Clinical investigation of bone graft substitutes in spinal fusion

Insights from a randomized intrapatient controlled trial

A.M. Lehr

Clinical investigation of bone graft substitutes in spinal fusion
Insights from a randomized intrapatient controlled trial

A.M. Lehr, 2022
Doctoral thesis, Utrecht University, Utrecht, The Netherlands

Layout: W.V. Potters
Printing: Gildeprint
ISBN: 978-90-393-7466-5
DOI: 10.33540/1313

The clinical trial described in this thesis (ClinicalTrials.gov Identifier: NCT01982045) was supported by a research grant from NuVasive Inc. Printing of this thesis was financially supported by the Nederlandse Orthopaedische Vereniging, Dutch Spine Society and NuVasive Inc.

Clinical investigation of bone graft substitutes in spinal fusion

Insights from a randomized intrapatient controlled trial

Klinisch onderzoek van botvervangende materialen voor wervelkolomfusie

Inzichten uit een gerandomiseerde intrapatiënt gecontroleerde studie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de rector magnificus,
prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 9 juni 2022 des ochtends te 10.15 uur

door

Anne Mechteld Lehr

geboren op 14 juni 1987
te Assen

Promotoren

Prof. dr. M.C. Kruyt
Prof. dr. F.C. Oner
Prof. dr. R.M. Castelein

Table of contents

Abbreviations

AI	Artificial intelligence
ANOVA	Analysis of variance
AP	Anterior-posterior
AR	Autoregressive
ASA	American Society of Anesthesiologists
BCP	Biphasic calcium phosphate
BMI	Body mass index
CaP	Calcium phosphate
cc	Cubic centimeter
CDSR	Cochrane Database of Systematic Reviews
CENTRAL	Cochrane Central Register of Controlled Trials
CI	Confidence interval
CT	Computed tomograpy
DEXA	Dual energy x-ray absorptiometry
FDA	Food and Drug Administration
GEE	Generalized Estimating Equations
HA	Hydroxyapatite
IBF	Interbody fusion
ICC	Intraclass correlation coefficient
IQR	Interquartile range
L	Lumbar (vertebrae)
MCID	Minimal clinically important difference
MDR	Medical device regulation
mg	Milligram
mm	Millimeter
MRI	Magnetic resonance imaging
n	Number
Ncm	Newton centimeter
NPV	Negative predictive value
NRS	Numeric rating scale
ODI	Oswestry Disability Index
OP	Osteogenic protein
OR	Odds ratio
PEEK	Polyetheretherketone
PET	Positron emission tomography
PLF	Posterolateral fusion
PLIF	Posterior lumbar interbody fusion
PMCF	Post-market clinical follow-up
PPV	Positive predictive value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient reported outcome
PROM	Patient reported outcome measure
r	Pearson correlation coefficient

RCT	Randomized controlled trial
S	Sacral (vertebrae)
SAE	Serious adverse event
SAVES	Spine AdVerse Events Severity
SD	Standard deviation
SPECT	Single photon emission computed tomography
T	Thoracic (vertebrae)
TCP	Tricalcium phosphate
TLIF	Transforaminal lumbar interbody fusion
US	Ultrasound; United States
USA	United States of America
VAS	Visual analogue scale
WMO	Wet mensgebonden onderzoek (Medical Research Involving Human Subjects Act)
X	Radiograph

CHAPTER 1

Introduction and aims

Introduction

Spinal fusion is an established surgical treatment for various spinal disorders that require stabilization of the vertebral column. The aimed outcome of this treatment is a solid arthrodesis between two or more vertebrae. In case of posterolateral spinal fusion, a bone graft is placed between the transverse processes to achieve a continuous bony bridge. Since the introduction of spinal instrumentation around 1960, pedicle screw fixation systems are often added to provide immediate biomechanical stabilization and thereby facilitate intertransverse bone formation.[1–5]

Although autologous bone harvested from the iliac crest is considered the gold standard for bone grafting, major disadvantages of this procedure are recognized. These include the need for an additional procedure to harvest the bone graft and potential donor site morbidity. Moreover, bone quality might be insufficient and the amount that can be harvested is limited.[6,7] Reinforced by the increasing demand for bone grafts, numerous alternatives that mimic autograft have been explored. Currently available strategies can be roughly categorized as allografts, growth factors and synthetic ceramics, all of which have their strengths and weaknesses.[8–10] Despite decades of intensive research, cell-based tissue engineering techniques that have the potential for unlimited amounts of viable bone graft never reached the market. However, these research efforts strongly enhanced the understanding of bone regeneration and delineated the requirements for alternatives to autologous bone graft. The ideal bone graft substitute should be biocompatible, bioresorbable, osteoconductive, osteoinductive, osteogenic, unlimited in supply, available off the shelf, inexpensive and should facilitate 100% fusion rate.[11–13]

Calcium phosphate (CaP) based synthetic ceramics, including hydroxyapatite (HA), β-tricalcium phosphate (β-TCP) and biphasic calcium phosphate (BCP), have been investigated and applied extensively over the past five decades. Main reasons are their close resemblance to the mineral component of natural bone (biocompatibility) and ability to provide a scaffold for bone formation (osteoconductivity). Moreover, this group of bone graft substitutes is relatively easy to produce, sterilize and store, which is reflected by lower costs compared to for example growth factor-based substitutes. Optimization of the physicochemical and microstructural properties of BCP ceramics resulted in a controlled bioresorption rate and the ability to induce bone formation in preclinical (animal) models, without the direct presence of osteogenic factors like mesenchymal stem cells (osteoinductive potential). These microporous β-TCP/HA composites with submicron surface topography (BCP$_{<\mu m}$) hold great promise for application as a bone graft substitute, as supported by their performance in repairing critical-sized bone defects.[14–22]

Although several synthetic ceramic bone graft substitutes are commercially available for instrumented spinal fusion, evidence of their clinical efficacy stays behind.[23–25] This can be explained by several factors. First of all, despite the apparently clear aim of spinal fusion surgery, there is still no generally accepted method for the radiological assessment of solid bony fusion.[26,27] This makes it difficult to compare the outcomes of different studies and bone grafts, and to establish acceptable fusion rates. Secondly, the conduct of a high-quality randomized controlled trial (RCT) comparing a bone graft substitute with autograft in a representative study population can be very challenging.[28–30] Thirdly, evidence of clinical efficacy is not a strict requirement for market introduction of a new bone graft substitute. Most bone graft substitutes are registered based on claims of equivalence to already registered competitors. As a consequence of these shortcomings, clinicians have to rely on poorly conducted RCTs or cohort studies.[31,32] The randomized intrapatient controlled trial described in this thesis was actually initiated as the first prospective clinical trial to assess the efficacy of a standalone ceramic bone graft substitute (i.e. not mixed with autograft) for posterolateral spinal fusion.

Solid clinical evidence is essential to demonstrate the added value of bone graft substitutes and to guide the appropriate application for specific indications. In our opinion, such evidence should be gathered through well-designed prospective comparative studies with the performance of the bone graft as primary outcome. In addition, a critical attitude should be exercised with respect to the real need for bone grafting and its substitutes.

Aim of this thesis

The general aim of this thesis is to advance the clinical investigation and application of bone graft substitutes in posterolateral spinal fusion by:

1. Systematically reviewing criteria used for the radiological assessment of posterolateral lumbar fusion;
2. Investigating the efficacy of a microporous BCP$_{<\mu m}$ ceramic bone graft substitute in a randomized intrapatient controlled noninferiority trial;
3. Appraising the methodological aspects of the intrapatient controlled trial design.

Outline of this thesis

In order to advance the assessment of spinal fusion as main outcome of clinical trials investigating bone grafts, a systematic literature review of criteria used for radiological fusion assessment after posterolateral lumbar fusion was performed (**Chapter 2**).

Chapters 3 and **4** present the results of an investigator initiated, multicenter, randomized intrapatient controlled, noninferiority trial comparing a commercially available microporous $BCP_{<\mu m}$ ceramic (AttraX® Putty, NuVasive Inc., San Diego, CA, USA) with autograft in instrumented posterolateral (thoraco)lumbar fusion. This trial was performed between 2013 and 2018 by the Dutch Clinical Spine Research Group. **Chapter 3** reports on the efficacy of the grafts in promoting bony fusion at 1-year follow-up. Secondary, clinical outcomes and adverse events were evaluated. **Chapter 4** describes the 2-year outcomes of this trial, with a specific focus on the progression of bony fusion and the role of bone grafting. Subsequently, methodological aspects of the applied intrapatient controlled trial design are described and quantified in **Chapter 5**, with the aim to further establish this design in clinical (spinal) research.

In **Chapter 6**, the long-term clinical outcomes of instrumented posterolateral lumbar fusion, as well as the relationship with the short-term fusion status, among patients of a previous RCT on bone grafting are evaluated.

The last chapter adds a critical note to the (industry driven) use of bone graft substitutes. Whereas donor site pain is frequently reported as the main reason to use a bone graft substitute, recent studies have debated the incidence and severity of donor site pain in lumbar fusion surgery.[33–36] In light of this controversy, we investigated whether posterolateral lumbar fusion patients, blinded to the donor site, were actually able to identify the iliac crest used for bone graft harvesting (**Chapter 7**).

This thesis concludes with a general discussion and future directions of the clinical investigation and application of bone graft substitutes (**Chapter 8**).

CHAPTER 2

Systematic review of criteria used for radiological fusion assessment after instrumented posterolateral lumbar fusion

A.M. Lehr
A.A.A. Duits
M.R.L. Reijnders
D. Nutzinger
R.M. Castelein
F.C. Oner
M.C. Kruyt

How to assess posterolateral lumbar fusion?
A systematic review of radiological fusion criteria
Submitted for peer review

Introduction

Aimed to obtain a solid arthrodesis between vertebrae, spinal fusion surgery is nowadays an established treatment for numerous lumbar spine disorders, including mechanical instability, degenerative disease and deformity. Over the past century, the surgical strategy evolved from posterior fusion with autologous bone graft to rigid pedicle screw systems and/or interbody fusion devices in combination with bone graft substitutes.[1–5,12,37]

Reliable and accurate assessment of the postoperative fusion status is essential to timely diagnose patients with symptomatic pseudoarthrosis that require additional treatment and to evaluate the performance of spinal fusion procedures. However, in striking contrast to the advancement of surgical techniques, implant designs and diagnostic imaging technologies, researchers and clinicians are still faced with the lack of an established noninvasive method to ascertain solid lumbar spinal fusion. Numerous combinations of imaging modalities, criteria and cut-off values to diagnose solid fusion or pseudoarthrosis are applied in practice. Moreover, evidence of the diagnostic accuracy of radiographic fusion assessments is still limited. This partly explains the wide range of fusion rates reported in literature, and complicates the interpretation and comparison of different studies.

Whereas multiple (review) articles investigated the pros and cons of available imaging modalities to assess the fusion status after posterolateral fusion surgery and thin-slice computed tomography (CT) scanning with multiplanar reconstructions is becoming the recommended standard, there is clearly no consensus on the criteria for successful fusion.[26,27,38–41] To address this knowledge gap, the current systematic review aimed to summarize and evaluate the criteria used for radiological fusion assessment after posterolateral lumbar fusion.

Materials and methods

This systematic review is part of a combined endeavor for both posterolateral and interbody fusions of the lumbar spine and followed a two-step approach. The first step aimed to identify reproducible criteria used for radiological fusion assessment and determine which are used most frequently. The second step focused on the accuracy (with surgical exploration as a reference) and reliability (in terms of inter- and intraobserver agreement) of these criteria. This article solely reports on the results for posterolateral fusion.

Search strategies

For the first step, the electronic databases Medline, Embase, Cochrane Central Register of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR) were searched for literature published through November 2018. The second search on accuracy and reliability was performed in July 2021, in Medline and Embase. Key search terms are listed in **Table 2.1**. The Medline search string of both searches, adapted for the other databases, can be found in **Appendix 1**.

Table 2.1: Key search terms of search 1 and search 2. The full search strings are given in **Appendix 1**.

Search 1 - Identification of reproducible criteria for radiological fusion assessment	
Population	Lumbar spine, degenerative (disc) disease, spondylolisthesis, spinal canal stenosis, spinal deformity
Intervention	Spinal fusion, posterolateral fusion, interbody fusion, bone grafts
Outcome	Fusion, nonunion
Search 2 - Accuracy and reliability of radiological fusion criteria	
Population	Lumbar spine, spinal fusion, posterolateral fusion, interbody fusion
Index test	Radiography, CT, DEXA, SPECT, PET, MRI
Reference test	Surgical exploration
Diagnosis	Fusion, nonunion
Outcome	Accuracy measures, reliability measures

CT = computed tomography, DEXA = dual-energy x-ray absorptiometry, SPECT = single photon emission computed tomography, PET = positron emission tomography, MRI = magnetic resonance imaging

Selection of articles

After removal of duplicates, the identified references were assessed for eligibility based on title and abstract by 2 reviewers (AD and ML) independently using Rayyan QCRI.[42] Next, the full-text of potential articles was retrieved and checked for inclusion by the same reviewers using Zotero Version 5. Disagreements between the reviewers were resolved by discussion. In case consensus could not be reached, a third reviewer was consulted (CO).

For the first search, primary human studies, including (randomized) controlled trials, observational studies and multiple case reports/case series, that described radiological criteria for fusion assessment after primary instrumented posterolateral fusion between T10 and S1 in adult patients were considered. Studies with more than 50% of the patients meeting one or more of the following criteria were excluded: <18 years of age, cervical or main thoracic fusion, revision of instrumented spinal fusion, traumatic fractures, pathological conditions like tumors or infections. In addition, studies unclear about the target population or method of fusion assessment, as well as studies with less than 10 patients meeting the inclusion criteria, were excluded.

For the second search, inclusion criteria were: primary human studies on the accuracy and/or reliability of reproducible radiological criteria for fusion assessment of the (thoraco)lumbar spine. Accuracy studies were only considered when the reference standard was surgical exploration. Exclusion criteria were studies about cervical fusion, comparing different imaging modalities, or with reliability as secondary outcome measure. In addition, studies that did not describe how spinal fusion was assessed were excluded. Systematic reviews on the searched topic were only used for reference checking, to identify additional eligible articles.

Data extraction from included articles

For step 1, the following data were extracted from the included articles using a predefined electronic form (Microsoft Excel Version 2016): year of publication, first and last author, imaging modality, description and reference of the used fusion criteria, cut-off value for successful fusion, and whether accuracy and/or reliability was reported as secondary outcome. Data extraction was divided between four reviewers (AD/DN/ML/MR).

For step 2 articles, data extraction (by AD/ML) additionally included: description of the study design and population and as far as applicable description of the surgical fusion assessment method, measures of accuracy (including percentage agreement between radiological fusion and surgical exploration, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)) and measures of reliability (interobserver and/or intraobserver agreement). All data extractions were checked by another reviewer (AD/ML) and discrepancies were discussed to reach consensus.

Data analysis

For both steps, interreviewer reliability of the inclusion of full-text articles was measured by percentage agreement. The characteristics of the included articles were summarized using descriptive statistics. For the outcome of step 1, it was calculated how many articles applied a specific classification or criterion (absolute frequency, n) and how this related to the total number of articles (relative frequency, %). Further frequency analyses included the used imaging modality, combinations of criteria and cut-off values for successful fusion. In case step 2 articles reported raw data instead of percentage agreement, sensitivity, specificity, PPV and/or NPV, these measures of accuracy were calculated to facilitate the comparison between studies.

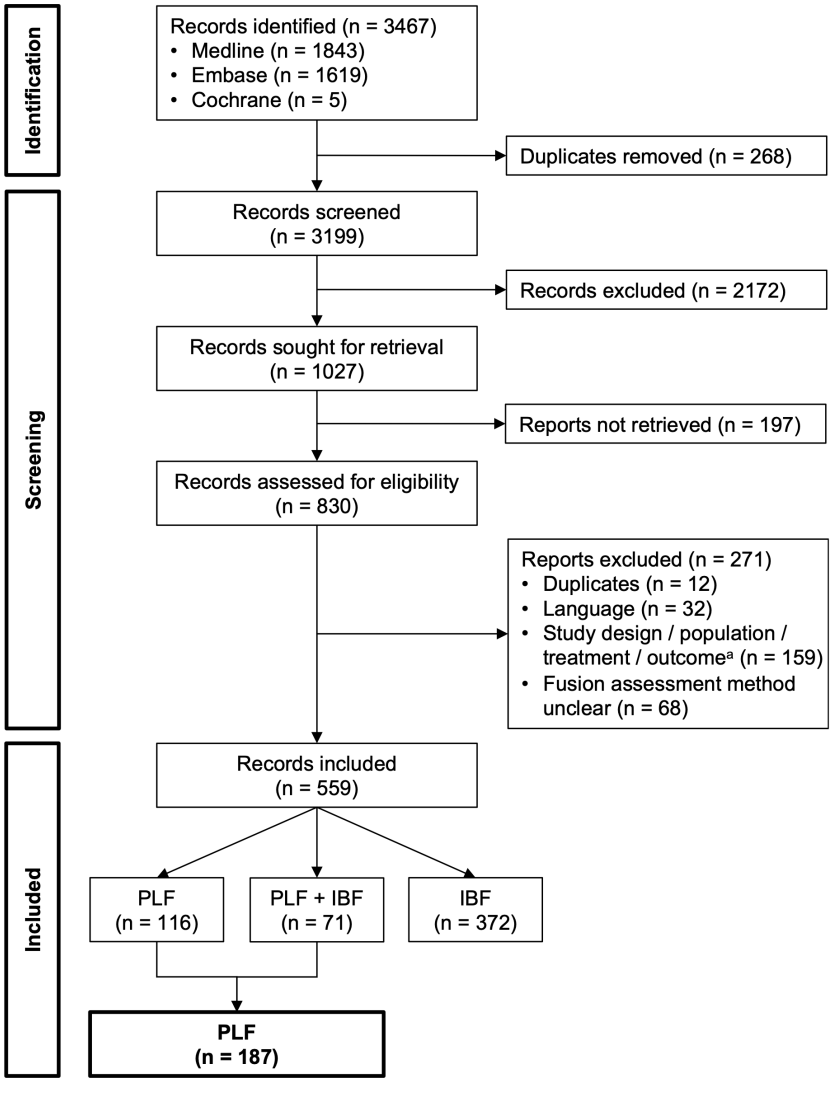
Results step 1 - identification of criteria

Selection of articles

As shown in **Figure 2.1**, the first search yielded 3199 unique references. A total of 830 full-text articles were assessed for eligibility and 559 articles were included, of which 187 involved posterolateral fusion. It is noteworthy that 69 full-text articles (8%) were excluded because they did not describe how spinal fusion was assessed. Agreement between the reviewers on the title-abstract and full-text selection of the combined search was 89% and 85% respectively. Seventy-one of the 187 articles (38%) also discussed interbody fusion, i.e. posterolateral fusion in combination or comparison with interbody fusion.

Fusion assessment method

Study characteristics including year of publication, applied imaging modality and the classification or criteria for fusion assessment are summarized in **Table 2.2**. Fusion was most commonly assessed using static radiographs (134 out of 187 articles; 72%), followed by dynamic radiographs (95; 51%) and CT scans (66; 35%). More than half (55%) of the articles used a combination of imaging modalities. Of the articles that used a single modality, 53 (62%) assessed fusion on static radiographs, 8 (9%) on dynamic radiographs and 22 (26%) on CT scans. Looking per decade, the percentage of articles using CT increased from 14% in the 90s to 48% in the 10s, while the percentage of articles using radiographs decreased from 100% to 83%. A total of 88 (47%) articles used a (previously published) classification or grading system, while 118 (61%) articles used one or more descriptive criteria like bony bridging and absence of signs of nonunion. Nineteen (10%) articles used a combination of a classification and descriptive criteria.



[a] Hierarchical stratification of exclusion reasons: 36 articles wrong study design, 40 wrong population, 55 wrong treatment and 28 wrong outcome.

Figure 2.1: PRISMA flow diagram of identification, screening and inclusion of articles from search 1.

Table 2.2: Study characteristics of the 187 articles included in step 1.

	Frequency of articles, n (%)
Decade of publication	
1990 - 1999	29 (16%)
2000 - 2009	71 (38%)
2010 - 2019	87 (47%)
Imaging modality	
Static radiographs	134 (72%)
Dynamic radiographs	95 (51%)
Computed tomography	66 (35%)
Unclear	4 (2%)
Fusion assessment method	
Classification	88 (47%)
Descriptive criteria	118 (63%)

n = number of articles

Classifications for fusion assessment

Among the 88 articles that used a classification or grading system, including modifications of the original version, the Lenke classification[43] was reported most frequently (30; 34%), followed by the Christensen classification[44] (17; 19%). The frequency of other classifications is given in **Table 2.3**.

A total of 22 (25%) articles used a defined but unnamed grading system. Reference checking revealed that 4 of these articles used the grading system described by Singh et al.[45] and 4 articles used the assessment method of Suk et al.[46]. The remaining 14 (16%) classifications were only reported in a single article and therefore not further analyzed. Analysis of the used imaging modality revealed that the classifications by Jorgenson et al. and Suk et al. were solely used on radiographs, whereas all other classifications were applied to both imaging modalities. Only the grading systems by Glassman et al. and Singh et al. were originally described for CT.

All classifications evaluate continuity of bony bridging on a 3 to 5 point scale. Interestingly, the quality of bridging is named differently without a clear reference including: trabecular, mature, dense, solid, amorphous and cortical edges. Some classifications focus on whether the posterolateral fusion is unilateral or bilateral.[43,47,48] The method by Singh et al.[45] also includes absence of signs of nonunion as criterion for successful fusion, as does Suk et al.[46] by additional assessment of intersegmental motion on dynamic radiographs (cut-off value 4°). Which grades were actually considered successful fusion, for example Lenke A (bilateral solid fusion mass) or Lenke A and B (bilateral

and unilateral solid fusion mass), as well as whether only intertransverse fusion or also facet fusion was assessed, varied widely between studies.

Table 2.3: Classifications (including modified versions) identified in step 1, absolute and relative frequency of reporting (based on 88 articles) and stratification to the used imaging modality.

Classification	Frequency of articles, n (%)		
	Overall[a]	Stratified to imaging modality[b]	
		Radiography	CT
Brantigan and Steffee[49]	3 (3%)	1	2
Bridwell[47]	7 (8%)	6	1
Christensen[44]	17 (19%)	11	7
Glassman[48]	7 (8%)	1	7
Jorgenson[50]	4 (5%)	4	
Lenke[43]	30 (34%)	24	8
Miscellaneous	22 (25%)		
Singh[45]	4 (5%)	2	3
Suk[46]	4 (5%)	4	
Miscellaneous[c]	14 (16%)	11	3

n = number of articles, CT = computed tomography

[a] Based on the 88 articles that used a classification for fusion assessment.

[b] Excluding articles unclear about the used imaging modality (n = 2).

[c] Classifications reported in a single article.

Descriptive criteria for fusion assessment

The variety of descriptive criteria for fusion assessment, extracted from 118 articles, are summarized in **Appendix 2**. For further analyses, these were categorized into three types of criteria: 1) continuity of bony bridging, 2) absence of motion and 3) absence of static signs of nonunion (**Table 2.4**). Criteria related to continuity of bony bridging were most frequently used (104 articles, 88%), followed by absence of motion (78, 66%) as assessed by dynamic radiographs, and absence of static signs of nonunion like radiolucency around the screws (39, 33%). Combinations of criteria as reported by 78 (66%) articles are summarized in **Table 2.5**. Twenty-nine (25%) articles solely considered criteria from the category continuity of bony bridging, of which 20 articles even used a single criterion to determine the fusion rate. While the absence of intervertebral motion is logically assessed by dynamic radiographs, the criteria related to bony bridging and other signs of nonunion were applied to both CT and radiographs, without a clear preference for one of these imaging modalities.

Table 2.4: Categorized descriptive criteria identified in step 1, absolute and relative frequency of reporting (based on 118 articles) and stratification to the used imaging modality.

Classification	Frequency of articles, n (%)		
		Stratified to imaging modality[b]	
	Overall[a]	Radiography	CT
Continuity of bony bridging	104 (88%)	80	37
Absence of motion	78 (66%)	78	
Absence of static signs of nonunion	39 (33%)	31	11
Miscellaneous	3 (3%)	3	1

n = number of articles, CT = computed tomography

[a] Based on the 118 articles that used descriptive criteria for fusion assessment.

[b] Excluding articles unclear about the used imaging modality (n = 2).

Table 2.5: Combinations of descriptive criteria, indicated by 'x', and the absolute frequency of reporting (based on 78 articles).

Continuity of bony bridging	Absence of motion	Absence of static signs of nonunion	Absolute frequency of articles
x	x	x	25
x	x		39
x		x	9
	x	x	5

Reported cut-off values for the absence of rotational (1.5 - 10°) and translational motion (2 - 4.5mm) in flexion-extension radiographs varied considerably (**Appendix 2**). In addition, the exact measurement method of intersegmental motion was often not described. Static signs of nonunion were defined as implant failure or loosening (21 articles), radiolucency around the implant (14), and a radiographic gap, cleft or line in the fusion mass (9).

Further exploration of the predominant continuity of bony bridging category revealed that half (50%) of the articles particularly assessed bony bridging between the transverse processes and 9 (9%) articles included facet fusion. Interestingly, bony bridging between vertebral bodies (i.e. interbody fusion) was also reported as criterion for posterolateral fusion by 7 articles that included both posterolateral and interbody fusion.

The wide variety of terms that were used for the quality of bony bridging is comparable to the terminology of the classifications described above. Sixteen articles only considered bilateral fusion successful, 6 articles both unilateral and bilateral fusion, and 5 articles specifically demanded all levels to be fused in case of multilevel fusions.

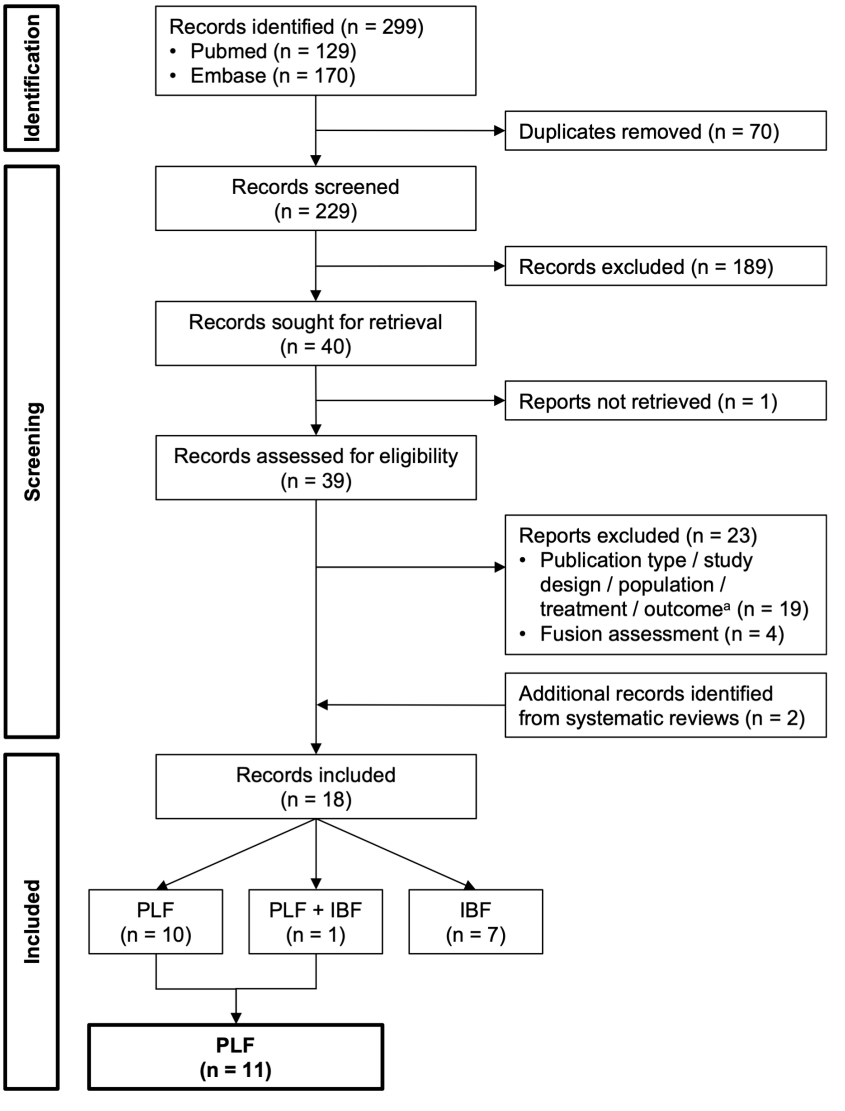
Combination of classification and descriptive criteria

Of the 19 articles that used both a classification and descriptive criteria, 5 (26%) reported separate fusion rates based on the different assessment methods. A total of 13 (68%) articles used continuity of bony bridging (assessed with a classification on CT scans or radiographs) and absence of motion (assessed by dynamic radiographs) as combined criteria for successful fusion. Three of these articles used absence of implant failure as third criterion. One article used a classification to assess the fusion status on radiographs, in combination with continuity of bony bridging on CT.

Results step 2 - accuracy and reliability

Article selection

The flow diagram of the second search is shown in **Figure 2.2**. Agreement between the 2 reviewers on the 229 title-abstract and 39 full-text selections was 82% and 90% respectively. Checking the reference list of the 6 systematic reviews that were identified by this search yielded 2 additional inclusions (both on interbody fusion). Overall, 18 articles were included of which 11 reported on posterolateral fusion. The study design, method of fusion assessment and measures of accuracy and/or reliability of these 11 articles are summarized in **Table 2.6**.



[a] Hierarchical stratification of exclusion reasons: 9 articles wrong publication type (abstract or review), 7 wrong study design, 1 wrong treatment and 2 wrong outcome.

Figure 2.2: PRISMA flow diagram of identification, screening and inclusion of articles from search 2.

Table 2.6: Summary of studies investigating the diagnostic accuracy and/or reliability of the radiological assessment of posterolateral fusion (in order of publication year, with accuracy studies first).

Article	Study design	Fusion assessment	Fusion outcome	Accuracy measures
Kant et al. (1995)[51]	Retrospective cohort study of 75 patients (126 levels) who had unintentional hardware removal (mean 51 weeks, range 1 month-212 weeks) after instrumented lumbar PLF.	Radiographic fusion: solid bone between transverse processes or obliteration and fusion of the facet joint (oblique views) on static X. Surgical exploration: no motion (as determined by 4 different methods) and visualization of a solid bone mass from ala to transverse process, or total obliteration of the facet joint.	Radiographic fusion rate = 78%. Upon surgical exploration, 87 levels (69%) were fused.	Poor overall agreement (70.6%) between X and surgical exploration (mean kappa = 0.26 (95% CI 0.07-0.44, range 0.03-0.59)). Accuracy for detecting fusion: Sensitivity = 85%, specificity = 38%. PPV = 76%, NPV = 54%.
Larsen et al. (1996)[52]	Prospective cohort study of 25 symptomatic patients (41 levels) who underwent hardware removal (> 1 year) after instrumented lumbar PLF.	Radiological fusion: presence of bridging bony trabeculae on static X (AP, lateral and oblique views; n = 21), < 3° of motion on dynamic X (n = 11), presence of bridging bony trabeculae on thick-slice CT (5mm; sagittal and coronal reformations; n = 24), lack of increased uptake on bone scintigraphy (n = 20). Surgical exploration: inspection of the fusion mass (solid fusion or pseudoarthrosis).	Radiological fusion rate = 29% (static X), 91% (dynamic X), 42% (CT), 80% (bone scintigraphy). At surgery, 16 patients were classified as fused (64%), whereas 9 patients were diagnosed with pseudoarthrosis.	Agreement between imaging and surgical exploration = 62% (static X), 55% (dynamic X), 63% (CT), 60% (bone scintigraphy). Accuracy for detecting fusion: Sensitivity = 42% (static X), 86% (dynamic X), 53% (CT), 83% (scintigraphy). Specificity = 89% (static X), 0% (dynamic X), 78% (CT), 25% (scintigraphy). PPV = 83% (static X), 60% (dynamic X), 80% (CT), 63% (scintigraphy). NPV = 53% (static X), 0% (dynamic X), 50% (CT), 50% (scintigraphy).
Jacobson et al. (1997)[53]	Prospective cohort study of 10 symptomatic patients (20 sites) who underwent second-look surgery ($\geq$ 9 months) after (thoraco)lumbar PLF.	Solid fusion on US: presence of an echogenic and shadowing interface that bridged the transverse processes, facets, laminae or spinous processes. Surgical exploration: visual presence of bridging bone. In case of an cleft or indentation, pressure was applied to asses for motion.	US depicted solid fusion in 6 of 20 sites. At surgery, 10 sites were considered fused.	Agreement between US and surgical exploration = 80%. Accuracy for detecting pseudoarthrosis: Sensitivity = 100%, specificity = 60%. PPV = 71%, NPV = 100%.
Kanayama et al. (2006)[54]	RCT of 19 patients with single-level instrumented lumbar PLF and per protocol implant removal when the radiological fusion criteria were met.	Radiological fusion: < 5° of angular motion and < 2mm of translation at the operative level on dynamic X and evidence of posterolateral bridging bone on CT. Surgical exploration: intraoperative manipulation after instrumentation removal.	Radiological fusion rate = 84%. At surgical exploration, 11 of the 16 patients with radiological fusion were considered to have a solid fusion (69%).	Overall agreement between imaging and surgical exploration = 69%.

Table 2.6: *(continued)*

Article	Study design	Fusion assessment	Fusion outcome	Accuracy measures	Reliability measures
Carreon et al. (2007)[55]	Retrospective cohort study of 93 patients (163 levels) who had revision surgery (mean 49 ± 38 months, range 1-148 months) after instrumented lumbar PLF.	Radiological fusion: obliteration of the facet joint space (facet fusion) and continuous trabeculated bone connecting the transverse processes (posterolateral gutter fusion) on fine-cut CT scans (1mm; sagittal and coronal reconstruction). Assessment by 3 observers. Surgical exploration: absence of bony continuity on inspection of the posterolateral gutters and facets and the presence of motion under distraction were defined as a nonunion.	Radiological fusion rates: bilateral facet = 71%, unilateral facet = 8%, bilateral posterolateral gutter = 43%, unilateral = 14%. On surgical exploration, the presence of nonunion was 20%.	Likelihood ratio for solid facet fusion = 2.90 (bilateral), 0.55 (unilateral), 0.19 (no fusion). Likelihood ratio for solid posterolateral gutter fusion = 8.31 (bilateral), 5.37 (unilateral), 0.35 (no fusion). PPV facet fusion = 74% (bilateral), 35% (unilateral), 16% (no fusion). PPV posterolateral gutter fusion = 89% (bilateral), 84% (unilateral), 26% (no fusion).	Moderate interobserver reliability for facet fusion (kappa = 0.42) and substantial for posterolateral gutter fusion (kappa = 0.62)
Fogel et al. (2008)[56]	Retrospective cohort study of 90 patients (172 levels) who had unintentional surgical exploration (mean 27 months, range 12-65 months) after combined lumbar IBF and instrumented PLF.	Radiological fusion: Lenke classification grade A on static X (AP and lateral views; 2 observers) and thin-section helical CT (1mm; sagittal and coronal reconstructions). Surgical exploration: observation of a solid cortical posterolateral bone bridge and absence of visible motion.	Radiological fusion rate = 75% (static X), 68% (CT). Surgical fusion rate = 97%.	Agreement imaging and surgical exploration = 77% (static X), 72% (CT) Accuracy for detecting pseudoarthrosis: Sensitivity = 100% (static X and CT). Specificity = 77% (static X), 70% (CT). PPV = 9% (static X), 11% (CT). NPV = 100% (static X and CT).	Overall interobserver agreement for IBF and PLF on static X = 99%.
Spirig et al. (2019)[57]	Prospective cohort study of 41 patients (159 pedicle screws) who underwent revision surgery (mean 3.0 ± 3.1 years, range 0-13.5 years) after instrumented lumbar PLF.	Radiological pseudoarthrosis: presence of peri-screw edema on MRI (2 observers) and presence of peri-screw osteolysis on static X (AP and lateral views) and thin-slice CT (2mm; 2 observers). Surgical exploration: screw loosening was defined as an unscrewing torque ≤ 60Ncm.	Radiological screw loosening ranged between 31% and 36%. At surgical exploration, 35% of the screws were loose.	Accuracy for screw loosening: Sensitivity = 34.5-43.9% (MRI), 54.2% (static X), 52.4-64.8% (CT). Specificity = 77.4-92.1% (MRI), 83.5% (static X), 93.8-96.7% (CT).	Poor interobserver reliability for MRI (kappa = 0.289). Intraclass correlation analysis showed good agreement for CT (ICC = 0.860)

Table 2.6: *(continued)*

Article	Study design	Fusion assessment	Fusion outcome	Reliability measures
Christensen et al. (2001)[44]	RCT of 43 patients (53 levels) with instrumented vs. 36 patients (50 levels) with non-instrumented lumbar PLF (1 year follow-up).	Radiological fusion: Christensen classification (uni- or bilateral fusion at all intended levels) on static X (AP and lateral views). Assessment by 4 observers and repeated after 8 weeks.	Mean fusion rate = 81% (instrumented fusion 70%-82%, noninstrumented fusion 81%-92%). 68% of the patients were classified as fused at all levels by all 4 observers.	Mean interobserver agreement = 87% (range 83-93%) with mean kappa = 0.58 (range 0.44-0.70; fair to good). Mean intraobserver agreement = 93% (mean kappa = 0.76, excellent).
Tokuhashi et al. (2008)[58]	Retrospective cohort study of 190 patients with instrumented lumbar PLF, with or without additional IBF (mean 5.7 years, range 3-13 years follow-up).	Radiological pseudoarthrosis: clear-zone ($\geq$ 1 mm circumferential lucency around a screw confirmed in $\geq$ 2 directions by >2 of 3 observers) on static X. Assessment was repeated after 4-weeks.	The clear-zone positive rate decreased from 41% (n = 78) at 6 months follow-up to 15% (n = 28) at $\geq$ 3 years follow-up.	Mean interobserver agreement = 96.2% (range 95.1%-96.8%) with mean kappa = 0.90 (range 0.87-0.91, excellent). Mean intraobserver agreement = 97.4% (range 96.8%-97.9%) with mean kappa = 0.95 (range 0.93-0.96, excellent).
Dakhil-Jerew et al. (2009)[59]	Prospective cohort study of 50 patients with dynamic posterolateral pedicle screw stabilization (260 screws; mean 40.9 months, range 20-74 months follow-up).	Screw loosening: assessed on static X (AP and lateral) based on a "halo zone sign" (7 observers) and a "double-halo sign" (4 observers).	"Halo zone sign": 3-44% screw loosening. "Double-halo sign": 7-9% screw loosening.	Poor interobserver agreement for "halo zone sign": kappa = 0.1462 (95% CI 0.0332-0.2592). Substantial interobserver agreement for "double-halo sign": kappa = 0.666 (95% CI 0.496-0.836).
Gotfryd et al. (2014)[60]	Prospective cross-sectional study of 20 patients with instrumented lumbar PLF ($\geq$ 24 months, mean 32 months follow-up).	Radiological fusion: Christensen classification (uni- or bilateral fusion) on static X (AP and lateral) and <5° difference in Cobb's angle on dynamic X. Assessment by 6 observers; 4 observers repeated the assessment after 8 weeks.	(Not reported)	Based on 1st rating, mean interobserver agreement = 76 $\pm$ 7.8% (kappa = 0.07-0.50, poor-moderate) for static X and 78 $\pm$ 9.1% (kappa = -0.08- 0.50, poor-moderate) for dynamic X. Mean intraobserver agreement = 63 $\pm$ 10% (kappa = 0.06-0.26, poor-reasonable) for static X and 84 $\pm$ 10% (kappa = 0.20-0.73, poor-substantial) for dynamic X.

PLF = posterolateral fusion X = radiograph, AP = anterior-posterior, n = number of patients, CT = computed tomography, CI = confidence interval, PPV = positive predicting values, NPV = negative predicting value, US = ultrasound, RCT = randomized controlled trial, IBF = interbody fusion, MRI = magnetic resonance imaging, Ncm = Newton centimeter, ICC = intraclass correlation coefficient.

Diagnostic accuracy

Diagnostic accuracy of the Lenke classification on plain radiographs and thin-slice CT scans was assessed by 1 retrospective cohort study, including 90 patients who underwent combined interbody fusion and instrumented posterolateral fusion.[56] Considering bilateral fusion masses (Lenke A) as successful posterolateral fusion, the fusion rate determined with surgical exploration (97%) was underestimated by both radiography (75%) and CT (68%). Agreement with surgical exploration was 77% for radiography and 72% CT respectively. The sensitivity for detecting pseudoarthrosis was 100% for both imaging modalities. Specificity was 77% for radiographs and 70% for CT scans.[56] There were no studies available on the accuracy of other classifications identified in step 1.

The following 5 accuracy studies evaluated continuity of bony bridging, of which 2 also included absence of motion as criterion for successful fusion.[51–55] Jacobson et al. used ultrasound (US) to evaluate thoracic and/or posterolateral lumbar fusion (between the transverse processes, facets laminae or spinous processes) in 10 patients who underwent second-look surgery for clinically suspected pseudoarthrosis. The accuracy in terms of agreement between US and surgical exploration was 80%, with 100% sensitivity and 60% specificity for detecting nonunion.[53] Kant et al. considered solid bone from one transverse process to the adjacent transverse process or obliteration and fusion of the facet joint on plain radiographs as solid fusion. Overall agreement between radiographs and surgical exploration after hardware removal in 75 patients was poor with mean kappa = 0.26 (95% CI 0.07 - 0.44, range 0.03 - 0.59). Sensitivity to detect solid fusion was 85% while specificity was 38%.[51] Carreon et al. investigated the accuracy of facet fusion (obliteration of the facet joint space) and posterolateral gutter fusion (continuous trabeculated bone connecting the transverse processes) using fine-cut CT scans in 93 patients (163 levels) who underwent revision surgery after instrumented lumbar fusion. The likelihood ratio for bilateral facet fusion (2.90) was higher than for unilateral fusion (0.55). For posterolateral gutter fusion these likelihood ratios were better (8.31 and 5.37 respectively).[55]

Larsen et al. compared the presence of bridging bony trabeculae on static radiographs and on thick-slice CT with operative findings after hardware removal in 25 symptomatic patients. In addition, they reported on the diagnostic accuracy of dynamic radiographs (with < 3° of motion as cut-off value) and bone scintigraphy (lack of increased uptake). Depending on the imaging modality, the radiographic fusion rate ranged between 29% (static radiographs) and 91% (dynamic radiographs), while the surgical fusion rate was 64%. Bridging bony trabeculae on static radiographs had the lowest sensitivity (42%), but the highest specificity (89%) to detect fusion. In contrast, dynamic radiographs had

the highest sensitivity (86%) but 0% specificity. Agreement with surgical exploration was highest for CT (63%).[60] Kanayama et al. evaluated posterolateral bridging bone on CT in combination with $<5°$ angular motion and <2mm translation on dynamic radiographs. In this randomized controlled trial (RCT) including 19 single-level lumbar fusions, pedicle screw instrumentation was removed when the radiological fusion criteria were met at 1 year follow-up (in 16 patients). The surgical fusion rate among these patients was only 69%.[54]

Spirig et al. evaluated the accuracy of pedicle screw loosening on radiographs, CT and magnetic resonance imaging (MRI) in 41 patients (159 screws) by intraoperative quantification of screw hold using a torque meter. At surgical exploration, 34% of the screws were considered loose. Sensitivity for detecting screw loosening was 54% for radiographs, 52 - 65% for CT, and 35 - 44% for MRI. The specificity of these imaging modalities was 84%, 94 - 97% and 77 - 92% respectively.[57]

Inter- and intraobserver reliability

Two studies investigated the reliability of the Christensen classification on static radiographs.[44,60] The RCT on instrumented (n = 43) vs. non-instrumented (n = 36) posterolateral lumbar fusion with 1-year follow-up by Christensen et al. reported fair to good agreement among 3 observers (mean 87% (range 83 - 93%), kappa = 0.63 (range 0.44 - 0.70)) and excellent intraobserver agreement based on a 8-week interval (mean 93%, kappa = 0.81).[44] Based on a prospective cross-sectional study including 20 patients with at least 2-year follow-up after instrumented posterolateral lumbar fusion, mean agreement between 6 observers with different levels of expertise was $76 \pm 7.8\%$ (kappa = 0.07 - 0.50). Mean intraobserver agreement, based on 4 observers and an interval of 8 weeks, was $63 \pm 10\%$ (kappa = 0.06 - 0.26). This study also reported on the interobserver agreement ($78 \pm 9.1\%$ (kappa = -0.08 - 0.50)) and intraobserver agreement ($84 \pm 10\%$ (kappa = 0.2 - 0.73)) for the fusion criterion $<5°$ difference in Cobb's angle on dynamic radiographs.[60]

The previously described CT-based accuracy study by Carreon et al. reported moderate agreement between 3 observers for facet fusion (kappa = 0.42) and substantial agreement for posterolateral gutter fusion (kappa = 0.62).[55] The 3 studies that investigated radiolucency or osteolysis around the pedicle screws as sign of pseudoarthrosis showed poor to excellent reliability (**Table 2.6**).[57–59]

Reliability measures reported as secondary outcome by 13 articles from step 1, ranging from good to excellent, are summarized in **Appendix 3**.

Discussion

Based on the intended outcome of spinal fusion surgery, postoperative assessment of the bony fusion status is imperative. Although open surgical exploration is considered the gold standard for fusion assessment, this method is obviously not feasible for routine follow-up, so clinicians and researchers have to rely on noninvasive diagnostic imaging. There is, however, no consensus on the definition and assessment of successful fusion. As a consequence, reported fusion rates are based on a variety of imaging modalities, criteria and cut-off values, often with limited clinical evidence. In fact, 4 frequently cited studies on the diagnostic accuracy of radiological fusion assessment could not be included in the current review, as they did not describe criteria for successful fusion.[61–64] While in clinical practice the diagnosis of symptomatic nonunion is most relevant, clinical trials on spinal fusion surgery largely rely on fusion rates. These applications generally ask for a different approach, as absence of fusion does not imply nonunion and vice versa. In an attempt to advance the determination of successful posterolateral lumbar fusion, we systematically reviewed the employed radiological criteria, as well as their diagnostic accuracy and reliability.

Not surprisingly, the presence of a continuous bony bridge between adjacent vertebrae was the most commonly used criterion for successful fusion, but the terminology for the quality of bony bridging and assessed anatomical location varied greatly. Although this review focused on instrumented fusions, many authors (42%) assessed angular or translational motion using dynamic radiographs, which has been shown to be an inaccurate method.[52,63] Besides that absence of any intervertebral movement does not necessarily correspond with solid fusion, it is unclear how dynamic radiographs can identify nonunion in the presence of rigid instrumentation. Other recognized pitfalls of dynamic assessment include the lack of normative data, the wide range of used cut-off values, and measurement error (during both image capture and analysis).[65–68] Interestingly, the upper limits of <3mm translational and $<5°$ angular motion are accepted for successful lumbar fusion by the United States Food and Drug Administration (US FDA) guidance for the evaluation of spinal systems, without any reference.[69]

The frequent use of the Lenke classification and Christensen classification (for both radiographs and CT scans) is supported by their good to excellent reliability for experienced observers, although evidence of diagnostic accuracy is very limited. While both classifications look for a continuous bony bridge, the Lenke classification also includes quality of the fusion mass and distinction between unilateral and bilateral fusion.

The ideal method for noninvasive fusion assessment is relevant, reliable, accurate, cost-effective and without radiation. One of the challenges in the search for the most

optimal method concerns the investigation of diagnostic accuracy. Since rigid instrumentation is typically not removed, surgical exploration of the fusion mass is often limited to symptomatic patients that qualify for revision surgery. This implies selection bias. Nevertheless, it would be interesting to apply different classifications or criteria to the same set of radiological images and compare the outcomes with surgical exploration, to determine which assessment method is most appropriate.[55] So far, such studies mainly focused on the comparison of different imaging modalities.[26]

While this systematic review showed that the use of CT increased over time, static radiographs remained the predominant imaging modality for fusion assessment. This might be explained by the retrospective design of many included articles. Although CT is suggested to be more adequate than plain radiographs, superiority of this imaging modality has not been established. This is because sensitivity and specificity values (for detecting fusion or nonunion) in the relatively old accuracy studies (published more than a decade ago) vary widely, and agreement between radiological fusion and surgical exploration is generally below 80%.[26,27,38,51,52,56,57,61,63,64] Image quality has been enhanced greatly by the rapid advancement of CT techniques, such as helical scanning, improved resolution, multiplanar reconstructions and reduced artifacts from implants. Therefore we believe that modern CT scanning is currently the most optimal tool to assess the structural integrity of the posterolateral fusion mass.

In spite of the extensiveness of this systematic review, some limitations are recognized. The full-text of many non-English articles could not be retrieved, resulting in the sole inclusion of English articles. Thereby, fusion assessment criteria that are only used in specific world regions such as Asia might have been missed. Given the large number of articles, data extraction was limited to the fusion assessment method, i.e. data about the study design, sample size and population were not considered.

Recommendations

For evaluation of posterolateral bony bridging and exclusion of signs of nonunion in daily practice, plain radiographs remain the first choice. In the presence of rigid instrumentation, dynamic radiographs seem to have no added value. In case of suspected symptomatic pseudoarthrosis or the need for a better understanding of the quality of fusion, modern CT is recommended as it offers much more detail.

Despite superiority in fusion assessment could not be demonstrated by this systematic review, thin-slice CT with multiplanar reconstructions is considered most appropriate for clinical studies with radiographic fusion as primary outcome. For systematic assessment of posterolateral bony bridging, classification systems like the Lenke and Chris-

tensen classification showed good reliability, but the terminology that is widely used for bony bridging can be subjective. Moreover, discrimination between intertransverse and facet fusions, and whether uni- or bilateral fusion is considered successful, has been shown to be relevant. Therefore, we propose to use a systematic approach that directs specific anatomical locations to be assessed in multiple planes and allows grading of the quality of the fusion mass at each side of each fusion level. Whether this should be done in perpendicular or reconstructed planes and what terminology is most appropriate remains to be studied. We also recommend to include signs of nonunion in the classification as these preclude solid fusion. The findings from this systematic review are, however, not conclusive which signs and assessment method are most predictive.

As long as there is no reliable and generally accepted method, multiple experienced and trained assessors and consensus meetings are needed. Whenever relevant and possible, observers should be blinded to the treatment aspects under investigation. Interobserver reliability should be reported by means of percentage of agreement and kappa statistics. Following tremendous technological advancements in the automated classification of degenerative discs, spinal deformities and spinal fractures, we also expect the assessment of spinal fusion to profit from artificial intelligence.[70–72]

Further research

This systematic review has demonstrated that none of the available criteria for noninvasive assessment of the fusion status after instrumented posterolateral fusion have both sufficient accuracy and reliability. Further elaboration of a well-defined and detailed systematic assessment method is a prerequisite for both the clinical and research field. Recognizing the limited feasibility of diagnostic studies with surgical exploration as reference standard, further research should be directed at the accuracy of promising fusion assessment methods using the current generation of CT scanning and implants, as well as the location of radiological fusion that correlates best with a true solid fusion. Although it seems logical to assess the intended location of fusion, i.e. often between the intertransverse processes, several studies reported on the relevance of facet joint fusions.[55,73,74] Another aspect that is underexposed, related to the quality of the fusion mass, is the distinction between (potential) ongoing bone formation and nonunion. Next to CT, the applicability of radiation-free imaging modalities like MRI and ultrasound should be continued to explored.[53,75]

Appendix 1: Medline search strings

Search 1 (24 November 2018):

(Back[Mesh] OR "lumbar vertebrae"[Mesh] OR lumbosacral [Tiab] OR lumbar [Tiab] OR intervertebral[Tiab]) AND ("spinal fusion" [Mesh] OR "spinal fusion"[Tiab] OR "spine surgery"[Tiab] OR "spine procedure"[Tiab] OR "spine procedures"[Tiab] OR "lumbar surgery"[Tiab] OR "lumbar spinal fusion"[Tiab] OR "spine fusion"[Tiab] OR spondylodesis[Tiab] OR "spinal arthrodesis"[Tiab] OR "posterolateral fusion"[Tiab] OR"posterior lumbar fusion"[Tiab] OR "posterior fusion"[Tiab] OR "posterior spinal fusion"[Tiab] OR "posterolateral lumbar spine fusion"[Tiab] OR "posterior spine fusion"[Tiab] OR "posterolateral spine fusion"[Tiab] OR "interbody fusion"[Tiab] OR "anterior spinal fusion"[Tiab] OR "anterior spine fusion"[Tiab] OR "anterior interbody fusion"[Tiab] OR "anterior lumbar interbody fusion"[Tiab] OR "transforaminal lumbar interbody fusion"[Tiab] OR "transforaminal interbody fusion"[Tiab] OR "lateral lumbar interbody fusion"[Tiab] OR "lateral interbody fusion"[Tiab] OR "posterior interbody fusion"[Tiab] OR "posterior lumbar interbody fusion"[Tiab] OR "extreme lateral interbody fusion"[Tiab] OR "extreme lateral lumbar interbody fusion"[Tiab] OR ALIF [Tiab] OR TLIF[Tiab] OR LLIF[Tiab] OR PLIF[Tiab] OR XLIF[Tiab]) AND (Radiculopathies[Tiab] OR Radiculopathy[Tiab] OR "Nerve root"[Tiab] OR radiculitis[Tiab] OR radiculitides[Tiab] OR radiating[Tiab] OR Radicular[Tiab] OR "spinal stenosis"[Mesh] OR "spinal stenosis"[Tiab] OR stenosis[Tiab] OR spondylolisthesis[Mesh] OR spondylolisthesis[Tiab] OR "low back pain"[Mesh] OR "low back pain"[Tiab] OR "back pain"[Tiab] OR sciatica[Mesh] OR sciatica[Tiab] OR sciaticas[Tiab] OR sciatic[Tiab] OR spondylosis [Mesh] OR spondylosis[Tiab] OR "intervertebral disc degeneration"[Mesh] OR "intervertebral disc degeneration"[Tiab] OR "degenerative disc disease"[Tiab] OR "degenerative disc disorder"[Tiab] OR "spinal degeneration"[Tiab] OR facetarthrosis[Tiab]) OR "bone substitutes"[Mesh] OR "Bone transplantation"[Mesh] OR "graft rejection"[Mesh] OR "graft survival"[Mesh] OR transplantation, autologous"[Mesh] OR allograft[Mesh] OR "absorbable collagen sponge[Tiab] OR "autologous iliac crest bone"[Tiab] OR "bone marrow aspirate"[Tiab] OR coralline [Tiab] OR "compression-resistant matrix" [Tiab] OR "demineralized bone matrix"[Tiab] OR DBM[Tiab] OR "femoral ring allograft"[Tiab] OR hydroxyapatite[Tiab] OR "iliac crest bone graft[Tiab] OR "osteogenic protein"[Tiab] OR bonegraft[Tiab] OR "bone substitute"[Tiab] OR "bone morphogenic protein"[Tiab] OR "recombinant human bone morphogenetic protein"[Tiab] OR "tricalcium phosphate"[Tiab] OR autograft[Tiab] OR "bone graft substitutes[Tiab] OR "bone graft alternatives"[Tiab] OR "fusion extenders"[Tiab] OR ceramics[Tiab] OR "calcium sulphate"[Tiab] OR "tricalcium sulphate"[Tiab] OR "autologous growth factors"[Tiab] OR "AGF peptides"[Tiab] OR "stem cells"[Tiab] OR rhBMP-2[Tiab] OR rhBMP-7 OR OP-1[Tiab] OR "synthetic peptides"[Tiab]) AND (Osseointegration[Mesh] OR osseointegration[Tiab] OR "bony fusion"[Tiab] OR "spinal fusion/adverse effects"[Mesh] OR "spinal fusion/classification"[Mesh] OR fusing[Tiab] OR "fusion rate"[Tiab] OR "fusion rates"[Tiab] OR "radiographic fusion"[Tiab] OR "radiological fusion"[Tiab] OR "successful fusion"[Tiab] OR nonunion[Tiab] OR pseudoarthrosis[Tiab] OR pseudoarthrosis[Tiab] OR pseudarthrosis[Tiab] OR "bone union rate"[Tiab] OR "fused"[Tiab]) NOT (animals [Mesh] NOT humans [Mesh])

Search 2 (19 July 2021):

(Lumbar[Tiab] OR “Lumbar vertebrae”[Mesh] OR “lumbosacral region”[Mesh]) AND (“spinal fusion”[Mesh] OR “spinal fusion”[Tiab] OR “spine fusion”[Tiab] OR spondylodesis[Tiab] OR “spinal arthrodesis”[Tiab] OR “posterolateral fusion”[Tiab] OR “posterior lumbar fusion”[Tiab] OR “posterior fusion”[Tiab] OR “posterior spinal fusion”[Tiab] OR “posterolateral lumbar fusion”[Tiab] OR “posterior spine fusion”[Tiab] OR “posterolateral spine fusion”[Tiab] OR “posterolateral spinal fusion” OR “interbody fusion”[Tiab] OR “anterior spinal fusion”[Tiab] OR “anterior spine fusion”[Tiab] OR “anterior interbody fusion”[Tiab] OR “anterior lumbar interbody fusion”[Tiab] OR “transforaminal lumbar interbody fusion”[Tiab] OR “transforaminal interbody fusion”[Tiab] OR “lateral lumbar interbody fusion”[Tiab] OR “lateral interbody fusion”[Tiab] OR “posterior interbody fusion”[Tiab] OR “posterior lumbar interbody fusion”[Tiab] OR “extreme lateral interbody fusion”[Tiab] OR “extreme lateral lumbar interbody fusion”[Tiab] OR ALIF[Tiab] OR TLIF[Tiab] OR LLIF[Tiab] OR PLIF[Tiab] OR XLIF[Tiab]) AND (“lumbar vertebrae/diagnostic imaging”[Mesh] OR “Lumbosacral Region/diagnostic imaging”[Mesh] OR “Spine/diagnostic imaging”[Mesh] OR “pseudarthrosis/diagnostic imaging”[Mesh] OR radiography[MesH] OR “Tomography, X-ray”[Mesh] OR “Tomography, X-ray Computed”[Mesh] OR “X-ray”[Tiab] OR radiograph*[Tiab] OR “computed tomography”[Tiab] OR CT[Tiab] OR DEXA[Tiab] OR “Dual Energy X-ray Absorptiometry”[Tiab] OR SPECT[Tiab] OR “Single Photon Emission Computed Tomography”[Tiab] OR PET[Tiab] OR “Positron Emission Tomography”[Tiab] OR MRI[Tiab] OR “Magnetic Resonance Imaging”[Tiab] OR Absorptiometry, Photon[MesH] OR Tomography, Emission-Computed, Single-Photon[MesH] OR Positron Emission Tomography Computed Tomography[MesH] OR Magnetic Resonance Imaging[MesH]) AND (“Predictive value of tests”[Mesh] OR “Sensitivity and Specificity”[Mesh] OR “reference standards”[Mesh] OR “reproducibility of results”[Mesh] OR “observer variation”[Mesh] OR reliability[Tiab] OR “diagnostic accuracy”[Tiab] OR agreement[Tiab] OR “predictive value”[Tiab] OR “sensitivity”[Tiab] OR “specificity”[Tiab] OR “false positive”[Tiab] OR “false negative”[Tiab] OR “Surgical exploration”[Tiab] OR “direct observation”[Tiab] OR “surgically explored”[Tiab]) AND (solidity[Tiab] OR Osseointegration[Mesh] OR osseointegration[Tiab] OR “bony fusion”[Tiab] OR “spinal fusion/classification”[Mesh] OR fusing[Tiab] OR “fusion rate”[Tiab] OR “fusion rates”[Tiab] OR “radiographic fusion”[Tiab] OR “radiological fusion”[Tiab] OR “successful fusion”[Tiab] OR “solid fusion”[Tiab] OR nonunion[Tiab] OR pseudoarthrosis[Tiab] OR pseudarthrosis[Tiab] OR “bone union rate”[Tiab] OR “fused”[Tiab])

Appendix 2: Descriptive criteria from step 1

Table 2.7: Descriptive criteria identified in step 1 (based on 118 articles) and the absolute frequency of reporting. For the criteria related to absence of motion, also the used cut-off values are summarized.

Category	Criteria	Absolute frequency of articles
Continuity of bony bridging	Facet joint	9
	Interbody	7
	Intertransverse	52
	Posterolateral	19
	Not specified	35
Absence of motion	Cut-off values for angular motion	
	$\leq 1.5°$	4
	$\leq 2°$	6
	$\leq 3°$	4
	$\leq 4°$	7
	$\leq 5°$	22
	Miscellaneous cut-off values[a]	3
	Cut-off values for translational motion	
	≤ 2mm	9
	≤ 3mm	11
	$\leq 3°$	2
	Miscellaneous cut-off values[a]	2
	Cut-off value not specified	2
Absence of static signs of nonunion	Implant failure / loosening / migration	21
	Radiolucency around the implant	14
	Radiographic gap / cleft / line in fusion mass	9
Miscellaneous		3

[a] Cut-off values reported in a single article.

Appendix 3: Summary of reliability measures from step 1

Table 2.8: Reliability measures reported as secondary outcome by articles included from step 1.

Article	Fusion assessment	Reliability measures
Thomsen et al. (1997)[76]	Static X: continuous intertransverse bony bridges at one of the two sides, at all intended levels (2 observers).	Interobserver reliability = 82% (kappa = 0.55)
Molinari et al. (1999)[77]	Static X: Bridwell classification (2 observers).	Interobserver agreement = 99%
Möller et al. (2000)[78]	Static X: Lenke classification, grade A (2 observers).	Interobserver agreement = 88%
Korovessis et al. (2005)[79]	Static X: Christensen classification, grade 3 (2 observers, reassessment within 3 weeks).	Interobserver reliability: r = 0.71 Intraobserver reliability: r = 0.67
Sengupta et al. (2006)[80]	Presence of a solid fusion mass (static X), absence of halo around the implant (static X) and absence of motion (dynamic X). Assessed twice by 1 observer with 3 months interval.	Intraobserver reliability: kappa = 0.7 (good)
Yu et al. (2008)[81]	Static X: continuity in the fusion mass between the cephalad and caudal transverse process (2 observers).	Interobserver agreement = 92% (kappa = 0.623)
Acebal-Cortina et al. (2011)[82]	Static X: unnamed classification, type A and/or B (2 observers). Type A: Bilateral, uniform and continuous intertransverse mass; B: Unilateral uniform and continuous intertransverse mass with discontinuous, irregular or absent intertransverse counterlateral mass; C: Discontinuous, irregular or absent bilateral intertransverse mass.	Interobserver reliability: kappa = 0.80 (excellent)
Korovessis et al. (2012)[83]	CT: Christensen classification, grade 2 and 3 (2 observers).	Interobserver reliability: kappa = 0.85-0.89 Intraobserver reliability: kappa = 0.88-0.91
Yamada et al. (2012)[84]	CT: presence of continuous bone connecting the transverse processes, found by 2 observers.	Interobserver reliability: kappa = 0.69 (good)

Table 2.8: *(continued)*

Article	Fusion assessment	Reliability measures
Hurlbert et al. (2013)[85]	Bridging trabecular bone in both fusion masses (left and right intertransverse spaces) (static X), absence of traversing radiolucent lines (static X), $\leq$3mm translation (dynamic X) and $<5°$ angulation (dynamic X). 2 principle investigators assessed 50 selected cases to determine concordance with the 2 observers.	Interobserver agreement = 88-92%
Park et al. (2013)[86]	CT and dynamic X: continuous intertransverse bony bridging at the target level (assessed twice by 1 observer with 1 month interval).	Intraobserver reliability: ICC = 0.908 (good)
Guerado et al. (2016)[87]	CT and static X: unnamed classification (3 observers). Bone continuity: Full central trabecular continuity and partial trabecular continuity (Grade I and II of criteria by Tan et al.); No bone continuity: unipolar non-union denoting superior or inferior non-union of the central allograft with partial trabecular discontinuity centrally, and bipolar pseudoarthrosis. (Grade III and IV of criteria by Tan et al.).	Interobserver reliability (CT): kappa > 0.95 Intraobserver reliability (CT): kappa > 0.95
Cho et al. (2017)[88]	CT and static X: Glassman classification, grades 2-5 (2 observers, reassessment after 1 month).	Interobserver reliability 1st rating: ICC = 0.785 (moderate) Interobserver reliability 2nd rating: ICC = 0.748 (moderate) Intraobserver reliability: ICC = 0.802-0.836 (good)
Choi et al. (2017)[89]	Dynamic X: presence of bone trabeculation, no observed motion between the graft and instrumentation and lack of instrumentation loosening or migration (3 observers).	Interobserver reliability: Fleiss kappa = 0.78

X = radiograph, CT = computed tomography, r = Pearson correlation coefficient, ICC = intraclass correlation coefficient.

CHAPTER 3

1-year efficacy of a standalone microporous biphasic calcium phosphate ceramic vs. autograft in instrumented posterolateral (thoraco)lumbar fusion

A.M. Lehr
F.C. Oner
D. Delawi
R.K. Stellato
E.A. Hoebink
D.H.R. Kempen
J.L.C. van Susante
R.M. Castelein
M.C. Kruyt
on behalf of the Dutch Clinical Spine Research Group

Efficacy of a standalone microporous ceramic vs. autograft in instrumented posterolateral spinal fusion - A multicenter, randomized, intrapatient controlled, noninferiority trial
Spine 2020;45:944-951

doi: 10.1097/BRS.0000000000003440

Introduction

Spinal fusion is one of the most commonly performed surgical treatments for various conditions requiring stabilization of the vertebral column. Over the past two decades, the annual number of spinal fusions in the United States increased almost threefold to about 500,000.[5,90] Autologous iliac crest bone graft is considered the gold standard to establish a bony fusion, as it possesses natural osteoconductive, osteoinductive and osteogenic properties. However, the need for an additional surgical procedure to harvest the bone graft and relatively limited availability are recognized as the main drawbacks of using autograft.[7] To overcome these shortcomings, numerous biological and synthetic bone graft extenders and substitutes have been developed and marketed.[8,9,23]

Since the 1970s, calcium phosphate (CaP) based synthetic ceramics including hydroxyapatite (HA), β-tricalcium phosphate (β-TCP) and biphasic calcium phosphate (BCP) have been investigated extensively as their composition and properties are similar to the inorganic component of bone. Moreover, these materials are nonimmunogenic, unlimited in supply, easy to sterilize and store and relatively cheap. In addition to excellent osteoconductivity and modifiable bioresorbability, a small subclass of CaP biomaterials with specific physicochemical properties have been demonstrated to possess intrinsic osteoinductive properties in different animal models.[11,13,19,20,91] Orthotopically, these osteoinductive CaP's have been shown to perform superior to noninductive materials due to the stimulation of osteogenic differentiation and enhanced osteoconduction.[17] The latter is important for bony bridging in posterolateral spinal fusions. Continued development resulted in a microporous BCP (>90% β-TCP/<10% HA) with a high specific surface area and controlled resorption rate that showed favorable bone formation comparable to autograft in multiple preclinical studies.[14,16,18,22] Mixed with a fast resorbing polymer carrier to improve surgical handling, this product (BCP$_{<\mu m}$) is commercially available as AttraX® Putty (NuVasive Inc., San Diego, CA, USA). However, no clinical studies evaluating efficacy in spinal fusions are available.

The present clinical study aimed to determine noninferiority of BCP$_{<\mu m}$ as a bone graft substitute for autograft in instrumented posterolateral fusion in the thoracolumbar spine. This article reports on the efficacy and safety of BCP$_{<\mu m}$ compared to autograft in promoting fusion at 1 year follow-up.

Materials and methods

Study design

This study is a patient and observer blinded, multicenter, randomized, noninferiority trial with intrapatient comparisons (ClinicalTrials.gov NCT01982045). After approval by the Medical Ethics Review Committee of the University Medical Center Utrecht, Utrecht, The Netherlands, and local Institutional Review Boards, it was conducted in four Dutch hospitals in accordance with the principles of the Declaration of Helsinki (Version October 2008) and the Medical Research Involving Human Subjects Act. Based on computerized simple randomization, one side of each fusion trajectory was grafted with BCP$_{<\mu m}$. The contralateral side was treated with autograft, so each patient received both treatments and served as his/her own control. The primary efficacy outcome was posterolateral fusion at 1 year follow-up assessed on CT scans. Fusion performance of BCP$_{<\mu m}$ was tested with a noninferiority margin of 15%. Safety was evaluated by analysis of (serious) adverse events.

Patients

Patients between 18 and 80 years of age scheduled for a primary single or multilevel instrumented posterolateral fusion between T10 and S1 were considered eligible for this study. Indications for surgery were deformity, structural instability and/or expected instability (for example as a result of decompression for spinal stenosis). Additional inclusion and exclusion criteria are provided in **Table 3.1**.

Surgical technique

All patients underwent a single or multilevel posterolateral fusion with pedicle screw instrumentation through a posterior midline approach. When indicated, decompression and/or an additional interbody fusion procedure with local bone were performed. After placement of all instrumentation and thorough preparation of the grafting side by decortication, the randomized allocation side (left/right) of BCP$_{<\mu m}$ was disclosed by opening a sealed opaque envelope.

For autografting, corticocancellous bone was harvested from a single posterior iliac crest.[92] Both local decompression bone and iliac crest bone were morselized. To match the unilateral use of 10cc of BCP$_{<\mu m}$, a volume of 8 - 10cc autograft per fusion level was intended. The autograft condition consisted of a mixture of available local bone and at least 50% iliac crest bone. In case a total volume of 8cc per fusion level could not be reached, this was accepted as a consequence of autografting. Graft volumes were assessed by slight compression in a 20cc syringe.

Both grafts were placed at the allocated side around the posterior instrumentation and in the decorticated lateral gutters, bridging the dorsal surfaces of the transverse processes, facets and laminae. The wound was closed in layers, followed by standard postoperative care.

Table 3.1: Inclusion and exclusion criteria.

Inclusion criteria
1. Eligible for single or multilevel instrumented posterolateral thoracolumbar spinal fusion in the T10 to S1/ilium region, with or without additional posteriorly inserted interbody devices (PLIF, TLIF), because of deformity, structural instability and/or expected or potential instability
2. Nonresponsive to $\geq$6 months of nonoperative treatment
3. Between 18 and 80 years of age
4. Informed consent
Exclusion criteria
1. Previous fusion attempt(s) at indicated level(s)
2. Previous treatments that compromise fusion surgery
3. Previous autologous bone grafting procedures that compromise the amount of iliac crest bone graft
4. Traumatic instability
5. Active local and/or systemic infection
6. Spinal metastasis at indicated level(s)
7. Systemic disease or condition affecting the ability to participate in study requirements or to evaluate graft efficacy
8. Risk for noncompliance
9. Participation in clinical trials evaluating investigational devices, pharmaceuticals or biologics <3 months of enrollment
10. Intended pregnancy <1.5 year of enrollment
11. Body mass index (BMI) >35
12. Expected to require additional surgery to the same spinal region <6 months
13. Current or recent (<1 year) corticosteroid use equivalent to prednisone $\geq$5 mg/day, prescribed for >6 weeks

Outcome measures

Clinical and radiographic assessments were done preoperatively and at 6 weeks, 3 months, 6 months and 1 year postoperative for evaluation. Patient reported outcomes measures (PROMs) included the Visual Analogue Scale (VAS) for back pain, ranging from 0 (“no pain at all”) to 100 (“intolerable pain”), Oswestry Disability Index (ODI) and EQ-5D-5L. The condition-specific ODI ranges from 0% to 100%, with higher scores indicating more functional disability related to low back pain.[93] A score of $\leq$22% indicates a satisfactory symptom state.[94] Generic health status was measured with the EQ-5D-5L and converted into a single index value ranging from -0.329 (worst health state) to 1.000 (full health).[95]

Fusion assessment

For the primary efficacy outcome, CT scans with a slice thickness of $\leq$ 1mm and multiplanar reconstructions were obtained at 1 year follow-up. Posterolateral fusion was evaluated individually by two spine surgeons blinded to the treatment sides using a protocol based on Christensen et al.[44] and Carreon et al.[55] (**Table 3.2**). Each side of each fused segment was assessed in three planes for intertransverse fusion and/or fusion around the rod including facet fusion and scored as fusion, doubtful fusion or nonunion. Interbody fusion was scored similarly in two planes. CT scans with disagreements were reassessed to reach consensus. For statistical analyses, the posterolateral fusion scores of each segment and side, as well as the scores for interbody fusion, were dichotomized into "fused" (fusion) and "not fused" (doubtful fusion or nonunion).

Table 3.2: Fusion assessment method.

Scoring	Definition
Fusion	Continuous bony bridge from one vertebra to the other, in the absence of any secondary signs of nonunion such as fracture or loosening of the screws or rods
Doubtful fusion	Doubts about continuity or quality of the bony bridge
Nonunion	Definite discontinuity or lack of a fusion mass, as well as obvious indications of mobility like material failure or apparent pseudoarthrosis

Safety evaluation

To evaluate safety, adverse events were registered until last follow-up and evaluated for any potential relation with BCP$_{<\mu m}$. Adverse events were defined as any unexpected, undesirable medical experience occurring to a subject during the study, whether or not considered related to BCP$_{<\mu m}$. Events were classified as serious when they resulted in death, were life-threating, required hospitalization or prolongation of existing hospitalization and/or resulted in persistent or significant disability or incapacity.

Statistical methods

This study was powered based on an estimated unilateral fusion rate of 50% and 70% concordance between the left and right side of the fusion trajectory.[50,84,96,97] Weighing the disadvantages of autografting against the consequences of less successful fusions at the BCP$_{<\mu m}$ side, the noninferiority margin was set at an absolute difference of 15%. With a desired power of 80% and one-sided significance level of 0.05, a minimum

sample size of 84 patients was estimated. Assuming that approximately 15% of the patients would not be evaluable for the primary efficacy analysis, e.g. due to revision surgery or lost to follow-up, the total number of patients to be treated was set at 100.

Study data were processed in Research Online for Researchers (Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands) and analyzed using SPSS Statistics Version 22 (IBM Corp., Armonk, NY, USA). Baseline characteristics, surgical details, PROMs and fusion rates on segment level were summarized using descriptive statistics. The VAS for back pain and ODI at baseline and 1 year follow-up were compared with paired samples t-test ($p < 0.05$) and a minimal clinically important difference (MCID) of 15 points was adopted for both outcome measures.[93,98,99] Missing values were handled by pairwise deletion of cases.

Interobserver reliability of fusion assessment was measured by percentage agreement and Cohen's kappa. To examine fusion on segment level, while accounting for clustering of fusion scores within segments and within patients, a three-level Generalized Estimating Equations (GEE) model with an independent correlation structure and treatment condition as predictor was used. The relation between successful interbody fusion and posterolateral fusion on either or both sides was analyzed using a similar two-level GEE model with spinal level and interbody fusion as predictors. For both models, the significance level was $p = 0.05$. Odds Ratios (OR) along with their 95% confidence interval (CI) are reported.

For the primary efficacy analyses, a posterolateral fusion performance score per treatment condition was calculated to correct for multilevel fusions. This score was based on the number of fused segments compared to the contralateral side. Noninferiority of BCP$_{<\mu m}$ vs. autograft was tested against the upper limit of the two-sided 90% CI around the difference in paired proportions for successful posterolateral fusion performance, corresponding to a one-sided significance level of 0.05.

Results

Patient characteristics

Between November 2013 and July 2016, 108 patients gave written informed consent (minimal 18 patients per center). Patients withdrawn or excluded prior to randomization (n = 8) were replaced to reach the target sample size 100 treated patients (**Figure 3.1**).

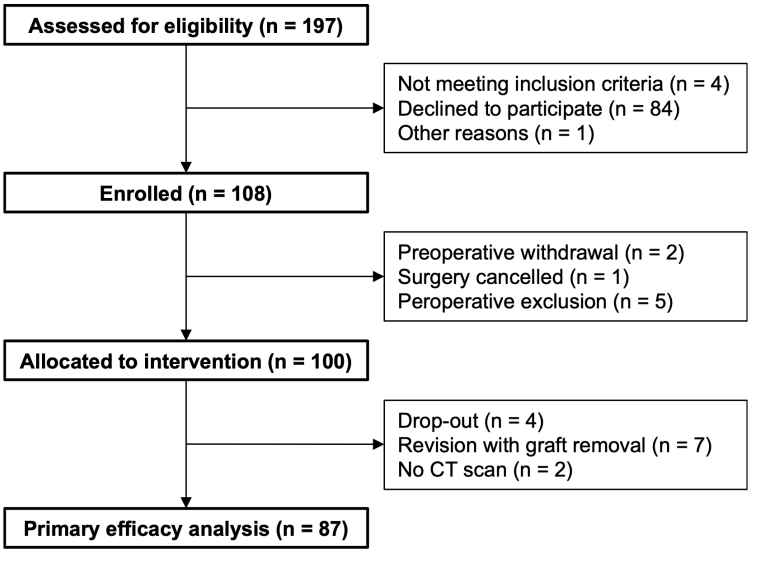


Figure 3.1: Flowchart of the progress of patients through each stage of the study.

Baseline characteristics and surgical details are summarized in **Table 3.3**. There were 49 males and 51 females with a mean age of 55.4 ± 12.0 (range 27 - 79) years. The majority of the patients (66%) underwent a single-level fusion. The total number of instrumented segments was 172 and 71 additional interbody fusion procedures were performed in 62 patients. The intended mixture and volume of local bone and iliac crest bone for autografting was reached in 93 patients.

Table 3.3: Baseline characteristics and surgical details (n = 100).

Age, mean ± SD (range), years	55.4 ± 12.0 (27 - 79)
Sex, n (%)	
Male	49 (49%)
Female	51 (51%
Smokers, n (%)	34%
Indication(s) for surgery, n (%)	
Deformity	12 (12%)
Degenerative lumbar scoliosis	8
Post-traumatic kyphosis	3
Post-laminectomy deformity	1
Structural instability	45 (45%)
Expected instability	60 (60%)
Missing	7 (7%)

ASA classification, n (%)		
I	28 (28%)	
II	62 (62%)	
III	7 (7%)	
Missing	3 (3%)	
Number of segments fused, n (%)		
1	66 (66%)	
2	20 (20%)	
>2	14 (14%)	
Median number of segments fused (range)	1 (1 - 8)	
Spinal region fused, n (%)		
Thoracic	1 (1%)	
Thoracolumbar	7 (7%)	
Lumbar	92 (92%)	
Decompression, n (%)	95 (95%)	
Interbody device, n (%)	62 (62%)	
Level and type of interbody device, n	PLIF	TLIF
L3-L4	5	6
L4-L5	21	7
L5-S1	26	6
Iliac crest bone graft, median (range), cc	8 (5 - 40)	
Operative time, n (%)		
<2 hours	25 (25%)	
2 - 4 hours	61 (61%)	
>4 hours	14 (14%)	
Blood loss, median (range), cc	450 (50 - 1500)	
Length of stay, median (range), days	5 (2 - 35)	

n = number of patients, PLIF = posterior lumbar interbody fusion, SD = standard deviation, TLIF = transforaminal lumbar interbody fusion

The main clinical outcomes are presented in **Figure 3.2**. At 1 year follow-up, both the VAS for back pain and ODI improved from baseline, with a mean decrease of 28 ± 30 and 21 ± 19 points respectively ($p < 0.001$). In more than half of the patients the improvement was above the MCID (VAS 60%, ODI 61%). Moreover, 60% reached an ODI $\leq$ 22% at 1 year follow-up. The EQ-5D index value improved from median 0.53 (IQR 0.39 - 0.68) at baseline to 0.78 (IQR 0.69 - 0.87) at 1 year follow-up.

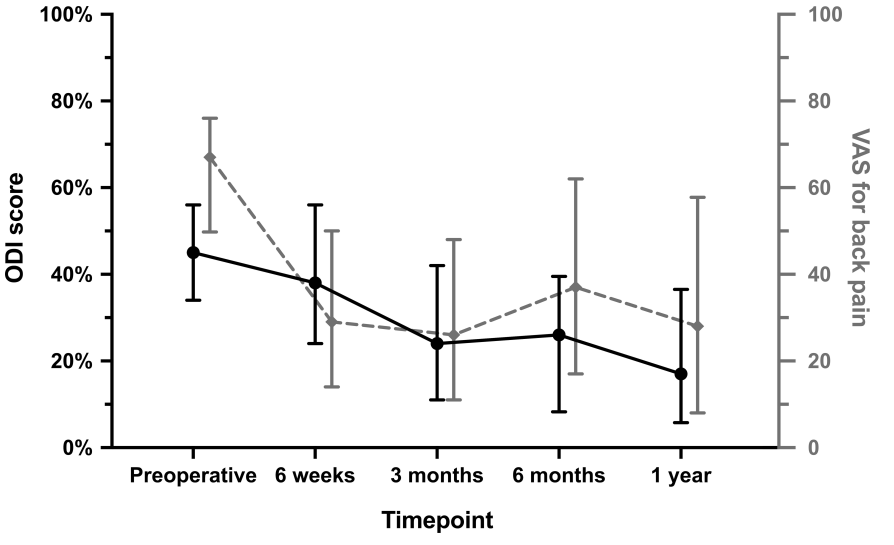


Figure 3.2: Oswestry Disability Index (ODI) (0 - 100%, in black) and Visual Analogue Scale (VAS) for back pain (0 - 100, in grey) scores at baseline and each follow-up. Median values along with their interquartile range are given as the data are not normally distributed.

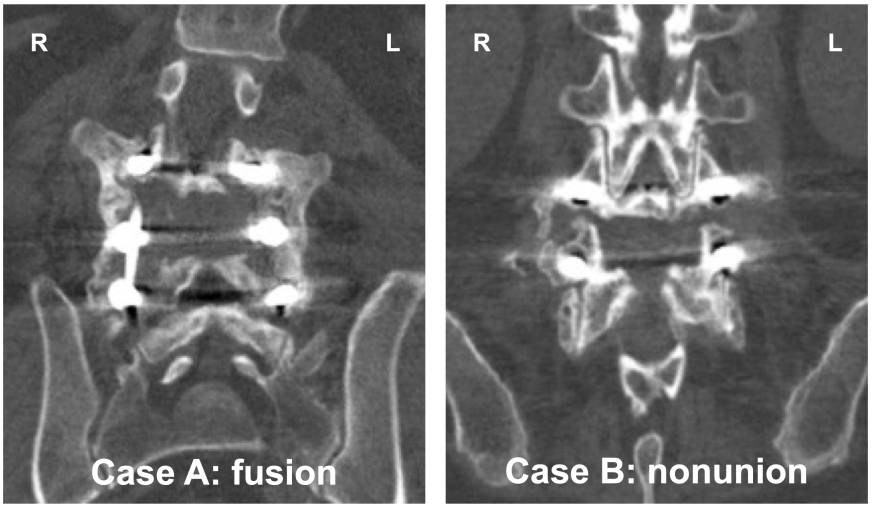


Figure 3.3: Coronal CT image at 1 year follow-up demonstrating a bilateral continuous bony bridge between the transverse processes (case A) and a nonunion (case B). In both cases, $BCP_{<\mu m}$ was applied to the left side.

Fusion assessment

Efficacy analyses of the grafts are based on 87 out of 100 patients due to circumstances mentioned in **Figure 3.1**. These included 28 multilevel fusions and 63 inter-

body fusions in 55 patients. Interobserver agreement was 72% for posterolateral fusion (kappa = 0.45) and 78% for interbody fusion (kappa = 0.56). **Figure 3.3** shows an example of a successful bilateral posterolateral fusion (case A) and a nonunion (case B).

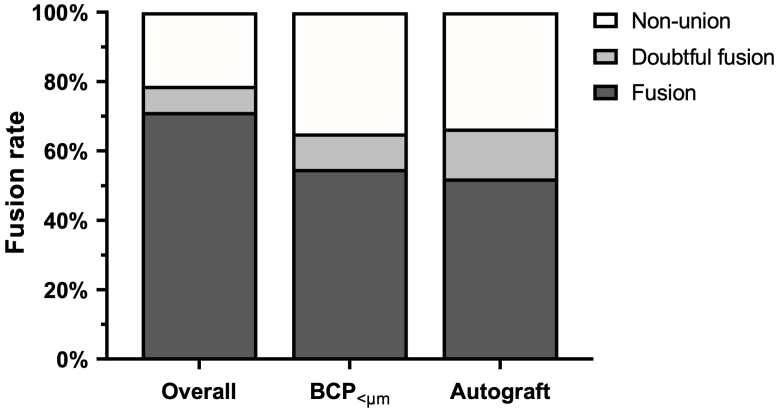


Figure 3.4: Posterolateral fusion rates on segment level. The overall fusion rate (i.e. either or both sides fused) and unilateral fusion rates per treatment condition are shown.

Of the 146 segments assessed for posterolateral fusion, 104 (71%) were scored as fused on either or both sides (**Figure 3.4**). The posterolateral fusion rate at the BCP$_{<\mu m}$ side was 55% vs. 52% at the autograft side (OR = 1.1, 95% CI = 0.7 to 1.7, p = 0.617). Concordance between left and right was 64%; 36% of the segments showed bilateral fusion whereas 29% were not fused. The interbody rate was 62%. Secondary GEE-analyses on segment level showed a positive relation between successful interbody fusion and posterolateral fusion fusion (OR = 7.3, 95% CI = 2.0 to 27.0, p = 0.003). After correction for multilevel fusions, resulting in a single posterolateral fusion performance score per treatment condition (**Table 3.4**), the CI for the absolute difference between the treatments excluded the predetermined noninferiority margin (difference = 2.3%, 90% CI = -9.1% to +13.7%).

Table 3.4: Posterolateral fusion performance per treatment condition, after correction for multilevel fusions (n = 87). The difference in paired proportions of successful fusion performance was 2.3% with a 90% confidence interval of -9.1% to +13.7%.

		Autograft		
		Not fused	**Fused**	**Total**
BCP$_{<\mu m}$	**Not fused**	21	19	40
	Fused	17	30	47
	Total	38	49	87

Safety evaluation

During the first year after surgery, 32 serious adverse events were reported in 26 patients (**Table 3.5**). Indications for resurgery with graft removal were surgical site infection (n = 4), persistent cerebrospinal fluid leakage (n = 1), screw malposition (n = 1) and cage dislocation (n = 1). Two patients had screw loosening and symptoms of pseudoarthrosis at 1 year follow-up and where therefore indicated for revision surgery. One patient died 4.5 months after surgery due to progressive amyotrophic lateral sclerosis. Furthermore, 78 adverse events were registered, ranging from wound complications to unrelated events like hip bursitis. None of the (serious) adverse events could be directly related to BCP$_{<\mu m}$.

Table 3.5: Number of serious adverse events (n = 100).

Surgical site infection	6	Gastrointestinal complications	5
Instrumentation failure	4	Malignancy	3
Symptomatic dural tear	3	Cardiovascular complications	2
Neurological complications	3	Respiratory complications	2
Prolonged wound leakage	1	Miscellaneous	3

Discussion

Over the past decades, calcium phosphate based synthetic ceramics gained popularity as alternative for autograft in spinal fusion surgery since they closely resemble natural bone. In addition to biocompatibility and osteoconductivity, third-generation CaPs have been shown to possess intrinsic osteoinductive properties due to specific physicochemical and microstructural properties. However, the clinical evidence for the standalone use of these materials as bone graft substitute is limited.[8,11,23] This patient and observer blinded, multicenter, randomized, intrapatient controlled trial demonstrated noninferiority of a microporous BCP$_{<\mu m}$ (AttraX® Putty) vs. autograft in terms of fusion performance 1 year after instrumented posterolateral (thoraco)lumbar fusion in 87 patients.

Fusion rates reported in literature vary widely, depending on the surgical technique, number of levels fused, criteria for radiographic fusion assessment and follow-up period, as well as patient factors like smoking.[1] This complicates the comparison between studies and graft materials. For the primary outcome of this study, the fusion status of both grafts was assessed on CT scans using a detailed radiographic classification system. Interobserver reliability was moderate and comparable to previous radiologi-

cal studies.[44,55,100,101] Although the observed unilateral fusion rates on segment level seem at the lower end (BCP$_{<\mu m}$ 55%, autograft 52%), the overall posterolateral fusion rate (i.e. either or both sides fused) of 71% is in accordance with literature.[44,84,96,102,103] There are indications that the process of bony fusion continues after 1 year, which may advocate a minimum follow-up of 2 years.[96,104] However, as this is most likely a result of surgical immobilization instead of graft related fusion, the primary outcome of this study was assessed 1 year after surgery. Despite this focus on graft related bony fusion it is noteworthy that, for both conditions, solid intertransverse fusions that would be undoubtedly related to grafting were limited. Many bony bridges were observed more medial, around the implants and facet joints, and after 1 year both grafts had been resorbed. Whether the grafts resorbed too fast compared to the rate of new bone formation will be the subject of further investigations. In contrast to most studies in this research field, patients with multilevel fusions (34%) were included because the real value of bone graft substitutes is mainly for those more extensive surgical procedures where bone graft volume is a limitation. Indeed in 7 patients, all of whom underwent a multilevel fusion, the intended mixture and volume of autograft could not be reached; in the 3 patients included in the primary analysis autograft (4 - 6 cc per level) performed inferior to BCP$_{<\mu m}$. To avoid overrepresentation of patients with multilevel fusions, the test for noninferiority with a margin of 15% was based on a fusion performance score per treatment condition instead of absolute fusion rates. Although more recent insights recommend a 95% CI for noninferiority tests, we followed the original statistical plan in the study protocol as registered in ClinicalTrials.gov prior to the start of the study.

The scientific investigation of the efficacy of a bone graft substitute in the challenging patient group that will benefit most from it is impeded by the variation in patient conditions, diagnosis and treatment strategies. To overcome this, we employed an intrapatient study design with each patient serving as their own control. The major advantage of this design is the elimination of interpatient variability and its numerous confounders. The concordance of 64% between the left and right side of the fusion trajectory nicely confirms this patient factor. Other factors that might play a role in the deformity cases are eliminated by randomization. Also from an ethical point of view the intrapatient design is advantageous as the clinical consequences of unexpected inferior performance of the bone graft substitute are minimized when each patient also receives the gold standard contralaterally. In the presence of rigid instrumentation, the process of bone formation on one side of the spine is not expected to be affected by the fusion status at the contralateral side.[84,96,105,106]

An obvious limitation of an intrapatient design is that clinical outcomes like PROMs and adverse events cannot be attributed separately to the treatment conditions. These outcomes were therefore mainly collected to confirm a general treatment effect as

expected based on control populations.[107–109] In an effort to evaluate safety, all unexpected, undesirable medical experiences, whether or not considered related to the spinal fusion, were registered prospectively. Based on this broad definition a total of 110 (serious) adverse events were registered in 48 patients. Two-third (68%) of these events occurred in 28 of the 37 patients treated in the academic tertiary referral spine center, reflecting its complex patient population. The types and frequencies of complications were in accordance with previous reports from prospective studies and local complication registries.[110]

In conclusion, the results of this randomized intrapatient controlled trial support the clinical use of BCP$_{<\mu m}$ as a standalone bone graft substitute for autograft in instrumented posterolateral (thoraco)lumbar fusion.

CHAPTER 4

Increasing fusion rate between 1 and 2 years after instrumented posterolateral (thoraco)lumbar fusion and the role of bone grafting

A.M. Lehr
F.C. Oner
D. Delawi
R.K. Stellato
E.A. Hoebink
D.H.R. Kempen
J.L.C. van Susante
R.M. Castelein
M.C. Kruyt
on behalf of the Dutch Clinical Spine Research Group

Increasing fusion rate between 1 and 2 years after instrumented posterolateral spinal fusion and the role of bone grafting
Spine 2020;45:1403-1410

doi: 10.1097/BRS.0000000000003558

Introduction

Since the first description by Hibbs in 1911, spinal fusion surgery has evolved into an established treatment of various spinal disorders including deformity, trauma and degenerative conditions. Over the past decades, the surgical technique has shifted from noninstrumented procedures to rigid instrumentation including pedicle screws and interbody cages.[2,3,5,111] Moreover, numerous biological and synthetic alternatives for the use of autologous bone graft have been developed.[8,9,23] Although the primary goal of spinal fusion is to obtain a solid arthrodesis, bony fusion is not for granted and success rates reported in literature vary widely.[1,3,112] Outcomes are affected by surgical factors and patient factors, as well as the method and timing of radiographic fusion assessment.[1,3,26,100,113,114] Nonunion (pseudoarthrosis) is commonly defined as a failure of bony bridging $\leq$1 year after surgery, but there are indications that bony fusion proceeds.[103,104,115] This argues for prolonged follow-up to definitely evaluate fusion status. However, whether this delayed or mid-term fusion can still be ascribed to the bone graft or results from a process of facet ankylosis due to immobilization is not yet clear.[116] Fusion mass exclusively related to the graft can most likely only be assumed in the intertransverse grafting area.

This article describes the 2-year radiographic and clinical outcomes of an intrapatient controlled noninferiority trial investigating the efficacy of a microporous biphasic calcium phosphate ceramic bone graft substitute (BCP$_{<\mu m}$; AttraX® Putty, NuVasive Inc., San Diego, CA, USA) vs. autograft for instrumented posterolateral fusion of the (thoraco)lumbar spine. We aimed to 1) compare posterolateral fusion rates between 1 and 2 years of follow-up and between graft types, and 2) explore the role of the grafts based on the location of the posterolateral fusion mass at both timepoints. Moreover, the mid-term fusion potential of additional interbody fusion cages and the relationship between radiographic and clinical outcomes were analysed.

Methods

Study design

This double-blind, multicentre, randomized, intrapatient controlled, noninferiority trial including 2-year follow-up was designed to investigate the 1-year efficacy of BCP$_{<\mu m}$ as a standalone bone graft substitute for instrumented fusion of the (thoraco)lumbar spine. This product is made of a microporous biphasic calcium phosphate.[117] The study was approved by the Medical Ethics Review Committee of the University Medical Center Utrecht, Utrecht, The Netherlands, and local board of each participating hos-

pital. The study was registered in ClinicalTrials.gov under NCT01982045. At 1 year follow-up, noninferiority of BCP$_{<\mu m}$ vs. autograft in terms of posterolateral fusion performance based on a margin of 15% was demonstrated.[117] The current study focuses on predefined secondary analyses of the fusion status and clinical outcomes at 2-year follow-up with specific attention to the role of the bone grafts.

Study population

The study population consisted of 100 adult patients treated with a primary single or multilevel instrumented posterolateral fusion between T10 and S1/ilium after at least 6 months of unsuccessful nonoperative treatment. Indications for surgery were deformity, structural instability and/or expected instability (for example as a result of decompression for spinal stenosis). In- and exclusion criteria are listed in the 1-year article.[117]

Intervention

The surgical technique comprised a standard open posterolateral fusion via a midline approach. When indicated, additional interbody fusion with a titanium or polyetheretherketone (PEEK) cage (based on surgeon preference) filled with local bone was performed. After instrumentation and thorough preparation of the fusion bed, including the posterior surfaces of the transverse processes and laminae, the randomized allocation side (left/right) of the two different grafts was disclosed. The decorticated gutters at one side of each fusion trajectory were grafted with 10cc BCP$_{<\mu m}$ per level, whereas a mixture of iliac crest bone and available local bone was applied to the other side. A volume of 8 - 10cc autograft ($\geq$ 50% iliac crest bone) per fusion level was intended.

Fusion assessment

The posterolateral fusion rate was assessed after 1 year using thin slice ($\leq$ 1mm) computed tomography (CT) scans and multiplanar reconstructions. Each side of each instrumented segment, as well as each interbody cage, was scored in three planes by two blinded spine surgeons using a detailed three-point classification system as given in **Appendix 1**.[44,55,117] To gain further insight into the contribution of the grafts, intertransverse fusion (lateral to the rod) and facet/lamina fusion (at/medial to the rod) were scored separately. Any disagreements between the reviewers were resolved by reassessment and consensus. Patients without fusion at all instrumented segments at 1 year were pursued for an additional CT assessment at 2-year follow-up.

Clinical assessment

To evaluate the clinical effect of the fusion surgery, patient reported outcome measures (PROMs) including the Oswestry Disability Index (ODI), 0 - 100 Visual Analogue Scale (VAS) for back pain and EQ-5D-5L were collected preoperatively and at 6 weeks, 3 months, 6 months, 1 year and 2 years postoperatively. To evaluate safety, unexpected (serious) adverse events, whether or not considered related to the use of $BCP_{<\mu m}$, were documented until last follow-up.

Statistical analyses

Study data were collected using paper case report forms, processed in Research Online for Researchers (Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands) and analysed with SPSS Statistics Version 25 (IBM Corp., Armonk, NY, USA). Baseline characteristics, surgical details, fusion rates and locations as well as PROMs are described by descriptive statistics.

Fusion scores were dichotomized into “fused” (fusion) and “not fused” (doubtful fusion or nonunion). Interobserver reliability of the fusion assessments was evaluated by percentage agreement and Cohen’s kappa. Differences in posterolateral fusion rates on segment level between 1 and 2 years and between treatment conditions were analysed using a logistic regression Generalized Estimating Equations (GEE) model with an independent correlation structure to account for clustering of fusion scores within segments and within subjects ($p < 0.05$). Similar GEE models were used to compare the interbody fusion rates between 1 and 2 years and between titanium and PEEK cages, as well as the relation between successful interbody fusion and posterolateral fusion on either or both sides. Odds ratio’s (OR) along with their 95% confidence interval (CI) are reported.

PROMs at each timepoint were described as median and interquartile range (IQR) based on an intention to treat principle. Changes from baseline to 2-year follow-up were analyzed using the paired samples t-test ($p < 0.05$) and the minimal clinically important difference (MCID) was set to 15 points for both the ODI and VAS back pain.[98,99] Cases with missing values were omitted by pairwise deletion.

A mixed model for repeated measures with a random intercept was used to analyze the relationship between radiographic fusion and ODI ($p < 0.05$). Fixed effects were timepoint (1 and 2 years), preoperative ODI and fusion status. Successful fusion was defined as posterolateral and/or interbody fusion at all instrumented segments.

Results

Patient characteristics

As illustrated by **Figure 4.1**, 96 of the 100 operated patients reached the 1 year follow-up and 87 were included in the primary efficacy analysis. During the second year, 3 patients underwent revision surgery and 7 patients (including 3 revisions) dropped out, resulting in a final follow-up rate of 89%.

Baseline characteristics and surgical details of both the entire study population and patients included in the fusion analysis at 2 years are presented in **Table 4.1**. The main indication for surgery was structural and/or expected instability and two-third of the patients underwent a single level fusion. The additional titanium and PEEK cages had a ratio of 1:2.

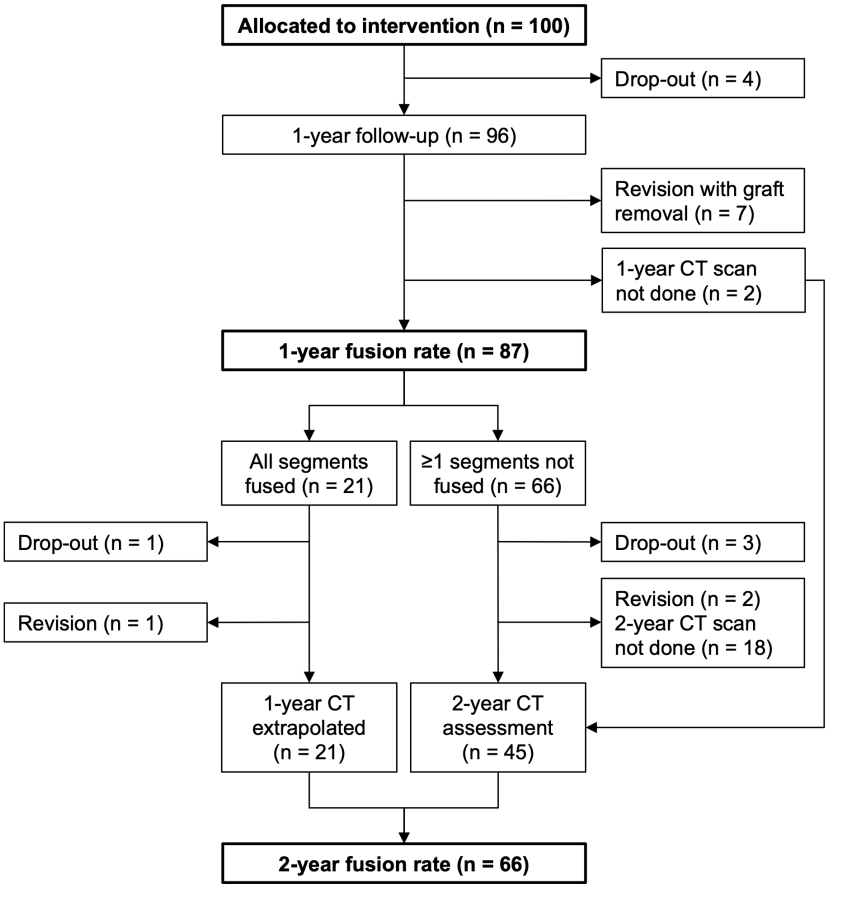


Figure 4.1: Flowchart of the progress of patients through each stage of the study.

Table 4.1: Baseline characteristics and surgical details of entire study population at baseline (n = 100) and patients included in the fusion analysis at 2-year follow-up (n = 66).

	Baseline (n = 100)	**2-year follow-up (n = 66)**
Age, mean ± SD (range), years	55.4 ± 12.0 (27 - 79)	54.9 ± 11.5 (27 - 79)
Sex, n (%)		
Male	49 (49%)	33 (50%)
Female	51 (51%)	33 (50%)
Smokers, n (%)	34 (34%)	19 (29%)
Indication(s) for surgery, n (%)		
Deformity	12 (12%)	8 (12%)
Structural instability	45 (45%)	26 (39%)
Expected instability	60 (60%)	41 (62%)
Missing	7 (7%)	6 (9%)
Number of segments fused, n (%)		
1	66 (66%)	45 (68%)
2	20 (20%)	11 (17%)
> 2	14 (14%)	10 (15%)
Median number of segments fused (range)	1 (1 - 8)	1 (1 - 8)
Interbody device(s), n (%)	62 (62%)	48 (73%)
Type of interbody device, n (%)		
Titanium	26 (37%)	18 (33%)
PEEK	45 (63%)	37 (67%)

n = number of patients, SD = standard deviation, PEEK = polyetheretherketone

Radiographic fusion

In 21 patients, all 26 segments assessed for posterolateral fusion and all 14 interbody cages were scored as fused at 1 year. Of the remaining patients that were not considered completely fused, 43 underwent an additional CT scan at 2-year follow-up. Furthermore, 2 patients were only assessed at 2 years (**Figure 4.1**). Interobserver agreement of the 2-year CT scans was 83% (kappa = 0.65) for posterolateral fusion and 88% (kappa = 0.75) for interbody fusion, which appeared slightly better than the 1-year assessments (72% (kappa = 0.45) and 78% (kappa = 0.56) respectively). Extrapolating the successful fusions at 1 year, the 2-year posterolateral fusion rate was based on 113 segments and the interbody fusion rate on 55 segments. Fusion rates at 1 and 2 years of follow-up are presented in **Figure 4.2**.

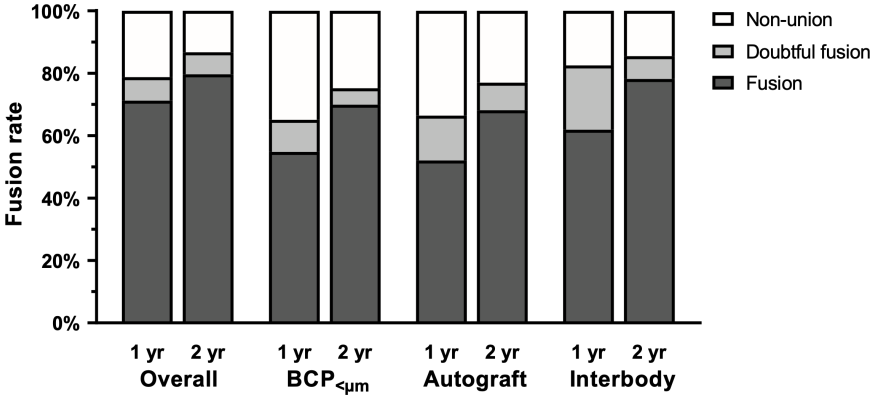


Figure 4.2: Fusion rates on segment level at 1 and 2 years (yr) of follow-up. From left to right: overall posterolateral fusion rate (i.e. either or both sides fused), unilateral posterolateral fusion rate at the BCP$_{<\mu m}$ or autograft side, and interbody fusion rate.

The overall posterolateral fusion rate, i.e. left and/or right side of a segment scored as fused, increased from 71% to 80%. At 2-year follow-up, the fusion rate at the BCP$_{<\mu m}$ side was 70% and 68% at the autograft side, compared to 55% and 52% at 1 year. GEE-analysis demonstrated a significant increase in unilateral posterolateral fusion rate between 1 and 2 years (OR = 2.0, 95% CI = 1.5 - 2.7, $p < 0.001$), but no difference between the treatment conditions (OR = 0.9, 95% CI = 0.6 - 1.3, $p = 0.595$).

After exclusion of the 2 patients with only a 2-year CT scan, further analyses of the posterolateral fusion location (intertransverse vs. facet fusion) in time were based on 64 patients and 111 segments. **Table 4.2** demonstrates that for both grafts the number of intertransverse fusions at 1 and 2 years of follow-up was very similar, whereas the posterolateral fusion rate (either intertransverse or interfacet) increased from 58% to 70% indicating an increase of facet fusions. Of the additional fusions at 2-year follow-up, 59% were scored as nonunion at 1 year and 41% as doubtful fusion.

Table 4.2: Number of segment sides scored as posterolateral fusion and specifically as intertransverse fusion (n = 464 patients, with 111 spinal segments and 222 segment sides).

Timepoint	Posterolateral fusions			Intertransverse fusions	
	Fusion rate	BCP$_{<\mu m}$	Autograft	BCP$_{<\mu m}$	Autograft
1 year	129/222 (58%)	67/111 (60%)	62/111 (56%)	28/67	33/62
2 years	156/222 (70%)	79/111 (71%)	77/111 (69%)	28/79	31/77

The interbody fusion rate (**Figure 4.2**) increased as well, from 62% to 78% (OR = 2.2, 95% CI = 1.3 - 3.7, $p = 0.002$). Breakdown by cage type showed that 91% of the titanium cages were fused at 1 year and 100% at 2 years, whereas the fusion rate for PEEK increased from 48% to 68%. This difference was highly significant (OR = 17.8, 95% CI = 3.8 - 82.8, $p < 0.001$). In line with the 1-year results, a positive relation between successful interbody fusion and posterolateral fusion was found (OR = 8.5, 95% CI = 1.8 - 39.9, $p = 0.006$).

Patient reported outcome measures

Clinical outcomes up to 2 years are illustrated by **Figure 4.3**. Both the ODI (**Figure 4.3**a) and VAS back pain (**Figure 4.3**b) improved above the MCID with a mean difference of -20 $\pm$ 19 and -31 $\pm$ 27 respectively ($p < 0.001$). At 2-year follow-up, 58% of the patients achieved the MCID of the ODI. The MCID for the VAS back pain was reached by 66%. The EQ-5D Dutch utility index (**Figure 4.3**c) increased from median 0.529 (IQR 0.394 - 0.683) to 0.805 (IQR 0.651 - 0.874). The mixed model analysis, adjusted for baseline scores, revealed that patients with a bony bridge at all instrumented segments had a lower ODI (estimated difference 8.9 points, 95% CI 2.4 = 15.4, $p = 0.008$), indicating a relationship between successful fusion and improved clinical outcome.

Adverse events

In addition to the events described in the 1-year article[117], 8 serious adverse events were registered between 1 and 2 years of follow-up. Two patients were diagnosed with failed back surgery syndrome, one patient underwent revision surgery for pseudoarthrosis and screw loosening, and another patient had a deep wound infection after revision surgery. The remaining serious events were unrelated to the fusion surgery, but required hospitalization: cardiovascular disease (n = 2), humerus fracture (n = 1) and gastric bypass (n = 1). Of the 15 adverse events, 6 described back and/or leg pain. The total reoperation rate was 13%, including 3 revisions for pseudoarthrosis.

Discussion

This study examined the progression of posterolateral and interbody fusions between 1 and 2 years as part of a randomized, intrapatient controlled trial investigating BCP$_{<\mu m}$ vs. autograft. Currently, there is no consensus on the method and timing of radiographic fusion assessment. This impedes the comparison of many treatment outcomes. Moreover, little is known on the progression of bone formation over time and especially to what extent this can be related to bone grafting.

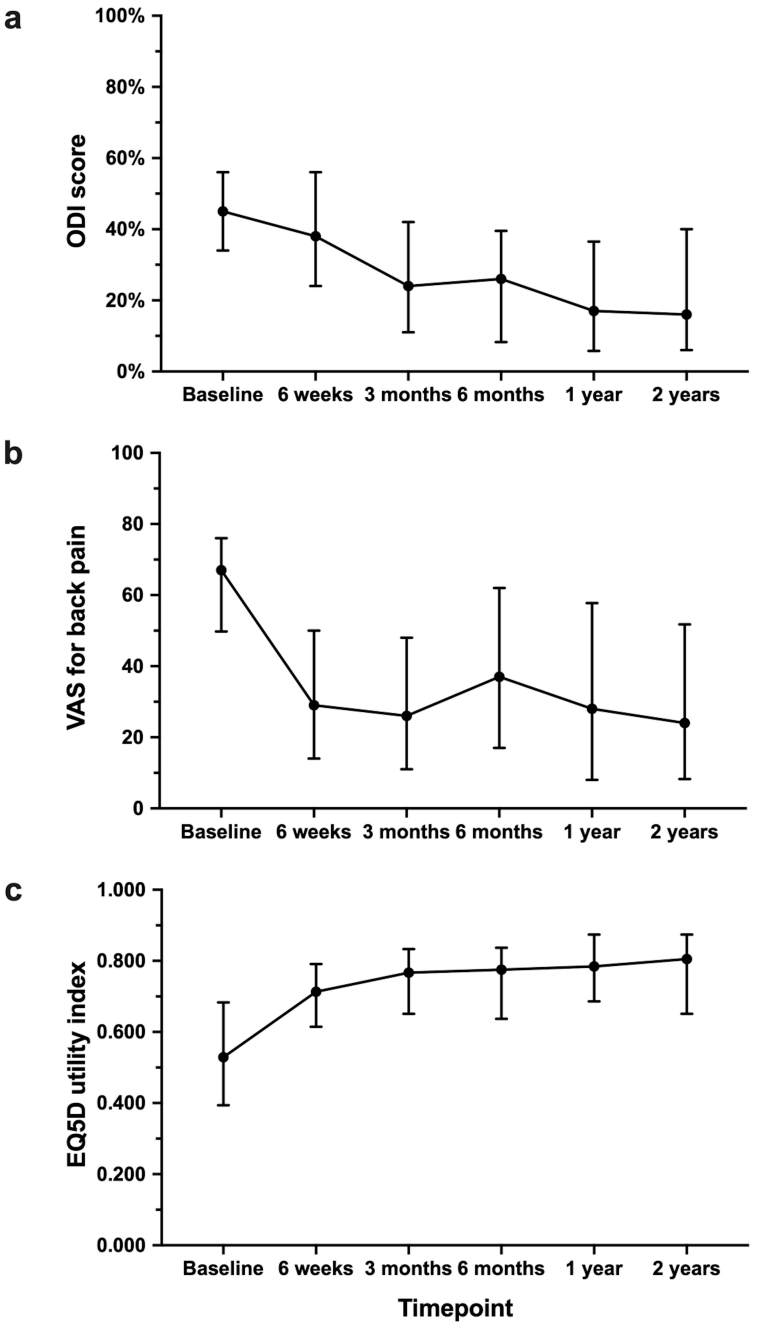


Figure 4.3: a) Oswestry Disability Index (ODI) (0 - 100%), **b)** Visual Analogue Scale (VAS) for back pain (0 - 100), and **c)** EQ-5D utility index (-0.329 to 1.000) at baseline and each follow-up. Median values along with their interquartile range are displayed as the data are not normally distributed.

The CT-based posterolateral fusion rate of both $BCP_{<\mu m}$ and autograft, as well as the additional interbody fusion rate, increased between 1 and 2 years of follow-up. Interestingly, ongoing bone formation was not observed in the intertransverse fusion area, but only between the immobilized facet joints and in/around the interbody cages. Based on the location of the grafts and the fact that both grafts were completely resorbed on the 1-year CT scans, the increase in posterolateral fusion rates is unlikely the result of grafting. This is in agreement with other studies that have shown that bone graft induced fusion by creeping substitution mainly occurs during the first 6 months.[118,119] The observations that resorbable bone graft is particularly effective within 1 year, and that facet fusions most likely occur as a result of immobilization, has important consequences for research in this field. We believe that true assessment of bone graft (substitutes) should be no later than 1 year after surgery and preferably limited to the area where this graft is most likely crucial, i.e. the intertransverse process area.

Fusion rates depend on many factors including the modality and method of fusion assessment itself.[26,100,114] The detailed classification system used in this study resulted in an interobserver agreement for both posterolateral fusion and interbody fusion that was substantial based on Cohen's kappa.[44] Between 1 and 2 years, the overall posterolateral fusion rate (i.e. uni-/bilateral fusion) had increased from 71% to 80% and the unilateral fusion rate from 52 - 55% to 68 - 70%. These fusion rates seem to be higher than the results of a similar intrapatient controlled trial by Cammisa et al., but they only assessed intertransverse fusion.[96] In contrast, Dimar et al. reported 1 and 2 years after single-level instrumented fusion with autograft a bilateral intertransverse fusion rate of 72% and 84% respectively.[103] In a randomized trial comparing two bone graft substitutes in combined posterolateral and interbody fusion, the uni- and/or bilateral fusion rate increased from 53 - 56% at 1 year to 80% at 2-year follow-up.[120] Recently, Kim et al. demonstrated the significance of facet joint fusion and increase of these fusions between 6 and 12 months after posterolateral fusion. An additional interbody fusion procedure negatively influenced posterolateral fusion, probably due to the associated facetectomies.[73] This effect was not observed in the current study, possibly because we mostly performed posterior instead of transforaminal interbody fusion. Contrary, a positive relation between successful interbody fusion and posterolateral fusion at both 1 and 2-year follow-up was found. This may be related to patient factors or increased stability. Despite the challenges to compare the radiographic outcomes of different studies, the current study adds to the evidence that spinal fusion is an ongoing process and radiological nonunion after 1 year should not be regarded as definitive failure.

In line with comparable study populations, improvements in clinical outcomes continued up to 2 years and were clinically relevant for both the ODI and VAS for back pain.[93,103,107,108,121] Despite the low median ODI of 16% (IQR 6 - 40) at final follow-up

indicating minimal disability, 42% of the patients did not reach the MCID of 15 points. Further exploration revealed that one-third of these patients had an ODI $\leq$20 at baseline and/or 2-year follow-up.

The relationship between radiographic and clinical outcomes is still controversial.[122] Several studies have shown increased fusion rates by the addition of instrumentation, but no difference in clinical outcomes, whereas others have demonstrated the long-term clinical benefits of arthrodesis over pseudoarthrosis.[123–126] The current study indicated a positive relationship between radiographic fusion and ODI. However, the estimated difference in ODI (8.9 points, 95% CI 2.4 - 15.4) was below the assumed MCID.

Strength of this study is the excellent follow-up rate of 89% at 2-year follow-up. Nonetheless, we do recognize some limitations. To limit the exposure to ionizing radiation, only patients without fusion at all of the instrumented segments were scheduled for an additional CT scan at 2 years. For logistical reasons this decision was made by the treating physician. Unfortunately, 14 patients were not reassessed as the treating physician, unlike the blinded observers, qualified these as complete fusion. Another limitation is the assumption that successful fusions can be extrapolated. However, of the 43 patients that were reassessed, only 6.5% of the segment sides that were scored as fused at 1 year were scored differently at 2 years. This is most likely the result of variance in (re)assessment, as also reflected in the 72% interobserver agreement at 1 year. Furthermore, the contribution of the bone grafts to the fusion process during the first and second year after surgery was only explored visually based on the location of the posterolateral fusion mass. Imaging-based quantification of bone (graft) resorption and remodelling over time is still in its infancy.[74,118,119] Last, the intrapatient design limits the separate attribution of adverse events to the treatment conditions. Nevertheless, the observed adverse events were not likely related to BCP$_{<\mu m}$ and the reoperation rate is in accordance with literature.[127–130]

In conclusion, this intrapatient controlled trial comparing two bone grafts demonstrated an increase in fusion rates between 1 and 2 years after instrumented posterolateral fusion in the (thoraco)lumbar spine. Moreover, there was no difference between BCP$_{<\mu m}$ and autograft. During the second year after surgery, bony fusion around the facet joints and additional interbody cages continued, whereas the number of intertransverse fusions that can be fully ascribed to the grafts remained unchanged. This indicates that bone graft induced fusion occurs within the first year and mid-term progression of bony fusion is most likely the result of immobilization. Further research is needed to elucidate the mechanisms behind spinal fusion over time, to guide optimal material (resorption) characteristics and fusion assessment.

Appendix 1: Fusion assessment on CT scans

Data quality:

- Volume (helical) CT scan to make planar reconstructions
- Slice thickness $\leq$ 1 mm
- Raw volume datasets (DICOM scans) and preferably reconstructions in the coronal, sagittal and axial plane stored and exported

Software:

- Horos medical imaging viewer for Mac OS X

Method of fusion assessment and scoring:

- Each side of each segment was assessed in three planes (sagittal, coronal and axial)
- Two locations were assessed separately: intertransverse fusion (lateral to the rod) and facet fusion (at or medial to the rod)
- According to the definitions in **Table 4.3** and using the scoring sheet in **Table 4.4** (columns 2 to 7), fusion in each plane was scored as nonunion (0), doubtful fusion (1) or fusion (2)
- Based on the sum score of all three planes (**Table 4.4**, columns 8 and 9), a fusion decision per location (intertransverse or facet) was calculated (**Table 4.4**, columns 10 and 11)
  - A sum score of $\leq$ 3 with at least score 2 in one plane was considered a fusion (2); otherwise it was doubtful fusion (1)
  - A sum score of 2 was regarded as doubtful fusion (1) and a sum score of 0 or 1 as nonunion (0)
- The final decision of posterolateral fusion of the segment side (**Table 4.4**, column 12) was based on the highest score for intertransverse or facet fusion
- Interbody fusion was scored similarly in two planes (sagittal and coronal) using the scoring sheet in **Table 4.5**
  - A sum score of 3 or 4 was considered fusion, 2 doubtful fusion and 0 or 1 nonunion

Table 4.3: Applied three-point classification system for posterolateral and interbody fusion.

Scoring	Definition
2 = fusion	Continuous bony bridge from one vertebra to the other, in the absence of any secondary signs of nonunion such as fracture or loosening of the screws or rods.
1 = doubtful fusion	Doubts about continuity or quality of the bony bridge.
0 = nonunion	Definite discontinuity or lack of a fusion mass, as well as obvious indications of mobility like material failure or apparent pseudoarthrosis.
NE = not evaluable	When the quality of the scan does not allow assessment of the bone mass (for example due to scattering) or in case of an osteotomy.

Table 4.4: Example scoring sheet for posterolateral fusion at L4-L5.

	Sagittal plane		Coronal plane		Axial plane		Sum score		Decision		
Segment and side	Inter-transverse	Facet joint	Inter-transverse	Facet joint	Inter-transverse	Facet joint	Inter-transverse	Facet joint	Inter-transverse	Facet joint	Segment
L4-L5 right	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 - 6	0 - 6	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2
L4-L5 left	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2	0 - 6	0 - 6	0 / 1 / 2	0 / 1 / 2	0 / 1 / 2

Table 4.5: Example scoring sheet for interbody fusion at L4-L5.

Segment	Sagittal plane	Coronal plane	Sum score	Decision
L4-L5	0 / 1 / 2	0 / 1 / 2	0 - 4	0 / 1 / 2

CHAPTER 5

Methodological aspects of a randomized intrapatient controlled design for clinical trials in spine surgery

A.M. Lehr
W.C.H. Jacobs
R.K. Stellato
R.M. Castelein
F.C. Oner
M.C. Kruyt

Methodological aspects of a randomized within-patient concurrent controlled design for clinical trials in spine surgery
Clinical Trials 2022

doi: 10.1177/17407745221084705

Introduction

Emerged in the field of pharmacological interventions, the randomized controlled trial (RCT) has become the highest grade of evidence in the hierarchy of research designs. The main advantage of this design is the elimination of known and unknown confounders, as randomization balances prognostic differences between the experimental and control group. Whenever possible, double- or triple-blinding is added to reduce detection and performance bias. Although this approach is very effective for many non-surgical comparative studies, the feasibility of a classic RCT in surgical research is challenged by several factors.[28–30,131–134]

First of all, given the invasiveness and irreversibility of surgery, recruitment might be impeded and biased by patient preference. Second, randomization between a novel and established surgical treatment might be biased by surgeon experience. Cook et al. and Houwert et al. also described the influence of differing pre-, peri- and postoperative procedures (between hospitals or patients) on the observed effect size in surgical trials.[28,132] Third, depending on the choice of comparator, patient blinding might not be possible. This increases the risk of cross-overs or drop-outs. Moreover, awareness of the received treatment can impact (subjective) patient reported outcomes, while these have gained importance as outcome measure of clinical trials. Fourth, as blinding of the treating physician is impossible, blinding in surgical RCTs is often limited to the outcome assessor. Last but not least, in study populations with a high variability in diagnoses or treatment, or with nonparametric outcomes, it might be difficult to recruit a representative and especially a large enough sample of patients due to budget and time limitations.

To guarantee the scientific quality of RCTs nonetheless, tremendous investments are required. These investments are closely related to the number of patients that have to be enrolled. As a consequence, underpowered and poorly conducted trials that may produce misleading results (type II errors) are frequently produced.[31,32,135,136]

To overcome some of the challenges of a classic RCT, a design where the patient can be his/her own control is very effective. Such an intrapatient concurrent controlled design, in which the experimental and control condition are applied concurrently, is also known as a side-by-side (comparison), self-controlled or within-patient design and commonly applied in dentistry, ophthalmology and dermatology.[137] In surgical research this specific design is used for obvious applications such as bilateral arthroplasty, but since the 1990s it has also been applied in the field of posterolateral spinal fusion.[50,84,96,97,105,106,138–145] However, to the best of our knowledge, the methodological and statistical aspects and consequences of the intrapatient concurrent controlled design are not yet described in depth.

We recently opted for the randomized intrapatient concurrent controlled trial (intrapatient RCT) design to investigate noninferiority of a ceramic bone graft substitute compared to autologous bone graft in 100 patients undergoing instrumented lumbar posterolateral spinal fusion surgery (**Chapter 3** and **Figure 5.1**).[117] In brief, the ceramic bone graft was applied to one side of each fusion trajectory and the contralateral side was treated with autograft. The current article describes the theoretical advantages as well as disadvantages of this design with a specific focus on posterolateral spinal fusion research, followed by a quantitative analysis of its added value based on our collected study data.

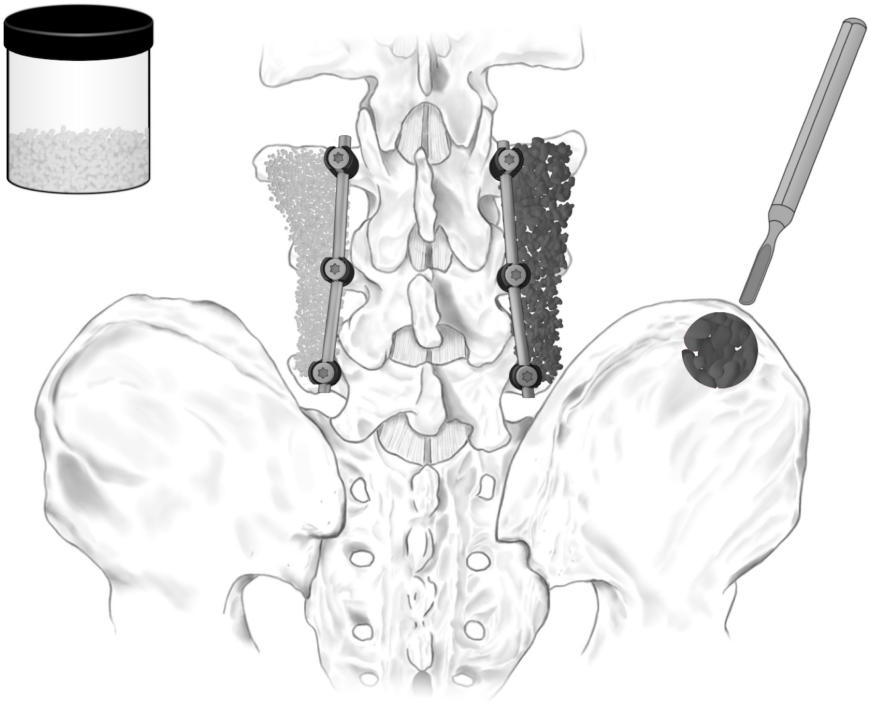


Figure 5.1: Illustration of the design and surgical procedure of the intrapatient RCT described in **Chapter 3**. In this case, the ceramic was randomized to the left side. Autologous bone graft was harvested from the right iliac crest and applied to the right side of the spine.

Rationale for intrapatient RCT

In posterolateral fusion surgery, investigation of the efficacy of a bone graft substitute in a representative patient population is specifically challenged by the variation in characteristics, diagnosis and treatment of the target population. These factors include but

are not limited to age, gender, smoking status, indication for surgery, length of the fusion trajectory, type of instrumentation, graft bed preparation, additional procedures like interbody fusion, and surgeon experience. Therefore, the obvious rationale to employ an intrapatient controlled trial design with side-by-side comparison between the experimental and control condition is the elimination of interpatient variability.

Theoretical pros of intrapatient RCT

Despite proper randomization and strict eligibility criteria, imbalance between treatment groups that may affect the outcome can still occur.[102] An intrapatient concurrent controlled trial design eliminates confounders directly, which reduces outcome variance and increases the validity of the estimated treatment effect. It also avoids bias due to cross-over issues.

The required number of patients is at most half the sample size of a classic RCT with parallel groups, as each patient receives both the experimental and control condition. The efficiency of a paired design over an unpaired design further increases with increasing intrapatient correlation between the treatment outcomes of the two allocation groups (i.e. increasing effect of patient-related factors).[146–148]

Both the reduced variance and sample size has several important implications. From an ethical point of view, fewer patients need to be exposed to a not-yet-proven intervention. Moreover, each patient also benefits from a proven control condition. This may mitigate the potential undesired effect of the experimental condition. In case of posterolateral fusion surgery, a high unilateral fusion rate at the control side minimizes the risk of symptomatic bilateral nonunions.[50,140] This knowledge might also contribute to patients' willingness to participate in the study and prevent drop-outs. From a practical point of view, due to the reduced variability concerns more patients become eligible, as less strict inclusion and exclusion criteria can be applied, and a multi-center design is more feasible. All together this dramatically reduces the duration of the study, which improves the motivation of the research team and heavily reduces associated costs.

Last but not least, outcome assessment of the experimental and control condition can be done concurrently by a single assessor, which improves intraobserver reliability.

Theoretical cons of intrapatient RCT

First of all, an intrapatient concurrent controlled trial design is only an option when interventions can be applied concurrently to separate but similar pairs of sites.

A frequently cited objection of intrapatient comparisons is the possibility that the interventions influence each other's effects, as they are applied in parallel. Specific concerns in posterolateral fusion surgery are carry-across effects of the applied grafts, or that bone formation at one side of the spine can influence the outcome at the other side due to the stabilizing effect of fusion.[84,105,106,138,140,141] Also, the occurrence of an adverse event could be prevented by the concurrent treatment condition.

Another disadvantage is that clinical and functional outcomes on patient level, including some adverse events (AEs), cannot be attributed separately to the investigated interventions.[96,106,140,145] In addition, secondary assessment of possible factors that influence the observed difference in treatment effect is precluded.

Obviously, loss to follow-up of one patient affects both treatment conditions. The same holds for revision surgery.

Applicability of intrapatient RCT

Taking together these pros and cons, we believe that the intrapatient RCT design is possible for efficacy studies of non-bioactive bone graft substitutes in instrumented posterolateral spinal fusion, when the primary outcome is based on a local parameter like radiological bony fusion or bone mineral density. When investigating non-bioactive grafts, carry-across effects are not expected as the fusion area at the left and right side of the spine are physically separated and there are no systematic effects. Rigid instrumentation will largely eliminate the risk that the fusion process at one side of the spine is affected by the other side. The boundary conditions for using the intrapatient concurrent controlled design are summarized in **Table 5.1**.[149–151]

Table 5.1: Boundary conditions for the application of an intrapatient concurrent controlled trial design.

Aspect	Boundary condition
Intervention	Identical but physically separated investigational sites
	Concurrent interventions
	Uncorrelated treatment effects
Outcome assessment	Local treatment effects
	Independent assessment of treatment effects
Statistics	(Power) analysis for paired design
	Account for clustering within patients

Quantitative analyses of intrapatient RCT

To investigate some of the theoretical advantages expounded above, we analyzed the data of our completed intrapatient RCT. This trial is described in detail in **Chapter 3**.

Measures of interpatient variability

To assess the contribution of patient-related factors to the variance in fusion rate, we determined interpatient variability by two statistical approaches: the intraclass correlation coefficient (ICC) and concordance between the left and right side of the fusion trajectory (SPSS Statistics Version 26. IBM Corp., Armonk, NY, USA).

For the first approach, a two-level logistic mixed model with a random intercept per patient was used. The ICC is known as a measure for the correlation of outcomes within a cluster (in our case patient) and defined as the ratio of variance between patients to the total variance (sum of variance between and within patients). A higher ICC reflects a stronger effect of clustering, indicating that the fusion scores within the same patient are more likely to be similar than between different patients. Using an estimate of between-patient variance of $\pi^2/3$,[152] the ICC of the intercept-only logistic model was 17%. In other words, 17% of the variance in fusion rate could be explained by patient-related factors. Inclusion of age, sex, smoking status, ASA classification, prior treatment to included spinal segments, number of instrumented segments and treating hospital as fixed effects increased the ICC to 32%. However, none of these individual factors was significant.

In addition, the concordance in outcome between the left and right side of the fusion trajectory was calculated. Without the effect of patient-related factors, fusion rates between the two sides would be independent (i.e. concordance of 50%). However, as expected, it appeared that fusion on one side of a spinal segment correlated to the outcome on the other side. At 1-year follow-up, 36% of the segments were fused bilaterally and 29% were not fused, resulting in 64% concordance between sides. These findings confirm our ICC calculations and indicate that some patients are more likely to achieve fusion than others. Moreover, these findings are in line with a comparable intrapatient RCT by Cammisa et al., as well as the assumed concordance underlying our sample size calculation.[145]

Sample size of intrapatient vs. parallel-group RCT

To determine the combined effect of side-by-side comparisons and interpatient variability, we compared the adopted sample size with the sample size of a classic RCT

design with two parallel groups (PASS 2008 Version 8.0.16. NCSS, LLC. Kaysville, Utah, USA). Based on the assumptions described in **Chapter 3**, the required sample size for the employed intrapatient RCT with 1:1 randomization was 84 patients.[153] Repeating this calculation for a parallel-group RCT using a one-sided unpooled Z-test resulted in a sample size of 138 patients per treatment condition and thereby 276 patients in total.[154] This means that over 3 times more patients need to be enrolled if parallel groups were used.

To further investigate the effect of clustering within patients in terms of agreement in fusion outcome between the left and right side of the spine, we also repeated the intrapatient sample size calculation with assumed concordance ranging from 50% (no patient effect) to 90%. As expected, this yielded a linear inverse association from 138 to 38 patients respectively. In other words, the higher the concordance between sides, the more the standard error of the difference in outcome probabilities, and thereby the sample size, decreases.[146,147,153,155]

Duration and costs of intrapatient vs. parallel-group RCT

The duration of a clinical trial obviously depends on the required sample size and the recruitment rate. In turn, study costs are related to both the duration and number of sites. A multicenter design facilitates faster enrolment, as well as generalizability and acceptance of the study outcomes, but heterogeneity among the participating sites may introduce bias. In addition, multicenter trials are more costly in terms of logistic, administrative and personnel costs. The increased efficiency and feasibility of an intrapatient RCT is quantified in **Table 5.2**, by extrapolating our case to a classic parallel-group RCT.[156,157]

Table 5.2: Comparison of a randomized intrapatient concurrent controlled trial (intrapatient RCT) and classic RCT with parallel groups, by extrapolation of the completed intrapatient RCT.[117]

	Intrapatient RCT	**Classic RCT**	
Number of patients[a]	100	325	
Number of sites	4	4	13
Duration[b]	4.5 years	10 years	4.5 years
Costs[c]	€505.000	€890.000	€815.000

[a] Based on sample size calculation and 15% drop-out rate

[b] First patient in until last patient last visit

[c] Based on KCE Trials Budgeting tool V5.0; including 15% margin, excluding overhead costs and VAT[107,158]

Discussion

This article evaluated the application of an intrapatient concurrent controlled trial design as a pragmatic and efficient alternative to a classic RCT. Key aspects are the elimination of interpatient variability and reduction in sample size due to side-by-side comparison of the experimental and control condition. That makes this design particularly suitable when the feasibility of a parallel-group RCT is challenged by the recruitment of a representative and large enough sample. Whether an intrapatient concurrent controlled trial design is appropriate, however, depends on several factors including the clinical condition, anatomical location, intervention and primary endpoint. Boundary conditions for this approach include the concurrent application of the experimental and control condition to identical but physically separated sites. Moreover, the outcome of interest should be local, uncorrelated and independently assessable. Although the described intrapatient methodology is suitable for different surgical models, we specifically focused on posterolateral spinal fusion. Distinct differences with for example bilateral arthroplasty are that both treatment conditions are applied within the same surgical procedure and loaded identically, contributing to a further reduction of outcome variance.

Based on the data of a recently published intrapatient RCT on the efficacy of a bone graft substitute in instrumented lumbar posterolateral spinal fusion, we found that the intrapatient concurrent controlled design considerably reduced the sample size (in this case to less than one-third of a parallel-group RCT) and we demonstrated that the posterolateral fusion rate highly depends on patient-related factors. The latter has been shown extensively for comorbidities and smoking.[66,159] Also, this design allowed us to successfully include a representative case mix of patients including multilevel fusions that benefit most from bone graft substitutes.

Whereas the ICC is a recognized parameter in cluster RCTs, to quantify the correlation of outcomes within a cluster and calculate the required sample size, the effect of clustering is strongly underexposed in studies with intrapatient comparisons. Moreover, even in research fields where this design is more common, many studies applied inappropriate statistical analysis.[160,161] As a consequence, studies might be underpowered or estimated treatment effects invalid.[137,162–164] The report of methodological and statistical aspects of the intrapatient concurrent controlled design, especially regarding the correlation of outcomes within subjects, can further demonstrate its value and assist the set-up of future studies. Another limitation of several previous intrapatient controlled trials is the lack of randomization, which unnecessarily limited their level of evidence.[97,143,145,165] In our case, randomization between the left and right side of the fusion trajectory just before application of the grafts confined the introduction of bias during surgery and contributed to the blinding of the outcome assessors.

We expected that the application of both the experimental and control condition would enhance patients' willingness to participate, but our enrollment rate (55%) was comparable to other surgical RCTs.[107,158] Of the 197 patients that were assessed for eligibility and approached for the study, 43% declined to receive the patient information letter or to participate. Reasons for refusal were not registered. Notable differences in enrollment rate between the participating sites were observed, suggesting that enrollment success was affected by patient or organizational factors as well.

Although very unlikely when inert non-bioactive grafts are applied to the lateral gutters of rigidly instrumented spinal segments, we do recognize that an effect of one condition on the other (carry-across effect) could not be completely ruled out nor be assessed from our study data. Nevertheless, bone formation across the midline was not identified in any of the radiological evaluations. Another factor that could not be investigated is the potential impact of surgeon handedness on the surgical technique applied to the left and right side of the fusion trajectory.

While it is undisputable that outcomes on patient level, like patient-reported outcomes (PROs), cannot be attributed separately to the investigated sites in an intrapatient controlled trial, the relevance of this limitation depends on the application. PROs are often not the primary research aim when evaluating surgical techniques or procedures. That is not because they are not the most important, but because these outcomes often do not reflect essential results at short-term. In spinal fusion research, radiographic fusion is the most frequently used primary endpoint as the aim of the treatment is to obtain a solid arthrodesis between vertebrae. Moreover, even in a classic RCT it is difficult to establish a direct relation between for example the investigated bone graft and PROs. The use of PROs as secondary outcome is encouraged, also in an intrapatient RCT, as they give valuable information about the general treatment effect as experienced by the patients and they signal deviations from control populations. In other applications, like bilateral arthroplasty, PROs of functioning or pain can be assessed per treatment condition.[166,167]

In conclusion, the randomized intrapatient concurrent controlled trial design is an efficient but underappreciated design that provides a solution to some of the considerable challenges of a classic RCT. When suitable, this design with side-by-side instead of parallel group comparisons has a positive impact on feasibility and generalizability of results, thereby maximizing efficiency of resources and accelerating the implementation of evidence into the clinical practice. Based on our experiences in the field of spine surgery, we specifically recommend it for efficacy studies of non-bioactive bone grafts in instrumented posterolateral fusion surgery. We encourage further assessment of the methodological aspects and limitations of the intrapatient concurrent controlled trial design, to further establish this approach in clinical (spinal) research.

CHAPTER 6

Long-term (>10 years) clinical outcomes of instrumented posterolateral lumbar fusion for spondylolisthesis

A.M. Lehr
D. Delawi
J.L.C. van Susante
N. Verschoor
N. Wolterbeek
F.C. Oner
M.C. Kruyt

Long-term (>10 years) clinical outcomes of instrumented posterolateral fusion for spondylolisthesis
European Spine Journal 2021;30:1380-1386

doi: 10.1007/s00586-020-06671-6

Introduction

The rates of instrumented spinal fusion surgery increased markedly over the past decades, succeeded by growing evidence of especially short- and mid-term treatment effects for specific indications including lumbar spondylolisthesis associated with spinal stenosis.[2,108,111,158,168,169] Long-term clinical outcomes of spinal fusion are however still scarce and inconclusive.[170–174] Both deterioration and preservation of achieved clinical outcomes are reported, which can be partly explained by the heterogeneity in study designs and populations.

Another area of controversy is the relation between radiographic fusion and clinical outcomes.[122,126,175] Among patients with degenerative lumbar spondylolisthesis treated with decompression and uninstrumented posterolateral fusion, solid arthrodesis appeared only beneficial for long-term clinical outcomes.[124,125] However, in the presence of rigid instrumentation the necessity of a solid fusion within the first years can be debated. This emphasizes the need for long-term evaluations.

The current study investigated the long-term clinical outcomes of patients that were included in a randomized controlled trial (RCT) and who received instrumented posterolateral spinal fusion with autologous bone graft or osteogenic protein-1 for lumbar spondylolisthesis with neurological manifestations. The primary objective was to assess disability, as determined by the Oswestry Disability Index (ODI), at long-term follow-up compared to baseline and 1 year after surgery. In addition, the effect of diagnosis, graft type and fusion status at 1-year follow-up were investigated. Secondary outcomes included pain experience, quality of life, satisfaction with treatment and reoperation rate.[102]

Materials and Methods

Study design and population

We performed a cross-sectional long-term follow-up among the Dutch participants of the previously published international multicenter Osigraft RCT.[102] In this original study, 134 patients were randomized to osteogenic protein-1 (OP-1, also known as BMP-7) or autograft for posterolateral spinal fusion from 2004 to 2008; 113 patients were included in the primary analysis. All patients underwent single-level instrumented posterolateral fusion of the lumbar spine for degenerative or isthmic spondylolisthesis with symptoms of neurological compression. Patients in the OP-1 group received Osigraft (Stryker Biotech, Hopkinton, MA, USA) combined with local bone. Patients in the control group received autologous bone graft from the iliac crest combined with local bone (autograft

group). The primary outcome was overall success at 1-year follow-up, based on a combination of clinical outcomes and evidence of posterolateral fusion on computed tomography (CT) scans.

For the current study, patients were recruited from the Dutch study population with complete 1-year follow-up that consisted of 61 patients (**Figure 6.1**). In January 2018, available patients were invited to participate by mailing an information letter, informed consent form, set of questionnaires and return envelope. They were asked to return the blank questionnaires in case they declined to participate. Non-responders were sent a reminder after 4 weeks.

Clinical outcomes

To assess long-term clinical outcomes, a set of various disease specific and generic questionnaires as well as additional questions was compiled. In line with the assessments done at baseline and 1 year after surgery in the original study, patients received the following validated questionnaires: ODI, EQ-5D-3L and Visual Analogue Scale (VAS) for leg pain. Back pain was only assessed at long-term follow-up. The sum score of the disease specific ODI, defined as primary outcome, ranges from 0% (no disability) to 100% (maximum disability possible).[93] Responses to the EQ-5D-3L were converted into a single health state index score ranging from -0.329 (worst health state) and 1.000 (best possible health).[176,177] The VAS for pain runs from 0 (no pain) to 100 (terrible pain) and a score of ≤ 30 was considered as mild pain.[178,179]

Satisfaction with treatment at long-term follow-up was measured with a Numeric Rating Scale (NRS) ranging from 0 (very dissatisfied) to 10 (very satisfied). In addition, patients were asked 1) how their complaints of back pain and leg pain have changed since the index surgery, 2) for the main effect of surgery on their pain complaints and 3) if they would choose the same treatment if they had the same condition and complaints. Finally, patients were asked for any lumbar spine reoperations since the index surgery.

Statistics

Data were processed and analyzed in SPSS Statistics Version 24 (IBM Corp., Armonk, NY, USA). Patient characteristics and all patient reported outcome measures were evaluated using descriptive statistics. Differences in ODI over time (baseline, 1-year and long-term follow-up) and the effect of graft type (OP-1 vs. autograft) were analyzed using a mixed analysis of variance (ANOVA) model for repeated measures. In addition, a multiple regression (enter method) was run to predict the ODI score at long-term follow-up from graft type, diagnosis (degenerative vs. isthmic spondylolisthesis) and

fusion status at 1-year follow-up (fusion vs. doubtful fusion/nonunion). EQ-5D-3L index scores and VAS leg pain over time were analyzed with Friedman's test. For all statistical tests the threshold for significance was set to $p < 0.05$.

Ethical considerations

The Medical Ethical Committee of the University Medical Center Utrecht, Utrecht, The Netherlands, confirmed that this follow-up study did not fall under the Medical Research Involving Human Subjects Act and ethical approval was not required. Each study participant provided written informed consent.

Results

Study population

Since the 1-year follow-up, 5 of the 61 Dutch patients had died from causes unrelated to the index surgery, leaving 56 patients available for long-term follow-up. A total of 41 (73%) patients were enrolled, with a mean follow-up of 11.8 (range 10.1 - 13.7) years. Twelve patients did not respond to the questionnaire and 3 were not willing to participate. The distribution among treatment groups is shown in **Figure 6.1**.

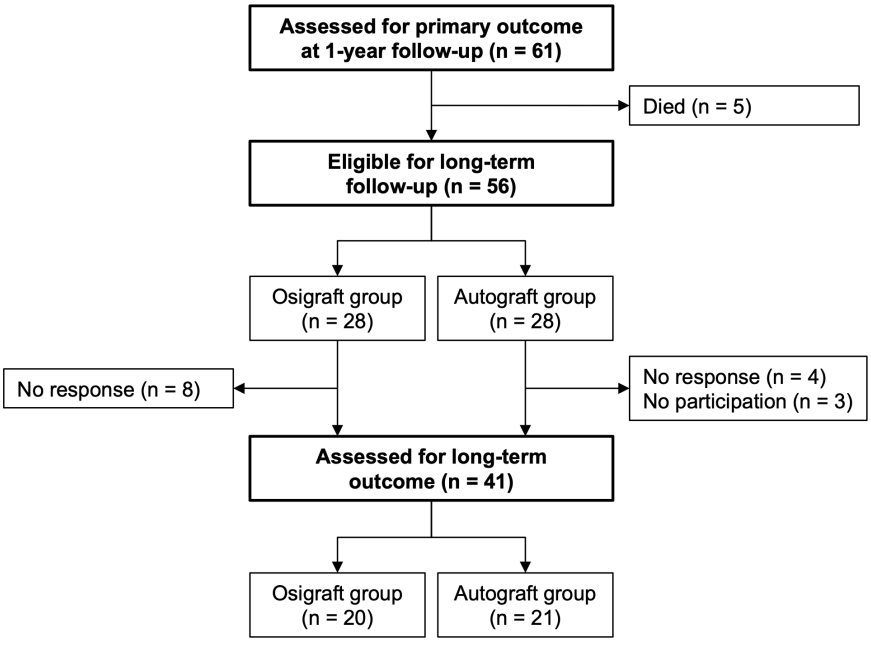


Figure 6.1: Flowchart of patients included in the long-term follow-up study.

Demographics, surgical details and 1-year fusion status on group level and per treatment condition are outlined in **Table 6.1**. The mean age of the 17 males and 24 females assessed at long-term follow-up was 62 ± 11 (range 30 - 91) years. The majority of the patients underwent surgery for isthmic spondylolisthesis (71%) and the overall 1-year fusion rate was 66%.

Table 6.1: Demographics, surgical details and 1-year fusion status on group level and per treatment group.

	Overall (n = 41)	**Osigraft (n = 20)**	**Autograft (n = 21)**
Follow-up (years), mean ± SD (range)	11.8 ± 1.0 (10.1 - 13.7)	11.8 ± 1.1 (10.1 - 13.7)	11.7 ± 1.0 (10.2 - 13.3)
Age (years), mean ± SD (range)	61.9 ± 10.6 (30 - 91)	61.6 ± 13.1 (30 - 91)	62.1 ± 7.7 (51 - 75)
Sex, n (%)			
Female	24 (59%)	10 (50%)	14 (67%)
Male	17 (42%)	10 (50%)	7 (33%)
Origin of instability, n (%)			
Degenerative spondylolisthesis	12 (29%)	6 (30%)	6 (29%)
Isthmic spondylolisthesis	29 (71%)	14 (70%)	15 (71%)
Level fused, n (%)			
L3-L4	5 (12%)	2 (10%)	3 (14%)
L4-L5	18 (44%)	8 (40%)	10 (48%)
L5-S1	18 (44%)	10 (50%)	8 (38%)
1-year fusion status, n (%)			
Fusion	27 (66%)	10 (50%)	17 (81%)
Doubtful fusion	7 (17%)	6 (30%)	1 (5%)
Nonunion	7 (17%)	4 (20%)	3 (14%)

n = number of patients, SD = standard deviation

Clinical outcomes

ODI, EQ-5D-3L index scores and VAS pain scores at each timepoint on group level are listed in **Table 6.2**. Both means ± standard deviation (SD) and medians along with their interquartile range (IQR) are reported as not all data are normally distributed.

The mean ODI improved from 43 ± 15 at baseline to 13 ± 16 at 1 year and slightly regressed to 20 ± 19 at final follow-up. The mixed ANOVA model for repeated measures showed no significant interaction between timing of follow-up and graft type on ODI ($F(2, 76) = 1.028$, $p = 0.363$). Tests of within-subjects effects and between-subjects effects of the mixed ANOVA indicated respectively a main effect of time ($F(2, 76) = 51.393$, $p < 0.001$), but no main effect of graft type ($F(1, 38) = 0.021$, $p = 0.884$). Post-hoc anal-

Table 6.2: Patient reported outcome measures at baseline, 1-year follow-up and long-term follow-up. Both means $\pm$ standard deviation (SD) and medians along with their interquartile range (IQR) are presented, as not all variables are normally distributed. VAS leg pain represents the maximum score for the left and right leg. VAS back pain is only measured at long-term follow-up.

	Baseline	**1-year follow-up**	**Long-term follow-up**
ODI	43 ± 15 47 (29 - 57)	13 ± 16 8 (0 - 20)	20 ± 19 16 (5 - 29)
EQ-5D-3L index score	0.448 ± 0.291 0.334 (0.183 - 0.693)	0.843 ± 0.210 0.843 (0.807 - 1.000)	0.784 ± 0.251 0.811 (0.750 - 1.000)
VAS leg pain	66 ± 24 68 (54 - 86)	16 ± 22 4 (0 - 32)	34 ± 33 23 (4 - 69)
VAS back pain			31 ± 28 21 (6 - 54)

ODI = Oswestry disability index, VAS = Visual analogue scale

ysis with Bonferroni correction confirmed a significant difference between baseline ODI and both postoperative timepoints ($p < 0.001$), but not between 1-year and long-term follow-up ($p = 0.075$).

Multiple regression showed that the ODI at long-term follow-up could not be predicted based on the independent variables: diagnosis, graft type or 1-year fusion status ($F(3, 37) = 1.033$, $p = 0.389$). The overall model fit was $R^2 = 0.077$. Based on these results and the sample size, all secondary outcomes are presented on group level.

As illustrated by **Table 6.2**, both the EQ-5D-3L index score and VAS leg pain regressed slightly between 1-year and long-term follow up. Friedman's test confirmed that the EQ-5D-3L index and VAS leg pain scores differed between timepoints (Friedman's $Q(2) = 36$, $p < 0.001$ and Friedman's $Q(2) = 28$, $p < 0.001$ respectively). Post-hoc testing with Dunn-Bonferroni correction showed however that for both outcomes the regression during follow-up was not significant (EQ-5D-3L $Z = 0.271$, $p = 0.769$ and VAS leg pain $Z = -0.485$, $p = 0.147$).

Satisfaction

Overall satisfaction with treatment was excellent, with a mean score of 8.0 ± 1.8 (range 3 - 10). The majority of the patients (76%) scored ≥ 8; only 5 patients scored < 6. Moreover, 78% would choose the same treatment again. The remaining patients answered this question with "I don't know".

Figure 6.2 shows that 78% of the patients reported improvement in back pain and 71% improvement in leg pain. Of the 6 patients who reported much worsening of back and/or leg pain, only 1 patient underwent revision surgery at the same level (case 3 in **Table 6.3**). Three of these 6 patients were scored as fused at 1-year follow-up, including the revised case.

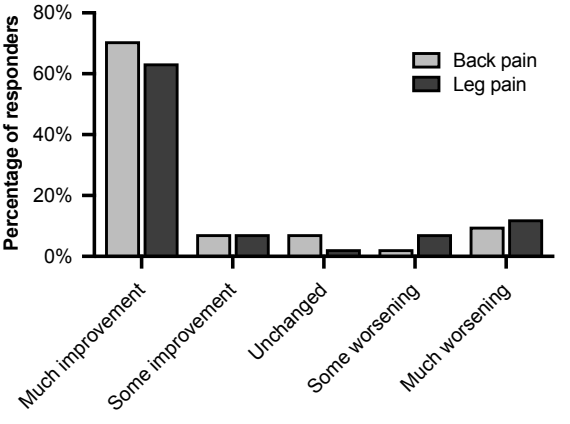


Figure 6.2: Effect of surgery on back pain (light grey) and leg pain (dark grey) at long-term follow-up.

To the question "On which complaint(s) had the surgery most effect?" 32% of the patients answered back pain, whereas 11% reported leg pain (**Figure 6.3**). More than half of the patients (53%) reported a combined effect. According to 2 patients, the surgery was not effective at all. These patients scored consistently low on all satisfaction questions (satisfaction 4 and much worsening of both back and leg pain). Moreover, they reported severe disability based on the ODI at both 1-year and long-term follow-up

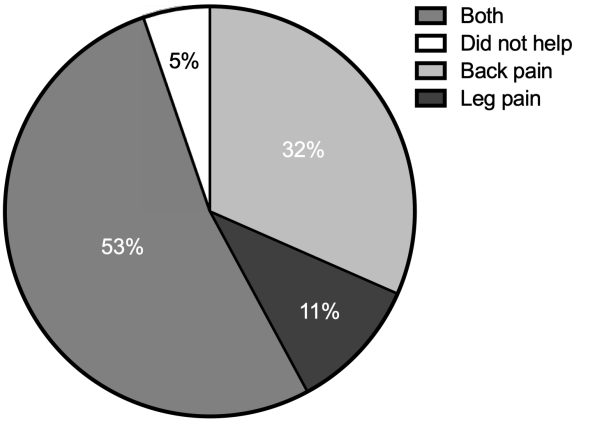


Figure 6.3: Main effect of surgery at long-term follow-up.

(ranging between 40 and 47), but only at long-term follow-up a severe VAS leg pain score (>80) and very low EQ-5D-3L index score (0.174). Their VAS back pain score at long-term follow-up was also >80. One of these unsatisfied patients was scored as fused at 1-year follow-up.

Additional surgery

As outlined in **Table 6.3**, 4 patients underwent additional lumbar spine surgery since the final follow-up of the initial study, but none of these surgeries were related to nonunion.

Table 6.3: Overview of additional lumbar spine surgeries since 1-year follow-up.

Case	Index surgery		Revision surgery		
	Level fused	Graft material	Timing (years)	Indication	Treatment
1	L4-L5	autograft	2.7	Discopathy L5-S1	Revision spondylodesis L4-S1
2	L3-L4	autograft	9.1	Stenosis L4-L5	Revision spondylodesis L3-L5
3	L5-S1	autograft	4.7	Back pain with radicular symptoms	Unknown
4	L4-L5	OP-1	9.4	Foraminal stenosis L3-L4	Decompression L3-L4

Discussion

This study showed excellent long-term (>10 years) clinical outcomes of instrumented posterolateral spinal fusion for degenerative and isthmic spondylolisthesis. Although the clinical success of spinal fusions is often debated, the quality of life and satisfaction outcomes of the current study are comparable to the most successful orthopaedic procedures, such as hip and knee arthroplasty.[180–183] Interestingly, patients reported not only clinical improvement for neurological symptoms, but at least as much for back pain. Only 11% indicated a main effect of surgery on leg pain. VAS back and leg pain scores at long-term follow-up were very similar. Apparently back pain is an important contributor to discomfort in spondylolisthesis cases; also in patients with neurological symptoms, which were a prerequisite for inclusion in the original study.

Although the clinical outcomes remained satisfactory for 10 years, a slight but non-significant deterioration in ODI, EQ-5D-3L index and VAS leg pain score compared to 1-year follow-up was observed. Such diminishment of the treatment effect was also ob-

served in a similar study by Ekman et al. and may be caused by adjacent segment degeneration or general effects of aging.[170] These effects cannot be further quantified as no radiographic or clinical assessment was performed and information on concomitant diseases was lacking. On the other hand, the clinical relevance of adjacent segment degeneration seems to be limited.[174,184,185] In the current study, only 3 patients (7%) underwent additional surgery at an adjacent level. Another explanation could be the diminishing of the placebo effect of surgery over time or the psychological phenomenon known as response shift.[186,187]

Recognizing the difficulty to compare our results with previous long-term follow-up studies of spinal fusion for spondylolisthesis, due to differences in indication, type of surgery, follow-up period and/or outcome measures, our patients reported relatively low ODI and relatively high EQ-5D-3L index scores at each timepoint.[109,170,171,173] Satisfaction with treatment falls well within the range reported in literature.[170,171,173,188,189] Contrary, the long-term VAS leg pain score was relatively high.[171,189] Interestingly, none of these previous studies had neurological manifestations as strict inclusion criterion. We and many others believe that these symptoms are an important indication for spinal surgery, as illustrated by the less favourable treatment effect achieved for patients with chronic low back pain without nerve root compression.[108,158,190] A recent meta-analysis on surgical treatment for degenerative spinal conditions indicated that lumbar radiculopathy was associated with the greatest mean change in health related quality of life from baseline.[191]

None of the participants underwent revision surgery for pseudoarthrosis, despite a substantial number of patients (34%) that were classified as “not fused” on the CT scan at 1-year follow-up. Also, based on the primary outcome measure ODI, no relationship was found between fusion status and long-term clinical outcome. Both patients classified as “fused” and “not fused” experienced a low level of disability at long-term follow-up (mean 21 ± 20 and 17 ± 16 respectively). Although a number of patients possibly developed further bony fusion in the course of the follow-up period, it is also possible that the combination of a fibrous union with pedicle screw instrumentation in situ offers sufficient stability in this patient population.

In line with the 1-year results of the original study, no difference in long-term ODI was seen between the patients who received OP-1 combined with local bone and solely autologous bone graft. This confirms the absence of a strong relationship between radiographic and clinical outcomes. Consecutive clinical trials failed to demonstrate noninferiority of OP-1 vs. autograft for spinal fusion and Osigraft was withdrawn from the market in 2015.[102,192]

The findings of this study add to the scarce literature on long-term clinical outcomes of spinal fusion and endorse the importance of appropriate surgical patient selection. However, we do recognize some limitations. First, this long-term follow-up was confined to only the Dutch participants of the original international multicenter study. Despite the acceptable follow-up rate of 73%, this resulted in a relatively small sample size.[193] Participants were however equally distributed among the randomized treatment groups and their baseline and 1-year clinical outcomes were comparable with the outcomes of both the total study population and the entire Dutch sample, reducing the risk of selection bias. Third, the outcomes of this study were limited to patient reported outcome measures. Radiological fusion was only evaluated at 1-year follow-up. Finally, back pain was only assessed at long-term follow-up and in relation to that, patients' preoperative main complaint was unknown.

In conclusion, this study showed favourable long-term clinical outcomes in patients who underwent instrumented posterolateral spinal fusion for spondylolisthesis with neurological symptoms. Diagnosis (degenerative vs. isthmic spondylolisthesis), graft type (OP-1 vs. autograft) and 1-year fusion status were not predictive for the ODI >10 years after surgery. Comparison with available long-term follow-up studies stresses the necessity of established and strict indications for this procedure.

CHAPTER 7

Patients cannot reliably distinguish the iliac crest bone graft donor site from the contralateral side after posterolateral lumbar fusion

A.M. Lehr
F.C. Oner
E.A. Hoebink
D.H.R. Kempen
J.L.C. van Susante
R.M. Castelein
M.C. Kruyt

Patients cannot reliably distinguish the iliac crest bone graft donor site from the contralateral side after lumbar spine fusion - A patient-blinded randomized controlled trial
Spine 2019;44:527-533

doi: 10.1097/BRS.0000000000002882

Introduction

Spinal fusion is a surgical procedure that frequently involves a bone grafting procedure to establish a permanent bony fusion between vertebral segments. It is most commonly performed in the lumbar spine, followed by the cervical spine. Over the last decades, the number and costs of spinal fusion as treatment for various spinal conditions increased significantly.[5,194,195] With more than 500,000 procedures performed annually in the United States (US)[90], spinal fusion is likely the largest single indication for bone grafting.

Autologous bone graft is considered to be the gold standard to promote a bony fusion and the iliac crest is the most common harvest site due to its easy access and adequate bone stock.[8,12] However, potential donor site morbidity is generally considered a serious disadvantage of the harvesting procedure, with donor site pain being the most commonly reported problem.[6,196–202] This has driven the development and popularity of many bone graft substitutes like allografts, bone morphogenetic proteins and synthetic bone grafts, with a global market value of more than US$2.5 billion.[203–206]

Interestingly, some recent studies have debated the incidence and severity of donor site pain. Especially in lumbar fusion surgery, where the pain related to the primary surgical site may interfere with perceived iliac crest pain and a less traumatic harvesting technique avoiding separate incisions can be used.[33–35,207,208] However, most of these studies are limited by their retrospective design, non-blinded assessment of subjective outcomes or small sample size, which contributes to the controversy about donor site morbidity in this specific group.

We investigated the relevance of pain attributed to the donor site after instrumented posterolateral lumbar fusion in a prospective cohort of patients blinded to the bone graft harvest site. These data are collected as part of a multicenter, blinded, randomized, intrapatient controlled trial to investigate the noninferiority of a ceramic bone graft substitute compared to autograft. The current study investigated 1) whether patients can identify the iliac crest that is used for bone harvesting and 2) whether this iliac crest is more painful than the contralateral side.

Materials and methods

Study design

The study involves a cohort of 92 adult patients who underwent an open, instrumented, posterolateral fusion between T10 and S1, as part of a multicenter, randomized, intra-

patient controlled trial on safety and efficacy of a microporous biphasic calcium phosphate ceramic bone graft substitute (AttraX® Putty, NuVasive Inc., San Diego, CA, USA; ClinicalTrials.gov NCT01982045). One side of the fusion trajectory was grafted with the ceramic bone graft substitute; the contralateral side was grafted with autologous bone from the iliac crest combined with local decompressive bone (autograft). Unilateral iliac crest bone graft harvesting was done through the same midline incision as used for the primary fusion surgery. Allocation of the graft type (left vs. right), and therefore the iliac crest bone graft harvest site, was based on computerized simple randomization. So each patient was blinded to both the allocation of the graft types and the iliac crest donor site. Radiological and clinical assessments were performed preoperatively and at 6 weeks, 3 months, 6 months and 1 year after surgery.

Patients

Patients between 18 and 80 years of age were enrolled and treated in four Dutch hospitals between November 2013 and July 2016. Indications for fusion surgery were deformity, structural instability and/or expected instability (for example as a result of decompression for spinal stenosis). Exclusion criteria included recent traumatic spinal fracture, primary or metastatic spinal tumor, active local or systemic infection, any previous surgical attempt for spinal fusion or previous autologous bone grafting procedures that comprised the quality and amount of iliac crest bone graft.

Surgical technique

All patients underwent a standard single or multilevel, instrumented, posterolateral fusion, with or without an additional interbody fusion device filled with local bone, at any of the segments between T10 and S1 via a posterior midline approach. The allocation side of the bone graft substitute was disclosed peroperatively after placement of all spinal implants, by opening a sealed envelope.

Iliac crest bone graft was harvested through the primary midline skin incision if possible, from the posterior iliac crest at the same side as placement of the autograft was allocated. Typically a separate fascia incision was made and cancellous bone was harvested through a unicortical window using osteotomes and surgical spoons or gouges. The use of hemostatic sponges was according the surgeon's preference. For the autograft side, a total of 10cc autograft per fusion level, comprising of at least 50% iliac crest bone, was intended. The volume was assessed by gentle pressure of the graft in a 20cc syringe.

After placement of both grafts around the posterior instrumentation and in the decorticated lateral gutters, the wound was closed in layers according to the surgeon's preference. No local anesthetics in the wound were used.

Outcome measures

Patients received a set of patient reported outcome measures (PROMs) preoperatively and at 6 weeks, 3 months, 6 months and 1 year after surgery. Pain in the lower back, left iliac crest area and right iliac crest area was measured using a continuous Visual Analogue Scale (VAS) ranging from 0 ("no pain at all") to 100 ("intolerable pain"). In addition, patients were asked to identify the iliac crest used for bone graft harvesting ("left" or "right") at all postoperative timepoints.

Statistical methods

Data were processed in Research Online for Researchers (Julius Center, University Medical Center Utrecht, Utrecht, The Netherlands) and analyzed using SPSS Statistics Version 22 (IBM). Baseline characteristics, surgical details and the identified harvest site were summarized using descriptive statistics.

To investigate whether patients could identify the iliac crest used for bone graft harvesting during the first year after surgery, their responses at each timepoint were summarized into three categories: "consistently correct" (harvest site identified correctly at each follow-up), "consistently incorrect" (contralateral iliac crest identified as harvest site at each follow-up) and "no idea/alternating responses" (no idea or alternating responses between follow-up visits). These categories were analyzed with descriptive statistics.

The VAS pain scores were analyzed using mixed effects models, to take into account the correlation of repeated measurements within patients. Postoperative pain scores for the iliac crest donor site were compared with the pain scores for the contralateral iliac crest. Fixed effects were timepoint (in weeks), preoperative iliac crest pain score and iliac crest side. In addition, the relation between donor site pain and amount of bone graft harvested was assessed with timepoint, preoperative iliac crest pain score and amount of bone graft harvested as fixed effects. In both models, a random intercept per patient and a first-order autoregressive correlation structure (AR(1)) were used to account for the correlation of measurements within patients over time. A third mixed model with a random intercept per patient was built to assess the relation between iliac crest pain and back pain, by subsequently adding the variables timepoint and back pain as fixed effects and calculating the R^2. The robustness of the mixed models was

verified by a log transformation of iliac crest pain scores to correct for the right-skewed distribution. Statistical significance was set at $p < 0.05$ and a minimal clinically important difference (MCID) value of 15mm for the VAS was adopted.[99]

Patients who completed less than two postoperative questionnaires were excluded from all analyses. Missing values were handled by pairwise deletion of cases, except for the mixed models. Normally distributed data are presented as mean and standard deviation (SD), whereas non-normally distributed data are presented as median and range or interquartile range (IQR).

Ethical considerations

This investigator-initiated study was approved by the Medical Research Ethics Committee of the University Medical Center Utrecht, Utrecht, The Netherlands, and the Institutional Review Board of each participating hospital. Study procedures were performed in accordance with the World Medical Association Declaration of Helsinki (version October 2008) and the Medical Research Involving Human Subjects Act (WMO). Written informed consent was obtained from each individual patient included in the study.

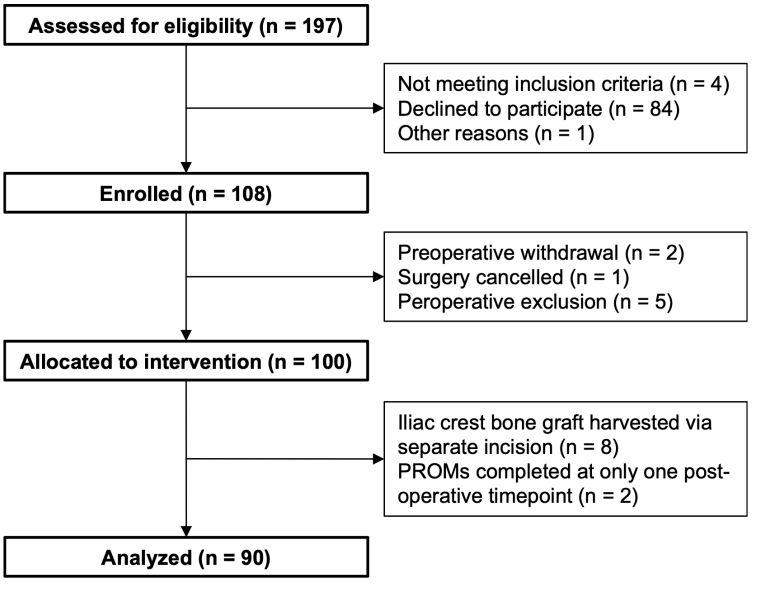


Figure 7.1: Flowchart of the progress of patients through each stage of the study.

Results

Patient characteristics

In 92 of the 100 patients operated on, the iliac crest bone graft was harvested through the primary midline incision and they were blinded to the donor site. The remaining 8 patients were excluded as a separate incision was used. Two other patients were excluded, because they completed the PROMs at only one postoperative timepoint (**Figure 7.1**).

Baseline characteristics and surgical details of the 90 patients included in this study are presented in **Table 7.1**. There were 44 males and 46 females with a mean age of 54.4 ± 12.1 years (range 27 - 79). The majority of the patients underwent a single level (67%) or two-level (20%) fusion. All fusions involved the lower lumbar area (L3 or more caudally), whereas 89% underwent a low lumbar fusion between L3 and S1. Additional

Table 7.1: Baseline characteristics and surgical details.

Number of patients	90	
Age, mean ± SD (range), years	54.4 ± 12.1 (27 - 79)	
Sex, n (%)		
Male	44 (49%)	
Female	46 (51%)	
Smokers, n (%)	31 (34%)	
Number of segments fused, median (range)	1 (1 - 8)	
Spinal region fused, n (%)		
Thoracolumbar	5 (6%)	
Lumbar	85 (94%)	
Decompression, n (%)	87 (97%)	
Interbody device, n (%)	57 (63%)	
Level and type of interbody device, n	PLIF	TLIF
L3-L4	6	4
L4-L5	18	6
L5-S1	25	6
Donor site, n (%)		
Left iliac crest	45 (50%)	
Right iliac crest	45 (50%)	
Obtained iliac crest bone graft, median (range), cc	6 (5 - 40)	

n = number of patients, PLIF = posterior lumbar interbody fusion, SD = standard deviation, TLIF = transforaminal lumbar interbody fusion

posterior lumbar interbody fusion (PLIF) and transformational lumbar interbody fusion (TLIF) procedures were performed in 49% and 14% of the patients respectively. The left/right distribution of the iliac crest donor site was 50/50. The amount of obtained iliac crest bone graft ranged from 5 to 40cc, with a median of 6cc. A total of 73 patients (81%) completed all PROMs. Two patients dropped-out after 3 months and 6 months of follow-up respectively. The remaining 15 patients completed follow-up, but had one or more missing values.

Harvesting procedure

Iliac crest bone graft harvesting was uneventful in all patients and there were no deviations from the assigned iliac crest based on bone quality. One patient was unblinded after 3 months because of concerns about pain in the iliac crest area. The pain was indeed on the donor site and no further treatment was required.

To gain some insight in the extra surgical time required for harvesting, this was recorded in the last 36 patients. The median time to obtain median 5 (5 - 40)cc bone graft was 7.5 (5 - 25) minutes.

Donor site identification

To the question “Is bone graft harvested from your left or right iliac crest?”, 46 patients (51%) consistently reported the same iliac crest site at each follow-up (**Figure 7.2**). Of these patients, only 22 (48%) identified the harvest site correctly (“consistently correct”); the other 24 (52%) patients identified the non-operated iliac crest (“consistently incorrect”). The remaining 44 (49%) patients had no idea which iliac crest was used for bone graft harvesting or their responses alternated between follow-up visits. Looking at the iliac crest reported most frequently by these patients, 20 (45%) guessed the harvest site correctly.

Post-hoc subgroup analyses using Mann Whitney U-tests showed no difference in number of segments fused ($p = 0.444$) or amount of iliac crest bone graft harvested ($p = 0.471$) between the 22 patients that identified the harvest site correctly and the other 68 patients.

VAS pain scores

The results of the VAS pain scores are summarized in the Tukey boxplots in **Figure 7.3**, each representing at least 84 patients. Although the pain scores varied widely, the VAS for back pain was higher than the iliac crest pain scores at all timepoints measured. The highest scores for both back pain and iliac crest pain were observed preoperatively.

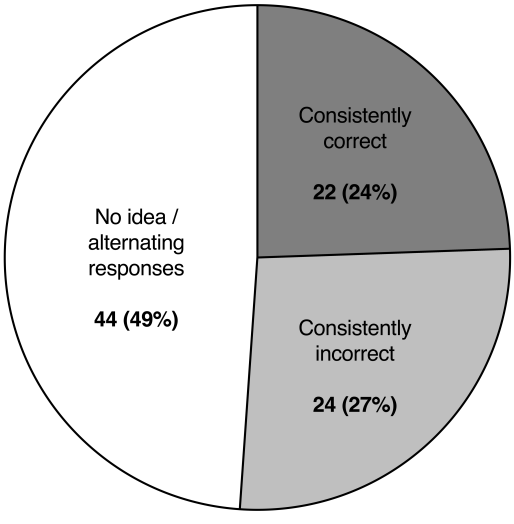


Figure 7.2: Number and percentage of patients that identified the iliac crest used for bone graft harvesting at each follow-up correctly (consistently correct), incorrectly (consistently incorrect), or had no idea/ alternating responses between follow-up visits (no idea/alternating responses).

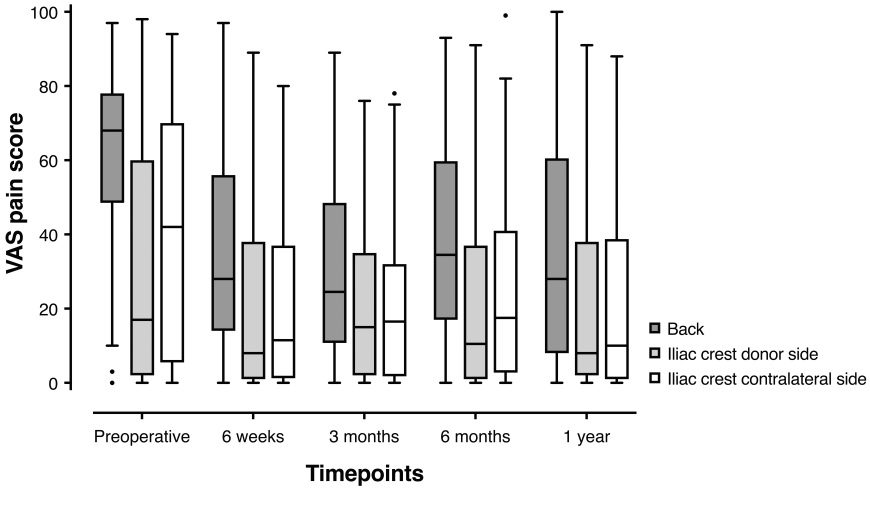


Figure 7.3: VAS scores (0 - 100) for back pain (back), pain at the iliac crest used for bone graft harvesting (iliac crest donor side) and pain at the contralateral iliac crest (iliac crest contralateral side) at each timepoint. The Tukey boxplots illustrate the median, interquartile range (IQR), 1.5 IQR (whiskers) and outliers (dots).

Postoperatively, the median VAS for iliac crest pain ranged between 8 and 18mm (IQR 1 - 41), whereas the median VAS for back pain ranged between 28 and 35mm (IQR 8 - 61). Based on a MCID of 15mm, no relevant changes in pain scores over time were observed, except for a mean decrease of 28mm (SD 32) in the VAS for back pain at 6 weeks follow-up compared to the preoperative score.

The mixed model for iliac crest pain, corrected for preoperative pain scores, showed no difference between the VAS pain scores for the donor site compared to the contralateral iliac crest (estimated difference 0.57mm, 95% confidence interval -3.57 to 4.71, $p = 0.787$). In addition, there was no effect of time after surgery ($p = 0.807$) and no relation between the VAS for donor site pain and the amount of bone graft harvested ($p = 0.574$). There was a strong relation between iliac crest pain and back pain ($p < 0.0005$). Analysis of the estimates of covariance parameters showed that 34% of the variance in iliac crest pain can be explained by back pain.

Discussion

Despite numerous articles that highlight the morbidity of iliac crest bone graft harvesting, there is actually controversy regarding the true morbidity of this procedure in spinal surgery. Especially in lumbar fusion surgery, where donor site pain might be overestimated due to the close proximity of the primary surgical site. The incidence and severity of iliac crest donor site pain reported in literature varies widely, probably explained by poor (unblinded) study designs, as well as different study populations and surgical techniques.

In this current prospective, patient-blinded, randomized, controlled trial, the perceived pain during the first year after instrumented posterolateral fusion of the lumbar spine via a single midline incision approach could not be related to the actual donor site. Half of the patients had no idea from which iliac crest the bone graft was harvested and of the other half, only 48% identified the donor site correctly. In other words, the correct identification of the donor site followed a perfect random distribution. Moreover, the patient reported VAS pain scores for the donor site and untouched contralateral iliac crest did not differ and were at each timepoint lower than the scores for back pain. Remarkably, the highest iliac crest pain scores were observed preoperatively. These findings are actually in line with some recently published studies, which underlined the difficulty to distinguish iliac crest pain from other potential sources of back pain after lumbar fusion surgery.[33–36,196,201] In a similar, patient-blinded, randomized study with 32 patients and 12 months follow-up, also no difference in iliac crest VAS pain scores was identified and highest pain scores were measured preoperatively.[36] Delawi et al. found that pain attributed to the donor site at mean 7.3 (range 2.3 - 11.6) years follow-up

was significantly lower in patient who underwent a high lumbar fusion compared with low fusions, suggesting an interference from the primary surgical site or residual back pain.[33] By investigating the incidence of iliac crest pain in patients with and without unilateral iliac crest harvesting for lumbar spinal fusion, Howard et al. demonstrated that iliac crest pain is a poor marker for donor site morbidity.[34]

Bone graft harvesting from the iliac crest, whether or not via a separate incision, is an invasive procedure, inherent to donor site morbidity. Especially the harvest of large structural grafts via a separate incision may be problematic. However, the vast majority of iliac bone graft is used for lumbar spinal fusion surgery and can often be obtained via the primary skin incision. As shown in the current study including 90 patients, this iliac crest bone graft harvesting technique does not lead to clinically relevant pain that can be attributed to the actual donor site. This knowledge is important since the presumed donor site pain is the main reason for many surgeons to use expensive commercially available bone graft alternatives. We believe that, until alternatives have demonstrated clear superiority over autologous bone graft, the added value of these materials in lumbar fusion surgery is limited to specific cases where the quality or quantity of the autologous bone graft is limited or to shorten the surgical procedure. The knowledge from this study may also help to manage patient's expectations.

Strengths of this study include the prospective and blinded design with multiple measurements including at baseline, as well as the relatively large sample size. Blinding is known to reduce the risk of bias, especially in trials with subjective outcome measures like pain scores.[209,210] In addition, pain in both iliac crests and the back were measured separately. However, there are also limitations. First of all, this study was a priori powered to determine noninferiority of a bone graft substitute and not to detect differences between iliac crest pain scores. However, as the 95% confidence interval around the estimated difference in iliac crest pain is much narrower than the MCID of 15mm on the VAS, it is unlikely that we missed a clinically relevant effect. Second, the exact locations of perceived pain in the iliac crest/hip region and lower back region were not assessed. In addition, the type of pain (e.g. sharp, aching or tingling) and interference with functioning in the direct postoperative phase and daily activities were not measured. Pain sensation may also be influenced by satisfaction with treatment.[199] Regarding donor site identification, patients were not asked for the level of confidence in their responses. To account for guessing, the corresponding question was repeated at each follow-up and responses were summarized into three categories. Finally, the bone graft harvesting procedure was not standardized, although all surgeons applied a comparable cortical window technique without heat necrosis.

In conclusion, patients surgically treated for lumbar spine fusion could not reliably identify the iliac crest used for bone graft harvesting and this iliac crest was not more painful than the untouched contralateral iliac crest. Therefore, donor site pain should not be the main reason to use bone graft alternatives for lumbar spinal fusion surgery.

CHAPTER 8

General discussion and future directions

This thesis focused on the clinical investigation of bone graft substitutes as an alternative to autologous bone graft in posterolateral spinal fusion surgery. Both the methodological aspects of a randomized intrapatient controlled trial (intrapatient RCT) in this field and the efficacy of a promising biphasic calcium phosphate ceramic (BCP$_{<\mu m}$; AttraX® Putty, NuVasive Inc., San Diego, CA, USA) were examined. Moreover, in order to advance the assessment of posterolateral lumbar fusion, an extensive systematic review of radiological fusion criteria was performed. In this chapter, the findings of the previous chapters are summarized and critically discussed to advance the clinical evaluation and application of bone graft substitutes.

Key findings

The key findings of this thesis are:

- A wide variety of classifications and criteria has been used to assess the fusion status after instrumented posterolateral lumbar fusion, mainly on radiographs and computed tomography scans, but none of these have both sufficient diagnostic accuracy and reliability (**Chapter 2**).
- Based on a noninferiority margin of 15% difference, the 1-year fusion performance of BCP$_{<\mu m}$ was noninferior to autograft in instrumented posterolateral (thoraco)lumbar fusion (**Chapter 3**).
- The fusion rate of both BCP$_{<\mu m}$ and autograft increased between 1 and 2 years of follow-up. Ongoing bone formation was, however, mainly observed between the immobilized facet joints, rather than in the intertransverse area where the grafts were applied (**Chapter 4**).
- Analysis of the correlation of fusion outcomes within patients, both between the left and right side of a spinal segment and between adjacent segments, showed that the posterolateral fusion rate highly depends on patient-related factors (**Chapter 5**).
- The applied intrapatient RCT design eliminated interpatient variability and reduced the required sample size considerably to less than half the sample size of a classic RCT (**Chapter 5**).
- More than 10 years after single-level instrumented posterolateral lumbar fusion for spondylolisthesis with neurological symptoms, patients reported favorable clinical outcomes, unrelated to the fusion status at 1-year follow-up (**Chapter 6**).
- Although bone graft harvesting from the iliac crests involves an extra procedure, patients who underwent a posterolateral (thoraco)lumbar fusion via a single midline incision could not reliably identify which iliac crest was used and the donor site was not more painful than the untouched contralateral iliac crest (**Chapter 7**).

Discussion

Due to both the increasing prevalence of spinal disorders among the aging population and advanced surgical techniques, the number of (lumbar) spinal fusions continues to rise.[3,4,111,211] Moreover, in comorbid or elderly patients with less bone (healing) quality, multilevel fusions or revision surgeries, autologous bone graft is often insufficient. The wide variety of bone graft substitutes that have been marketed may offer a solution to the increasing demand for bone grafts, but evidence of their clinical performance and benefit is still limited.[8,212,213] For ceramic bone graft substitutes, we know from in vitro and in vivo studies that performance strongly depends on their distinctive physicochemical and microstructural properties, whereas market introduction is largely based on substantial equivalence to an already marketed competitor as demonstrated through technical documentation.[15,22,214,215] This is in sharp contrast to pharmaceutical products that can only be registered when clinical efficacy and safety have been demonstrated (phase III trial). To guide the selection and proper application of bone graft substitutes for specific indications, we urge to invest in clinical investigation of these materials, preferably in comparison with the current gold standard, with bony fusion as primary efficacy outcome. Especially in lumbar fusion patients where autologous bone graft can be harvested quite harmlessly via the primary incision (**Chapter 7**), the risks of autologous bone grafting should be carefully weighed against the presumed benefits and costs of a bone graft substitute.

Although an RCT is considered to generate the highest level of clinical evidence, the applicability of this design has been criticized in several non-pharmaceutical research fields including spine surgery. Alternative designs that overcome some of the major challenges of a conventional RCT, without compromising on validity, are gradually gaining ground.[28,132,137,216–218] The intrapatient RCT design discussed in this thesis holds specific promise as an efficient alternative to investigate the clinical performance of non-bioactive bone graft substitutes in instrumented posterolateral spinal fusion (**Chapter 5**). We thoroughly explored this design to provide a guideline for appropriate use. Key feature of the concurrent comparison of the experimental and control condition within each patient is the elimination of interpatient variability as confounder, resulting in a more direct comparison between the grafts and fusion outcomes. This yields immediate ethical, practical and financial advantages due to the reduction in sample size. To further develop and gain acceptance for this design, we advocate explicit reporting of the applied methodology, potential sources of bias and quantification of the correlation of outcomes within patients. The latter has a substantial impact on the required sample size and statistical methods that are appropriate. Neglecting this aspect can lead to under-/overpowered studies or invalid estimates of the treat-

ment effect.[160,162] Hence, we strongly recommend to involve both an epidemiologist and statistician in the design and analysis phase of this type of clinical trial.

Another step forward for the clinical investigation of bone graft substitutes would be the standardization of outcomes, both to improve the quality of clinical studies and to facilitate the comparison between different studies. Computed tomography (CT) is becoming the recommended standard for noninvasive assessment of the fusion status, but to the best of our knowledge none of the classifications or criteria that have been used to determine solid posterolateral fusion have sufficient reliability and accuracy (**Chapter 2**). In addition to the rather subjective terminology, many articles do not specify the location of bony bridging (intertransverse or interfacet) and whether uni- and/or bilateral fusion is considered successful. Last but not least, in spite of the introduction of rigid instrumentation, absence of intervertebral motion based on flexion-extension radiographs remains a dominant criterion. As a consequence of both the variety and limited diagnostic evidence of applied fusion assessment methods, it is difficult to draw solid conclusions about the efficacy of bone graft (substitutes) and to compare the outcomes of different clinical studies. This possibly also contributes to the ongoing controversy about the correlation between radiographic fusion status and clinical outcomes.[1,122] Although impractical and also considered unethical for routine follow-up and as primary clinical study outcome, thorough surgical exploration of the fusion mass and intervertebral motion persists as most valid method to determine the fusion status. The findings of **Chapter 2** direct towards the development of a systematic CT-based approach that allows grading of the quality of the fusion mass at each side of each fusion level, at specified anatomical locations, and includes the assessment of signs of nonunion.

In line with previous studies, this thesis demonstrates the significant contribution of facet fusions to a solid arthrodesis (**Chapter 3** and **Chapter 4**).[26,55,73,219] Ankylosis of the facet joints most likely occurs as a result of immobilization due to rigid instrumentation, with or without destruction of the joint surface. This is supported by studies reporting spontaneous facet fusion after percutaneous fixation of the spine.[220,221] Progression of fusion between 1 and 2 years of follow-up was only identified at the facet joints (**Chapter 4**), which corresponds with the observation that both the ceramic and autograft were completely resorbed on the 1-year CT scans and therefore no longer present as template for bone formation (**Chapter 3**). Based on these findings, we believe that the true efficacy of bone grafting is best determined by bony bridging between the transverse processes, at the latest 1 year after the surgical procedure. For routine follow-up, assessment of both intertransverse and interfacet fusion is considered relevant. Although still in its infancy, repeated assessment of the volume or bone mineral density of the grafted area during the first year after surgery could provide valuable in-

sight in the dynamics of graft resorption, bone deposition, and remodeling of the fusion mass.[118,222]

Recognizing the limitations of radiographic fusion assessment, the relatively low inter-transverse fusion rate of both the ceramic (25%) and autograft (30%) observed at 1 year follow-up may suggest that the grafts resorbed too fast compared to the rate of bone formation (**Chapter 3**). Recent developments of microporous BCP$_{<\mu m}$ ceramic bone graft substitutes focused on a slower resorption rate and accelerated bone formation by, on the one hand decreasing the ratio of fast resorbing β-tricalcium phosphate (β-TCP) and slow resorbing hydroxyapatite (HA), and on the other hand improvement of the submicron surface structure.[214,223,224] The Dutch Clinical Spine Research Group is currently investigating the clinical efficacy of this successor BCP$_{<\mu m}$ ceramic (commercially available as MagnetOs™ Granules, Kuros Biosciences BV, Bilthoven, The Netherlands) in a replica intrapatient RCT.

In addition to radiological fusion as logical primary outcome measure of the intended use of bone grafts (i.e. to establish a solid fusion between vertebrae), relevant secondary outcome measures of clinical trials investigating bone grafts are patient reported outcome measures (PROMs) and (serious) adverse events ((S)AEs). While we acknowledge that the main limitation of an intrapatient controlled trial design is that these outcomes cannot be attributed separately to the investigated grafts, the relevance of this limitation can be debated (**Chapter 5**). Also in two-arm RCTs it is difficult to establish a direct relation between the fusion performance of the investigated grafts and patient reported outcomes. A standard set of disease specific, domain specific and generic PROMs mainly gives valuable information about the overall effect of spinal fusion surgery as experienced by the patient, adds to the interpretation of for example radiological outcomes and can signal deviations from control populations, both on the short and long term.

The relevance of (S)AE registrations largely relies on their reliability in terms of unbiased and consistent reporting. Potential sources of bias are observer and reporting bias. Using the inclusive definition of AEs*, we experienced the distinction between AEs and the usual postoperative course as a major challenge. Especially in comorbid patients and patients who underwent multilevel fusions or additional surgical procedures like an osteotomy for whom AEs are a daily observation. Similar to the conduct of RCTs, this methodological practice of registering all AEs is much easier in pharma-

*Adverse events are defined as any undesirable experience occurring to a subject during the study, whether or not considered related to [the investigational product / trial procedure/ the experimental intervention]. Source: https://www.ccmo.nl/onderzoekers/publicaties/formulieren/2018/09/19/standaardonderzoeksdossier-c1-model-onderzoeksprotocol

ceutical trials, where patients are not subject to major interventions. To objectify (S)AE registration wherever possible, we therefore advise to limit and specify in the study protocol which (S)AEs (do not) need to be registered, and to discuss encountered (S)AEs low-threshold within the research team. Assessment tools that have been developed and validated for specific patient populations or interventions like for example the Spine AdVerse Events Severity system (SAVES) might be helpful, as they provide clinically relevant universal definitions.[110,225]

Current developments and future directions

Supported by the clinical data from this thesis, BCP ceramics have great potential as standalone bone graft substitute for instrumented posterolateral spinal fusion when autologous bone graft is insufficient. Main advantages with respect to other strategies, like cell-based implants and bioactive glasses, are the relatively low costs, no safety concerns and established efficacy as standalone product. However, the fusion rate is still suboptimal, as is the case with autologous bone graft.[15,24]

Although great scientific achievements have been made, especially with regard to surface features, unraveling the effect of distinct physicochemical properties on fusion performance and the underlying interaction with the body's own cells may further improve the regenerative capacity of BCP ceramics. Advancement of BCP$_{<\mu m}$ ceramics is currently directed at the interplay between graft resorption and bone formation.[15,214,224] Future preclinical and clinical investigations have to reveal whether superior fusion performance compared to autograft can be achieved, without compromising on safety and benefits like ease of manufacturing and storage.

In our opinion, the findings of this thesis also give reason to take one step back and further examine the position of bone grafting in the light of current surgical techniques. Intertransverse process arthrodesis with autograft stems from the era before the application of rigid spinal instrumentation. At that time, patients were grafted and subsequently immobilized in a plaster cast for half a year or longer to achieve bony bridging. With the current surgical standard for posterolateral fusion, it might not be necessary to pursue a true lateral intertransverse process arthrodesis for a successful outcome. On the other hand, we have the impression that surgeons increasingly rely on the performance of spinal implants and bone graft substitutes, at the expense of the surgical techniques that are crucial for the biological process of bone healing like thorough decortication and careful positioning of sufficient amounts of bone graft.[226–228] Although facet fusions alone are presumed to be not sufficient, a solid construct can likely be achieved by a more posteromedial bony bridging involving the facet joints.[55,73,219] However, to the best of our knowledge, the long-term outcomes of such a fusion are not

yet known. From a fundamental scientific perspective, ethical considerations aside, an intrapatient comparison of bone graft vs. no bone graft with long-term follow-up would be a very interesting approach to elucidate the value of bone grafting for obtaining a solid arthrodesis in the presence of rigid instrumentation.

Following the establishment of more objective and accurate criteria for noninvasive assessment of the postoperative fusion status, we expect benefit from artificial intelligence (AI) as a powerful tool for automated quantification and classification. This will increase both the reliability and reproducibility of fusion assessment. Another potential application of AI is the prediction of outcomes, based on the current fusion status in combination with other parameters, and thereby support decision-making for further treatment.[70–72] Furthermore, current developments directed at the visualization of bony structures using magnetic resonance imaging and ultrasound in other spinal research fields, including chronic low back pain and scoliosis, may open doors for radiation-free assessment of bony fusion.[53,75,229,230]

With the application of the European Medical Device Regulation (MDR) in 2021, clinical evidence of the safety and efficacy of medical devices such as bone graft substitutes becomes the norm. Predominantly, the requirements for the technical documentation of medical devices to obtain the CE-mark are more strict, the equivalence pathway is practically impossible and post-market clinical follow-up (PMCF) is now mandatory. These developments ask for reallocation and more efficient use of research and development resources of medical device manufacturers. Otherwise, medical device companies will drastically reduce their portfolio and/or cover the additional study costs through higher sales prices. Specifically for the clinical evaluation of ceramic bone graft substitutes in instrumented posterolateral spinal fusion, the intrapatient concurrent controlled design described in this thesis holds great opportunities as it reduces both study duration and costs.

The elaboration of the MDR also fuels the discussion on sponsorship, conflicts of interest and publication bias of clinical trials. While the availability of clinical trial registries, stricter criteria for scientific publications, open access journals and public data repositories are great steps forward, important dilemmas remain unsolved.[231] For example: How can the independence of research, that potentially bears commercial benefits for the medical industry, be warranted when sponsored by industry? Or conversely, should clinical trials in this field be done without industry funding and consequently consume great amounts of public resources? Possible directions could be investigator initiated trials funded by industry grants in combination with unrestricted research grants, to advance related research fields, or backflow of commercial profits to public funds dedicated for clinical investigations of medical devices. The idea of an independent medical

device research fund fueled by a small global tax on sold products has been proposed a decade ago, but such a funding strategy is not (yet) embraced.[232] Similarly, such funds should be applied for independent and ideally standardized PMCF.

Although the practical implementation of especially PMCF is yet to be seen, the MDR requirements are expected to constrain the expansion of the bone graft substitutes market. Combined with a critical attitude towards the growing body of clinical data on efficacy and real benefit, and towards the underlying fusion assessment methods, this development will contribute to a more rational selection and application of bone graft substitutes. In the end, that is the essence of evidence based medicine.

References

1. Bono CM and Lee CK. Critical analysis of trends in fusion for degenerative disc disease over the past 20 years: influence of technique on fusion rate and clinical outcome. Spine 2004;29:455–63. DOI: 10.1097/01.BRS.0000090825.94611.28.
2. Deyo RA, Gray DT, Kreuter W, Mirza S, and Martin BI. United States trends in lumbar fusion surgery for degenerative conditions. Spine 2005;30:1441–5. DOI: 10.1097/01.brs.0000166503.37969.8a.
3. Makanji H, Schoenfeld AJ, Bhalla A, and Bono CM. Critical analysis of trends in lumbar fusion for degenerative disorders revisited: influence of technique on fusion rate and clinical outcomes. European Spine Journal 2018;27:1868–76. DOI: 10.1007/s00586-018-5544-x.
4. Martin BI, Mirza SK, Spina N, Spiker WR, Lawrence B, and Brodke DS. Trends in Lumbar Fusion Procedure Rates and Associated Hospital Costs for Degenerative Spinal Diseases in the United States, 2004 to 2015. Spine 2019;44:369–76. DOI: 10.1097/BRS.0000000000002822.
5. Rajaee SS, Bae HW, Kanim LEA, and Delamarter RB. Spinal fusion in the United States: Analysis of trends from 1998 to 2008. Spine 2012;37:67–76. DOI: 10.1097/BRS.0b013e31820cccfb.
6. Dimitriou R, Mataliotakis GI, Angoules AG, Kanakaris NK, and Giannoudis PV. Complications following autologous bone graft harvesting from the iliac crest and using the RIA: A systematic review. Injury 2011;42:S3–S15. DOI: 10.1016/j.injury.2011.06.015.
7. Myeroff C and Archdeacon M. Autogenous bone graft: Donor sites and techniques. The Journal of Bone & Joint Surgery 2011;93:2227–36. DOI: 10.2106/JBJS.J.01513.
8. Kadam A, Millhouse PW, Kepler CK, et al. Bone substitutes and expanders in spine surgery: A review of their fusion efficacies. International Journal of Spine Surgery 2016;10:33. DOI: 10.14444/3033.
9. Kurien T, Pearson RG, and Scammell BE. Bone graft substitutes currently available in orthopaedic practice. The Bone & Joint Journal 2013;95-B:583–97. DOI: 10.1302/0301-620X.95B5.30286.
10. Morris MT, Tarpada SP, and Cho W. Bone graft materials for posterolateral fusion made simple: a systematic review. European Spine Journal 2018;27:1856–67. DOI: 10.1007/s00586-018-5511-6.
11. Barradas AMC, Yuan H, van Blitterswijk CA, and Habibovic P. Osteoinductive biomaterials: current knowledge of properties, experimental models and biological mechanisms. European cells & Materials 2011;21:407–29. DOI: 10.22203/ecm.v021a31.
12. Campana V, Milano G, Pagano E, et al. Bone substitutes in orthopaedic surgery: from basic science to clinical practice. Journal of Materials Science: Materials in Medicine 2014;25:2445–61. DOI: 10.1007/s10856-014-5240-2.
13. Habibovic P and de Groot K. Osteoinductive biomaterials - properties and relevance in bone repair. Journal of Tissue Engineering and Regenerative Medicine 2007;1:25–32. DOI: 10.1002/term.5.

14. Barbieri D, Yuan H, Ismailolu AS, and de Bruijn JD. Comparison of two moldable calcium phosphate-based bone graft materials in a noninstrumented canine interspinous implantation model. Tissue Engineering Part A 2017;23:1310–20. DOI: 10.1089/ten.TEA.2016.0347.

15. Bouler JM, Pilet P, Gauthier O, and Verron E. Biphasic calcium phosphate ceramics for bone reconstruction: A review of biological response. Acta Biomaterialia 2017;53:1–12. DOI: 10.1016/j.actbio.2017.01.076.

16. Duan R, Barbieri D, Luo X, et al. Variation of the bone forming ability with the physicochemical properties of calcium phosphate bone substitutes. Biomaterials Science 2018;6:136–45. DOI: 10.1039/c7bm00717e.

17. Habibovic P, Yuan H, van den Doel M, Sees TM, van Blitterswijk CA, and de Groot K. Relevance of osteoinductive biomaterials in critical-sized orthotopic defect. Journal of Orthopaedic Research 2006;24:867–76. DOI: 10.1002/jor.20115.

18. Habibovic P, Kruyt MC, Juhl MV, et al. Comparative in vivo study of six hydroxyapatite-based bone graft substitutes. Journal of Orthopaedic Research 2008;26:1363–70. DOI: 10.1002/jor.20648.

19. LeGeros RZ. Calcium phosphate-based osteoinductive materials. Chemical Reviews 2008;108:4742–53. DOI: 10.1021/cr800427g.

20. Yuan H, Yang Z, Li Y, Zhang X, de Bruijn JD, and de Groot K. Osteoinduction by calcium phosphate biomaterials. Journal of Materials Science: Materials in Medicine 1998;9:723–6. DOI: 10.1023/A:1008950902047.

21. Yuan H, van Blitterswijk CA, de Groot K, and de Bruijn JD. Cross-species comparison of ectopic bone formation in biphasic calcium phosphate (BCP) and hydroxyapatite (HA) scaffolds. Tissue Engineering 2006;12:1607–15. DOI: 10.1089/ten.2006.12.1607.

22. Yuan H, Fernandes H, Habibovic P, et al. Osteoinductive ceramics as a synthetic alternative to autologous bone grafting. Proceedings of the National Academy of Sciences of the United States of America 2010;107:13614–9. DOI: 10.1073/pnas.1003600107.

23. Buser Z, Brodke DS, Youssef JA, et al. Synthetic bone graft versus autograft or allograft for spinal fusion: a systematic review. Journal of Neurosurgery: Spine 2016;25:509–16. DOI: 10.3171/2016.1.SPINE151005.

24. Cook RW and Hsu WK. Ceramics: Clinical evidence for ceramics in spine fusion. Seminars in Spine Surgery 2016;28:217–25. DOI: 10.1053/j.semss.2016.08.004.

25. Hsu WK, Nickoli MS, Wang JC, et al. Improving the Clinical Evidence of Bone Graft Substitute Technology in Lumbar Spine Surgery. Global Spine Journal 2012;2:239–48. DOI: 10.1055/s-0032-1315454.

26. Choudhri TF, Mummaneni PV, Dhall SS, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 4: Radiographic assessment of fusion status. Journal of Neurosurgery: Spine 2014;21:23–30. DOI: 10.3171/2014.4.SPINE14267.

27. Goldstein C and Drew B. When is a spine fused? Injury 2011;42:306–13. DOI: 10.1016/j.injury.2010.11.041.

28. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. Trials 2009;10:9. DOI: 10.1186/1745-6215-10-9.

29. McCulloch P, Taylor I, Sasako M, Lovett B, and Griffin D. Randomised trials in surgery: problems and possible solutions. British Medical Journal 2002;324:1448–51. DOI: 10.1136/bmj.324.7351.1448.

30. Sibai T, Carlisle H, and Tornetta III P. The darker side of randomized trials: Recruitment challenges. The Journal of Bone & Joint Surgery 2012;94:49–55. DOI: 10.2106/JBJS.L.00240.

31. Briel M, Speich B, von Elm E, and Gloy V. Comparison of randomized controlled trials discontinued or revised for poor recruitment and completed trials with the same research question: a matched qualitative study. Trials 2019;20:800. DOI: 10.1186/s13063-019-3957-4.

32. Ioannidis JP. Why most published research findings are false. PLoS Medicine 2005;2:696–701. DOI: 10.1371/journal.pmed.0020124.

33. Delawi D, Dhert WJA, Castelein RM, Verbout AJ, and Oner FC. The incidence of donor site pain after bone graft harvesting from the posterior iliac crest may be overestimated: A study on spine fracture patients. Spine 2007;32:1865–8. DOI: 10.1097/BRS.0b013e318107674e.

34. Howard JM, Glassman SD, and Carreon LY. Posterior iliac crest pain after posterolateral fusion with or without iliac crest graft harvest. The Spine Journal 2011;11:534–7. DOI: 10.1016/j.spinee.2010.09.001.

35. Pirris SM, Nottmeier EW, Kimes S, O'Brien M, and Rahmathulla G. A retrospective study of iliac crest bone grafting techniques with allograft reconstruction: do patients even know which iliac crest was harvested? Journal of Neurosurgery: Spine 2014;21:595–600. DOI: 10.3171/2014.6.SPINE13902.

36. Sheha ED, Meredith DS, Shifflett GD, et al. Postoperative pain following posterior iliac crest bone graft harvesting in spine surgery: a prospective, randomized trial. The Spine Journal 2018;18:986–92. DOI: 10.1016/j.spinee.2017.10.011.

37. Vaccaro AR, Chiba K, Heller JG, et al. Bone grafting alternatives in spinal surgery. The Spine Journal 2002;2:206–15. DOI: 10.1016/S1529-9430(02)00180-8.

38. Chun DS, Baker KC, and Hsu WK. Lumbar pseudarthrosis: a review of current diagnosis and treatment. Neurosurgery Focus 2015;39:E10. DOI: 10.3171/2015.7.FOCUS15292.

39. Hilibrand AS and Dina TS. The use of diagnostic imaging to assess spinal arthrodesis. Orthopedic Clinics of North America 1998;29:591–601. DOI: 10.1016/S0030-5898(05)70033-X.

40. Resnick DK, Choudhri TF, Dailey AT, et al. Guidelines for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 4: radiographic assessment of fusion. Journal of Neurosurgery: Spine 2005;2:653–7.
DOI: 10.3171/spi.2005.2.6.0653.

41. Peters MJM, Bastiaenen CHG, Brans BT, Weijers RE, and Willems PC. The diagnostic accuracy of imaging modalities to detect pseudarthrosis after spinal fusion - a systematic review and meta-analysis of the literature. Skeletal Radiology 2019;48:1499–510.
DOI: 10.1007/s00256-019-03181-5.

42. Ouzzani M, Hammady H, Fedorowicz Z, and Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. Systematic Reviews 2016;5:210.
DOI: 10.1186/s13643-016-0384-4.

43. Lenke LG, Bridwell KH, Bullis D, Betz RR, Baldus C, and Schoenecker PL. Results of In Situ Fusion for Isthmic Spondylisthesis. Journal of Spinal Disorders 1992;5:433–42.
DOI: 10.1097/00002517-199212000-00008.

44. Christensen FB, Laursen M, Gelineck J, Eiskjr SP, Thomsen K, and Bünger CE. Interobserver and intraobserver agreement of radiograph interpretation with and without pedicle screw implants: The need for a detailed classification system in posterolateral spinal fusion. Spine 2001;26:538–43. DOI: 10.1097/00007632-200103010-00018.

45. Singh K, Smucker JD, Gill S, and Boden SD. Use of recombinant human bone morphogenetic protein-2 as an adjunct in posterolateral lumbar spine fusion: a prospective CT-scan analysis at one and two years. Journal of Spinal Disorders & Techniques 2006;19:416–23. DOI: 10.1097/00024720-200608000-00008.

46. Suk SI, Lee CK, Kim WJ, Lee JH, Cho KJ, and Kim HG. Adding Posterior Lumbar Interbody Fusion to Pedicle Screw Fixation and Posterolateral Fusion After Decompression in Spondylolytic Spondylolisthesis. Spine 1997;22:210–9.
DOI: 10.1097/00007632-199701150-00016.

47. Bridwell KH, Lenke LG, McEnery KW, Baldus C, and Blanke K. Anterior Fresh Frozen Structural Allografts in the Thoracic and Lumbar Spine: Do they work if combined with posterior fusion and instrumentation in adult patients with kyphosis or anterior column defects? Spine 1995;20:1410–8. DOI: 10.1097/00007632-199520120-00014.

48. Glassman SD, Dimar JR, Carreon LY, Campbell MJ, Puno RM, and Johnson JR. Initial fusion rates with recombinant human bone morphogenetic protein-2/compression resistant matrix and a hydroxyapatite and tricalcium phosphate/collagen carrier in posterolateral spinal fusion. Spine 2005;30:1694–8. DOI: 10.1097/01.brs.0000172157.39513.80.

49. Brantigan JW and Steffee AD. A Carbon Fiber Implant to Aid Interbody Lumbar Fusion: Two-Year Clinical Results in the First 26 Patients. Spine 1993;18:2106–17.
DOI: 10.1097/00007632-199310001-00030.

50. Jorgenson SS, Lowe TG, France J, and Sabin J. A prospective analysis of autograft versus allograft in posterolateral lumbar fusion in the same patient. A minimum of 1-year follow-up in 144 patients. Spine 1994;19:2048–53. DOI: 10.1097/00007632-199409150-00005.

51. Kant AP, Daum WJ, Dean SM, and Uchida T. Evaluation of lumbar spine fusion. Plain radiographs versus direct surgical exploration and observation. Spine 1995;20:2313–7.

52. Larsen JM, Rimoldi RL, Capen DA, Nelson RW, Nagelberg S, and Thomas JC. Assessment of pseudarthrosis in pedicle screw fusion: A prospective study comparing plain radiographs, flexion/extension radiographs, CT scanning, and bone scintigraphy with operative findings. Journal of Spinal Disorders 1996;9:117–20. DOI: 10.1097/00002517-199604000-00005.

53. Jacobson JA, Starok M, Pathria MN, and Garfin SR. Pseudarthrosis: US evaluation after posterolateral spinal fusion (Work in progress). Radiology 1997;204:853–8. DOI: 10.1148/radiology.204.3.9280271.

54. Kanayama M, Hashimoto T, Shigenobu K, Yamane S, Bauer TW, and Togawa D. A Prospective Randomized Study of Posterolateral Lumbar Fusion Using Osteogenic Protein-1 (OP-1) Versus Local Autograft With Ceramic Bone Substitute. Spine 2006;31:1067–74. DOI: 10.1097/01.brs.0000216444.01888.21.

55. Carreon LY, Djurasovic M, Glassman SD, and Sailer P. Diagnostic accuracy and reliability of fine-cut CT scans with reconstructions to determine the status of an instrumented posterolateral fusion with surgical exploration as reference standard. Spine 2007;32:892–5. DOI: 10.1097/01.brs.0000259808.47104.dd.

56. Fogel GR, Toohey JS, Neidre A, and Brantigan JW. Fusion assessment of posterior lumbar interbody fusion using radiolucent cages : X-ray films and helical computed tomography scans compared with surgical exploration of fusion. The Spine Journal 2008;8:570–7. DOI: 10.1016/j.spinee.2007.03.013.

57. Spirig JM, Sutter R, Götschi T, Farshad-Amacker NA, and Farshad M. Value of standard radiographs, computed tomography, and magnetic resonance imaging of the lumbar spine in detection of intraoperatively confirmed pedicle screw loosening - a prospective clinical trial. The Spine Journal 2019;19:461–8. DOI: 10.1016/j.spinee.2018.06.345.

58. Tokuhashi Y, Matsuzaki H, Oda H, and Uei H. Clinical Course and Significance of the Clear Zone Around the Pedicle Screws in the Lumbar Degenerative Disease. Spine 2008;33:903–8. DOI: 10.1097/BRS.0b013e31816b1eff.

59. Dakhil-Jerew F, Jadeja H, Cohen A, and Shepperd J. Inter-observer reliability of detecting Dynesys pedicle screw using plain X-rays: a study on 50 post-operative patients. European Spine Journal 2009;18:1486–93. DOI: 10.1007/s00586-009-1071-0.

60. Gotfryd AO, Pomar FdM, Carneiro Neto NJ, Franzin FJ, Rodrigues LMR, and Poletto PR. Reliability analysis of radiographic methods for determination of posterolateral lumbossacral fusion. Einstein 2014;12:198–203. DOI: 10.1590/S1679-45082014AO2964.

61. Blumenthal SL and Gill K. Can lumbar spine radiographs accurately determine fusion in postoperative patients? Correlation of routine radiographs with a second surgical look at lumbar fusions. Spine 1993;18:1186–9. DOI: 10.1097/00007632-199307000-00010.

62. Bohnsack M, Gossé F, Rühmann O, and Wenger K. The value of scintigraphy in the diagnosis of pseudarthrosis after spinal fusion surgery. Journal of Spinal Disorders 1999;12:482–4.

63. Brodsky AE, Kovalsky ES, and Khalil MA. Correlation of radiologic assessment of lumbar spine fusions with surgical exploration. Spine 1991;16:S261–5.
DOI: 10.1097/00007632-199106001-00017.

64. Laasonen EM and Soini J. Low-back pain after lumbar fusion. Surgical and computed tomographic analysis. Spine 1989;14:210–3.
DOI: 10.1097/00007632-198902000-00011.

65. Bono CM, Khandha A, Vadapalli S, Holekamp S, Goel VK, and Garfin SR. Residual sagittal motion after lumbar fusion: A finite element analysis with implications on radiographic flexion-extension criteria. Spine 2007;32:417–22.
DOI: 10.1097/01.brs.0000255201.74795.20.

66. Lee KB, Johnson JS, Song KJ, Taghavi CE, and Wang JC. Use of autogenous bone graft compared with RhBMP in high-risk patients: A comparison of fusion rates and time to fusion. Journal of Spinal Disorders & Techniques 2013;26:233–8.
DOI: 10.1097/BSD.0b013e3182440162.

67. Nikkhoo M, Lu ML, Chen WC, et al. Biomechanical Investigation Between Rigid and Semirigid Posterolateral Fixation During Daily Activities: Geometrically Parametric Poroelastic Finite Element Analyses. Frontiers in Bioengineering and Biotechnology 2021;9:646079. DOI: 10.3389/fbioe.2021.646079.

68. Santos ERG, Goss DG, Morcom RK, and Fraser RD. Radiologic assessment of interbody fusion using carbon fiber cages. Spine 2003;28:997–1001.
DOI: 10.1097/01.BRS.0000061988.93175.74.

69. Center for Devices and Radiological Health; Orthopedic Devices Branch. Guidance Document for the Preparation of IDEs for Spinal Systems. Rockville, MD: U.S. Food and Drug Administration, 2000. URL: http://www.fda.gov/cdrh/ode/87.pdf.

70. Galbusera F, Casaroli G, and Bassani T. Artificial intelligence and machine learning in spine research. JOR SPINE 2019;2:e1044. DOI: 10.1002/jsp2.1044.

71. Huber FA and Guggenberger R. AI MSK clinical applications: spine imaging. Skeletal Radiology 2022;51:279–91. DOI: 10.1007/s00256-021-03862-0.

72. Rasouli JJ, Shao J, Neifert S, et al. Artificial Intelligence and Robotics in Spine Surgery. Global Spine Journal 2021;11:556–64. DOI: 10.1177/2192568220915718.

73. Kim DH, Hwang RW, Lee GH, et al. Potential significance of facet joint fusion or posteromedial fusion observed on CT imaging following attempted posterolateral or posterior interbody fusion. The Spine Journal 2020;20:337–43.
DOI: 10.1016/j.spinee.2019.10.010.

74. Kim KW, Ha KY, Moon MS, Kim YS, Kwon SY, and Woo YK. Volumetric change of the graft bone after intertransverse fusion. Spine 1999;24:428–33. DOI: 10.1097/00007632-199903010-00003.

75. Staartjes VE, Seevinck PR, Vandertop WP, van Stralen M, and Schröder ML. Magnetic resonance imaging-based synthetic computed tomography of the lumbar spine for surgical planning: a clinical proof-of-concept. Neurosurgical Focus 2021;50:E13. DOI: 10.3171/2020.10.FOCUS20801.

76. Thomsen K, Christensen FB, Eiskjaer SP, Hansen ES, Fruensgaard S, and Bünger CE. 1997 Volvo Award winner in clinical studies. The effect of pedicle screw instrumentation on functional outcome and fusion rates in posterolateral lumbar spinal fusion: a prospective, randomized clinical study. Spine 1997;22:2813–22. DOI: 10.1097/00007632-199712150-00004.

77. Molinari RW, Bridwell KH, Klepps SJ, and Baldus C. Minimum 5-Year Follow-up of Anterior Column Structural Allografts in the Thoracic and Lumbar Spine. Spine 1999;24:967–72. DOI: 10.1097/00007632-199905150-00007.

78. Möller H and Hedlund R. Instrumented and noninstrumented posterolateral fusion in adult spondylolisthesis. A prospective randomized study: Part 2. Spine 2000;25:1716–21. DOI: 10.1097/00007632-200007010-00017.

79. Korovessis P, Koureas G, Zacharatos S, Papazisis Z, and Lambiris E. Correlative radiological, self-assessment and clinical analysis of evolution in instrumented dorsal and lateral fusion for degenerative lumbar spine disease. Autograft versus coralline hydroxyapatite. European Spine Journal 2005;14:630–8. DOI: 10.1007/s00586-004-0855-5.

80. Sengupta DK, Truumees E, Patel CK, et al. Outcome of Local Bone Versus Autogenous Iliac Crest Bone Graft in the Instrumented Posterolateral Fusion of the Lumbar Spine. Spine 2006;31:985–91. DOI: 10.1097/01.brs.0000215048.51237.3c.

81. Yu CH, Wang CT, and Chen PQ. Instrumented Posterior Lumbar Interbody Fusion in Adult Spondylolisthesis. Clinical Orthopaedics & Related Research 2008;466:3034–43. DOI: 10.1007/s11999-008-0511-1.

82. Acebal-Cortina G, Suárez-Suárez MA, García-Menéndez C, Moro-Barrero L, Iglesias-Colao R, and Torres-Pérez A. Evaluation of autologous platelet concentrate for intertransverse lumbar fusion. European Spine Journal 2011;20:S361–S366. DOI: 10.1007/s00586-011-1904-5.

83. Korovessis P, Repantis T, Baikousis A, and Iliopoulos P. Posterolateral versus circumferential instrumented fusion for monosegmental lumbar degenerative disc disease using an expandable cage. European Journal of Orthopaedic Surgery & Traumatology 2012;22:639–45. DOI: 10.1007/s00590-011-0890-y.

84. Yamada T, Yoshii T, Sotome S, et al. Hybrid grafting using bone marrow aspirate combined with porous β-tricalcium phosphate and trephine bone for lumbar posterolateral spinal fusion: A prospective, comparative study versus local bone grafting. Spine 2012;37:E174–E179. DOI: 10.1097/BRS.0b013e3182269d64.

85. Hurlbert RJ, Alexander D, Bailey S, et al. rhBMP-2 for posterolateral instrumented lumbar fusion: A multicenter prospective randomized controlled trial. Spine 2013;38:2139–48. DOI: 10.1097/BRS.0000000000000007.

86. Park YS, Kim HS, Baek SW, Kong DY, and Ryu JA. The Effect of Zoledronic Acid on the Volume of the Fusion-Mass in Lumbar Spinal Fusion. Clinics in Orthopedic Surgery 2013;5:292–7. DOI: 10.4055/cios.2013.5.4.292.

87. Guerado E, Cervan AM, Bertrand ML, and Benitez-Parejo N. Allograft plus OP-1 enhances ossification in posterolateral lumbar fusion: A seven year follow-up. Injury 2016;47:S78–S82. DOI: 10.1016/S0020-1383(16)30611-8.

88. Cho JH, Lee JH, Yeom JS, et al. Efficacy of *Escherichia coli*-derived recombinant human bone morphogenetic protein-2 in posterolateral lumbar fusion: an open, active-controlled, randomized, multicenter trial. The Spine Journal 2017;17:1866–74. DOI: 10.1016/j.spinee.2017.06.023.

89. Choi MK, Kim SB, Park CK, Malla HP, and Kim SM. Cross-Sectional Area of the Lumbar Spine Trunk Muscle and Posterior Lumbar Interbody Fusion Rate: A Retrospective Study. Clinical Spine Surgery 2017;30:E798–E803. DOI: 10.1097/BSD.0000000000000424.

90. McDermott KW, Freeman WJ, and Elixhauser E. Overview of Operating Room Procedures During Inpatient Stays in U.S. Hospitals, 2014. Statistical Brief #233. Agency for Healthcare Research and Quality, 2017. URL: https://www.hcup-us.ahrq.gov/reports/statbriefs/statbriefs.jsp.

91. Yuan H, van Blitterswijk CA, de Groot K, and de Bruijn JD. A comparison of bone formation in biphasic calcium phosphate (BCP) and hydroxyapatite (HA) implanted in muscle and bone of dogs at different time periods. Journal of Biomedical Materials Research Part A 2006:139–47. DOI: doi.org/10.1002/jbm.a.30707.

92. Lehr AM, Oner FC, Hoebink EA, et al. Patients cannot reliably distinguish the iliac crest bone graft donor site from the contralateral side after lumbar spine fusion: A patient-blinded randomized controlled trial. Spine 2019;44:527–33. DOI: 10.1097/BRS.0000000000002882.

93. Fairbank JC and Pynsent PB. The Oswestry Disability Index. Spine 2000;25:2940–52. DOI: 10.1097/00007632-200011150-00017.

94. van Hooff ML, Mannion AF, Staub LP, Ostelo RW, and Fairbank JC. Determination of the Oswestry Disability Index score equivalent to a “satisfactory symptom state” in patients undergoing surgery for degenerative disorders of the lumbar spine - a Spine Tango registry-based study. The Spine Journal 2016;16:1221–30. DOI: 10.1016/j.spinee.2016.06.010.

95. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health 2012;15:708–15. DOI: 10.1016/j.jval.2012.02.008.

96. Cammisa FP, Lowery G, Garfin SR, et al. Two-year fusion rate equivalency between grafton® DBM gel and autograft in posterolateral spine fusion: A prospective controlled trial employing a side-by-side comparison in the same patient. Spine 2004;29:660–6. DOI: 10.1097/01.BRS.0000116588.17129.B9.

97. Niu CC, Tsai TT, Fu TS, Lai PL, Chen LH, and Chen WJ. A comparison of posterolateral lumbar fusion comparing autograft, autogenous laminectomy bone with bone marrow aspirate, and calcium sulphate with bone marrow aspirate: A prospective randomized study. Spine 2009;34:2715–9. DOI: 10.1097/BRS.0b013e3181b47232.

98. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, and Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. The Spine Journal 2008;8:968–74.
DOI: 10.1016/j.spinee.2007.11.006.

99. Ostelo RWJG, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: Towards international consensus regarding minimal important change. Spine 2008;33:90–4. DOI: 10.1097/BRS.0b013e31815e3a10.

100. Carreon LY, Glassman SD, and Djurasovic M. Reliability and agreement between fine-cut CT scans and plain radiography in the evaluation of posterolateral fusions. The Spine Journal 2007;7:39–43. DOI: 10.1016/j.spinee.2006.04.005.

101. Landis JR and Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33:159–74. DOI: 10.2307/2529310.

102. Delawi D, Jacobs W, van Susante JLC, et al. OP-1 compared with iliac crest autograft in instrumented posterolateral fusion: A randomized, multicenter non-inferiority trial. Journal of Bone & Joint Surgery 2016;98:441–8. DOI: 10.2106/JBJS.O.00209.

103. Dimar JR, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, and Carreon LY. Two-year fusion and clinical outcomes in 224 patients treated with a single-level instrumented posterolateral fusion with iliac crest bone graft. The Spine Journal 2009;9:880–5.
DOI: 10.1016/j.spinee.2009.03.013.

104. Kang J, An H, Hilibrand A, Yoon ST, Kavanagh E, and Boden S. Grafton and local bone have comparable outcomes to iliac crest bone in instrumented single-level lumbar fusions. Spine 2012;37:1083–91. DOI: 10.1097/BRS.0b013e31823ed817.

105. An HS, Lynch K, and Toth J. Prospective comparison of autograft vs. allograft for adult posterolateral lumbar spine fusion: differences among freeze-dried, frozen, and mixed grafts. Journal of Spinal Disorders 1995;8:131–5.
DOI: 10.1097/00002517-199504000-00007.

106. Lee JH, Hwang CJ, Song BW, Koo KH, Chang BS, and Lee CK. A prospective consecutive study of instrumented posterolateral lumbar fusion using synthetic hydroxyapatite (Bongros®-HA) as a bone graft extender. Journal of Biomedical Materials Research Part A 2009:804–10. DOI: 10.1002/jbm.a.32113.

107. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. New England Journal of Medicine 2008;358:794–810. DOI: 10.1056/NEJMoa0707136.

108. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical versus nonsurgical treatment for lumbar degenerative spondylolisthesis. New England Journal of Medicine 2007;356:2257–70. DOI: 10.1056/NEJMoa070302.

109. Fritzell P, Hägg O, Gerdham P, et al. Swespine 25 years - 2018 Annual report - Follow up of spine surgery performed in Sweden in 2017. Swedish Society of Spinal Surgeons, 2018.

110. Street JT, Lenehan BJ, DiPaola CP, et al. Morbidity and mortality of major adult spinal surgery. A prospective cohort analysis of 942 consecutive patients. The Spine Journal 2012;12:22–34. DOI: 10.1016/j.spinee.2011.12.003.

111. Deyo RA, Nachemson A, and Mirza SK. Spinal-Fusion Surgery - The Case for Restraint. New England Journal of Medicine 2004;350:722–6. DOI: 10.1056/NEJMsb031771.

112. Boden SD. Overview of the biology of lumbar spine fusion and principles for selecting a bone graft substitute. Spine 2002;27:S26–31. DOI: 10.1097/00007632-200208151-00007.

113. Bhalla A, Schoenfeld AJ, George J, Moghimi M, and Bono CM. The influence of subgroup diagnosis on radiographic and clinical outcomes after lumbar fusion for degenerative disc disorders revisited: a systematic review of the literature. The Spine Journal 2017;17:143–9. DOI: 10.1016/j.spinee.2016.09.021.

114. Dimar JR, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, and Carreon LY. Clinical and radiographic analysis of an optimized rhBMP-2 formulation as an autograft replacement in posterolateral lumbar spine arthrodesis. Journal of Bone & Joint Surgery 2009;91:1377–86. DOI: 10.2106/JBJS.H.00200.

115. Gruskay JA, Webb ML, and Grauer JN. Methods of evaluating lumbar and cervical fusion. The Spine Journal 2014;14:531–9. DOI: 10.1016/j.spinee.2013.07.459.

116. Chou P, Ma H, Wang S, Liu C, Chang M, and Yu W. Fusion may not be a necessary procedure for thoracolumbar and lumbar spines: A follow-up of at least ten years. Journal of Bone & Joint Surgery 2014;96:1724–31. DOI: 10.2106/JBJS.M.01486.

117. Lehr AM, Oner FC, Delawi D, et al. Efficacy of a standalone microporous ceramic versus autograft in instrumented posterolateral spinal fusion: A multicenter, randomized, intrapatient controlled, noninferiority trial. Spine 2020;45:944–51. DOI: 10.1097/BRS.0000000000003440.

118. Hagenmaier F, Kok D, Hol A, Rijnders T, Oner FC, and van Susante JLC. Changes in Bone Mineral Density in the Intertransverse Fusion Mass After Instrumented Single-Level Lumbar Fusion: A Prospective 1-Year Follow-up. Spine 2012;38:696–702. DOI: 10.1097/BRS.0b013e318276fa27.

119. Kong S, Park JH, and Roh SW. A prospective comparative study of radiological outcomes after instrumented posterolateral fusion mass using autologous local bone or a mixture of beta-tcp and autologous local bone in the same patient. Acta Neurochirurgica 2013;155:765–70. DOI: 10.1007/s00701-013-1669-1.

120. Coughlan M, Davies M, Mostert AK, et al. A prospective, randomized, multicenter study comparing silicated calcium phosphate versus BMP-2 synthetic bone graft in posterolateral instrumented lumbar fusion for degenerative spinal disorders. Spine 2018;43:E860–E868. DOI: 10.1097/BRS.0000000000002678.

121. Glassman SD, Carreon LY, Djurasovic M, et al. Lumbar fusion outcomes stratified by specific diagnostic indication. The Spine Journal 2009;9:13–21. DOI: 10.1016/j.spinee.2008.08.011.

122. Dhall SS, Choudhri TF, Eck JC, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 5: Correlation between radiographic outcome and function. Journal of Neurosurgery: Spine 2014;21:31–6. DOI: 10.3171/2014.4.SPINE14268.

123. Fischgrund JS, Mackay M, Herkowitz HN, Brower R, Montgomery DM, and Kurz LT. 1997 Volvo Award winner in clinical studies. Degenerative lumbar spondylolisthesis with spinal stenosis: a prospective, randomized study comparing decompressive laminectomy and arthrodesis with and without spinal instrumentation. Spine 1997;22:2807–12. DOI: 10.1097/00007632-199712150-00003.

124. Kornblum MB, Fischgrund JS, Herkowitz HN, Abraham DA, Berkower DL, and Ditkoff JS. Degenerative Lumbar Spondylolisthesis with Spinal Stenosis: A Prospective Long-Term Study Comparing Fusion and Pseudarthrosis. Spine 2004;29:726–33. DOI: 10.1097/01.BRS.0000119398.22620.92.

125. Tsutsumimoto T, Shimogata M, Yoshimura Y, and Misawa H. Union versus nonunion after posterolateral lumbar fusion: a comparison of long-term surgical outcomes in patients with degenerative lumbar spondylolisthesis. European Spine Journal 2008;17:1107–12. DOI: 10.1007/s00586-008-0695-9.

126. Djurasovic M, Glassman SD, Dimar JR, Howard JM, Bratcher KR, and Carreon LY. Does fusion status correlate with patient outcomes in lumbar spinal fusion? Spine 2011;36:404–9. DOI: 10.1097/BRS.0b013e3181fde2c4.

127. Irmola TM, Häkkinen A, Järvenpää S, Marttinen I, Vihtonen K, and Neva M. Reoperation Rates Following Instrumented Lumbar Spine Fusion. Spine 2018;43:295–301. DOI: 10.1097/BRS.0000000000002291.

128. Cho KJ, Suk SI, Park SR, et al. Complications in Posterior Fusion and Instrumentation for Degenerative Lumbar Scoliosis. Spine 2007;32:2232–7. DOI: 10.1097/BRS.0b013e31814b2d3c.

129. Fritzell P, Hägg O, Wessberg P, Nordwall A, and Swedish Lumbar Spine Study Group the. Chronic low back pain and fusion: A comparison of three surgical techniques. A prospective multicenter randomized study from the Swedish Lumbar Spine Study Group. Spine 2002;27:1131–41. DOI: 10.1097/00007632-200206010-00002.

130. Martin BI, Mirza SK, Comstock BA, Gray DT, Kreuter W, and Deyo RA. Reoperation rates following lumbar spine surgery and the influence of spinal fusion procedures. Spine 2007;32:382–7. DOI: 10.1097/01.brs.0000254104.55716.46.

131. Black N. Why we need observational studies to evaluate the effectiveness of health care. British Medical Journal 1996;312:1215–8. DOI: 10.1136/bmj.312.7040.1215.

132. Houwert RM, Beks RB, Dijkgraaf MG, et al. Study methodology in trauma care: towards question-based study designs. European Journal of Trauma and Emergency Surgery 2021;47:479–84. DOI: 10.1007/s00068-019-01248-5.

133. Jacobs WCH, Kruyt MC, Verbout AJ, and Oner FC. Spine surgery research: on and beyond current strategies. The Spine Journal 2012;12:706–13.
DOI: 10.1016/j.spinee.2012.08.424.

134. Jacobs WCH, Kruyt MC, Verbout AJ, and Oner FC. Effect of methodological quality measures in spinal surgery research: a metaepidemiological study. The Spine Journal 2012;12:339–48. DOI: 10.1016/j.spinee.2012.01.015.

135. Lochner HV, Bhandari M, and Tornetta III P. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. The Journal of Bone & Joint Surgery 2001:1650–5. DOI: 10.2106/00004623-200111000-00005.

136. Walters SJ, Bonacho dos Anjos Henriques-Cadby I, Bortolami O, et al. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. BMJ Open 2017;7:e015276. DOI: 10.1136/bmjopen-2016-015276.

137. Pandis N, Chung B, Scherer RW, Elbourne D, and Altman DG. CONSORT 2010 statement: extension checklist for reporting within person randomised trials. BMJ 2017;357:j2835. DOI: 10.1136/bmj.j2835.

138. Alexander DI, Manson NA, and Mitchell MJ. Efficacy of calcium sulfate plus decompression bone in lumbar and lumbosacral spinal fusion: preliminary results in 40 patients. Canadian Journal of Surgery 2001;44:262–6.

139. Bansal S, Chauhan V, Sharma S, Maheshwari R, Juyal A, and Raghuvanshi S. Evaluation of hydroxyapatite and beta-tricalcium phosphate mixed with bone marrow aspirate as a bone graft substitute for posterolateral spinal fusion. Indian Journal of Orthopaedics 2009;43:234–9. DOI: 10.4103/0019-5413.49387.

140. Boden SD, Grob D, and Damien C. Ne-Osteo bone growth factor for posterolateral lumbar spine fusion: results from a nonhuman primate study and a prospective human clinical pilot study. Spine 2004;29:504–14.
DOI: 10.1097/01.BRS.0000101446.26071.EB.

141. Chen WJ, Tsai TT, Chen LH, et al. The fusion rate of calcium sulfate with local autograft bone compared with autologous iliac bone graft for instrumented short-segment spinal fusion. Spine 2005;30:2293–7. DOI: 10.1097/01.brs.0000182087.35335.05.

142. Gunzburg R and Szpalski M. Use of a novel β-tricalcium phosphate-based bone void filler as a graft extender in spinal fusion surgeries. Orthopedics 2002;25:s591–s595. DOI: 10.3928/0147-7447-20020502-08.

143. Katayama Y, Matsuyama Y, Yoshihara H, et al. Clinical and radiographic outcomes of posterolateral lumbar spine fusion in humans using recombinant human bone morphogenetic protein-2: an average five-year follow-up study. International Orthopaedics 2009;33:1061–7. DOI: 10.1007/s00264-008-0600-5.

144. Kitchel SH. A preliminary comparative study of radiographic results using mineralized collagen and bone marrow aspirate versus autologous bone in the same patients undergoing posterior lumbar interbody fusion with instrumented posterolateral lumbar fusion. The Spine Journal 2006;6:405–11. DOI: 10.1016/j.spinee.2005.09.013.

145. Vadalà G, Di Martino A, Russo F, et al. Autologous bone marrow concentrate combined with platelet-rich plasma enhance bone allograft potential to induce spinal fusion. Journal of Biological Regulators and Homeostatic Agents 2016;30:165–72.

146. Nam JM. Establishing equivalence of two treatments and sample size requirements in matched-pairs design. Biometrics 1997;53:1422–30.

147. Nam JM and Blackwelder WC. Analysis of the ratio of marginal probabilities in a matched-pair setting. Statistics in Medicine 2002;21:689–99. DOI: 10.1002/sim.1017.

148. Zhu H, Zhang S, and Ahn C. Sample size considerations for split-mouth design. Statistical Methods in Medical Research 2017;26:2543–51.
DOI: 10.1177/0962280215601137.

149. Block JE. Potential effectiveness of novel bone graft substitute materials. Spine 1995;20:2261–2.

150. Gandhi R, Smith HN, Mahomed NN, Rizek R, and Bhandari M. Incorrect use of the student t test in randomized trials of bilateral hip and knee arthroplasty patients. The Journal of Arthroplasty 2011;26:811–6. DOI: 10.1016/j.arth.2010.06.007.

151. Schimandle JH and Boden SD. Spine update: the use of animal models to study spinal fusion. Spine 1994;19:1998–2006. DOI: 10.1097/00007632-199409000-00023.

152. Hox JJ. Multilevel Analysis Techniques and Applications. Mahwah, NJ: Lawrence Erlbaum Associates Inc., 2002.

153. NCSS. Non-Inferiority Tests for the Difference Between Two Correlated Proportions. In: *PASS statistical software training documentation.* Chapter 160: 160-1–160-11.

154. NCSS. Non-inferiority Tests for the Difference Between Two Proportions. In: *PASS statistical software training documentation.* Chapter 210: 210-1–210-25.

155. Liu JP, Hsueh HM, Hsieh E, and Chen JJ. Tests for equivalence or non-inferiority for paired binary data. Statistics in Medicine 2002;21:231–45. DOI: 10.1002/sim.1012.

156. KCE Trials. KCE Trials 2020 investigator-led call. Accessed July 10, 2020. 2020. URL: https://kce.fgov.be/en/kce-trials-2020-investigator-led-call.

157. Nevens H, Harrison J, Vrijens F, et al. Budgeting of non-commercial clinical trials: development of a budget tool by a public funding agency. Trials 2019;20:714. DOI: 10.1186/s13063-019-3900-8.

158. Ghogawala Z, Dziura J, Butler WE, et al. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. New England Journal of Medicine 2016;374:1424–34. DOI: 10.1056/NEJMoa1508788.

159. Andersen T, Christensen FB, Laursen M, Høy K, Hansen ES, and Bünger C. Smoking as a predictor of negative outcome in lumbar spinal fusion. Spine 2001;26:2623–8. DOI: 10.1097/00007632-200112010-00018.

160. Lesaffre E, Philstrom B, Needleman I, and Worthington H. The design and analysis of split-mouth studies: What statisticians and clinicians should know. Statistics in Medicine 2009;28:3470–82. DOI: 10.1002/sim.3634.

161. Pandis N, Walsh T, Polychronopoulou A, Katsaros C, and Eliades T. Split-mouth designs in orthodontics: an overview with applications to orthodontic clinical trials. The European Journal of Orthodontics 2013;35:783–9. DOI: 10.1093/ejo/cjs108.

162. Cook JA, Bruckner T, MacLennan GS, and Seiler CM. Clustering in surgical trials - database of intracluster correlations. Trials 2012;13:2. DOI: 10.1186/1745-6215-13-2.

163. Stuart B, Becque T, Moore M, and Little P. Clustering of continuous and binary outcomes at the general practice level in individually randomised studies in primary care - a review of 10 years of primary care trials. BMC Medical Research Methodology 2020;20:83. DOI: 10.1186/s12874-020-00971-7.

164. Wu S, Crespi CM, and Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. Contemporary Clinical Trials 2012;33:869–80. DOI: 10.1016/j.cct.2012.05.004.

165. Hsu CJ, Chou WY, Teng HP, Chang WN, and Chou YJ. Coralline hydroxyapatite and laminectomy-derived bone as adjuvant graft material for lumbar posterolateral fusion. Journal of Neurosurgery: Spine 2005;3:271–5. DOI: 10.3171/spi.2005.3.4.0271.

166. Kim YH, Park JW, and Kim JS. The Clinical Outcome of Computer-Navigated Compared with Conventional Knee Arthroplasty in the Same Patients: A Prospective, Randomized, Double-Blind, Long-Term Study. The Journal of Bone & Joint Surgery 2017:989–96. DOI: 10.2106/JBJS.16.00791.

167. Shapiro SA, Kazmerchak SE, Heckman MG, Zubair AC, and O'Connor MI. A Prospective, Single-Blind, Placebo-Controlled Trial of Bone Marrow Aspirate Concentrate for Knee Osteoarthritis. The American Journal of Sports Medicine 2017;45:82–90. DOI: 10.1177/0363546516662455.

168. Herkowitz HN and Kurz LT. Degenerative lumbar spondylolisthesis with spinal stenosis: A prospective study comparing decompression with decompression and intertransverse process arthrodesis. Journal of Bone & Joint Surgery 1991:802–8. DOI: 10.2106/00004623-199173060-00002.

169. Möller H and Hedlund R. Surgery Versus Conservative Management in Adult Isthmic Spondylolisthesis. Spine 2000;25:1711–5. DOI: 10.1097/00007632-200007010-00016.

170. Ekman P, Möller H, and Hedlund R. The long-term effect of posterolateral fusion in adult isthmic spondylolisthesis: a randomized controlled study. The Spine Journal 2005;5:36–44. DOI: 10.1016/j.spinee.2004.05.249.

171. Andersen T, Videbæk TS, Hansen ES, Bünger C, and Christensen FB. The positive effect of posterolateral lumbar spinal fusion is preserved at long-term follow-up: a RCT with 11-13 year follow-up. European Spine Journal 2008;17:272–80. DOI: 10.1007/s00586-007-0494-8.

172. Abdu WA, Sacks OA, Tosteson AN, et al. Long-term results of surgery compared with nonoperative treatment for lumbar degenerative spondylolisthesis in the spine patient outcomes research trial (SPORT). Spine 2018;43:1619–30. DOI: 10.1097/BRS.0000000000002682.

173. Turunen V, Nyyssönen T, Miettinen H, et al. Lumbar instrumented posterolateral fusion in spondylolisthetic and failed back patients: a long-term follow-up study spanning 11-13 years. European Spine Journal 2012;21:2140–8. DOI: 10.1007/s00586-012-2320-1.

174. Endler P, Ekman P, Ljungqvist H, Brismar TB, Gerdhem P, and Möller H. Long-term outcome after spinal fusion for isthmic spondylolisthesis in adults. The Spine Journal 2019;19:501–8. DOI: 10.1016/j.spinee.2018.08.008.

175. Sengupta DK and Herkowitz HN. Degenerative spondylolisthesis: Review of current trends and controversies. Spine 2005;30:s71–s81. DOI: 10.1097/01.brs.0000155579.88537.8e.

176. The Euroqol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health policy 1990;16:199–208. DOI: 10.1016/0168-8510(90)90421-9.

177. Janssen B and Szende A. Population Norms for the EQ-5D. In: *Self-Reported Population Health: An International Perspective based on EQ-5D*. Ed. by Szende A, Janssen B, and Cabases J. Dordrecht: Springer, 2014: 19–30. DOI: 10.1007/978-94-007-7596-1_3.

178. Carragee EJ. The increasing morbidity of elective spinal stenosis surgery: Is it necessary? JAMA 2010;303:1309–10. DOI: 10.1001/jama.2010.402.

179. Boonstra AM, Preuper HR, Balk GA, and Stewart RE. Cut-off points for mild, moderate, and severe pain on the visual analogue scale for pain in patients with chronic musculoskeletal pain. Pain 2014;155:2545–50. DOI: 10.1016/j.pain.2014.09.014.

180. Hansson T, Hansson E, and Malchau H. Utility of Spine Surgery: A Comparison of Common Elective Orthopaedic Surgical Procedures. Spine 2008;33:2819–30. DOI: 10.1097/brs.0b013e31818e2914.

181. Jansson KA and Granath F. Health-related quality of life (EQ-5D) before and after orthopedic surgery. Acta Orthopaedica 2011;82:82–9.
DOI: 10.3109/17453674.2010.548026.

182. Lindgren JV, Wretenberg P, Kärrholm J, Garellick G, and Rolfson O. Patient-reported outcome is influenced by surgical approach in total hip replacement: a study of the Swedish Hip Arthroplasty Register including 42 233 patients. The Bone & Joint Journal 2014;96B:590–6. DOI: 10.1302/0301-620X.96B5.32341.

183. Shan L, Shan B, Suzuki A, Nouh F, and Saxena A. Intermediate and long-term quality of life after total knee replacement: A systematic review and meta-analysis. The Journal of Bone & Joint Surgery 2015:156–68. DOI: 10.2106/JBJS.M.00372.

184. Park P, Garton HJ, Gala VC, Hoff JT, and McGillicuddy JE. Adjacent segment disease after lumbar or lumbosacral fusion: Review of the literature. Spine 2004;29:1938–44. DOI: 10.1097/01.brs.0000137069.88904.03.

185. Ekman P, Möller H, Shalabi A, Yu YX, and Hedlund R. A prospective randomised study on the long-term effect of lumbar fusion on adjacent disc degeneration. European Spine Journal 2009;18:1175–86. DOI: 10.1007/s00586-009-0947-3.

186. Schwartz CE, Sajobi TT, Lix LM, Quaranto BR, and Finkelstein JA. Changing values, changing outcomes: the influence of reprioritization response shift on outcome assessment after spine surgery. Quality of Life Research 2013;22:2255–64.
DOI: 10.1007/s11136-013-0377-x.

187. Finkelstein JA and Schwartz CE. Patient-reported outcomes in spine surgery: past, current, and future directions. Journal of Neurosurgery: Spine 2019;31:155–64.
DOI: 10.3171/2019.1.SPINE18770.

188. Glaser J, Stanley M, Sayre H, Woody J, Found E, and Spratt K. A 10-year follow-up evaluation of lumbar spine fusion with pedicle screw fixation. Spine 2003;28:1390–5. DOI: 10.1097/01.BRS.0000067112.15753.AD.

189. Wenger M, Sapio N, and Markwalder TM. Long-term outcome in 132 consecutive patients after posterior internal fixation and fusion for Grade I and II isthmic spondylolisthesis. Journal of neurosurgery: Spine 2005;2:289–97. DOI: 10.3171/spi.2005.2.3.0289.

190. Hedlund R, Johansson C, Hägg O, Fritzell P, Tullberg T, and Swedish Lumbar Spine Study Group. The long-term outcome of lumbar fusion in the Swedish lumbar spine study. The Spine Journal 2016;16:579–87. DOI: 10.1016/j.spinee.2015.08.065.

191. Nayak NR, Stephen JH, Piazza MA, Obayemi AA, Stein SC, and Malhotra NR. Quality of Life in Patients Undergoing Spine Surgery: Systematic Review and Meta-Analysis. Global Spine Journal 2019;9:67–76. DOI: 10.1177/2192568217701104.

192. Vaccaro AR, Lawrence JP, Patel T, et al. The safety and efficacy of OP-1 (rhBMP-7) as a replacement for iliac crest autograft in posterolateral lumbar arthrodesis: A long-term (>4 years) pivotal study. Spine 2008;33:2850–62. DOI: 10.1097/BRS.0b013e31818a314d.

193. OME Cleveland Clinic Orthopaedics. Value in Research: Achieving Validated Outcome Measurements While Mitigating Follow-up Cost. The Journal of Bone & Joint Surgery 2020:419–27. DOI: 10.2106/JBJS.19.00531.

194. Harris IA and Dao ATT. Trends of spinal fusion surgery in Australia: 1997 to 2006. ANZ Journal of Surgery 2009;79:783–8. DOI: 10.1111/j.1445-2197.2009.05095.x.

195. United States Bone and Joint Initiative. The Burden of Musculoskeletal Diseases in the United States, Third Edition. Rosemont, IL, 2014. URL: https://www.boneandjointburden.org/resources/archives.

196. Fernyhough JC, Schimandle JJ, Weigel MC, Edwards CC, and Levine AM. Chronic donor site pain complicating bone graft harvesting from the posterior iliac crest for spinal fusion. Spine 1992;17:1474–80. DOI: 10.1097/00007632-199212000-00006.

197. Banwart JC, Asher MA, and Hassanein RS. Iliac crest bone graft harvest donor site morbidity. A statistical evaluation. Spine 1995;20:1055–60. DOI: 10.1097/00007632-199505000-00012.

198. Goulet JA, Senunas LE, DeSilva GL, and Greenfield ML. Autogenous iliac crest bone graft. Complications and functional assessment. Clinical Orthopaedics and Related Research 1997;339:76–81. DOI: 10.1097/00003086-199706000-00011.

199. Summers BN and Eisenstein SM. Donor site pain from the ilium. A complication of lumbar spine fusion. The Journal of Bone & Joint Surgery (British volume) 1989;71:677–80. DOI: 10.1302/0301-620X.71B4.2768321.

200. Younger EM and Chapman MW. Morbidity at bone graft donor sites. Journal of Orthopaedic Trauma 1989;3:192–5. DOI: 10.1097/00005131-198909000-00002.

201. Robertson PA and Wray AC. Natural history of posterior iliac crest bone graft donation for spinal surgery: A prospective analysis of morbidity. Spine 2001;26:1473–6. DOI: 10.1097/00007632-200107010-00018.

202. Kurz LT, Garfin SR, and Booth RE. Harvesting autogenous iliac bone grafts. A review of complications and techniques. Spine 1989;14:1324–31. DOI: 10.1097/00007632-198912000-00009.

203. Allied Market Research. Bone Grafts and Substitutes Market by Product (Allografts, Bone Grafts Substitutes, and Cell-based Matrices), by Application (Spinal Fusion, Long Bone, Foot & Ankle, Craniomaxillofacial, Joint Reconstruction, and Dental Bone Grafting). Accessed January 15, 2018. 2016. URL: https://www.alliedmarketresearch.com/bone-graft-substitutes-market.

204. Transparency Market Research. Bone Grafts and Substitutes Market - Global Industry Analysis, Size, Share, Growth, Trends and Forecast 2015 - 2023. Accessed January 11, 2018. 2015. URL: https://www.transparencymarketresearch.com/bone-grafts-substitutes-market.html.

205. GlobalData. MediPoint: Bone Grafts and Substitutes - Global Analysis and Market Forecasts. Accessed January 11, 2018. 2014. URL: https://www.marketresearch.com/product/sample-8021855.pdf.

206. GlobalData. MediPoint: Bone Grafts & Substitutes - Global Analysis and Market Forecasts. Accessed January 11, 2018. 2017. URL: https://www.businesswire.com/news/home/20170620005822/en/Bone-Grafts-Substitutes—Global-Analysis-Market.

207. Bezer M, Kocaolu B, Aydin N, and Güven O. Comparison of traditional and intrafascial iliac crest bone-graft harvesting in lumbar spinal surgery. International Orthopaedics 2004;28:325–8. DOI: 10.1007/s00264-004-0594-6.

208. David R, Folman Y, Pikarsky I, Leitner Y, Catz A, and Gepstein R. Harvesting Bone Graft From the Posterior Iliac Crest by Less Traumatic, Midline Approach. Journal of Spinal Disorders & Techniques 2003;16:27–30. DOI: 10.1097/00024720-200302000-00005.

209. Jacobs WCH, Kruyt MC, Moojen WA, Verbout AJ, and Oner FC. No evidence for intervention-dependent influence of methodological features on treatment effect. Journal of Clinical Epidemiology 2013;66:1347–1355.e3. DOI: 10.1016/j.jclinepi.2013.06.007.

210. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601. DOI: 10.1136/bmj.39465.451748.AD.

211. United States Bone and Joint Initiative. The Burden of Musculoskeletal Diseases in the United States, Fourth Edition. Rosemont, IL, 2018. URL: https://www.boneandjointburden.org/fourth-edition.

212. Feng Jt, Yang Xg, Wang F, He X, and Hu Yc. Efficacy and safety of bone substitutes in lumbar spinal fusion: a systematic review and network meta-analysis of randomized controlled trials. European Spine Journal 2020;29:1261–76.
DOI: 10.1007/s00586-019-06257-x.

213. Plantz MA, Gerlach EB, and Hsu WK. Synthetic Bone Graft Materials in Spine Fusion: Current Evidence and Future Trends. International Journal of Spine Surgery 2021;15:104–12. DOI: 10.14444/8058.

214. Duan R, van Dijk LA, Barbieri D, de Groot F, Yuan H, and de Bruijn JD. Accelerated bone formation by biphasic calcium phosphate with a novel sub-micron surface topography. European cells & Materials 2019;37:60–73. DOI: 10.22203/eCM.v037a05.

215. Habibovic P, Sees TM, van den Doel MA, van Blitterswijk CA, and de Groot K. Osteoinduction by biomaterials - physicochemical and structural influence. Journal of Biomedical Materials Research Part A 2006;77:747–62. DOI: 10.1002/jbm.a.30712.

216. Couwenberg AM, Burbach JPM, May AM, Berbee M, Intven MPW, and Verkooijen HM. The trials within cohorts design facilitated efficient patient enrollment and generalizability in oncology setting. Journal of Clinical Epidemiology 2020;120:33–9.
DOI: 10.1016/j.jclinepi.2019.12.015.

217. Ergina PL, Cook JA, Blazeby JM, et al. Surgical Innovation and Evaluation 2: Challenges in evaluating surgical innovation. Lancet 2009;374:1097–104.
DOI: 10.1016/S0140-6736(09)61086-2.

218. Zuidgeest MGP, Goetz I, Groenwold RHH, Irving E, van Thiel GJMW, and Grobbee DE. Series: Pragmatic trials and real world evidence: Paper 1. Introduction. Journal of Clinical Epidemiology 2017;88:7–13. DOI: 10.1016/j.jclinepi.2016.12.023.

219. Kim KW, Ha KY, Moon MS, Kim YS, Kwon SY, and Woo YK. Fate of the facet joints after instrumented intertransverse process fusion. Clinical Orthopaedics and Related Research 1999;366:110–9. DOI: 10.1097/00003086-199909000-00015.

220. Izeki M, Fujio K, Ota S, Soga S, and Matsuda S. Radiological follow-up of the degenerated facet joints after lateral lumbar interbody fusion with percutaneous pedicle screw fixation: Focus on spontaneous facet joint fusion. Journal of Orthopaedic Science 2021. DOI: 10.1016/j.jos.2021.06.018.

221. Meleis A, Larkin MB, de Almeida Bastos DC, et al. Single-center outcomes for percutaneous pedicle screw fixation in metastatic spinal lesions: can spontaneous facet fusion occur? Neurosurgical Focus 2021;50:E9. DOI: 10.3171/2021.1.FOCUS20671.

222. Ha KY, Lee JS, and Kim KW. Bone graft volumetric changes and clinical outcomes after instrumented lumbar or lumbosacral fusion: A prospective cohort study with a five-year follow-up. Spine 2009;34:1663–8. DOI: 10.1097/BRS.0b013e3181aacab5.

223. van Dijk LA, Duan R, Luo X, et al. Biphasic calcium phosphate with submicron surface topography in an Ovine model of instrumented posterolateral spinal fusion. JOR Spine 2018;1:e1039. DOI: 10.1002/jsp2.1039.

224. van Dijk LA, Barrère-de Groot F, Rosenberg AJWP, et al. MagnetOs, Vitoss, and Novabone in a Multi-endpoint Study of Posterolateral Fusion: A True Fusion or Not? Clinical Spine Surgery 2020;33:E276–E287. DOI: 10.1097/BSD.0000000000000920.

225. Spinal Adverse Events Severity System, version 2 (SAVES-V2): inter- and intraobserver reliability assessment. Journal of Neurosurgery: Spine 2016;25:256–63. DOI: 10.3171/2016.1.SPINE14808.

226. Hibbs RA and Peltier LF. A report of fifty-nine cases of scoliosis treated by the fusion operation. Clinical Orthopaedics and Related Research 1988;229:4–19. DOI: 10.1097/00003086-198804000-00002.

227. Boden SD and Sumner DR. Biological factors affecting spinal fusion and bone regeneration. Spine 1995;20:102S–112S.

228. Slappey G, Toribatake Y, Ganey TM, Ogden JA, and Hutton WC. Guidelines to decortication in posterolateral spine fusion. Journal of Spinal Disorders 1998;11:102–9.

229. Pang H, Wong Ys, Yip BHk, et al. Using Ultrasound to Screen for Scoliosis to Reduce Unnecessary Radiographic Radiation: A Prospective Diagnostic Accuracy Study on 442 Schoolchildren. Ultrasound in Medicine & Biology 2021;47:2598–607. DOI: 10.1016/j.ultrasmedbio.2021.05.020.

230. Zheng YP, Lee TTY, Lai KKL, et al. A reliability and validity study for Scolioscan: a radiation-free scoliosis assessment system using 3D ultrasound imaging. Scoliosis and Spinal Disorders 2016;11:13. DOI: 10.1186/s13013-016-0074-y.

231. McCulloch P, Feinberg J, Philippou Y, et al. Progress in clinical research in surgery and IDEAL. The Lancet 2018;392:88–94. DOI: 10.1016/S0140-6736(18)30102-8.

232. Oner FC and Dhert WJA. Challenging the medico-industrial-administrative complex. The Spine Journal 2011;11:698–9. DOI: 10.1016/j.spinee.2011.08.010.

Summary in Dutch (Nederlandse samenvatting)

Inleiding en doel

Een spondylodese (wervelfusie) operatie is een behandeling voor diverse aandoeningen die instabiliteit van de wervelkolom veroorzaken. Bijvoorbeeld een fractuur, vervorming of slijtage van de wervelkolom. Het doel van deze operatie is dat twee of meerdere wervels aan elkaar vastgroeien, zodat ze onderling niet meer kunnen bewegen. Bij een geïnstrumenteerde posterolaterale wervelfusie worden de wervels aan elkaar vastgezet met schroeven en staven (instrumentatie). Dit zorgt direct voor stabiliteit. Vervolgens wordt tussen de dwarsuitsteeksels van de wervels een bottransplantaat gelegd. Het bottransplantaat stimuleert het lichaam om bot aan te maken en wordt langzaam vervangen door levend bot. Zo ontstaat na verloop van tijd een solide brug van bot (botfusie) tussen de wervels.

Van oudsher wordt bij een wervelfusie operatie lichaamseigen bottransplantaat (autograft) uit de bekkenkam van de patiënt gebruikt, maar deze methode heeft een aantal nadelen. Allereerst is er een extra ingreep nodig om het bot te oogsten. Deze ingreep kan complicaties met zich mee brengen, zoals chronische pijn of een infectie in het donorgebied. Daarnaast kan maar een beperkte hoeveelheid bot worden geoogst en kan de kwaliteit van dit bottransplantaat onvoldoende zijn. Om deze nadelen te voorkomen is er de afgelopen decennia veel onderzoek gedaan naar botvervangende materialen die autograft zo goed mogelijk imiteren. Kunstmatige keramieken op basis van calciumfosfaat zijn een interessante botvervanger, omdat ze veel lijken op de mineraalcomponent van bot. Daarnaast zijn dit soort keramieken biocompatibel en relatief goedkoop te fabriceren, steriliseren en bewaren. De eerste generatie calciumfosfaat keramieken diende vooral als matrix om het proces van botvorming te ondersteunen (osteoconductie). De huidige bifasische calciumfosfaat keramieken met microporiën en een submicron oppervlaktestructuur (BCP$_{<\mu m}$) zijn daarentegen ook in staat om omliggende cellen te stimuleren en activeren tot botvormende cellen (osteoinductie). Bovendien worden ze geleidelijk afgebroken door het lichaam (bioresorptie).

Hoewel meerdere botvervangende materialen commercieel verkrijgbaar zijn voor geïnstrumenteerde wervelfusies, is er weinig klinisch bewijs van hun werkzaamheid. Hierdoor weten we niet of ze daadwerkelijk botfusie tussen wervels bevorderen en daarmee voor stabiliteit op de lange termijn zorgen. Dit proefschrift gaat in op een aantal uitdagingen binnen dit onderzoeksveld en beschrijft de resultaten van een klinisch studie naar een BCP$_{<\mu m}$. Hiermee beogen we het klinisch onderzoek naar en de toepassing van botvervangende materialen voor wervelfusies te bevorderen.

Methoden en resultaten

In de klinische praktijk wordt botfusie tussen wervels beoordeeld met radiologische beeldvorming, zoals een röntgenfoto of computertomografie (CT) scan. Er is echter geen eenduidige standaard om de mate van botfusie vast te stellen. Dit bemoeilijkt de vergelijking van verschillende klinische studies op dit gebied en draagt waarschijnlijk bij aan de voortdurende discussie over de relatie tussen de radiologische en klinische uitkomsten van een wervelfusie operatie. **Hoofdstuk 2** geeft een literatuuroverzicht van 1) gebruikte posterolaterale wervelfusie criteria en 2) de diagnostische nauwkeurigheid en betrouwbaarheid van deze criteria. Bijna twee derde (63%) van de 187 geïncludeerde artikelen gebruikte één of meerdere criteria om de continuïteit van de botbrug, bewegelijkheid tussen wervels of tekenen van gebrek aan fusie (pseudoartrose) te beoordelen. De helft (47%) van de artikelen scoorde botfusie aan de hand van een classificatiesysteem. De literatuur bevat veel variatie in de gebruikte beeldvorming, beoordeelde anatomische locatie, terminologie voor de continuïteit en kwaliteit van een botbrug, en afkapwaarden voor succesvolle wervelfusie. Bovendien is geen van de gevonden criteria en classificatiesystemen bewezen nauwkeurig en betrouwbaar voor het niet-invasief vaststellen van een wervelfusie. Op basis van dit literatuuroverzicht adviseren we een classificatiesysteem te ontwikkelen en valideren, waarmee de botfusie uitkomst per wervelniveau en per kant wordt gescoord op basis van de volgende aspecten: de kwaliteit van de botmassa tussen de dwarsuitsteeksels, de kwaliteit van de botmassa tussen het facetgewricht en tekenen van pseudoartrose (zoals radiolucentie rond de schroeven). CT lijkt hiervoor de meest geschikte beeldvormingstechniek.

Hoofdstuk 3 en **Hoofdstuk 4** beschrijven respectievelijk de 1 en 2 jaar resultaten van de multicenter gerandomiseerde intrapatiënt gecontroleerde studie (intrapatiënt RCT) die we hebben uitgevoerd. In deze studie hebben we de effectiviteit van een BCP$_{<\mu m}$ (AttraX® Putty van het bedrijf NuVasive Inc.) vergeleken met autograft in patiënten die een primaire geïnstrumenteerde posterolaterale wervelfusie ondergingen. Het hoofddoel was aantonen dat de botfusie uitkomst van BCP$_{<\mu m}$ niet inferieur is aan autograft op basis van 1 jaar CT-scans. Daarnaast is de toename van botfusie tussen 1 en 2 jaar na de operatie onderzocht, alsmede complicaties en patiëntgerapporteerde uitkomstmaten (PROM's). Bij 100 volwassen patiënten is, na instrumentatie van de wervels en volgens randomisatie, aan één kant van de wervelkolom BCP$_{<\mu m}$ geplaatst en aan de andere kant autograft uit de bekkenkam. Elke patiënt heeft dus beide onderzochte behandelingen ontvangen en was zijn/haar eigen controle. Eén jaar na de operatie was het botfusiepercentage aan de BCP$_{<\mu m}$ kant 55% en aan de kant met autograft 52% (**Hoofdstuk 3**). Na correctie voor fusies tussen meer dan twee wervels, resulterend in één fusiescore per kant per patiënt, was het absolute verschil in succesvolle botfusie

tussen BCP$_{<\mu m}$ en autograft slechts 2,3% met een 90% betrouwbaarheidsinterval van -9,1% tot 13,7%. Daarmee tonen we aan dat BCP$_{<\mu m}$ in termen van radiologische botfusie 1 jaar na geïnstrumenteerde posterolaterale wervelfusie niet inferieur is aan autograft.

In **Hoofdstuk 4** hebben we het proces van botvorming na het eerste jaar en de rol van botmateriaal onderzocht. Hiervoor hebben de patiënten die na 1 jaar geen succesvolle botfusie hadden een extra CT-scan bij de 2 jaar controle gehad. Het botfusiepercentage van beide behandelingen nam significant toe, tot 70% aan de BCP$_{<\mu m}$ kant en 68% aan de autograft kant. De toename van botfusie vond echter alleen plaats tussen de facetgewrichten, niet tussen de dwarsuitsteeksels waar het botmateriaal is neergelegd. Dit komt overeen met de bevinding dat zowel het BCP$_{<\mu m}$ als autograft tussen de dwarsuitsteeksels na 1 jaar was geresorbeerd. Waarschijnlijk is de fusie tussen de facetgewrichten het gevolg van immobilisatie van de wervelkolom met starre implantaten. Op basis van deze resultaten adviseren we voor onderzoek naar de effectiviteit van (lichaamseigen of kunstmatig) botmateriaal de vorming van een botbrug 1 jaar na de operatie te evalueren.

De secundaire resultaten van **Hoofdstuk 3** en **Hoofdstuk 4** laten zien dat er tijdens de eerste 2 jaar na de wervelfusie operatie geen onverwachte complicaties zijn opgetreden en dat geen van de geregistreerde complicaties direct gerelateerd was aan het gebruik van BCP$_{<\mu m}$. Overeenkomstig met vergelijkbare patiëntpopulaties verbeterden de klinische uitkomsten, gemeten met de gevalideerde Oswestry Disability Index (ODI), EQ-5D-5L en Visuele Analoge Schaal (VAS) voor rugpijn, tot de 2 jaar controle en met name tijdens de eerste 3 maanden.

Hoofdstuk 5 gaat in op de uitdagingen ten aanzien van de haalbaarheid van een chirurgische RCT en onze keuze voor een alternatieve studieopzet waarbij elke patiënt zijn/haar eigen controle is. Daarnaast beschrijft en kwantificeert dit hoofdstuk de methodologische aspecten van zo'n intrapatiënt RCT. De voornaamste voordelen van een intrapatiënt RCT zijn de eliminatie van variabiliteit tussen patiënten en reductie van de steekproefgrootte. Nadere analyse van de studiedata uit Hoofdstuk 3 liet zien dat de mate van botfusie binnen patiënten meer overeenkomt dan tussen patiënten. Dit betekent dat patiëntfactoren (zoals bijvoorbeeld diagnose en roken) een grote rol spelen. Daarnaast was de steekproefgrootte van onze intrapatiënt RCT slechts een derde van de steekproefgrootte van een traditionele RCT met twee parallelle groepen. Dit betekent een aanzienlijke reductie van de duur en kosten van een studie. Potentiële nadelen van een intrapatiënt RCT zijn dat de ene behandeling de uitkomst van de andere behandeling kan beïnvloeden en dat klinische en functionele uitkomstmaten op patiëntniveau (zoals PROM's en complicaties) niet toegeschreven kunnen worden aan

één van de behandelingen. Op basis van deze voor- en nadelen hebben we de randvoorwaarden voor de toepassing van een intrapatiënt RCT gedefinieerd. Binnen de wervelkolomchirurgie achten we een intrapatiënt RCT in het bijzonder geschikt voor onderzoek naar de klinische werkzaamheid van niet-bioactieve botvervangende materialen voor geïnstrumenteerde posterolaterale wervelfusies, met botfusie als primaire uitkomst.

Hoewel het aantal wervelfusie operaties de afgelopen decennia sterk is toegenomen, is er nog weinig bekend over de lange termijn uitkomsten van deze behandeling. In **Hoofdstuk 6** hebben we de lange termijn klinische uitkomsten van een wervelfusie operatie onderzocht, bij de Nederlandse patiënten van een internationale RCT naar de effectiviteit en veiligheid van een botvervangend materiaal. Deze patiënten hebben een geïnstrumenteerde posterolaterale wervelfusie voor een afgegleden wervel (spondylolisthesis) met zenuwbeklemming ondergaan. Bij de ene helft van de patiënten is autograft gebruikt en bij de andere helft een groeifactor die botaanmaak stimuleert (bot morfogenetisch eiwit-7). Gemiddeld 11,8 (10,1-13,7) jaar na de wervelfusie operatie hebben 41 (73%) patiënten de volgende vragenlijsten beantwoord: ODI, EQ-5D-3L, VAS voor rugpijn, VAS voor beenpijn en aanvullende vragen over tevredenheid en eventuele heroperaties. De antwoorden van deze patiënten lieten zien dat een wervelfusie operatie ook op lange termijn gunstig is. Deze resultaten waren niet afhankelijk van het gebruikte botmateriaal of de radiologische botfusie uitkomst na 1 jaar. Bijna 80% van de patiënten was zeer tevreden over de operatie en er waren geen heroperaties in verband met pseudoartrose. Een opvallende bevinding was dat de operatie met name effect had op rugpijn, terwijl juist zenuwpijn (beenpijn) een voorwaarde voor inclusie was. Bovendien was het algehele behandeleffect beter dan in andere studies met chronische rugpijn patiënten zonder zenuwpijn. Onze resultaten benadrukken daarmee het belang van indicatiestelling voor wervelfusie operaties.

Het laatste hoofdstuk (**Hoofdstuk 7**) plaatst een kritische noot bij het gebruik van botvervangende materialen. Pijn in het donorgebied wordt vaak als voornaamste reden genoemd om een alternatief voor autograft te gebruiken, maar recente studies hebben onderbouwd dat de incidentie en ernst van bekkenkampijn bij laag lumbale wervelfusie operaties mogelijk wordt overschat door pijn in het primaire operatiegebied. Deze hypothese hebben we met een vragenlijst nader onderzocht bij 90 patiënten van de studie uit Hoofdstuk 3. Bij deze patiënten kon het bottransplantaat via de primaire incisie uit de bekkenkam worden geoogst, waardoor ze geblindeerd waren voor de donorkant. Bij elke nacontrole (6 weken, 3 maanden, 6 maanden en 1 jaar na de operatie) hebben we de patiënten gevraagd uit welke bekkenkam het bottransplantaat is geoogst. Slechts 24% van de patiënten heeft op elk meetmoment de juiste bekkenkam aangegeven. De VAS pijnscore voor rugpijn was op elk meetmoment hoger dan voor bekkenkampijn.

Daarnaast rapporteerden patiënten voorafgaand aan de operatie de hoogste pijnscores en was er geen verschil in pijnscore tussen de donorkant en de andere bekkenkam. Op basis van deze resultaten concluderen we dat het oogsten van bekkenbot voor lumbale wervelfusie via een enkele incisie niet leidt tot klinisch relevante pijn in het donorgebied.

Toekomstperspectief

De bevindingen van dit proefschrift geven aanleiding voor vervolgonderzoek op een aantal gebieden. Allereerst de (door)ontwikkeling en validatie van een gedetailleerd classificatiesysteem voor de radiologische beoordeling van botfusie. Nader onderzoek moet uitwijzen welke anatomische locatie, criteria en afkapwaarden de beste voorspellende waarde hebben. We verwachten dat artificiële intelligentie (AI) in de nabije toekomst een belangrijke rol gaat spelen in het betrouwbaar en reproduceerbaar beoordelen van botfusies. Hoewel CT momenteel de meest geschikte beeldvormingstechniek is, bieden ontwikkelingen op het gebied van onder andere magnetische resonantie (MRI) en echografie perspectief voor beoordeling zonder ioniserende straling.

Dit proefschrift laat zien dat BCP$_{<\mu m}$ een veelbelovend materiaal is voor geïnstrumenteerde posterolaterale wervelfusies en gebruikt kan worden als de hoeveelheid of kwaliteit van lichaamseigen bottransplantaat onvoldoende is. Het fusiepercentage is echter nog niet optimaal. Daarom is verdere verbetering van BCP$_{<\mu m}$ momenteel gericht op het (natuurlijke) proces van materiaalresorptie en botformatie. Daarnaast denken we dat de werkzaamheid van zowel botvervangende materialen als autograft vergroot kan worden door optimalisatie van de chirurgische techniek, waaronder de voorbereiding van het fusiegebied en zorgvuldige plaatsing van het botmateriaal. Een andere belangrijke onderzoeksvraag betreft de rol van botmaterialen in aanwezigheid van starre instrumentatie (schroeven en staven). Mogelijk is een meer posteromediale botbrug inclusief de facetgewrichten al voldoende en hoeven de dwarsuitsteeksels tijdens de operatie niet te worden vrijgelegd.

Naast deze kennishiaten zijn er ook belangrijke ontwikkelingen op het gebied van wet- en regelgeving rondom botvervangende materialen. Voorheen konden botvervangende materialen op basis van hun materiaaleigenschappen en preklinische onderzoeksresultaten toegelaten worden tot de markt, maar sinds 2021 is door de nieuwe Europese regelgeving voor medische hulpmiddelen (MDR) klinisch bewijs van hun werkzaamheid en veiligheid vereist. Daarnaast moeten fabrikanten het gebruik van toegelaten medische hulpmiddelen blijven controleren. Dit vraagt onder andere om doeltreffende investeringen in klinische studies en registraties. Voor de evaluatie van niet-bioactieve botvervangende materialen voor geïnstrumenteerde posterolaterale wervelfusies is bij-

voorbeeld een intrapatiënt RCT een zeer efficiënte studieopzet. Op basis van het toenemende klinisch bewijs van de werkzaamheid en toegevoegde waarde van botvervangende materialen kan een gedegen afweging worden gemaakt welk botmateriaal het meest geschikt is voor een specifieke behandeling en patiënt. Dat is de essentie van ‘evidence based medicine’.

List of publications

This thesis

1. Lehr AM, Oner FC, Delawi D, et al. Efficacy of a standalone microporous ceramic versus autograft in instrumented posterolateral spinal fusion: A multicenter, randomized, intrapatient controlled, noninferiority trial. Spine 2020;45:944–51. DOI: 10.1097/BRS.0000000000003440.
2. Lehr AM, Oner FC, Delawi D, et al. Increasing fusion rate between 1 and 2 years after instrumented posterolateral spinal fusion and the role of bone grafting. Spine 2020;45:1403–10. DOI: 10.1097/BRS.0000000000003558.
3. Lehr AM, Jacobs WCH, Stellato RK, Castelein RM, Oner FC, and Kruyt MC. Methodological aspects of a randomized within-patient concurrent controlled design for clinical trials in spine surgery. Clinical Trials 2022. DOI: 10.1177/17407745221084705.
4. Lehr AM, Delawi D, van Susante JLC, et al. Long-term (> 10 years) clinical outcomes of instrumented posterolateral fusion for spondylolisthesis. European Spine Journal 2020;45:1380–6. DOI: 10.1007/s00586-020-06671-6.
5. Lehr AM, Oner FC, Hoebink EA, et al. Patients cannot reliably distinguish the iliac crest bone graft donor site from the contralateral side after lumbar spine fusion: A patient-blinded randomized controlled trial. Spine 2019;44:527–33. DOI: 10.1097/BRS.0000000000002882.

Other peer-reviewed publications

1. Versteeg AL, van Tol FR, Lehr AM, Oner FC, and Verlaan JJ. Malnutrition in patients who underwent surgery for spinal metastases. Annals of Translational Medicine 2019;7:213. DOI: 10.21037/atm.2019.04.53.
2. Oner FC, Sadiqi S, Lehr AM, Schroeder GD, and Vaccaro AR. The Need of Validated Disease-Specific Outcome Instruments for Spine Trauma. Journal of Orthopaedic Trauma 2017;31:S33–S37. DOI: 10.1097/BOT.0000000000000945.
3. Sadiqi S, Lehr AM, Post MW, et al. Development of the AOSpine Patient Reported Outcome Spine Trauma (AOSpine PROST): A universal disease-specific outcome instrument for individuals with traumatic spinal column injury. European Spine Journal 2017;26:1550–7. DOI: 10.1007/s00586-017-5032-8.

4. Sadiqi S, Verlaan JJ, Lehr AM, et al. Universal disease-specific outcome instruments for spine trauma: A global perspective on relevant parameters to evaluate clinical and functional outcomes of thoracic and lumbar spine trauma patients. European Spine Journal 2017;26:1541–9. DOI: 10.1007/s00586-016-4596-z.

5. Sadiqi S, Verlaan JJ, Lehr AM, et al. Measurement of kyphosis and vertebral body height loss in traumatic spine fractures: An international study. European Spine Journal 2017;26:1483–91. DOI: 10.1007/s00586-016-4716-9.

6. Sadiqi S, Verlaan JJ, Lehr AM, et al. Surgeon reported outcome measure for spine trauma: An international expert survey identifying parameters relevant for the outcome of subaxial cervical spine injuries. Spine 2016;41:E1453–E1459. DOI: 10.1097/BRS.0000000000001683.

7. Sadiqi S, Lehr AM, Post MW, et al. The selection of core International Classification of Functioning, Disability, and Health (ICF) categories for patient-reported outcome measurement in spine trauma patientsresults of an international consensus process. The Spine Journal 2016;16:962–70. DOI: 10.1016/j.spinee.2016.03.050.

8. Oner FC, Jacobs WCH, Lehr AM, et al. Toward the development of a universal outcome instrument for spine trauma: A systematic review and content comparison of outcome measures used in spine trauma research using the ICF as reference. Spine 2016;41:358–67. DOI: 10.1097/BRS.0000000000001207.

9. Oner FC, Sadiqi S, Lehr AM, et al. Toward developing a specific outcome instrument for spine trauma: An empirical cross-sectional multicenter ICF-based study by AOSpine knowledge forum trauma. Spine 2015;40:1371–9. DOI: 10.1097/BRS.0000000000001009.

10. Oner FC, Sadiqi S, Lehr AM, et al. Towards the development of an outcome instrument for spinal trauma: An international survey of spinal surgeons. Spine 2015;40:E91–E96. DOI: 10.1097/BRS.0000000000000684.

11. Sadiqi S, Lehr AM, Post MW, Vaccaro AR, Dvorak MF, and Oner FC. Toward a specific outcome instrument for spinal trauma: How to measure function and health. Spine 2015;40:E578–E586. DOI: 10.1097/BRS.0000000000000852.

12. Boot W, Moojen DJF, Visser E, et al. Missed low-grade infection in suspected aseptic loosening has no consequences for the survival of total hip arthroplasty. Acta Orthopaedica 2015;86:678–83. DOI: 10.3109/17453674.2015.1086942.

13. Schlösser TPC, Stadhouder A, Schimmel JJP, Lehr AM, van der Heijden GJMG, and Castelein RM. Reliability and validity of the adapted Dutch version of the revised Scoliosis Research Society 22-item questionnaire. The Spine Journal 2014;14:1663–72. DOI: 10.1016/j.spinee.2013.09.046.

14. Homminga J, Lehr AM, Meijer GJM, et al. Posteriorly directed shear loads and disc degeneration affect the torsional stiffness of spinal motion segments: A biomechanical modeling study. Spine 2013;38:E1313–E1319.
DOI: 10.1097/BRS.0b013e3182a0d5fa.

An up-to-date list of all publications can be found on *ORCID*: orcid.org/0000-0001-8093-8768 or on *Google Scholar*: scholar.google.com/citations?user=pvmEDYcAAAAJ.

Conference presentations

1. Lehr AM, Kruyt MC. Assessment of bone grafts in spinal fusion: clinical trial design, pitfalls and lessons learned. Invited lecture Dutch Orthopaedic Association Spring Meeting 2021 (online).
2. Lehr AM, Oner FC, Hoebink EA, Kempen DHR, van Susante JLC, Castelein RM, Kruyt MC. Fusion performance of AttraX® Putty vs. autograft in posterolateral spinal fusion: a randomized intrapatient controlled noninferiority trial. EUROSPINE Annual Meeting 2019, Helsinki, Finland.
3. Lehr AM, Oner FC, Hoebink EA, Kempen DHR, van Susante JLC, Castelein RM, Kruyt MC. Fusion performance of AttraX® Putty vs. autograft in posterolateral spinal fusion: a randomized intrapatient controlled noninferiority trial. Global Spine Congress 2019, Toronto, Canada.
4. Lehr AM, Oner FC, Hoebink EA, Kempen DHR, van Susante JLC, Castelein RM, Kruyt MC. Posterior spinal fusion with a microporous ceramic: a randomized, intrapatient controlled trial. Invited lecture European Society for Biomaterials Annual Meeting 2018, Maastricht, The Netherlands.
5. Lehr AM, Oner FC, Hoebink EA, Kempen DHR, van Susante JLC, Castelein RM, Kruyt MC. Fusion performance of AttraX® Putty vs. autograft in posterolateral spinal fusion: a randomized intrapatient controlled noninferiority trial. Nordic Spinal Deformities Society Annual Meeting 2018, Kolding, Denmark.
6. Lehr AM, Oner FC, Hoebink EA, Kempen DHR, van Susante JLC, Castelein RM, Kruyt MC. Relevantie van bekkenkampijn na het oogsten van bekkenbot voor lumbale wervelkolomfusies: kunnen patiënten de donorzijde identificeren? Dutch Orthopaedic Association Annual Meeting 2018, Den Bosch, The Netherlands.
7. Lehr AM, Oner FC, Castelein RM, Kruyt MC. Relevance of iliac crest bone graft donor site pain after spinal fusion: can patients identify the harvest site? Nordic Spinal Deformities Society Annual Meeting 2017, Stockholm, Sweden.

Acknowledgements (Dankwoord)

Dit proefschrift is tot stand gekomen dankzij de inzet, betrokkenheid en steun van velen, waarvan ik een aantal graag in het bijzonder wil bedanken.

Allereerst mijn promotoren, prof. dr. Moyo Kruyt, prof. dr. Cumhur Oner en prof. dr. René Castelein. Ik vind het een enorme eer dat jullie alle drie mijn promotor zijn. Jullie hebben de afgelopen 10 jaar een belangrijke rol gespeeld in mijn ontwikkeling en ik heb veel bewondering voor wat ieder van jullie binnen de orthopedie heeft bereikt.

Moyo, ondanks en dankzij onze verschillende karakters vind ik ons een geweldig team. De AxA studie was voor ons beiden de eerste grote klinische studie en we kunnen trots zijn op het resultaat. Bedankt voor je vertrouwen om deze studie te coördineren en uiteindelijk een promotietraject te doorlopen. Ik heb veel geleerd van je kritische vragen en super snelle feedback op mijn manuscripten. In mijn ogen ben je bij uitstek een academisch orthopedisch chirurg; met je assertiviteit en eigenzinnigheid, technische blik, basale tot klinische onderzoekservaring en je enorme verantwoordelijkheidsgevoel richting patiënten. Ik vind het heel waardevol dat we op allerlei fronten kunnen sparren en elkaar verder kunnen helpen.

Cumhur, met het AOSpine trauma outcome project heb je me vertrouwd gemaakt met projectmanagement, patiëntencontact en samenwerken met verschillende professionals. Ik heb veel ontzag voor al je inzet voor het vakgebied wervelkolomchirurgie, zowel binnen het UMC Utrecht als nationaal en internationaal. Ook blijf ik onder de indruk van je enorme specialistische en algemene kennis, van gezondheidszorg in de breedste zin van het woord tot de geschiedenis van Sisi. Je hebt me vele wijze lessen geleerd. Bedankt dat ik altijd bij je terecht kan voor advies of een praatje.

René, dankzij jou is mijn carrière gestart bij de orthopedie. Vanaf dag 1 heb je gezegd dat een promotie een mooie kers op de taart zou zijn en na vele ontkenningen heb je gelijk gekregen. Met jouw vragen en opmerkingen stimuleerde je me om vanuit een andere hoek naar de verschillende onderzoeksresultaten te kijken. Daarnaast toonde je veel belangstelling voor mijn persoonlijke leven. Ik ben je heel dankbaar voor de kansen en ontwikkelmogelijkheden die je me de afgelopen 10 jaar hebt geboden, van financiële en management cursussen tot de nodige tijd voor dit proefschrift. Daarmee heb je me de ruimte gegeven om te ontdekken welke kant ik op kan of wil gaan.

Op de tweede plaats wil ik graag alle patiënten bedanken die hebben deelgenomen aan de klinische studies die in dit proefschrift zijn beschreven. Dankzij uw vertrouwen en bereidheid om onder andere vragenlijsten in te vullen hebben we de wetenschap en zorg een stukje vooruit kunnen helpen.

Collega's van de Dutch Clinical Spine Research Group, Diederik Kempen, Mark Altena, Job van Susante, Arnold Vreeling, Eric Hoebink, Hans Amelink, Willem-Bart Slooff, Diyar Delawi en in het bijzonder Nienke Willigenburg, Marit Sterenborg, Femke Hagenmaier, Annemiek Hol en Inge Wiljouw als drijvende kracht van het OLVG, Rijnstate Ziekenhuis en Amphia Ziekenhuis. Hartelijk dank voor jullie inzet voor de AxA studie, van protocol, toetsingsprocedure, inclusies, operaties, follow-ups en monitorvisites tot feedback op manuscripten, en de prettige samenwerking. Hopelijk kunnen we als multicenter team nog vele onderzoeksprojecten tot een succes maken.

Wilco Jacobs en Rebecca Stellato, als adviseurs op het gebied van methodologie en statistiek waren jullie onmisbaar bij dit promotietraject. Wilco, jouw orthopedische en methodologische expertise zijn een waardevolle combinatie waar we als afdeling graag op terugvallen. Rebecca, bedankt voor je tijd en geduld om ingewikkelde analyses zoals GEE op begrijpbare manier aan me uit te leggen.

Anneli Duits, waar zijn we in 2018 aan begonnen... Na ontelbare uren artikelen lezen, data analyseren en discussiëren is het ons gelukt om een prachtige systematic review neer te zetten. Ik heb met veel plezier met je samengewerkt en grote bewondering voor je kennis en doorzettingsvermogen. Maarten Reijnders en Don Nutzinger, bedankt voor de data extractie van honderden artikelen.

Nienke Wolterbeek, Babette van der Zwaard en Martijn Vrijkorte, bedankt voor jullie hulp bij de lange termijn follow-up van de Osigraft studie.

Rebecca Smith, thank you for the AxA study support from overseas and short lines of communication.

Irina Mancini, thank you for the professional illustration of the intrapatient controlled design.

Leden van de beoordelingscommissie, prof. dr. Gorter, prof. dr. Habibovic, prof. dr. de Jong, prof. dr. Rosenberg, prof. dr. van Royen en prof. dr. Verhaar (voorzitter), bedankt voor het lezen en beoordelen van mijn manuscript.

Graag wil ik ook de vele (oud-)collega's bedanken die me vanaf de zijlijn hebben gesteund en maken dat ik elke werkdag met veel plezier naar het UMC Utrecht fiets. Debby, bedankt dat je me als eerste kamergenoot een warm welkom hebt gegeven bij de orthopedie en op vele manieren op sleeptouw hebt genomen. Onderzoekers van D3 en Q3, bedankt voor de gezellige lunchpauzes en congresbezoeken, jullie kritische vragen tijdens de research meeting, jullie hulp bij de verschillende onderzoeksfases (van ondoorzichtige randomisatie enveloppen tot SPSS), jullie steun bij congrespre-

sentaties, de goedgevulde snoeptrommel etc. Willemijn, Anne en Jonneke, bedankt voor de vele koppen thee en cappuccino en jullie vriendschap. Esmee, Ruud en in het bijzonder Hilde en Pauline, bedankt voor het uit handen nemen van verschillende (onderzoeks)taken, waardoor ik me kon storten op het analyseren van data en schrijven van artikelen. Hilde en Pauline, met jullie komst hebben we het klinisch onderzoek van de orthopedie weer naar een hoger niveau kunnen tillen. Prachtig om te zien hoeveel klinische studies er (gaan) lopen en hoe jullie je functie hebben vormgegeven. Hilde, bedankt voor je kritische feedback op mijn Nederlandse samenvatting. Stafleden en dames van het secretariaat, in het bijzonder Simone, jullie zijn stuk voor stuk hele fijne collega's. Mattie, jij verdient een plekje in elk proefschrift van de orthopedie.

Lieve vrienden en vriendinnen, bedankt voor jullie belangstelling en sociale bijdrage. Van STER borrels en gezellige uitjes tot thuisdates met de kids, jullie vriendschap geeft me veel plezier en energie. Marloes, Christina en Nicole, het betekent heel veel voor mij dat we al zolang onderdeel van elkaars leven zijn.

Lieve paranimfen, bedankt dat jullie op 9 juni mijn grote steun zijn. Jonneke, jouw vrolijke en energieke verschijning doet me altijd goed. Onder het genot van vele cappuccino's hebben we heel wat afgekletst, gelachen, gespard, geregeld en noem maar op. Yvette, jij bent een hele lieve en betrokken vriendin die haarfijn aanvoelt wat de ander nodig heeft. Naast onze ontspannende koffie/lunch dates en vele wandelingen in COVID tijd ben ik je onder andere dankbaar voor je verfrissende inzichten bij carrière vraagstukken.

Lieve papa, mama en Frédérique, jullie interesse, steun en liefde zijn echt onvoorwaardelijk. Papa en mama, ik realiseer me steeds meer hoe bijzonder het is dat jullie altijd voor Frédérique en mij, en inmiddels ook de kleinkinderen, klaar staan. Fre, ondanks onze verschillende karakters en levens ben je altijd betrokken. Die fijne thuisbasis is in vele opzichten heel belangrijk voor me, dankjulliewel.

Lieve Thomas en Sophie, als het goed is hebben jullie zo min mogelijk meegekregen van de afronding van dit boekje. Thomas, al vroeg jij aan het einde wel "mama, wat is een review?". Bedankt voor de afleiding en ontspanning, maar vooral de vele dikke knuffels en kusjes. Elke dag met jullie is een feestje.

Lieve Wouter, jij hebt op vele manieren bijgedragen aan de totstandkoming van dit proefschrift. Heel veel verhalen over de orthopedie aanhoren, zenuwen voor congres presentaties bedaren, ICT support, figuren maken en bovenal in de avonduren de hele lay-out. Dat was gezien mijn perfectionisme niet altijd een even makkelijke klus, maar ik ben je er enorm dankbaar voor. Bovenal ben ik je dankbaar voor wie je bent en hoe je me tijdens de laatste loodjes hebt verrast en gesteund. Jij bent echt mijn thuis.

Curriculum Vitae

Mechteld Lehr was born on the 14th of June 1987 in Assen, The Netherlands. She grew up with her parents and older sister Frédérique in Rijswijk. In 2005, she graduated cum laude from high school (Interconfessioneel Makeblijde College) and moved to Enschede for the bachelor and master program Biomedical Engineering at the University of Twente. Next to several committees, she participated as student assistant in the courses Anatomy & Physiology and Medical & Sports Physiology, which increased her interest in the musculoskeletal system. During her master, she specialized in Human Function Technology and performed her graduation project in collaboration with the Department of Orthopedic Surgery of the University Medical Center Utrecht, under supervision of prof. dr. R.M. Castelein.

After obtaining her master's degree (cum laude) in 2011, she started her career in Utrecht as Clinical Research Coordinator at the Department of Orthopedic Surgery of the University Medical Center Utrecht. Together with multiple clinical investigators, she has set up and coordinated several (multicenter) clinical trials, both national and international. While her career developed towards staff advisor, in 2017 she embraced the opportunity to enhance her scientific expertise with a part-time PhD. Under supervision of prof. dr. M.C. Kruyt, prof. dr. F.C. Oner and prof. dr. R.M. Castelein, several research projects on spinal fusion resulted in this thesis.

Mechteld currently works as staff advisor at the Department of Orthopedic Surgery of the University Medical Center Utrecht, headed by prof. dr. L.W. van Rhijn. Business operations and quality assurance are her focus areas. She lives with her husband Wouter Potters and their children Thomas and Sophie in Utrecht.