Prognostic factors in lumbar spine fusion surgery: Protocol of a systematic review

Organisation, city, country: University of Birmingham, Birmingham, United Kingdom.

Prepared by: Andrew Powell

Supervisors: Alison Rushton and Bart Staal

Project lead: Andrew Powell

Research team members: Andrew Powell, Kostas Zoulas, Bart Staal and Alison Rushton
# Table of Contents

**ADMINISTRATIVE INFORMATION**

- Title ........................................................................................................................................... 1
- Registration .................................................................................................................................. 1
- PROSPERO international prospective register of systematic reviews. CRD42015023430. .......... 1
- Authors ....................................................................................................................................... 1
- Contact ....................................................................................................................................... 1
- Contributions .............................................................................................................................. 2
- Amendments ............................................................................................................................... 2
- Support ....................................................................................................................................... 2

**INTRODUCTION**

- Rationale ................................................................................................................................... 3
  - Background ................................................................................................................................. 3
  - Increasing use of lumbar fusion ............................................................................................... 3
  - Prognostic research .................................................................................................................. 5
  - Previous systematic reviews and research .............................................................................. 5
  - Summary ................................................................................................................................... 6
- Objectives ................................................................................................................................... 7

**METHODS**

- Eligibility criteria .......................................................................................................................... 7
- Information sources ...................................................................................................................... 10
- Search strategy ............................................................................................................................ 11
- Study records ............................................................................................................................... 12
  - Data management .................................................................................................................... 12
  - Selection Process ..................................................................................................................... 12
  - Data collection process .......................................................................................................... 13
- Data items ................................................................................................................................... 13
- Prognostic factors and prioritisations ......................................................................................... 13
- Risk of bias in individual studies ............................................................................................... 15
- Data synthesis ............................................................................................................................ 16
- Meta-bias(es) ............................................................................................................................. 17
- Confidence in cumulative evidence ............................................................................................ 18

**Acknowledgements** .................................................................................................................. 18

**References** .................................................................................................................................. 19


**ADMINISTRATIVE INFORMATION**

**Title**

Prognostic factors in lumbar spine fusion surgery: Protocol of a systematic review

**Registration**

PROSPERO international prospective register of systematic reviews. CRD42015023430.

**Authors**

**Contact**

Andrew Powell  
Affiliation - South Cambridge Physiotherapy & Advance Manipulative Physiotherapy MSc student at the University of Birmingham  
email - andrewjohnpowell@hotmail.co.uk.

Kostas Zoulas  
Affiliation - Corfu Physiotherapy, Greece & Advance Manipulative Physiotherapy MSc student at the University of Birmingham  
email - zoulask@gmail.com

J. Bart Staal  
Affiliation – Senior onderzoeker/epidemioloog  
Scientific Institute for Quality of Healthcare (IQ healthcare) Netherlands  
Email - Bart.Staal@radboudumc.nl

Alison Rushton  
Affiliation - School of Sport, Exercise and Rehabilitation Science, University of Birmingham.  
email - a.b.rushton@bham.ac.uk
Corresponding contributor: Dr Alison Rushton, School of Sport, Exercise and Rehabilitation Science, University of Birmingham, Edgbaston, Birmingham, West Midlands B15 2TT

Contributions

Andrew Powell: Administrative information, introduction, eligibility criteria, information sources, search strategy, data management, selection process, data collection process, data items, prognostic factors and prioritisation, risk of bias in individual studies data synthesis, meta bias(es), confidence in cumulative evidence, acknowledgements, references, appendices.

Kostas Zoulas: Data management, selection process, data collection process, data items, prognostic factors and prioritisation, risk of bias in individual studies.

Alison Rushton: Supervisor.

Bart Staal: Supervisor.

Amendments

Any amendments to this protocol will be documented and resubmitted to this registry.

Support

No specific funding (commercial, private or non-profit) was received for this research.
INTRODUCTION

Rationale

Background

Low back pain (LBP) is the global number one cause of disability (Hoy et al., 2014). In the United Kingdom (UK), approximately one-third of the adult population will experience low back pain, with 20% of that number going on to consult their General Practitioner (NICE, 2009).

There has been an increase in the number of fusion surgeries performed worldwide. The United States (US) has seen an increase from ~183,000 fusion operations in 1999 to over 407,000 in 2012 (HCUPnet), an increase of ~122% in 13 years. The UK has seen an increase in fusion operations (as primary reason for surgery) from 1,435 in 2000-2001, to 2,363 in 2013-14 (HES online); an increase of ~65% in 13 years. Spinal fusion procedures are responsible for the greatest cost of any hospital surgery in the US, at over $40 billion dollars (Deyo, 2015).

Increasing use of lumbar fusion

In the US, increases in the number of fusion surgeries for lumbar degenerative disc disease and lumbar stenosis have been identified as the two biggest factors for the total increase in fusion surgery between 1998 and 2008, with degenerative disc disease seeing the greatest increase (Rajaee et al., 2012). While diagnoses were not given for the cause of fusion in NHS data, lumbar spine fusion surgery (LSFS) was found to be responsible for a larger percentage of the rise in fusion surgery from 2001-02 to 2013-14 than cervical or thoracic fusion (HES online). This is despite randomised control trials (RCTs) suggesting that there is no advantage to LSFS over rehabilitation for degenerative disc disease (Brox et al., 2003; Fairbank et al., 2005) and that reoperation rates have also increased with increased numbers of LSFS (Martin et al., 2007b).

LSFS has been reported to be appropriate for the following pathologies of the lumbar spine (when criteria are met): infection; tumours; traumatic injury; deformity; stenosis; disc herniation; synovial facet cysts; discogenic pain; pseudarthrosis (North American Spine
Society Coverage Committee (NASSCC), 2014). The rationale is that fusion will prevent abnormal movement and so reduce pain (Martin et al., 2007a).

There have been wide variations reported in the rate of LSFS performed in US regions which were linked to the lack of evidence guiding the use of the procedure (Weinstein et al., 2006). Specific diagnostic indications for LSFS are poorly defined, with the exception of spondylolisthesis (Glassman et al, 2009). This is supported by a survey of surgeon members of the Dutch Spine Society that found a lack of consensus between surgeons as to the clinical reasoning and evidence base for the use of lumbar spinal fusion surgery (Willems et al., 2011). More recently, NASSCC (2014) published coverage policy recommendations which were considered ‘reasonable standard practice indications in the US’, based on the experience and expertise of the authors, and evidence-based where possible.

Successful outcomes reported for LSFS varies. Outcomes post LSFS have been measured by surgical outcome, pain, disability measures, work status, physical measures and patient satisfaction using prospective single centre (Ekman et al., 2009; Soriano et al., 2010), retrospective (Gum et al., 2013) and RCT (Hägg et al., 2003) designs. Worse and unchanged patient satisfaction has been reported in 26-50% of patients who undergo LSFS (Brox et al., 2006 - RCT multi centre design; Ekman et al., 2009 - prospective design). Significantly, the lower figure of dissatisfaction was in patients undergoing LSFS for spondylolisthesis; the pathology with the best diagnostic indications for fusion (Glassman et al., 2009 - prospective single centre design) as well as reliable benefit (Kornblum et al., 2004 - prospective single centre design; Weinstein et al., 2007 - randomised cohort and observational cohort - multi centre design).

There is conflicting evidence about the effectiveness of LSFS compared to conservative management. Phillips et al. (2013) performed a systematic review and reported that fusion surgery was a viable treatment option for reducing pain and improving function in patients with chronic low back pain, however Mannion et al. (2013) reported limited evidence that surgery may be more successful in patients who have failed to improve with conservative treatment, and that there was moderate evidence that surgery had similar outcomes to multidisciplinary cognitive-behavioural and exercise rehabilitation. Many studies commonly follow-up patients for 1-2 years, however the longest follow-up of patients receiving LSFS was 11 years (average) and was a multicentre randomised-control trial which showed no
difference between patient rated outcome between surgery and multidisciplinary cognitive-behavioural and exercise rehabilitation (Mannion et al., 2013).

A significant problem within the literature is the heterogeneity of patient populations and the wide variety of surgical intervention (anterior lumbar interbody fusion; posterior lumbar interbody fusion; posterolateral lumbar fusion with or without instrumentation; transforaminal interbody fusion; inter-laminar instrumented lumbar fusion; number of levels fused).

**Prognostic research**

Prognosis is the prediction of, or the estimation of, the probability of the outcome of a procedure or condition (Moons et al., 2009), and it is through prognosis research that informed decisions can be made about interventions (Huguet et al., 2013). This may include identification of modifiable prognostic factors to enhance the outcome of procedures or identification of individuals at risk of poor outcomes (Huguet et al., 2013). Prognostic factors can be vital in improving clinical outcomes, but the quality of prognostic factor research has been reported to be poor with studies often poorly designed, inappropriately analysed and poorly reported (Riley et al., 2013).

**Previous systematic reviews and research**

There have been three systematic reviews into prognostic factors for outcomes in LSFS, all published in a supplement for the journal ‘Spine’. The review topics were:

1. Sociodemographic factors affecting outcome (Mroz et al., 2011).
2. Comorbidity and general health factors affecting outcome (Choma et al., 2011).
3. Psychosocial factors affecting outcome (Daubs et al., 2011).

The search dates for these reviews was 1990 to December 2010 or January 2011, therefore there is the potential for 5 years of further research to add to these reviews. The literature searches were conducted in only 2 databases, Medline (which has been reported to only identify half of available randomised control trials (Glanville et al., 2006)) and the Cochrane Collaboration Library. It has been suggested that Medline and EMBASE should be searched at a minimum to ensure a comprehensive literature search (Furlan et al., 2009). These systematic reviews did not follow Preferred Reporting Items for Systematic Reviews and
Meta-Analyses (PRISMA) guidelines (Moher et al., 2009); perhaps most significantly by not providing search terms/strategy meaning the search could not be repeated.

A protocol has been registered with PROSPERO international prospective register of systematic reviews to investigate pain and psychological predictors of outcomes with LSFS (Cook et al., 2014) but results are yet to be published. Significant differences have been reported between systematic reviews reporting on the same clinical question (Dretzke et al., 2014). Different methodologies are the most likely cause of this disparity, and as this protocol methodology differs to Cook et al. (2014) it will potentially provide a different perspective to the prognostic factors associated with LSFS.

Within the literature there are conflicting results as to what are prognostic factors for LSFS. For example, there are conflicting results for the effect of age (Schoenfeld et al., 2013; Soriano et al., 2010; Hägg et al., 2003), gender (Ekman et al., 2009; Soriano et al., 2010), smoking (Anderson et al., 2006; Sorinao et al., 2010), depression (Hägg et al., 2003; LaCaille et al., 2005) and pain (Abbott et al., 2011; Anderson et al., 2006).

**Summary**

It is clear that LSFS is effective for some patients, but it has significant risks associated with it such as neurological, cardiac and respiratory complications, infection and mortality (Goz et al., 2013) and 1 in 5 requiring revision surgery in 10 years (Deyo, 2015). With patient dissatisfaction reported to be between 26-50%, evidence to suggest that conservative treatment can be just as effective and the significant financial cost of LSFS, the indications for surgery should be considered carefully.

With a century of debate regarding the effectiveness and appropriateness of LSFS in the treatment of LBP (Gibson and Waddell, 2005), there is still no consensus in the prognostic factors associated with LSFS. With increasing numbers of LSFS, higher reoperation rates, and poor satisfaction/success rates, it is important to identify prognostic factors; to improve clinical decision making in who should be considered for LSFS, reduce harm, improve healthcare efficiency, and identify those who would benefit with rehabilitation intervention pre and/or post LSFS.
**Objectives**

The objective of this review is to synthesise the current evidence (prospective studies) of prognostic factors (body functions, activity limitations, participation limitations, or environmental) in the outcome of adult patients (≥18 years) undergoing LSFS for LBP.

Objective 1. Do any body functions, activity limitations, participation limitations, or environmental factors predict outcome after LSFS?

Objective 2. If prognostic factors are found, do the data allow identification of prognostic factors for specific sub-groups undergoing LSFS?

This review aims to identify sub-groups of patients undergoing LSFS with positive and/or negative prognostic factors that can be used in future research and help inform clinician and patient choice.

**METHODS**

This protocol follows guidelines as described by the preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) (Moher et al., 2015).

**Eligibility criteria**

**Participants:**

Studies with adults (≥18 years) undergoing LSFS will be included.

Patients undergoing lumbar fusion for conditions defined by NASSCC (2014) (infection; tumours; traumatic injury; deformity; stenosis; disc herniation; synovial facet cysts; discogenic pain; pseudarthrosis) will be included in this review.

**Intervention:**

This review will synthesise studies that measure and analyse prognostic factors in LSFS. All surgical techniques will be included in analyses (anterior lumbar interbody fusion; posterior lumbar interbody fusion; posteriolateral lumbar fusion with or without instrumentation;
transforaminal interbody fusion; inter-laminar instrumented lumbar fusion; number of levels fused), but studies combining lumbar spine fusion with other surgery, such as decompression, will be excluded. No limitation will be imposed on number of levels fused.

Comparison:

LSFS will not be compared against other interventions.

Outcome (prognostic factors):

A standardised set of outcomes in LBP research were recommended by Deyo et al. (1998) and more recently updated (Delphi study) by Chiarotto et al. (2015) who described core outcome domains for non-specific LBP research as ‘physical functioning’, ‘pain intensity’, ‘health-related quality of life’ and ‘number of deaths’. Pincus et al. (2008) reported a consensus statement by the Multinational Musculoskeletal Inception Cohort Study, for a core set of factors for prospective cohorts in LBP, and WHO (2013) have published the International Classification of Functioning, Disability and Health framework. This systematic review will assess prognostic factors in-line with these statements and framework and include body functions, activity limitation, participation limitation and environmental prognostic factors.

The most appropriate measurement instruments for each outcome domain have not been described as yet (Chiarotto et al., 2015), therefore: body function prognostic factors will include pain intensity and psychological wellbeing measurements as well as patient satisfaction and global health outcome measures; activity limitation prognostic factors will include measurements such as the Oswestry Disability Index (ODI), Roland-Morris Disability Questionnaire (RMDQ) and any other low back pain specific questionnaire; participation restriction prognostic factors will be measured by measures of employment status and sick leave; environmental prognostic factors will be measured by socioeconomic measurements. These measures can be seen in table 1 below.
Table 1. Primary and secondary prognostic factor measures
(adapted from Chiarotto et al., 2015; WHO, 2013; Pincus et al., 2008)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Primary</th>
<th>Secondary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body function</td>
<td>Pain scales (e.g. VAS)</td>
<td>Patient satisfaction</td>
</tr>
<tr>
<td></td>
<td>Assessments of mental health</td>
<td>Global health outcomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Co-morbidities</td>
</tr>
<tr>
<td>Activity limitation</td>
<td>Measures of disability (e.g. ODI and RMDQ)</td>
<td>Quality of life measurements (e.g. SF-36)</td>
</tr>
<tr>
<td>Participation limitation</td>
<td>Employment status/sick leave</td>
<td></td>
</tr>
<tr>
<td>Environmental</td>
<td>Any measure of socioeconomic status</td>
<td></td>
</tr>
</tbody>
</table>

If feasible, outcome data will be grouped by indication for surgery (degenerative disc disease, stenosis, etc.) and intervention (type of fusion). This is dependent on the number of studies that meet inclusion criteria.

Study design:

Prospective cohort studies will be included if they define prospective prognostic factors \textit{a priori}. Prospective studies that are well designed and have no serious limitations can establish high quality evidence for prognosis (Huguet et al., 2013).

Table 2 summarises inclusion and exclusion criteria.
<table>
<thead>
<tr>
<th></th>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients</strong></td>
<td>≥18 years</td>
<td></td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td>Lumbar and lumbosacral fusion surgery secondary to infection; tumours;</td>
<td>Revision surgery</td>
</tr>
<tr>
<td></td>
<td>traumatic injury; deformity; stenosis; disc herniation; synovial facet</td>
<td>Any other lumbar surgical management with fusion</td>
</tr>
<tr>
<td></td>
<td>cysts; discogenic pain; pseudarthrosis</td>
<td></td>
</tr>
<tr>
<td><strong>Comparison</strong></td>
<td>Not appropriate as this review is not looking at comparisons.</td>
<td></td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td>Outcomes measure prior to surgery and at 1 year (minimum) post-surgery</td>
<td></td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td>Prospective cohort studies</td>
<td>Non-English report</td>
</tr>
</tbody>
</table>

**Information sources**

A comprehensive search will be performed through the following:

1. Electronic searches will be performed with the aid of an experienced librarian in the following databases: CINAHL, EMBASE (1974-2015 June 9), MEDLINE (1946-May Week 5 2015), PEDro, and ZETOC
2. Cochrane Central Register of Controlled Trials (CENTRAL)
3. Cochrane Back Review Group trials register
4. Current Controlled Trials website (York) Group, reference searches
6. PubMed
7. Screening reference lists by hand in papers that match the eligibility criteria

**Search strategy**

A health science librarian was consulted in the design of this search strategy.

Not including prognostic-related terms to narrow a search, can result in unnecessarily large numbers of results to review (Chatterley & Dennett, 2012). By combining a broad search filter with disease/intervention focussed search term, identification of appropriate studies can be done with excellent sensitivity (Geersing et al., 2012).

Prognosis search filters have been shown to improve the sensitivity of searches (Wilczynski et al., 2013). The search terms for prognosis will be taken from this research (the Haynes broad filter) with focussed disease/intervention specific search terms:

outcome:.sh. OR prognos*:.tw. OR predict*:.tw.

AND

spinal fusion:.mp. OR arthrodesis:.mp.

AND

Lumbar:.mp. OR lumbosacral:.mp.

AND

[degenerative disc disease OR degenerative disk disease OR disc herniation OR disk herniation OR discogenic pain OR stenosis OR spondylolisthesis OR spondylolysis OR spondylosis OR deformity OR scoliosis OR traumatic injur* OR fracture OR fracture-
dislocation OR dislocation OR infection OR tuberculosis OR discitis OR osteomyelitis OR epidural abscess OR facet cysts OR tumour OR pseudarthrosis].mp. OR low back pain:.sh.

Study records

Data management

Bibliographic management will be done using RefWorks software. Data collection will be performed with the use of pro forma, with guidelines as appropriate. These pro forma for data collection can be seen in Appendices A-C. Data will be managed and recorded throughout the review using Stata 14 software (StataCorp, College Station, TX: StataCorp LP).

Selection Process

Familiarisation will be performed with pro forma (appendices A) and training on two prognostic factor studies not related to LSFS completed (Keeley et al., 2008; Verkerk et al., 2015) with results compared to identify any problems prior to data collection. Two reviewers (AP & KZ) will independently perform the described search strategy and screen the titles and abstracts found by the initial search according to the eligibility criteria. Screening by more than one reviewer reduces the risk of excluding relevant studies (Edwards et al., 2002).

Full text articles will be obtained for the studies that satisfy the eligibility criteria or in any case of uncertainty for inclusion in the review. Full text articles will be independently screened by both reviewers. A study will be included when reviewers independently assess a study as satisfying the eligibility criteria. A third reviewer (AR) will be consulted if disagreements are not resolved through discussion and consensus at each stage. To investigate the agreement between reviewers, Cohen's kappa coefficient will be calculated. A Kappa coefficient of less than 0.01 will be considered to show poor agreement, 0.01-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 excellent agreement (Landis and Koch, 1977).
The process of study selection will be reported using the PRISMA flow diagram (Liberati et al., 2009).

**Data collection process**

Prior to data extraction authors (AP and KZ) will complete training which will involve familiarisation with pro forma (appendices A and B) and assessment of two prognostic factor studies not related to LSFS (previously referenced) with results compared to identify any problems prior to data collection. Any modifications to pro forma will be made *a priori* to minimise later disagreements or misinterpretations (van Tulder et al., 2003). Two authors (AP and KZ) will extract and compare data. If there is disagreement between authors not resolved by discussion, the third reviewer (AR) will be consulted. Investigators will be contacted to request additional information for any missing or unclearly reported data in included studies.

**Data items**

The data extracted from each selected study will include:

- **Identification features of the study**: Study title, author names and publication date.
- **Design of study and recruitment period**
- **Study population**: Sample size, age, gender, pathology.
- **Duration of follow-up**
- **Aims and key results of study**
- **Type of intervention**: Type of fusion surgery, number of levels.
- **Primary and secondary outcomes**
- **Conflict of interests**

**Prognostic factors and prioritisations**

The prognostic factors assessed in this review will be measures of body function, activity limitation, participation limitation and environmental factors, in-line with the statements and framework reported by Chiarotto et al. (2015), WHO (2013) and Pincus et al. (2008). Prognostic factors can be seen in table 1, with primary and secondary prognostic factor...
measures identified based on recommendations by the Cochrane Collaboration (O’Conner et al., 2011). Primary prognostic factor measures have been identified as those essential for decision-making on the prognostic factors associated with LSFS. Secondary prognostic factors have been identified as other main prognostic factors useful in exploring any effects seen.

Minimal important change (MIC) will be defined in-line with Ostelo et al. (2008) and is summarised in table 3. Any measures not described by Ostelo et al. (2008) that do not have MIC data published will have 30% improvement from baseline as a MIC. These measures of prognostic factor prioritisation and defining MIC will prevent reporting bias as described by Kirkham et al. (2010).

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Scoring range</th>
<th>MIC (absolute cutoff)</th>
<th>MIC (% improvement from baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS</td>
<td>0-100</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td>NRS</td>
<td>0-10</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>RDQ</td>
<td>0-24</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>ODI</td>
<td>0-100</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>QBPQ</td>
<td>0-100</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

Absolute values presented are intended for use anywhere in the range of the scale.

VAS indicates Visual Analogue Scale; NRS, Numerical Rating Scale; RDQ, Roland Morris Disability Questionnaire; ODI, Oswestry Disability Index; QBPQ, Quebec Back Pain Disability Questionnaire.
Risk of bias in individual studies

Data to assess risk of bias will include study participation; study attrition; prognostic factor measurement; outcome measurement; study confounding; statistical analysis and reporting as required for the Quality In Prognosis Studies (QUIPS) tool (Hayden et al., 2013). The QUIPS tool will be used to assess the risk of bias for each individual study. This tool has demonstrated acceptable inter-rater reliability (Hayden et al., 2013) and has been used in several studies of prognostic factors (Snodgrass et al., 2014; Bruls et al., 2015).

Each ‘risk of bias’ domain will be rated independently as ‘low’, ‘moderate’ or ‘high’, according to the responses to prompting items (Appendix B) with all domains weighted equally. In cases where the study does not provide sufficient information for judgment, the authors will be contacted for additional information. If authors cannot be contacted, the risk of bias domain will be rated as ‘unclear’.

Overall classification of risk of bias for individual studies will be defined as (Bruls et al., 2015):

- Low risk of bias/high quality when all domains are rated as low-moderate risk of bias.
- High risk of bias/low quality when one or more domains are rated as high risk of bias.

The ‘QUIPS risk of bias assessment instrument for prognostic factor studies’, with each of the 6 domains defined, can be seen in Appendix B, and is a formatted Excel spreadsheet available as an online resource (Hayden et al., 2006).

The risk of bias of each study will be critically evaluated by two reviewers (AP and KZ) independently using this standardised approach. Discrepancies will be resolved by discussion and if necessary a third author (AR) will resolve any disagreements. Inter-rater agreement will be evaluated with the Cohen’s kappa statistic (Carletta, 1996) for the independent ratings of the various domains. Sensitivity analyses will be performed as recommended by Hayden et al. (2006) to explore the impact of risk of bias decisions.

Reviewers will not be blinded to included studies, as true blinding in the risk of bias assessment is difficult to achieve because of involvement in the process of evaluation (Furlan et al., 2009), however, no association between blinded assessment of studies and
bias has been reported (Berlin et al., 1997; Verhagen et al., 1998). The QUIPS tool and the process of risk of bias assessment will be pilot-tested. Familiarisation will be performed with pro forma (appendices B) and training on two studies not related to LSFS done with results compared to identify any problems prior to data collection. In this systematic review the agreement between reviewers, Cohen's kappa coefficient will be calculated.

**Data synthesis**

The availability of appropriate data will determine if a meta-analysis can be performed. If three or more studies with sufficiently homogenous sub-groups are found, a meta-analysis will be conducted.

Clinical, methodological and statistical heterogeneity will be assessed. Clinical heterogeneity will be based on population and prognostic factor measurements, and methodological heterogeneity will be based on study design/biases. Statistical heterogeneity will be assessed using $I^2$ statistic, which gives the percentage of total variation across studies due to heterogeneity rather than chance (Higgins et al., 2003). Interpretation of $I^2$ will be based on the recommendation by the Cochrane Collaboration (Deeks et al., 2011):

- 0% to 40%: might not be important
- 30% to 60%: may represent moderate heterogeneity
- 50% to 90%: may represent substantial heterogeneity
- 75% to 100%: considerable heterogeneity

Meta-analyses, if appropriate, will be performed using Stata 14 software (StataCorp, College Station, TX: StataCorp LP). If there are sufficient data, sub groups will be defined by indication for surgery (degenerative disc disease, stenosis, etc.), intervention (type of fusion) and/or prognostic factor (pain, disability, etc.).

QUIPS will be used to include/exclude studies for meta-analysis, with studies found to have low risk of bias/high quality included in meta-analysis. A random effects statistical model will be used in the meta-analysis with results summarised by pooled estimates (estimate of
average effect), 95% confidence interval, estimates of Tau\(^2\) (between study variance) and a 95% prediction interval for the prognostic effect in a single population (Riley et al., 2011). Sensitivity analyses will be performed to explore the impact of the categorisation of study risk of bias using QUIPS. If feasible, sensitivity analyses may also be performed for study design and follow-up length.

If it is not possible or appropriate to combine results because of heterogeneity of studies or having fewer than three studies for meta-analysis, there will be a narrative critique of the data. Likelihood ratios have been described as one of the best measures for diagnostic accuracy (McGee, 2002). Where necessary, unpublished data will be requested and likelihood ratios with 95% confidence intervals calculated for prognostic factors as possible. Where appropriate stratified sampling will also be performed (e.g. workers’ compensation and other populations).

Meta-bias(es)

Publication bias is an important consideration for prognostic factor research as studies do not report data showing no effect on measures (Huguet et al., 2013). It is recommended that the ‘prudent default position’ to take is the assumption that there is a serious problem with publication bias in the prognostic research (Hemingway et al., 2009). If more than 10 studies are identified for any outcome measure, a funnel plot will be used to evaluate publication bias (Sterne et al., 2011).

For each study included in analysis, study protocols will be sought. Outcomes in the protocol and published study will then be compared as a measure of selective reporting within studies (Higgins et al., 2011). If no protocol is available, outcomes in the method section will be compared with outcomes in the results section. Any non-significant results not reported adequately will be considered as a risk of selective reporting within studies (Higgins et al., 2011).

Sensitivity analyses may be performed to identify any selective reporting within studies. No specific sensitivity analyses are described in this protocol as reasons for sensitivity analyses become apparent during the review process (Deeks et al, 2011). Any sensitivity analyses performed will be presented in a summary table.
Confidence in cumulative evidence

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (Guyatt et al., 2011) will be used to make a statement on the confidence in cumulative evidence and estimates of effects. The GRADE system criteria have been adapted to improve the application of the GRADE system to prognostic factor research (Huguet et al., 2013). There are five factors in the GRADE system that can reduce confidence in prognostic studies and two that could increase confidence in prognostic studies; these can be seen in table 4.

<table>
<thead>
<tr>
<th>Factors that may increase quality</th>
<th>Factors that may decrease quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate or large effect size</td>
<td>Phase of investigation - defined in Appendix C</td>
</tr>
<tr>
<td>Exposure-response gradient</td>
<td>Inconsistency of results across studies</td>
</tr>
<tr>
<td></td>
<td>Indirectness of evidence</td>
</tr>
<tr>
<td></td>
<td>Imprecision</td>
</tr>
<tr>
<td></td>
<td>Publication bias</td>
</tr>
</tbody>
</table>

An example of this adapted GRADE system can be seen in appendix C. Quality of evidence will be rated and reported as very low, low, medium or high.

Acknowledgements

The authors of this protocol would like to thank Lynne Harris from the University of Birmingham library academic services for her help in the designing of this search strategy.
References


enhance systematic reviews. PloS ONE, 7 (2); e32844. doi:10.1371/journal.pone.0032844.


# Data Collection Form

## REVIEW TITLE

<table>
<thead>
<tr>
<th>Study ID (surname of first author and year first full report of study was published e.g. Smith 2001)</th>
<th>PROGNOSTIC FACTORS IN LUMBAR SPINE FUSION SURGERY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notes</td>
<td></td>
</tr>
</tbody>
</table>

## General Information

<table>
<thead>
<tr>
<th>Date form completed (dd/mm/yyyy)</th>
<th>Name/ID of person extracting data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study author contact details</td>
</tr>
<tr>
<td></td>
<td>Study title</td>
</tr>
<tr>
<td></td>
<td>Publication type</td>
</tr>
<tr>
<td></td>
<td>Notes:</td>
</tr>
</tbody>
</table>


## Study Eligibility

<table>
<thead>
<tr>
<th>Study Characteristics</th>
<th>Eligibility criteria</th>
<th>Eligibility criteria met?</th>
<th>Location in text or source (pg/fig/table)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients</strong></td>
<td>&gt;18 years old</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fusion surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td>Lumbar or lumbosacral surgery performed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Secondary to infection; tumours; traumatic injury; deformity; stenosis; disc herniation; synovial facet cysts; discogenic pain; pseudarthrosis</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Outcome measures</strong></td>
<td>Measured prior to surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Measured at 1 year post surgery (minimum)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td>Prospective case controlled study</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**INCLUDE** □  **EXCLUDE** □

**Reason for exclusion**

- Reason for fusion:
  - Revision surgery □
  - Other surgery performed as well as lumbar fusion (e.g. decompression) □
  - Non-English □

**Notes:**

**DO NOT PROCEED IF STUDY EXCLUDED FROM REVIEW**
## Data Extraction Form

<table>
<thead>
<tr>
<th>Description</th>
<th>Location in text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author (year)</td>
<td></td>
</tr>
<tr>
<td>Type of study</td>
<td></td>
</tr>
<tr>
<td>Conflict of interest statement</td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td></td>
</tr>
<tr>
<td>Age/gender</td>
<td></td>
</tr>
<tr>
<td>Population size</td>
<td></td>
</tr>
<tr>
<td>Pathology</td>
<td></td>
</tr>
<tr>
<td>Enrolment period</td>
<td></td>
</tr>
<tr>
<td>Length of follow-up</td>
<td></td>
</tr>
<tr>
<td>Aims and key findings</td>
<td></td>
</tr>
<tr>
<td>Interventions</td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
</tr>
<tr>
<td>Include: baseline and change</td>
<td></td>
</tr>
<tr>
<td>Primary outcome</td>
<td></td>
</tr>
<tr>
<td>Secondary outcomes</td>
<td></td>
</tr>
<tr>
<td>Notes:</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B - QUIPS Risk of Bias Assessment Instrument for Prognostic Factor Studies

QUIPS Risk of Bias Assessment Instrument for Prognostic Factor Studies


<table>
<thead>
<tr>
<th>Author and year of publication</th>
<th>Study identifier</th>
<th>Reviewer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biases</th>
<th>Issues to consider for judging overall rating of &quot;Risk of bias&quot;</th>
<th>Study Methods &amp; Comments</th>
<th>Rating of reporting</th>
<th>Rating of &quot;Risk of bias&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Study Participation</td>
<td>Goal: To judge the risk of selection bias (likelihood that relationship between PF and outcome is different for participants and eligible non-participants).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source of target population</td>
<td>The source population or population of interest is adequately described for key characteristics (LIST).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method used to identify population</td>
<td>The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias (number and type used, e.g., referral patterns in health care)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruitment period</td>
<td>Period of recruitment is adequately described</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place of recruitment</td>
<td>Place of recruitment (setting and geographic location) are adequately described</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusion and exclusion criteria</td>
<td>Inclusion and exclusion criteria are adequately described (e.g., including explicit diagnostic criteria or “zero time” description).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate study participation</td>
<td>There is adequate participation in the study by eligible individuals</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instructions to assess the risk of each potential bias:

These issues will guide your thinking and judgment about the overall risk of bias within each of the 6 domains. Some 'issues' may not be relevant to the specific study or the review research question. These issues are taken together to inform the overall judgment of potential bias for each of the 6 domains.

Provide comments or text excerpts in the white boxes below, as necessary, to facilitate the consensus process that will follow. Click on each of the blue cells and choose from the drop-down menu to rate the adequacy of reporting as yes, partial, no or unsure. Click on the green cells; choose from the drop-down menu to rate potential risk of bias for each of the 6 domains as High, Moderate, or Low considering all relevant issues.
### Baseline characteristics
The baseline study sample (i.e., individuals entering the study) is adequately described for key characteristics (LIST).

### Summary Study participation
The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between PF and outcome.

### 2. Study Attrition
**Goal:** To judge the risk of attrition bias (likelihood that relationship between PF and outcome are different for completing and non-completing participants).

<table>
<thead>
<tr>
<th>Proportion of baseline sample available for analysis</th>
<th>Response rate (i.e., proportion of study sample completing the study and providing outcome data) is adequate.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attempts to collect information on participants who dropped out</td>
<td>Attempts to collect information on participants who dropped out of the study are described.</td>
</tr>
<tr>
<td>Reasons and potential impact of subjects lost to follow-up</td>
<td>Reasons for loss to follow-up are provided.</td>
</tr>
</tbody>
</table>

### Outcome and prognostic factor information on those lost to follow-up
- Participants lost to follow-up are adequately described for key characteristics (LIST).
- There are no important differences between key characteristics (LIST) and outcomes in participants who completed the study and those who did not.

### Study Attrition Summary
Loss to follow-up (from baseline sample to study population analyzed) is not associated with key characteristics (i.e., the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between PF and outcome.

### 3. Prognostic Factor Measurement
**Goal:** To judge the risk of measurement bias related to how PF was measured (differential measurement of PF related to the level of outcome).

<table>
<thead>
<tr>
<th>Definition of the PF</th>
<th>A clear definition or description of 'PF' is provided (e.g., including dose, level, duration of exposure, and clear specification of the method of measurement).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid and Reliable Measurement of PF</td>
<td>Method of PF measurement is adequately valid and reliable to limit misclassification bias (e.g.,</td>
</tr>
</tbody>
</table>
may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).

Continuous variables are reported or appropriate cut-points (i.e., not data-dependent) are used.

**Method and Setting of PF Measurement**

The method and setting of measurement of PF is the same for all study participants.

**Proportion of data on PF available for analysis**

Adequate proportion of the study sample has complete data for PF variable.

**Method used for missing data**

Appropriate methods of imputation are used for missing 'PF' data.

**PF Measurement Summary**

*PF is adequately measured in study participants to sufficiently limit potential bias.*

### 4. Outcome Measurement

**Goal:** To judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the baseline level of PF).

**Definition of the Outcome**

A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct.

**Valid and Reliable Measurement of Outcome**

The method of outcome measurement used is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and confirmation of outcome with valid and reliable test).

**Method and Setting of Outcome Measurement**

The method and setting of outcome measurement is the same for all study participants.

**Outcome Measurement Summary**

*Outcome of interest is adequately measured in study participants to sufficiently limit potential bias.*

### 5. Study Confounding

**Goal:** To judge the risk of bias due to confounding (i.e. the effect of PF is distorted by another factor that is related to PF and outcome).

**Important Confounders Measured**

All important confounders, including treatments (key variables in conceptual model: LIST), are measured.
<table>
<thead>
<tr>
<th><strong>Definition of the confounding factor</strong></th>
<th>Clear definitions of the important confounders measured are provided (e.g., including dose, level, and duration of exposures).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Valid and Reliable Measurement of Confounders</strong></td>
<td>Measurement of all important confounders is adequately valid and reliable (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).</td>
</tr>
<tr>
<td><strong>Method and Setting of Confounding Measurement</strong></td>
<td>The method and setting of confounding measurement are the same for all study participants.</td>
</tr>
<tr>
<td><strong>Method used for missing data</strong></td>
<td>Appropriate methods are used if imputation is used for missing confounder data.</td>
</tr>
<tr>
<td><strong>Appropriate Accounting for Confounding</strong></td>
<td>Important potential confounders are accounted for in the study design (e.g., matching for key variables, stratification, or initial assembly of comparable groups).</td>
</tr>
<tr>
<td></td>
<td>Important potential confounders are accounted for in the analysis (i.e., appropriate adjustment).</td>
</tr>
<tr>
<td><strong>Study Confounding Summary</strong></td>
<td>Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between PF and outcome.</td>
</tr>
</tbody>
</table>

6. **Statistical Analysis and Reporting**

<table>
<thead>
<tr>
<th><strong>Goal:</strong> To judge the risk of bias related to the statistical analysis and presentation of results.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Presentation of analytical strategy</strong></td>
</tr>
<tr>
<td><strong>Model development strategy</strong></td>
</tr>
<tr>
<td><strong>Reporting of results</strong></td>
</tr>
<tr>
<td><strong>Statistical Analysis and Presentation Summary</strong></td>
</tr>
</tbody>
</table>

## Appendix C – GRADE assessment form and guidance

Adapted Grading of Recommendations Assessment, Development and Evaluation (GRADE) table for systematic reviews with meta-analysis of prognostic studies. (Adapted from Huguet et al., 2013)

<table>
<thead>
<tr>
<th>Outcome:</th>
<th>GRADE factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prognostic factor</td>
<td>Number of participants</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------</td>
</tr>
</tbody>
</table>

GRADE factors: ✓, no serious limitations; X, serious limitations (or not present for moderate/large effect size); unclear, unable to rate item based on available information. For overall quality of evidence: +, very low; ++, low; ++++, moderate; ++++, high.

### GUIDE TO JUDGING THE QUALITY OF EVIDENCE FOR PROGNOSIS (from Hayden et al., 2014)

**Starting GRADE**

<table>
<thead>
<tr>
<th>Phase of investigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
</tr>
<tr>
<td>Phase 3 Explanatory Study: Explanatory research aimed to understand prognostic pathways; or Phase 2 Explanatory Study: Explanatory research aimed to confirm independent associations between potential prognostic factor and the outcome</td>
</tr>
<tr>
<td>Moderate</td>
</tr>
<tr>
<td>Phase 1 Explanatory Study: Explanatory research aimed to identify associations between potential prognostic factors and the outcome, or Outcome prediction research providing evidence about prognostic factor associations</td>
</tr>
</tbody>
</table>

**Downgrade if:**

- **Study limitations**: Serious limitations when most evidence is from studies with moderate or unclear risk of bias for most bias domains.
- **Inconsistency**: Unexplained heterogeneity or variability in results

**Upgrade if:**

- **Moderate or large effect**: For meta-analysis: pooled effect is moderate or large.
- **Exposure-gradient response**: For meta-analysis: gradient is present.
across studies with differences of results not clinically meaningful. This may be supported by: between analyses for factors measured at different doses

- For meta-analysis: significant heterogeneity detected by test of heterogeneity and large I² value. For narrative summary: possible gradient exists within and between primary studies

- For narrative summary: variations in effect estimates across studies with points of effect on either side of the line of no effect, and confidence intervals showing minimal overlap.

**Indirectness**
The study sample, the prognostic factor, and/or the outcome in the primary studies do not accurately reflect the review question.

**Imprecision**
For meta-analysis: (1) insufficient sample size and (2) no precise estimate of the effect size in the meta-analysis: confidence interval is excessively wide and overlaps the value of no effect and contain values implying that the factor plays an important role in protecting or putting the individual at risk.

For narrative summary: Within-study imprecision, (1) sample size justification is not provided and there are less than 10 outcome events for each prognostic variable (for dichotomous outcomes) OR there are less than 100 cases reaching endpoint (for continuous outcomes); and (2) no precision in the estimation of the effect size within each primary study, AND

Across study imprecision: there are few studies and small number of participants across studies.

**Publication bias**
We recommend downgrading unless the value of the risk/protective factor in predicting the outcome has been repetitively investigated, ideally by phase 2 and 3 studies.
Appendix D – Gantt chart for review

Spinal Fusion Systematic Review Timeline

Start Date: 26/01/2015

<table>
<thead>
<tr>
<th>Task</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>Completed</td>
</tr>
<tr>
<td>Protocol development</td>
<td>In progress</td>
</tr>
<tr>
<td>Meet librarian</td>
<td>Completed</td>
</tr>
<tr>
<td>Database search</td>
<td></td>
</tr>
<tr>
<td>Training/familiarization</td>
<td></td>
</tr>
<tr>
<td>Review abstracts 1st</td>
<td>1st and 2nd reviewer</td>
</tr>
<tr>
<td>and 2nd reviewer</td>
<td></td>
</tr>
<tr>
<td>Review articles 1st</td>
<td>1st and 2nd reviewer</td>
</tr>
<tr>
<td>and 2nd reviewer</td>
<td></td>
</tr>
<tr>
<td>Data analysis</td>
<td></td>
</tr>
<tr>
<td>Write up</td>
<td></td>
</tr>
<tr>
<td>Submit</td>
<td>20th October 2015</td>
</tr>
</tbody>
</table>