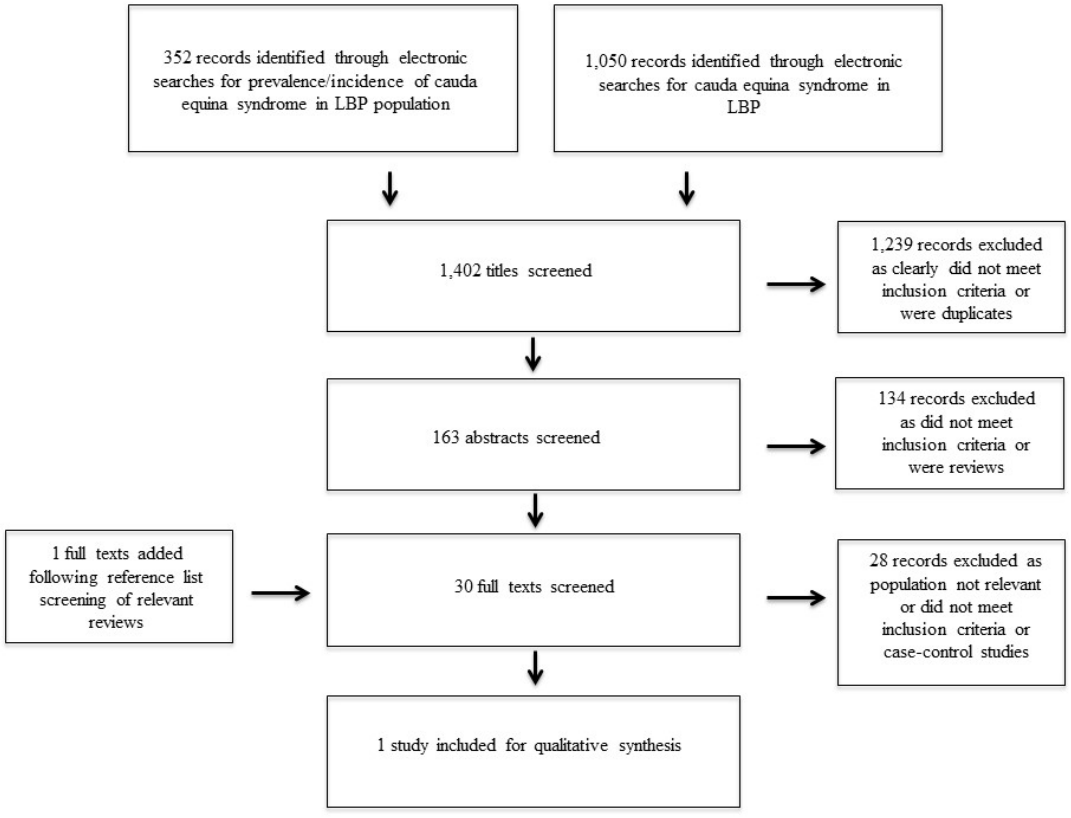


Figure C.2 Flow chart for cauda equina syndrome literature search

Table C.3 *Search terms for prevalence of and screening for spinal infection in the lumbar spine*

Search	Subject headings and search terms	Results
1	((lumbar OR lumbo* OR "low* back" OR spin*) AND pain)	130,041
2	infection* OR discitis OR diskitis OR osteomyelitis OR abscess OR spondylodiscitis OR "infective spondylitis"	1,636,752
3	prevalence OR incidence OR epidemiology	2,517,780
4	red flag* OR screening OR finding* OR "patient history" OR evaluation OR "medical history" OR "history taking" OR (clinical* N8 sign) OR (clinical* N8 symptom*) OR (clinical* N8 presentation)	295,012
Combine searches	1 AND 2 AND 3	884
	1 AND 2 AND 4	1,498

Note. * = truncation, N = proximity search for EMBASE, MEDLINE, CINAHL, SPORTDiscus (W/ used for proximity search in Scopus).

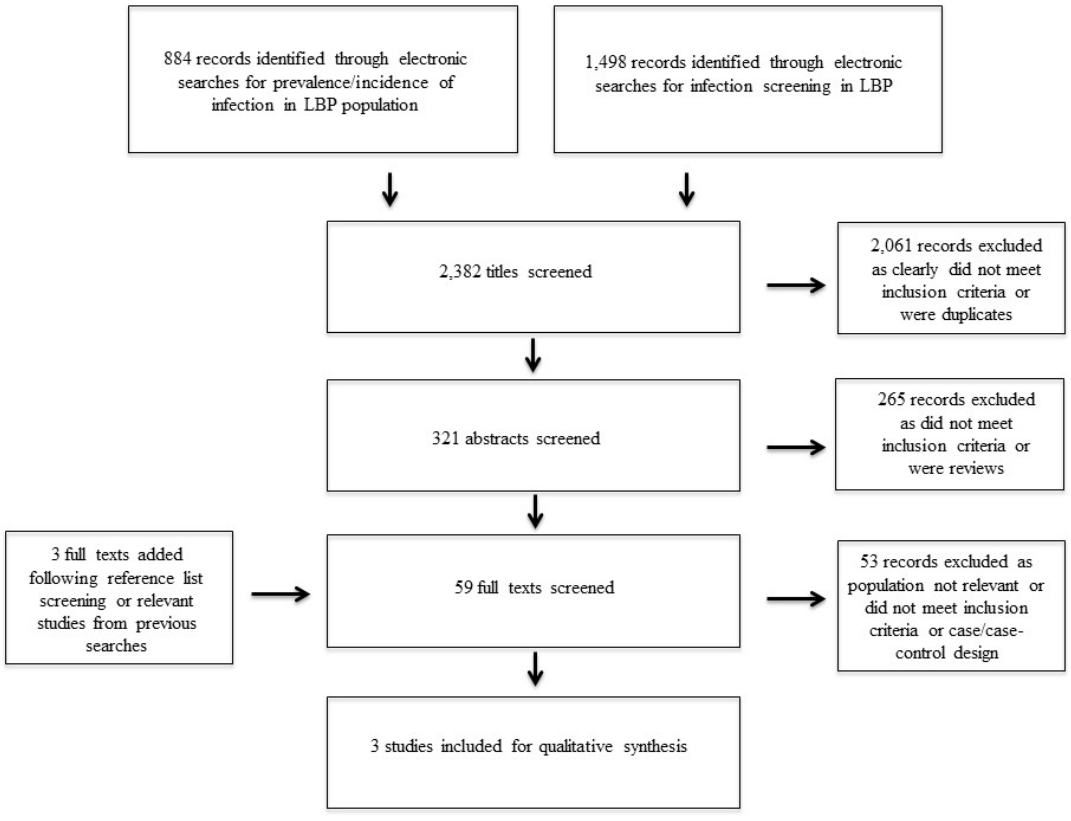


Figure C.3 Flow chart for spinal infection literature search

Appendix D Quality Assessment of Diagnostic Accuracy Studies

QUADAS-2

Phase 1: State the review question:

Patients (setting, intended use of index test, presentation, prior testing):
Index test(s):
Reference standard and target condition:

Phase 2: Draw a flow diagram for the primary study

Phase 3: Risk of bias and applicability judgments

QUADAS-2 is structured so that 4 key domains are each rated in terms of the risk of bias and the concern regarding applicability to the research question (as defined above). Each key domain has a set of signalling questions to help reach the judgments regarding bias and applicability.

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

CONCERN: LOW /HIGH/UNCLEAR

from the review question?

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

/HIGH/UNCLEAR

Appendix E Data Coding

E.1 Data coding manual

Disc
0 Normal – no age related changes
1 Mild degenerative changes - includes minimal bulging or mild loss of disc hydration (from 700 onwards)
2 Annular fissure
3 Degenerative change – dessication, fibrosis, narrowing, diffuse bulge, gas, fissuring, end plate sclerosis, inflammatory changes, defects, osteophytes of the vertebral apophyses, modic type 2, vacuum phenomenon
4 Herniation
5 Protrusion
6 Extrusion
7 Sequestration
8 Trauma – includes evidence of violent injury eg. fracture or dislocation
Nerve root
0 No compression
1 Contacts / abuts exiting nerve
2 Contacts / abuts descending nerve
3 Displaces/impinges exiting nerve
4 Displaces/impinges descending nerve
5 Flatten/Compress nerve exiting nerve
6 Flatten/Compress descending nerve
7 Compresses both exiting and descending nerves
Lumbar Spinal Stenosis
0 No stenosis
1 Mild central canal stenosis
2 Moderate to severe central stenosis
3 Foraminal stenosis (with nerve compression)
4 Lateral recess stenosis

Serious pathologies
0 No serious pathology
1 Tumour
2 Benign tumour eg. Haemangioma
3 Traumatic fracture
4 Osteoporotic, insufficiency fracture or pars defect / fracture
5 Cauda Equina Compression
6 Other eg. 6 lumbar vertebra, sacralisation, pseudoarticulation, DISH, spine surgery, congenital scoliosis
Degenerative changes
0 No changes
1 Facet joint changes – includes synovial cysts, chondral thinning, hypertrophy, OA
2 Mild degenerative spondylolithesis (Grade 1) – assume this includes facet jt changes
3 Significant degenerative spondyloithesis (Grade 2+)
4 Degenerative scoliosis (associated Dx changes, over 25)
5 Isthmic spondylolithesis
Inflammatory / Infection
0 No changes
1 Modic Type 1 changes
2 Spondyloarthropathy
3 Sacroilitis (record this at L5/S1 level)
4 Inflammatory discitis
5 Epidural abscess
6 Discitis
7 Paraspinal muscle infection /abscess (record at L5/S1 level)
8 Infection extending from spine into surrounding structures