CERVICAL AND LUMBAR TOTAL DISC REPLACEMENTS

Federaal Kenniscentrum voor de Gezondheidszorg
Centre Fédéral d'Expertise des Soins de Santé
Belgian Health Care Knowledge Centre

CERVICAL AND LUMBAR TOTAL DISC REPLACEMENTS

KIRSTEN HOLDT HENNINGSEN, NANCY THIRY, CHRIS DE LAET, SABINE STORDEUR, CÉCILE CAMBERLIN

.be

COLOPHON

Layout:	Ine Verhulst, Joyce Grijseels

■ TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATION	DEFINITION
95% CI	95% confidence interval
ACD	Anterior cervical discectomy (no fusion)
ACDF	Anterior cervical discectomy and fusion
ALIF	Anterior lumbar interbody fusion
ASD	Adjacent segment degeneration
ASP	Adjacent segment pathology
CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CLBP	Chronic low back pain
CMA	Cost-minimization analysis
Co	Cobalt
Cr	Chromium
CRD	Centre for Review and Dissemination
CTDR	Cervical total disc replacement
CUA	Cost-utility analysis
DCRA	Distortion compensated röntgen analysis
DDD	Degenerative disc disease
DGSIE – ADSEI	Direction générale Statistique et Information économique – Algemene Directie Statistiek en Economische Informatie - Statistics Belgium
EQ-5D	EuroQoL 5 dimensions
FA	Facet arthropathy
FDA	Food and Drug Administration (USA)
HCP	Health care payer
HRQoL	Health Related Quality of Life
HTA	Health Technology Assessment
HUI	Health Utilities Index

ICER	Incremental cost-effectiveness ratio
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification
INAMI – RIZIV	Institut national d'assurance maladie-invalidité - RijksInstituut voor ziekte en invaliditeits verzekering (Belgium)
INAHTA	International Network of Agencies for Health Technology Assessment
LBP	Low back pain
LOS	Length of stay
LTDR	Lumbar total disc replacement
LY	Life year
MD	Mean difference
MCID	Minimally clinically important difference
Mo	Molybdenum
NDI	Neck Disability Index
NHS EED	National Health Service Economic Evaluation Database
NICE	National Institute for Health and Care Excellence (UK)
NRS	Numeric Rating Scale
NSAID	Non-steroidal anti-inflammatory drug
ODI	Oswestry Disability Index
OR	Odds ratio
PICO	Patient, Intervention, Comparator, Outcomes
PLIF	Posterior lumbar interbody fusion
PLF	Posterolateral lumbar fusion
PMMA	Polymethylmetacrylat
QALY	Quality-adjusted life year
QoL	Quality of life
QWB	Quality of Well-Being scale
RCT	Randomised controlled trial
RHM – MZG	Résumé Hospitalier Minimum – Minimum Ziekenhuisgegevens
RMDQ	Rolland Morris Disability Questionnaire

ROM	Range of motion
RR	Relative risk
SA	Sensitivity analysis
SD	Standard deviation
SF-6D	Short Form-6 dimension
SF-12	Short Form-12 (items)
SF-36	Medical Outcome Study Short Form 36-Item Health Survey
SMD	Standardized mean difference
SPF – FOD	Service Public Fédéral Santé publique, Sécurité de la Chaîne alimentaire et Environnement – Federale Overheidsdienst Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu
TDR	Total disc replacement
Ti	Titanium
UHMWPE	Ultra-high molecular weight polyethylene
VAS	Visual Analog Scale
XLIF	Extreme lateral interbody fusion

■ SCIENTIFIC REPORT

1 INTRODUCTION

1.1 Background

The human spine comprises 33 vertebrae grouped according to their location: 7 cervical (numbered C1-C7), 12 thoracic (T1-T12), 5 lumbar (L1-L5), 5 sacral (S1-S5) and four coccygeal ones. The sacral and coccygeal vertebrae are fixed but the other vertebrae are said movable. Vertebrae C2 to S1 are separated by each other by an intervertebral disc made of a fibrocartilaginous annulus fibrosis surrounding a soft pulpy semi-gelatinous core: the nucleus pulposus. These intervertebral discs are flexible, play the important role of shock absorbers and ensure the spine movement and stability.

The natural curvature of the spine is convex in the cervical part (lordotic curve), concave in the thoracic part (kyphotic curve) and convex again in the lumbar part (Figure 1).

It is known that back pain is a major common public health problem. It impairs the quality of life, it represents one of the major cause of disability and is responsible for a large share of the healthcare costs. Approximatively 70 to 80% of the general population will encounter in its lifetime at least one episode of back pain. As a consequence, 1 to 2% of the gross national product will be engulfed in the healthcare and social costs, three times more than for cancer.[1]

The pain is said to be chronic when it lasts more than several weeks. The thresholds retained in the literature vary but are often set at 6 weeks for acute pain, between 6 and 12 weeks for subacute pain and 12 weeks for chronic pain.[2, 3] The survey results of the Belgian Health interview survey 2013 show that 21% of Belgians aged 15 years or more suffered from low back disorder or other chronic back defect in the past 12 months and 12% from neck disorder or other chronic neck defect in the same period.[4]

Figure 1 – Anatomy of the spine
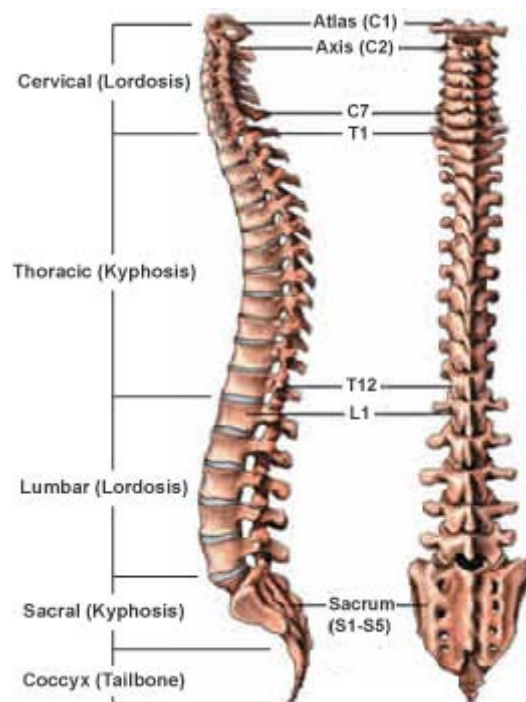
http://commons.wikimedia.org/wiki/File:Spine_Anatomy_Kisco.JPG

Although its exact cause remains unclear, degenerative disc disease (DDD) is often associated to chronic low back pain.[5] The first step of spinal degenerative disease is thought to be intervertebral disc disease, typically followed by osteophytes, disc narrowing and spinal stenosis.[6] The cervical and the lumbar spine are the most affected parts of the spine.[7] Degenerative disc disease of the cervical spine can result in significant pain, instability, radiculopathy, myelopathy or a combination of symptoms.[8] When lumbar spine is affected by DDD, the most common symptom is low back pain.

When the pain stays refractory to conservative treatment, surgery is often considered, consisting of discectomy with (as in the Belgian practice) or without fusion. After the removal of the diseased intervertebral disc (discectomy), the fusion consists in fastening the two vertebral bodies together, suppressing any spine mobility at this level. A possible alternative is a total disc replacement (TDR) during which the natural disc is replaced by a non-rigid artificial disc prosthesis without fastening the vertebrae together.

The aim of the present report is to clarify the claimed advantages and disadvantages of these alternatives based on the evidence, on cervical level as well as on lumbar level.

1.2 Scope and objectives

The research questions are the following:

1. What is the evidence of the short-term and long-term clinical effectiveness, safety and cost-effectiveness of total cervical disc replacement versus conservative treatment and/or (discectomy and) fusion in subacute/chronic radicular arm pain?

2. What is the evidence of the short-term and long-term clinical effectiveness, safety and cost-effectiveness of total lumbar disc replacement versus conservative treatment and/or (discectomy and) fusion in chronic lumbar pain due to intervertebral disc disorder?

The report also describes the Belgian use and reimbursement of cervical (CTDR) and lumbar (LTDR) total disc replacements. Organisational, legal, ethical or patient issues other than patient outcomes (patient satisfaction, quality of life) are not covered in this rapid HTA.

2 CERVICAL TOTAL DISC REPLACEMENT

2.1 Health problems

HTA CORE MODEL DOMAIN: CUR1

2.1.1 Population and condition

The main indication that declare patients refractory to conservative treatment likely to be candidates for cervical spine surgery - including cervical total disc replacement - is cervical radiculopathy either from disc herniation and/or osteophytes. In the natural process of aging, intervertebral discs begin to lose proteoglycans, leading to moisture loss.[9] They lose their flexibility, elasticity, and shock absorbing characteristics and begin to collapse. Osteophytes (bone spurs), disc herniation, kyphosis or instability begin to develop causing the narrowing of the foramen. Radiculopathy occurs when the nerve roots are compressed due to those changes, possibly causing neck pain, radiating pain in the shoulder, the arm or even the hand, muscle weakness and/or numbness or tingling in fingers or hands. Other symptoms may include lack of coordination, especially in the hands. In myelopathy, the compression involves the spinal cord. Because neural transmission is impaired, symptoms that can appear include altered gait of balance, sphincter disturbances, weakness, numbness or clumsiness of the hand. Over time, symptoms may diminish, stabilize or worsen.[9, 10]

The number of persons aged 40 or older presenting radiographic evidence of cervical DDD secondary to spondylosis amounts to 60%.[10] In the older age, some studies evoke a percentage by 65 years of nearly 95% of males and 70% of females.[11] Fortunately degenerations are most often asymptomatic.[9] When it is symptomatic, symptoms can include neck and arm pain associated with radiculopathy and myelopathy.[10]

According to a 2006 systematic review, the European one year prevalence of neck pain was estimated to be 26% (95% CI 13% to 39%). That means that one out of 4 European citizens experienced neck pain in the last 12 months. Adult population is more affected than children or elderly and women slightly more than men.[12] This is higher than the results of the Belgian Health Interview survey 2013 that show that about 12% of Belgians aged 15 years or more suffered from neck disorder or other chronic neck defect in the last year (14.3% women and 9.1% men).[4]

About 30% of symptomatic patients presenting cervical DDD do not respond to conservative treatments and may be considered to be candidates for cervical surgery.[7]

2.1.2 Existing treatments

The conservative therapeutic arsenal includes rest, physical rehabilitation, muscle-relaxants, analgesics or non-steroidal anti-inflammatory medication, epidural and selective nerve roots injections.[7, 9] When conservative treatment is not effective, generally after 6 weeks, cervical spine surgery may be performed.

Anterior cervical discectomy and fusion (ACDF) is commonly used to treat symptomatic cervical degenerative disc disease (DDD).[10, 13-15] During this procedure, the pressure on the spinal nerve roots is relieved (decompression), the diseased disc is anteriorly removed and the left empty space is fused using some material. This material can be pedicle screw fixations, plate systems or cages but can also include autologous bone (often from the iliac bone), allogeneic bone (from a human or animal donor) or bone graft substitutes and osteogenic factors.[16] This fusion relieves the pain supposedly by definitively suppressing the range of motion of the spine at the level of the disc disease.

2.2 Description and technical characteristics

HTA CORE MODEL DOMAIN: TEC

The primary rationale behind the insertion of a cervical artificial disc instead of a fusion is to reduce adjacent level disease incidence by maintaining motion of the spine and hence reducing loads in adjacent segments. Secondarily, there is no morbidity associated with an autograft site or due to allograft material. Limitations exist to total disc replacement like a known allergy to implant material, osteoporosis, fusion of an adjacent level, spondylosis or facet joint degeneration.[13] Total disc replacement may entail short-term well-known complications as implant migration, heterotopic ossification (ossification around the implant outside the vertebra body where it is fixated) and may also entail long-term complications on which less data are available such as implant subsidence, loosening or migration, material failure, allergic reactions, systemic release of metallic ions, visceral or neurologic injuries and infection.[13]

After a decade of experimentations in the fifties, the first real human implantation of a cervical device was performed by Ulf Fernstrom in 1966 (after several lumbar implantations of a similar device). The device was a stainless steel ball bearing a prosthesis. Others also experimented his type of device but the high failure rate (subsidence, migration and adjacent level hypermobility) put a curb on the early enthusiasm, surrendering to arthrodesis.[17] It was only in the eighties that the popularity of the lumbar prostheses put the cervical device in the forefront again, especially in Europe. Several prototypes were developed that gave birth to the currently used devices.

First introduced as the Cummins-Bristol disc in 1989, the Frenchay cervical disc (named after a hospital in the UK) demonstrated more favourable results in 2002. This disc was a metal-on-metal ball and socket device in stainless steel and the ancestor of the current Prestige Disc, manufactured by Medtronic. An anchoring screw was fixed into both pieces to avoid migration.[17, 18] A first RCT demonstrated promising results for the Prestige Disc compared to anterior cervical discectomy and fusion in 2004.[19]

Parallel to this UK story, the American Vincent Bryan developed the Bryan cervical disc in the nineties, consisting of two titanium alloy shells articulating with a plastic (polyurethane) core. Several European studies confirmed favourable results compared to fusion. The first Bryan cervical disc was implanted in Belgium in 2000 by Pr Jan Goffin.[20] In France, Dr Thierry Marnay conceived the ball and socket design of the Prodisc-C, as the cervical version of the lumbar one. The first implantation was performed in 2002.[17]

Compared to cervical discectomy and fusion, the aim of the cervical prosthesis is to maintain the natural disc mobility, to keep the original spine alignment, the cervical lordosis (the curve of the nape of the neck) and the intervertebral height to improve function. The maintenance of mobility also presents the advantage of potentially avoiding the adjacent segment degeneration that can occur after a fusion.[9] Like the cervical discectomy and fusion, the total cervical disc replacement is performed by orthopaedic surgeons or neurosurgeons.

The approach of a cervical prosthesis implantation is close to the anterior discectomy and fusion technique. Under general anaesthesia, the patient lies in supine position. Once the diseased intervertebral disc is removed by a small anterior incision in the neck and the neurological elements are decompressed if need be, the surgeon inserts the cervical disc prosthesis into the intervertebral space instead of fusing the vertebrae. The centring of the device via fluoroscopy is crucial to the good functioning, mobility and probably reduced wear of the prosthesis.[9, 21]

Cervical artificial prostheses differ by their anchoring system (e.g. screws), the material of the bearing surface (e.g. metal, ceramic), the coating (e.g. porous mineral compound to ensure anchoring), the constraint (the degree to which it allows movement other than uniaxial rotation)[16] their rotation centre (centre of disc, vertebral body) and their MRI compatibility (CoCrMo shows more artefacts on MRI rather than polymer or titanium).[16, 21] A same model of prosthesis comes in different sizes and heights to fit all patients. Most materials of cervical (and lumbar) artificial discs have been widely used for many years in other orthopaedic prostheses (e.g. artificial hip and knee devices).[7] Manufacturers also provide surgery kits dedicated to their branded cervical disc, those kits being different from the fusion material.

The motions available to the spinal column are flexion and extension, lateral flexion and rotation (translations are also possible, increasing in amplitude from C2 to C7). Motions are said to be coupled as movements around two axes are involved to obtain a certain motion. Hence, pure lateral flexion and pure rotation do not occur. Gliding and tilting of the vertebrae make motions possible. Intervertebral discs increase the possibility of motion by separating two vertebrae and transmit load from one vertebra to another. They are present between vertebrae of the C2-C7 region.[22] In this region, the surfaces of the vertebrae are not flat but rather curved in the sagittal plane.[23] Simplifying the complex motion process of the normal cervical intervertebral discs, the cervical prosthesis can present up to 6 degrees of freedom in its movements: 3 in translation and 3 in rotation (along the three axes: vertical, transverse and sagittal). This is the case for what is called non-constraint devices, whereas semi-constraint devices including a mobile core have 5 degrees of freedom (2 in translation and 3 in rotation). Constraint devices, on the other hand, have only three degrees of freedom (no translation and 3 degrees in rotation).[21] The appealing unconstrained character may present a disadvantage such as inducing kyphosis.[16, 21] The instantaneous centre of rotation of a cervical intervertebral disc is located somewhere in the superior part of the inferior vertebra (the lower the vertebra, the higher this centre is located in the vertebral body depending on the ratio between translation and

rotation).[23] This feature is mimicked by some artificial discs including those that have a superior convexity. Devices with an inferior convexity rotate from above the device instead of below. Other devices present a centre of rotation elsewhere: e.g. on the upper vertebral body, posteriorly or even in the centre of the device.[21]

The design of the prosthesis often includes teeth (cf. Prodisc-C NOVA hereunder) or keels (cf. Prodisc-C VIVO) to facilitate the short-term stability while the long-term stability is insured by osteo-integration (bony on growth) with surrounding vertebrae, facilitated by a porous coating of the external faces of the endplates.

Hereunder are the main cervical artificial discs available in Belgium. Characteristics and pictures of the main cervical artificial discs available in Belgium can be found in Table 1 and Table 2. Table 1 gives the year when the North American Food and Drug Administration (FDA) (USA) approved each device (PMA is the pre-market approval decision required for the prosthesis before US marketing) and when the CE mark was authorized (this CE mark is required for the launching on the European market).

2.2.1.1 Bryan

Probably the most implanted cervical disc prosthesis worldwide until now, the unconstrained Bryan cervical disc is made of two titanium pieces articulating with a polyurethane core. The porous titanium particles sprayed on the external convex plates insures a better bony on growth while anterior flanges prevent posterior migration of the device. The core that allows motion and shock absorption is surrounded by a flexible polyurethane membrane containing a sterile saline lubricant. The membrane prevents soft tissue on growth and wear debris migration. The centre of rotation is located in the centre of the prosthesis.

2.2.1.2 Prestige

The Prestige cervical disc is originally a semi-constrained stainless steel device made of two components articulated by a ball and socket design. The ball is on top and the inferior part is concave, which places the centre of rotation above the device. Since 1999, the flat side of each component that are in contact with the vertebrae is roughened through a grit blasting process to facilitate bony on growth. Anterior screws anchor the device to the two neighbouring vertebral bodies above and below the disc. In the more recent

version, the screws have been replaced by two anterior rails (Prestige STLP) and stainless steel has been replaced by titanium ceramic composite (Prestige LP).

2.2.1.3 Pro-Disc C

The Prodisc-C has two titanium plates coated by plasma-sprayed rough pure titanium particles to facilitate osteo-integration. The anchorage is also insured by three keels (model NOVA) or by multiple teeth (model VIVO). The hemispheric inlay is in ultra-high molecular weight polyethylene (UHMWPE) rotating on a cobalt chromium molybdenum (CoCrMo) alloy internal surface. The rotation centre is located below the inferior plate.

2.2.1.4 Mobi-C

The Mobi-C is a semi-constrained metal-on-polyethylene device. The metal plates are made of a cobalt chromium alloy, with a roughened titanium surface coated by hydroxyapatite. Naturally present in bone, hydroxyapatite is a porous calcium-based mineral compound that encourages bony on growth. The anchorage is also insured by the inclined teeth that also make the insertion easier. The motion is made possible by the polyethylene core, flat on the bottom and convex in its superior part to mould the concave surface of the superior plate. The centre of rotation is thus located on the lower vertebra surface. Two lateral stops on the inferior endplate limit the movement of the core. Until now the Mobi-C is the only device that has been approved for one-level and two-level use (two prostheses implanted at the same time) by the FDA (since August 2013).

2.2.1.5 Baguera C

Baguera C presents six degrees of freedom that are controlled in magnitude. Both titanium plates are coated with Diamolith, a diamond like-carbon titanium coating that avoids wear debris thanks to its hardness and increases the sliding of the nucleus. Anchoring is facilitated by the coating, the sloping anatomical shape of the plates and three fins on each of the plates. Shock absorption is ensured by the high density polyethylene mobile nucleus and the shape of the non-flat inferior endplate.

2.2.1.6 M6-C

This unconstrained prosthesis is made of two titanium endplates coated with a titanium plasma spray for bony on growth, presenting each three keels for anchoring. The core is a polyurethane nucleus wrapped in an annulus made of woven UHMWPE fibre. The annulus allows axial compression and the annulus controls the range of motion in the six degrees of freedom. The shock-absorbing annulus is surrounded by a polyurethane sheath designed to minimize soft tissue on growth and debris migration, like the membrane of the Bryan cervical disc described above.

Table 1 – List of main cervical artificial discs available in Belgium

Device	Manufacturer	Distributor in Belgium	Constraint	Bearing surface material	Centre of rotation	FDA Pre Market Approval	CE Mark approval
Bryan	Medtronic	Medtronic	Unconstrained	Metal (Ti) on polyurethane	Centre of the disc	2009	2000
Prestige	Medtronic	Medtronic	Semi-constrained	Metal on metal, lately Ti Ceramic	Upper vertebra surface, posterior	2007	2002
Prodisc-C	Depuy Synthes*	Depuy Synthes*	Constrained	Metal (CoCrMo) on polyethylene	Lower vertebra surface, posterior	2007	2002
Mobi-C	LDR Medical	LDR Medical	Semi-constrained	Metal (CoCr) on polyethylene	Lower vertebra surface	2013	2004
Baguera C	Spineart	HOSPITHERA	Semi-constrained	Metal (Ti) on polyethylene	Lower vertebra surface	No**	2007
M6 C	Spinekinetics	Cormed	Unconstrained	Metal (Ti) on polyurethane	Centre of the disc	No**	2005

Adapted from Sekhon.[16]
* Depuy and Synthes were merged in 2012, as a company by Johnson & Johnson.
** Currently not marketed in the USA.

Table 2 – Pictures of main cervical artificial discs widely available in Belgium

Bryan	Prestige LP	Prodisc-C: NOVA (left) and VIVO (right) models
Mobi-C	Baguera C	M6 C
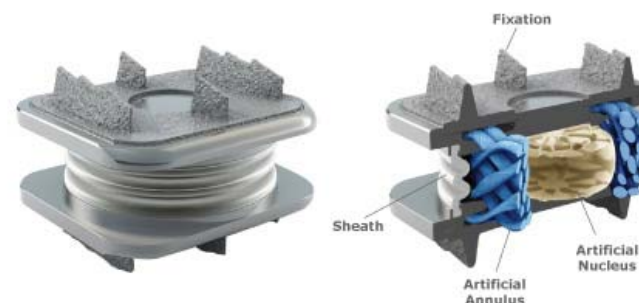
Note: scale is not preserved between models.

Many other models are currently available from manufacturers from all over the world but their sales are anecdotical or inexistent in Belgium (situation in November 2014). They include prostheses with a ball and socket design composed of two metal plates rotating thanks to a convex core such as Discocerv (Scient'X-Alphatech Spine) distributed by Orthogèse, activ C (aesculap) by B. Braun Medical or SECURE-C (Globus Medical); prostheses with a poly-ether-ethyl-ketone (PEEK) ball and socket design like the NuNec (rti surgical formerly Pioneer Surgical Technology) by Inspine Belgium; prostheses with a ceramic ball and socket design like the GRANVIA C (MEDICREA) by Cormed, prostheses designed with two metallic endplates fitted together such as CerviCore (Stryker), Mobile Cervical Disc (Osteon), but also non-constrained prostheses like the UFO (BIOMECH Paonan Biotech Co., Ltd.) by Cormed, consisting of two dome-shaped metal plates hemming a load absorbing core and the polymeric one piece prostheses Cadisc-C (Ranier) by Inspine Belgium and Almas (Novaspine).

2.2.1.7 *Belgian recommendations of good practice for cervical disc replacement*

The 2008 Belgian Society of Neurosurgery Recommendations of good practice considered the following situations to be acceptable for cervical disc replacement:[24]

- Age between 18 and 60, and
- radiculopathy due to soft disc herniation and/or moderate uncarthrosis;
- 1 or 2 levels maximum

Additionally, none of the following contraindications should be found

- Severe uncarthrosis.
- Severe facet arthritis.
- Clinical or radiological myelopathy with exception of myelopathy due to a big soft herniation in combination with a sufficiently large spinal canal.
- Spinal canal narrowing.
- Fracture.

According to the experts group accompanying this study (see colophon), other contra-indications are considered in practice: infection, osteoporosis, radiographic instability, segmental or global deformity.

2.2.1.8 *Belgian regulation and reimbursement*

Device

The artificial cervical disc is currently not reimbursed in Belgium. Currently prices of a complete artificial cervical disc turn around €2500.

Procedure

The cervical total disc replacement is billed under one of the following INAMI – RIZIV nomenclature codes:

- 281105 Cervical interbody arthrodesis, including the graft taking, or
- 281120 Surgical treatment of a cervical disc herniation.

In both cases, the reimbursed tariff amounts to €793.7 (N 625). These codes are not specific to the total disc arthroplasty, the first one can also be billed for a fusion. More details can be found in the appendix where one day surgery codes are also given (contrary to some surgeons abroad as in the USA for example, Belgian surgeons do not perform CTDR in one day hospitalisation setting).

2.3 Current use

HTA CORE MODEL DOMAIN: CUR2

2.3.1 *Methods*

2.3.1.1 *Data sources*

To estimate the number of cervical artificial discs implanted in Belgium, we used three different sources of data: two administrative databases supplemented by the input from the industry via UNAMEC.

The first administrative database is the Résumé Hospitalier Minimum – Minimum Ziekenhuisgegevens (RHM – MZG) managed by the Service Public Fédéral Santé publique, Sécurité de la Chaîne alimentaire et Environnement – Federale Overheidsdienst Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu (SPF – FOD), ensuing from the mandatory registration of all hospitalisations in every general non-psychiatric Belgian hospital since 1991. Patient information are recorded in this administrative database, such as year of birth, gender, residence as well as other information about the stay in the hospital such as length of stay,

ICD-9-CM diagnostic codes of relevant diagnoses present on admission or appearing during hospitalisation and ICD-9-CM procedure codes of diagnostic and therapeutic procedures performed during the stay. After stripping direct patient-identifying information, records have to be sent twice a year to the SPF – FOD. As CTDR is only performed in hospital setting, we extracted the hospitalisations discharged between 2008 and 2011 (last available year in November 2014) presenting an ICD-9-CM procedure code 84.62 Insertion of total spinal disc prosthesis, cervical or 81.02 Other cervical fusion, anterior technique for ACDF. Selecting the fusions also allowed us to compare both populations.

The INAMI – RIZIV houses the second administrative data we used. Since 1st of May 2009, every CE-marked implant and invasive medical device must be notified to the INAMI – RIZIV by the distributor, whether the product is reimbursable or not. Reimbursable devices must be notified to be reimbursed. A notified product receives an identification code. For non-reimbursable devices, such as cervical prostheses, this identification code must be written on the patient bill of the implanting hospital in order to be legally charged to the patient.

UNAMEC is the Belgian federation counting more than 200 firms in the medical technology field that covers 80% of the market. There are only a few firms that do not belong to UNAMEC and sell a few cervical prostheses a year. This federation launched a survey in November 2014 asking their members to send us the number of cervical prostheses sold in the last two years.

2.3.1.2 Analysis

We describe the current use of CTDR using simple descriptive statistics. After checking the normality, age at admission was compared between CTDR group and ACDF group using a t test, gender proportion using a χ^2 test and right skewed length of stay using a nonparametric Mann Whitney test.

2.3.2 Results

2.3.2.1 Number of cervical prostheses implanted in Belgium

Table 3 – Number of cervical artificial discs or cervical total disc replacements by data source

Source	Type of data	2008	2009	2010	2011	2012	2013
UNAMEC	Cervical artificial discs sales	-	-	-	-	635	615
RHM – MZG	Hospitalisations with CTDR coded in ICD-9-CM (84.62) and discharged in the year	570	479	485	443	-	-
INAMI – RIZIV	Cervical artificial discs notified on the patient bills	-	-	142(*)	548	471	481

(*) June-December.

Based on a survey organized by UNAMEC in November 2014, their members sold respectively 635 and 615 prostheses in 2012 and 2013 (Table 3). This number is higher than the number of the cervical prostheses notified to the INAMI – RIZIV and the procedures recorded in hospital discharges administrative database RHM – MZG.

Obviously, per definition, sales may differ from actually implanted prostheses and there can be different time windows between procedures recorded in the hospital records and number of cervical artificial discs notified on the patient bills. For example, a procedure performed in 2008 can be recorded in RHM – MZG 2009. Nevertheless, figures should be from the same order of magnitude but several other factors may explain those remaining apparent discrepancies.

Concerning the RHM – MZG, the number is probably underestimated. First, between 70 and 100 total disc replacements were coded each year using a nonspecific procedure ICD-9-CM code (84.60 *Insertion of spinal disc prosthesis, not otherwise specified*). Considering that total disc replacements in thoracic region are almost inexistent (± 3 a year), that partial disc replacements are much less frequent (33 cervical and lumbar

prostheses in 2011) and the ratio CTDR/LTDR increases over time (there were 240 lumbar procedures recorded for 2011 which meant 85% more cervical ones), most of those records can be attributed to the cervical region. Second, all hospitalisations during which at least one procedure was performed can be identified, but the number of times a procedure was performed, in the present matter the number of levels treated, is not a reliable variable and cannot be used. Hence, multi-level procedures are always counted once.

The costs of a cervical prosthesis that is implanted without being notified must be borne by the implanting hospital. It can thus not be excluded that some prostheses are implanted but not notified which could theoretically lead to an underestimation of the implanted prostheses. The 2011 high figure could be attributed to the implementation of the new notification procedure for the firms as well as for the hospitals.

In conclusion, we can estimate very roughly the number of cervical prostheses implanted in Belgium to range from 500 to 600 devices in 2013.

2.3.2.2 Characteristics of patients undergoing cervical total disc replacement and comparison with patients undergoing a fusion

Based on the hospitalisations recorded in the RHM – MZG database for the period 2008-2011, 56.8% of the 1977 implanted cases were female cases. On average patients were 45.8 year old (SD: 8.8, median: 46, range: 16-89). Figure 2 shows the age distribution per 5-year category per gender. Compared to three recent studies based on the USA Nationwide Inpatient Sample database using the same ICD-9-CM procedure codes, the proportion of women was higher in our data but the mean age was similar to their findings: 51.6% female cases and mean age of 46.4 year on 2002-2009 for Nandyala et al.,[11] 51.4% and 45.4 year on 2005-2008 for Nesterenko et al.,[25] and 51.9% and 49.5 year on 2004-2007 for Qureshi et al.[25, 26]

All patients were admitted at least one night in the hospital. The mean length of stay amounted to 3.7 days (SD: 3.2, median: 3, range: 1-73). Note that all patients were discharged alive. Length of stay appeared to be longer than what was observed in the USA (1.8,[11] 1.4,[25] and 1.67 days[26]) where surgeons also operate in one day hospitalisation setting.

Figure 2 – Cervical total disc replacement: Age distribution per gender
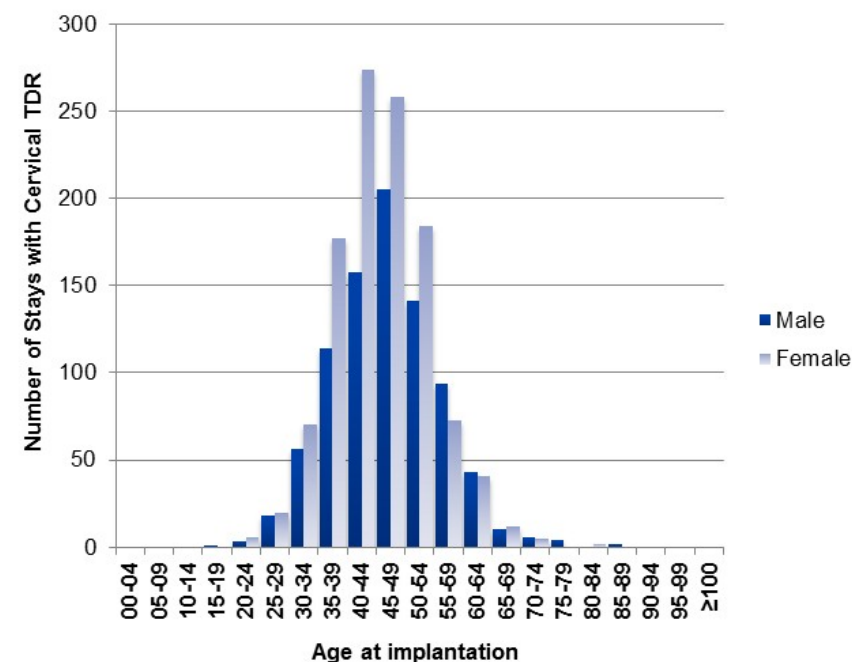
Source: RHM – MZG 2008-2011.

The most frequent principal diagnosis causing the admission was an intervertebral disc disorder (80.2% of the hospitalisations), followed by spondylosis (and allied disorders) (14.2%) or other disorders of the cervical region (3.8%). Inside the intervertebral disc disorder category, most of the cases suffered from a displacement of the cervical disc without myelopathy (50.6% of all hospitalisations). Others presented myelopathy (13%) and 13% had a cervical disc degeneration as principal diagnosis. The top 10 of the principal diagnosis in 3-digit ICD-9-CM codes, as well as the details of the first code split into 5-digit codes can be found in appendix.

We also extracted the data of the patients who underwent a cervical fusion in order to compare their patient characteristics with those of the CTDR group. Because some hospitals also register a code of fusion in case of disc

replacement, we selected the 14 917 cases of fusion for whom no disc replacement was coded at all (neither partial nor total, whatever the location) and with a principal diagnosis of cervical intervertebral disc disorder (degeneration, displacement or other or unspecified disorder of the cervical intervertebral disc: ICD-9-CM codes 722.0, 722.4, 722.71 or 722.91). The 14 917 cases of ACDF included a few more male cases than the CTDR cases (45.7% versus 45.2%; p=0.0418) and were significantly older (49.3 year versus 45.8 year for CTDR group; p<0.0001). Their length of stay was slightly lower than the CTDR cases (3.6 days versus 3.7 days for CTDR; p<0.0001). The three American studies cited above (Nandyala et al., Nesterenko at al., Qureshi et al.) also observed that fusion was performed to a higher age but during a longer hospital stay, associated with a higher comorbidity status. The gender proportion did not differ statistically between groups for none of the studies.[11, 25, 26]

2.3.2.3 Geographic variation of cervical total disc replacement use

Fifty out of the 105 Belgian acute hospitals performed at least one of the 443 CTDRs recorded in 2011. As shown on the bar chart (Figure 3), the number of procedures performed per hospital ranged from 1 to 37, with a yearly median of 4 procedures. The mean number was 4 (SD: 10.7). Most hospitals (n=33, 66%) did not perform 10 procedures, and 15 amongst them (30%) performed only one procedure. CTDRs are not concentrated in academic hospitals nor hospitals with university beds that are depicted in light shade in the bar chart.

Figure 3 – Number of cervical total disc replacement per hospital performing this technique in 2011
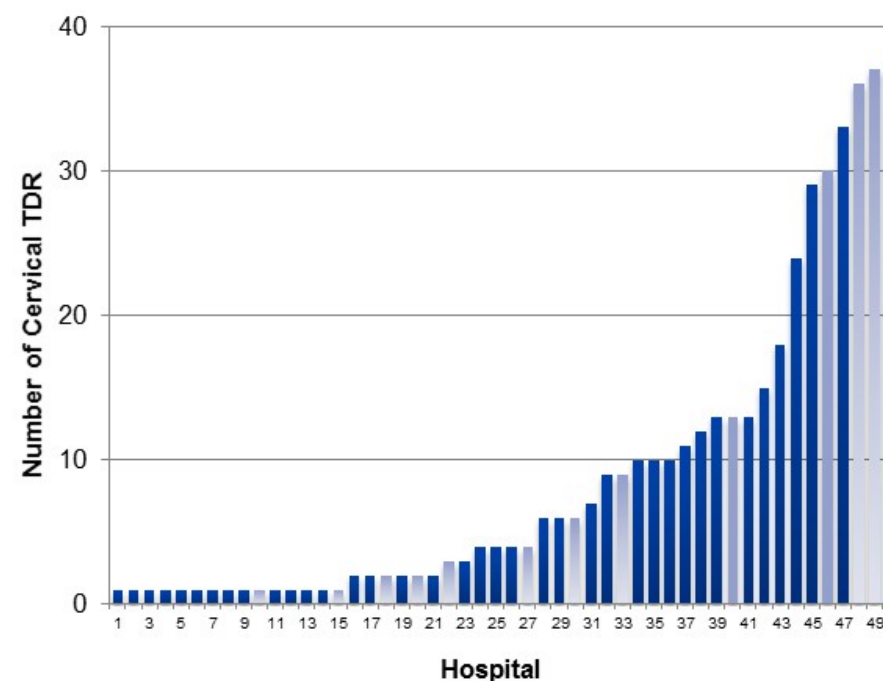
Source: RHM – MZG 2011 – Academic hospitals and hospitals with university beds are in light shade.

Figure 4 presents the number of procedures performed in the hospitals of a district per 100 000 inhabitants. Fifty CTDRs were done in the Capital Region, which represents 4.5 procedures per 100 000 inhabitants. The highest number of procedures reached 32.8 procedures per 100 000 inhabitants in a low-density arrondissement.

Figure 4 – Number of cervical total disc replacements performed in 2011 per hospital district ('arrondissement')
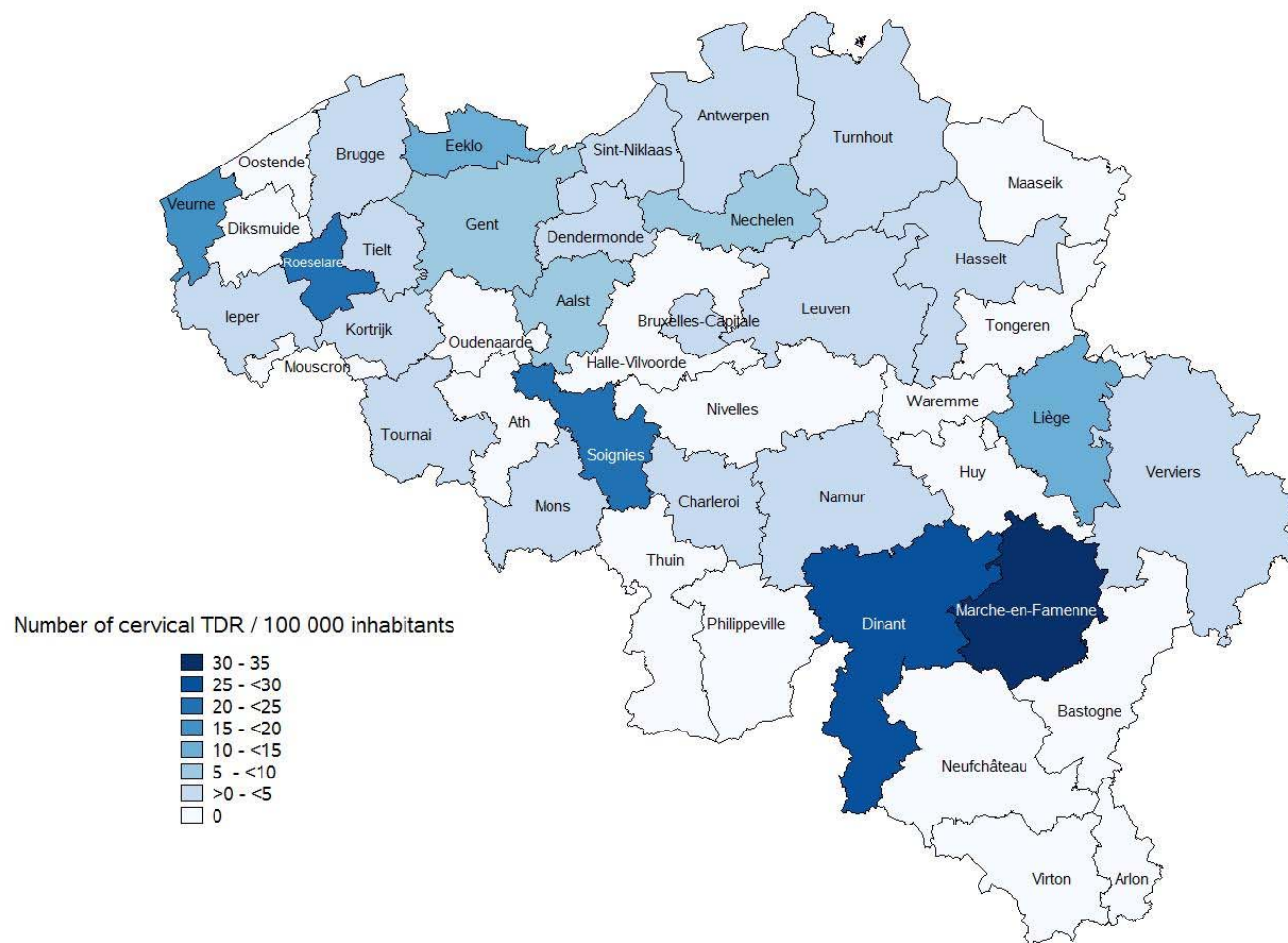
Source: RHM – MZG 2008-2011 - Population DGSIE – ADSEI 1 January 2011.

- **One out of eight Belgian adults declared that he/she suffered from neck disorder or other chronic neck defect in the last 12 months. Radiculopathy and myelopathy, that can be caused by intervertebral disc degeneration, can result in neck and arm pain, as well as numbness in the arm and fingers and neurological dysfunctions.**
- **About 30% of symptomatic patients suffering from cervical degenerative disc disease refractory to conservative treatment might be considered candidates for cervical surgery. Cervical total disc replacement is an alternative to fusion of vertebrae which is commonly used as surgery.**
- **More than a dozen of different cervical prostheses are available in Belgium. The full price of the cervical prosthesis, around €2500, is entirely borne by the patient as it is not reimbursed.**
- **The annual number of cervical prostheses ranges between 500 and 600 discs implanted in 50 Belgian hospitals. Patients are on average 46 years old and 57% of them are female.**

2.4 Clinical effectiveness and safety

HTA CORE MODEL DOMAIN: EFF - SAF

2.4.1 Methods

2.4.1.1 Search Strategy

A systematic search for HTA reports, systematic reviews and randomised controlled trials (RCT) on both cervical and lumbar total disc replacements (TDR) vs. other techniques was carried out in Medline Ovid, Embase and Cochrane (CDSR, DARE, HTA, Economic Evaluations and CENTRAL) on October 9th, 2014. This report is an update of a previous KCE report published in 2006 [27] in which a literature search for both cervical and lumbar total disc replacement was performed until 2006. At this moment, only one RCT on cervical prosthesis was available, and therefore the topic was not included in the report 2006 that focuses only on lumbar TDR (LTDR). The current literature search was consequently limited to publications from 2006 and onwards.

Additionally, we consulted INAHTA and the EUnetHTA POP database for recently published or on-going HTA reports.

The hits from the electronic databases were merged into a unique EndNote file and duplicates were removed. We then screened titles and abstracts to identify and exclude articles which did not fulfil the inclusion criteria (see Table 4). The remaining papers were retrieved and read in full for a final selection of studies to include in the review. No restrictions were imposed for language. While the systematic search for studies were performed on the cervical and lumbar disc replacement combined, the process of selecting potentially relevant articles was performed in two separate "rounds", one round for the cervical review and one round for the lumbar review. Bibliographies of selected studies have been hand-searched for possible relevant references. The flow chart of the selection process for the cervical and lumbar TDR is presented in appendix. For details on the search strategies and article selection process please see appendix. All systematic reviews finally included in our report were critically appraised by using the AMSTAR tool[28] and all RCTs by using the Cochrane grid for risk of bias.[29]

The PICO used for the selection of CTDR literature is presented in Table 4.

Table 5 describes the assessment instruments selected to measure the outcomes (for CTDR and/or LTDR).

Table 4 – PICO table and selection criteria for cervical total disc replacement

Selection criteria	Inclusion criteria	Exclusion criteria
Population	Adult patients (18 years of age or older) with chronic cervical indications, including degenerative disc disease and cervical disc hernia, refractory to conservative treatment for at least 6 weeks. Both patients with single or double-level disease will be included Literature that mixes cervical and lumbar patients will be considered only if outcomes are reported for each patient group separately	Metabolic bone disease (e.g. osteoporosis), malignancy (spinal metastasis, myeloma), haemangioma, osteonecrosis, trauma
Intervention	Cervical total disc replacement, all disc types/brands	Partial disc replacement
Comparators	• Conventional surgery, including discectomy, fusion (with plate, cage, autograft, allograft material) alone or in combination with conservative treatment • Conservative treatment alone	
Outcomes	• Neck related functional status (NDI or similar neck-specific scale) • Arm and neck pain (VAS, NRS or similar scale) • Mobility (ROM) • Neurological outcomes, compared to preoperative status • Patient satisfaction (Odom's outcome criteria or other patient satisfaction scale) • QoL (generic instrument e.g. SF-36, or generic utility instruments e.g. SF-6D, EQ-5D, QWB, HUI) • ASD (adjacent segment degeneration), includes radiographic adjacent segment pathology as well as clinical adjacent segment pathology • Complications, adverse events (wound infections, heterotopic ossification, medical and surgical adverse events e.g. cardiac, respiratory, dysphagia, intraoperative blood loss) • Revision surgery at operated segment/secondary surgery at adjacent segments (reoperation rate) • Mortality	• Working status • Length of stay/ operating time
Study design	HTAs, systematic reviews, randomised controlled trials (including non-inferiority RCTs)	Narrative reviews, observational studies, letters, editorials, notes, abstracts

Table 5 – Description of selected assessment instruments used in included studies

Measure index/ Measure scale	Purpose and description
Functional	
Neck Disability Index (NDI)	The Neck Disability Index (NDI) is a multi-dimensional self-report questionnaire used to determine how neck pain affects a patient's daily life. The NDI consists of ten questions in the following domains: Pain Intensity, Personal Care, Lifting, Reading, Headaches, Concentration, Work, Driving, Sleeping, and Recreation. Scoring: each question contains six answer choices, scored from 0 (no disability) to 5 (complete disability). All section scores are then totalled. Scoring is reported on a 0-50 scale, with 0 being the best possible score and 50 being the worst. Alternately, the score can be reported from 0-100. The score is often reported as a percentage (0-100%).
Oswestry Disability Questionnaire	The Oswestry Disability Index is a low back functional outcome tool used to measure a patient's permanent functional disability. The index consists of 10 sections. Each section consists of 6 possible answers that are translated into numerical values (0-6) with 0 being the best possible answer. Subsequently these are multiplied to percentage scores. If the whole index is used a score from 0-20% represents minimal disability, 21-40% moderate disability, 41-60% severe disability, 61-80% crippled and 81-100% bed-bound (or exaggerating their symptoms).
Roland-Morris Disability Questionnaire	The Roland Morris Disability Questionnaire (RMDQ) is a health status measure for low back pain. The RMDQ consists of 24 items each which has to be checked (or not checked) by the patient. Each statement checked adds one point to the total score (with the denominator kept at 24).
Range of Motion (ROM)	Range of motion is a unidimensional measure of mobility commonly on flexion extension x-rays, measured as segmental range of motion in the sagittal plane (sagittal ROM are often written as sROM), at target segment and adjacent segments. Scoring: range of motion is measured in degrees.
Odom's outcome criteria	Patient satisfaction measured according to Odom's outcome criteria is a unidimensional four-level assessment of success of surgery in relieving preoperative symptoms. Scoring: • Excellent outcome: all preoperative symptoms relieved. Abnormal findings improved. • Good outcome: minimal persistence of preoperative symptoms. Abnormal findings unchanged or improved. • Fair outcome: definite relief of some preoperative symptoms. Other symptoms unchanged or slightly improved. • Poor outcome: symptoms and signs unchanged or worse.
"Neurological success"	"Neurological success" was in the studies determined by measuring three objective clinical findings: motor function, sensory function, and reflex activity, and that overall neurological status success is determined by maintenance or improvement in all three clinical findings. It was noted by the Belgian experts that this outcome measure is primarily developed for study purposes and is not very relevant in clinical practice.
Kellgren-Lawrence Grading Scale	Grade 1: doubtful narrowing of joint space and possible osteophytic lipping Grade 2: definite osteophytes, definite narrowing of joint space Grade 3: moderate multiple osteophytes, definite narrowing of joints space, some sclerosis and possible deformity of bone contour Grade 4: large osteophytes, marked narrowing of joint space, severe sclerosis and definite deformity of bone contour

Measure index/ Measure scale	Purpose and description
Pain	
Visual Analogue Scale (VAS) for pain	The pain VAS is a unidimensional measure of pain intensity which has been widely used in diverse adult populations. The pain VAS is a continuous scale comprised of a horizontal (HVAS) or vertical (VVAS) axis, usually 10 centimetres (100 mm) in length, anchored by 2 verbal descriptors, one for each symptom extreme. For pain intensity, the scale is most commonly anchored by "no pain" (score of 0) and "pain as bad as it could be" or "worst imaginable pain" (score of 100 [100 mm scale]). Scoring: Using a ruler, the score is determined by measuring the distance (mm) on the 10 cm line between the "no pain" anchor and the patient's mark, providing a range of scores from 0–100.
Numeric Rating Scale (NRS) for pain	The NRS for pain is a unidimensional measure of pain intensity in adults, including those with chronic pain due to rheumatic diseases. Various iterations exist; however, the most commonly used is the 11-item NRS. The NRS is a segmented numeric version of the visual analogue scale (VAS) in which a respondent selects a whole number (0–10 integers) that best reflects the intensity of their pain. The common format is a horizontal bar or line. Similar to the pain VAS, the NRS is anchored by terms describing pain severity extremes. Scoring: The number that the respondent indicates on the scale to rate his pain intensity is recorded. Scores range from 0–10 with higher scores indicating greater pain intensity.
Health Related Quality of Life (HRQoL)	
Generic Medical Outcomes Study Short Form 36 (SF-36)	The SF-36 is a multi-dimensional questionnaire designed to measure self-reported HRQoL. It contains 36 questions. The SF-36 assesses eight health dimensions: physical functioning, role limitations owing to physical health, bodily pain, general health perceptions, vitality, social functioning, role limitations owing to emotional health, and mental limitations in physical activities because of health problems. Scoring: The SF-36 consists of eight scaled scores, which are the weighted sums of the questions in their section. Each scale is directly transformed into a 0-100 scale on the assumption that each question carries equal weight. The lower the score the more disability, the higher the score the less disability, i.e. a score of zero is equivalent to maximum disability and a score of 100 is equivalent to no disability. SF-36 PCS represents the SF-36 Physical Component summary Score, which is a summary of the four scaled scores representing physical domains (0-100). SF-36 MCS represents the SF-36 Mental Component summary Score, which is a summary of the four scaled scores representing mental domains (0-100).
Generic Medical Outcomes Study Short Form 12 (SF-12)	The SF-12 is a multi-dimensional questionnaire designed to measure self-reported HRQoL. It contains 12 questions. The SF-12 contains 12 questions and assesses eight health dimensions; physical functioning, physical role, bodily pain, general health, vitality, social functioning, emotional role and mental health.
Generic EuroQol 5 dimensions (EQ-5D)	The EQ-5D is a multi-dimensional questionnaire designed to measure self-reported HRQoL. It contains 5 questions. The EQ-5D assesses five health dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is assessed by a single question on a three-point ordinal scale: no problems, some problems, extreme problems. Scoring: The EQ-5D is combined and presented as a quasi-continuous outcome on a scale of 0-1, with 0 representing death and 1 'full health'. Sometimes negative numbers are used to represent health states valued as worse than death.

Determining clinical relevance

To determine clinical relevance of the results of the meta-analysis we used the pooled effect sizes as described in the review by Boselie.[2] The difference in effect size between the two treatments for each outcome at each time point was scored as a 'small', 'medium', or 'large' mean difference (MD), and were defined as follows:

- Small: MD < 10% of the scale, standardised mean difference (SMD) < 0.4, or as a risk ratio (RR) > 0.8 (or < 1.25, depending on the risk/benefit being for the investigational or the control group).
- Medium: MD 10% to 20% of the scale, SMD from 0.4 to 0.7, RR from 1.25 to 2.0 or from 0.5 to 0.8.
- Large: MD > 20% of the scale, SMD > 0.7, RR < 0.5 or > 2.0.

The majority of outcomes were suitable for a meta-analysis. Furthermore, no minimally clinically important difference (MCID) is established for cervical radicular pain in the literature which is one of the most important outcomes for the present comparison, and the expert group therefore agreed to utilize the above mentioned thresholds. However, it was noted by the experts that there are recent attempts to assess the MCID in pain, disability and quality of life after fusion.[30] No MCID was defined for the outcomes such as mobility or 'neurological success' which were deemed not clinically relevant by the experts (see 'neurological success' in Table 5).

2.4.2 Results on Clinical Effectiveness

This chapter presents a review of the evidence on the clinical benefits of cervical total disc replacement (CTDR) versus fusion alone or fusion in combination with conservative treatment or versus conservative treatment alone.

One relatively recent (search up to May 25th, 2011), very comprehensive and high quality Cochrane systematic review by Boselie et al.[2] was identified and included early in the selection process.[a] This systematic review included nine RCTs by Coric (2011), Heller (2009), Kelly (2011), Marzluff (2010), McAfee (2010), Mummaneni (2007), Nabhan (2007), Pettine (2010) and Porchet (2004). However, the Cochrane review only reported short-term (up to and including 3 months) and medium-term (one to two years) follow-ups and was restricted to single level CTDR compared to single level fusion. It was therefore important to assess whether we could identify additional systematic reviews with:

- A more recent search date (in particular if the systematic review had included primary studies that were not already included in the Cochrane review).
- A longer follow-up period.
- An inclusion of multi-level CTDR studies.
- A comparison between CTDR and prolonged conservative treatment (> 6 weeks).

Several more recent systematic reviews were identified. However, after a quality assessment it was decided to retain only one additional systematic review:

- A review by Ren[31] with a search date up to March 2013, because this review reports on trials with a minimum of 48 months follow-up, and has a large number of outcomes identified as important for this review.

In this process we also considered a review by Luo[32] because this review has a very recent search date (April 2014) and therefore includes RCTs that are not included in other systematic reviews. However, there were considerable inconsistencies in the numbers reported in some of the meta-analyses within this review to an extent where it was deemed necessary to exclude this paper.

The two systematic reviews retained[2, 31] both reported results for clinical effectiveness but at different time-points.

---

a   On 20 May 2015, this systematic review was withdrawn due to non-compliance with The Cochrane Collaboration's Commercial Sponsorship Policy. The impact of this decision on our own work can be considered as minor, since no publication bias was identified and because we also included an additional systematic review with longer follow-up period (i.e. 4 years; Ren 2014) and recent RCTs published after the search date of Boselie. Results reported by original RCTs included in both systematic reviews (Boselie, 2012; Ren 2014) and by recent RCTs were pooled for each outcome under study. Our conclusions are based on these meta-analyses performed by the author of this chapter (KH).

Results from the two systematic reviews are presented below by outcome (domain) and subsequently by time-period (Cochrane review at outcomes after 3 months and at 12-24 months, and the Ren review on outcomes after ≥ 48 months).

Please note that the review by Ren has reported mean difference results favouring CTDR as positive values in all their meta-analyses (e.g. a lower score in the CTDR group on the neck disability index is reported as a positive mean difference favouring CTDR, whereas the Cochrane review reports a similar mean difference favouring CTDR as a negative mean difference). For comprehensibility and readability we decided to report results in a similar way (practically this means that the result signs from the Ren review are reversed for outcomes such as functional status and pain).

In cases where p-values were not provided but could be calculated or in cases where the statistics used to calculate p-values were doubtful, we used the Fischer Exact test to calculate the p-values ourselves.

After the first experts meeting it was requested by some of the experts to add some more information on ASD, because they believe this is a very important outcome, and because the literature we initially included did not have this outcome as their main focus. We therefore searched for SRs specifically on this topic and added results from the most recent meta-analysis by Verma[33] looking at the rate of ASD related surgery at 2 to 5 years to complement the information from the other SRs.

Additionally, we searched for all randomised controlled trials that were not included in the retained systematic reviews with an aim to:

- update the results from the systematic reviews,
- search for RCTs on multi-level disease as well as RCTs comparing CTDR with prolonged conservative treatment.

This process resulted in the selection of the following RCTs:

Single level surgery (3 additional RCTs)

- An RCT by Vaccaro et al. from 2013[34] providing up to two-year results on clinical outcomes with selective constrained SECURE-C Cervical Disc TDR versus fusion. The trial included 380 patients. Unfortunately the trial had a non-inferiority design and the presentation of data and statistical methods used for assessing superiority could not be used to update meta-analysis (no MD, SD etc.).

- An RCT by Phillips et al. from 2013[35] with 416 patients providing up to two-year results on effectiveness and safety for Porous Coated Motion (PCM) Cervical Disc Arthroplasty versus fusion.

- An RCT by Zhang from 2012[36] with 120 patients providing up to two-year results on outcomes comparing Bryan Cervical Disc Arthroplasty with fusion.

We also considered but finally not retained 2 other RCTs:

- An RCT by Cheng.[37] However, this small RCT with a total of 83 patients was of low quality, mixed single level surgery patients with multi-level (2 and 3 levels) patients without reporting outcomes for the two groups separately, and we therefore decided not to include this trial.

- An RCT by Skeppholm from 2013[38] with 136 patients providing up to two-year results on a comparison of dysphagia between cervical disc replacement (with the Discover Artificial Disc) and fusion. Similarly, this RCT included data from both single-level and two-level surgery patients without properly reporting outcomes for the two groups separately, and we therefore also decided not to include this trial.

Multiple level surgery (2 additional RCTs)

- An RCT by Davis et al.[39] with 330 patients comparing CTDR (with the Mobi-C disc) with fusion for 2-level symptomatic disc disease (C3-C7). Patients were randomised in a 2:1 ratio (CTDR patients to fusion patients). The trial had a follow-up at 6 weeks, 3, 6, 12, 18, 24 and 48 months. The results after 3 months, 12 months, 24 months and 48 months are presented by outcome below.

- An RCT by Cheng et al.[40] with 65 patients suffering from two-level cervical disc disease (C3-4 to C6-7) comparing CTDR (with the Bryan Cervical Disc Replacement) with fusion. The trial had a follow-up at 1 week, 3 weeks, 6 weeks, 3 months, 6 months, 12 months and 24 months after surgery. However, statistical group differences were only reported for functional outcomes after 12 and 24 months. The results are presented by outcome below.

Total disc replacement versus prolonged conservative treatment

- No RCTs were identified for this comparison.

Additional meta-analyses

For the meta-analyses, we used fixed effects meta-analysis except for the range of motion (ROM) mean difference at 12-24 months (CTDR versus fusion), where a random effects meta-analysis was preferred due to the high heterogeneity between the studies (I²=95%).

Quality of the evidence

Boselie et al. provided a GRADE score by outcome: the quality of evidence was considered to be very low (patient satisfaction), low (arm pain, global mental score SF-36 MCS) or moderate (neck pain, functional status, neurological success and global physical score SF-36 PCS).[2] Due either to the non-blinded nature of the supplementary RCTs, their small size and/or the incomplete outcome data for some of them we judged the quality of the supplementary body of evidence as low for CTDR versus fusion. Consequently, we did not provide a GRADE score for each outcome separately in the scope of this rapid review. Besides, if there are very severe problems for any factor of the GRADE scoring tool, RCT evidence may fall by two levels due to that factor alone anyway.[41, 42]

2.4.2.1 Results single-level disease

Functional Status

Both systematic reviews and the additional RCTs reported functional status using the Neck Disability Index (NDI), using a 0-100 scale (lower is better).

3 months

The Cochrane systematic review found a <u>significant difference</u> between CTDR and fusion in favour of CTDR (MD -5.14; 95% CI -6.94 to -3.34). Clinical relevance was low, since the pooled effect size was small (< 10% of the scale).

12-24 months

The Cochrane systematic review found a <u>significant difference</u> between CTDR and fusion in favour of CTDR (MD -2.79; 95% CI -4.73 to -0.85). Clinical relevance was low, since the pooled effect size was small (< 10% of the scale).

For this outcome and time-point it was possible to update this meta-analysis with data from one RCT by Zhang.[36] As displayed in Figure 5 the effect remained with a significant difference between CTDR and fusion in favour of CTDR (MD -1.89; 95% CI -3.44 to - 0.35) but with low clinical relevance.

Figure 5 – Functional status: 12-24 months NDI mean difference (cervical total disc replacement versus fusion)

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Coric 2011	22.1	21.6	119	22.1	20.7	115	8.1%	0.00 [-5.42, 5.42]
Heller 2009	15.1	16.5	234	18.8	19.2	197	20.4%	-3.70 [-7.11, -0.29]
Marzluff 2010	13.3	17.11	140	15	17.16	124	13.8%	-1.70 [-5.84, 2.44]
Mummaneni 2007	20.6	20.9	263	23.4	21.3	223	16.7%	-2.80 [-6.57, 0.97]
Pettine 2010	22.6	19.2	28	34.7	19.2	25	2.2%	-12.10 [-22.45, -1.75]
Porchet 2004	17.2	18.7	19	19.2	16.8	18	1.8%	-2.00 [-13.44, 9.44]
Zhang 2012	14.89	5.8	56	15.25	7.54	53	36.9%	-0.36 [-2.90, 2.18]
Total (95% CI)			**859**			**755**	**100.0%**	**-1.89 [-3.44, -0.35]**

Heterogeneity: Chi² = 6.91, df = 6 (P = 0.33); I² = 13%
Test for overall effect: Z = 2.41 (P = 0.02)

Two additional recent RCTs by Philips and Vaccaro did not provide mean and SD values and could therefore not be incorporated in the update of the meta-analysis. However, the RCT by Philips[35] with a total of 416 patients found that NDI scores significantly favoured CTDR with a mean change in CTDR group of 21.8 vs. 25.5 in the fusion group (p=0.029). The other RCT by Vaccaro[34] with 380 patients did not find superiority for one group vs. another according to their protocol-specified success criteria (≥ 25% improvement in NDI scores).

≥48 months

The systematic review by Ren found a significant difference between CTDR and fusion in favour of CTDR (MD 5.49; 95% CI 2.79 to 8.20; p<0.0001). Clinical relevance was low, since the pooled effect size was small (< 10% of the scale).

Pain
Both systematic reviews reported on arm pain and neck pain separately. The Cochrane review reported on results from the Visual Analogue Scale (VAS) or the Numeric Rating Scale (NRS) both measured on a 0-100 scale

and combined results from the two scales in the meta-analysis. The review by Ren reported for the VAS scale only. The additional RCTs all used the VAS scale for reporting pain.

Arm Pain - 3 months

The Cochrane review reported a significant difference between CTDR and fusion in favour of CTDR (MD -2.18; 95% CI -3.68 to -0.68). Clinical relevance was low, since the pooled effect size was small (< 10% of the scale).

Arm Pain - 12-24 months

Significant difference between CTDR and fusion in favour of CTDR (MD - 1.54; 95% CI -2.86 to -0.22). Clinical relevance was low, since the pooled difference in effect size was small (< 10% of the scale). For this outcome and time-point it was possible to update this meta-analysis with data from one RCT by Zhang.[36] As displayed in Figure 6 the effect remained with a significant difference between CTDR and fusion in favour of CTDR (MD - 1.48; 95% CI -2.71 to -0.26).

Figure 6 – Arm Pain: 12-24 months NRS/VAS mean difference (cervical total disc replacement versus fusion)
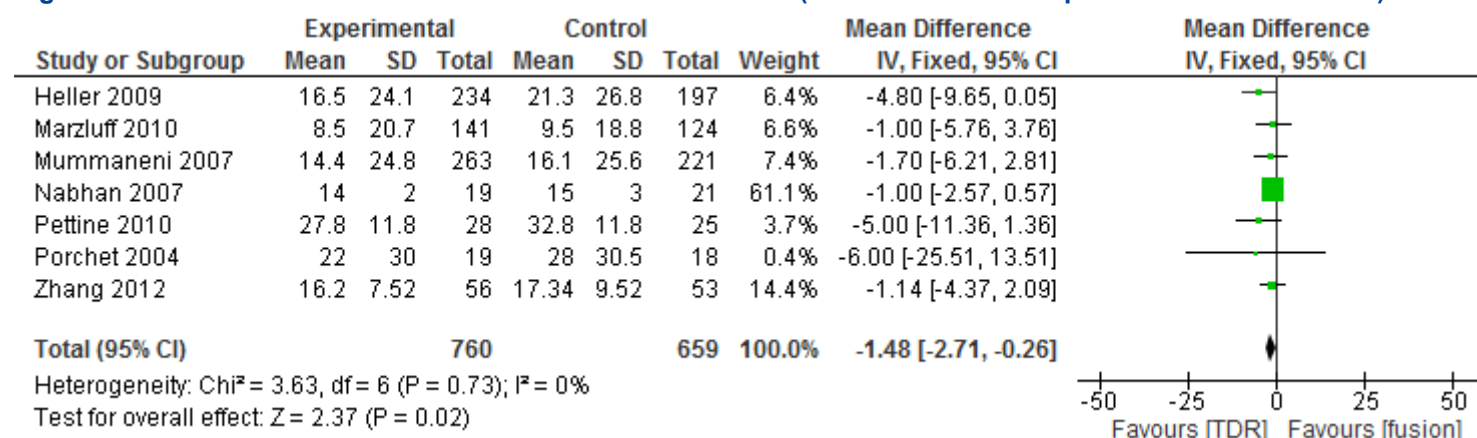
Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Heller 2009	16.5	24.1	234	21.3	26.8	197	6.4%	-4.80 [-9.65, 0.05]
Marzluff 2010	8.5	20.7	141	9.5	18.8	124	6.6%	-1.00 [-5.76, 3.76]
Mummaneni 2007	14.4	24.8	263	16.1	25.6	221	7.4%	-1.70 [-6.21, 2.81]
Nabhan 2007	14	2	19	15	3	21	61.1%	-1.00 [-2.57, 0.57]
Pettine 2010	27.8	11.8	28	32.8	11.8	25	3.7%	-5.00 [-11.36, 1.36]
Porchet 2004	22	30	19	28	30.5	18	0.4%	-6.00 [-25.51, 13.51]
Zhang 2012	16.2	7.52	56	17.34	9.52	53	14.4%	-1.14 [-4.37, 2.09]
Total (95% CI)			**760**			**659**	**100.0%**	**-1.48 [-2.71, -0.26]**

Heterogeneity: Chi² = 3.63, df = 6 (P = 0.73); I² = 0%
Test for overall effect: Z = 2.37 (P = 0.02)

Favours [TDR]   Favours [fusion]

Two additional recent RCTs did not provide mean and SD values and could therefore not be incorporated in the update of the meta-analysis. One RCT by Philips[35] with a total of 416 patients did not find any significant between groups difference for arm pain on the VAS scale. Another RCT by Vaccaro[34] with a total of 380 patients did not find any significant between groups difference according to their protocol-specified success criteria (VAS ≥ 20mm improvement).

Arm Pain - ≥48 months

The Ren review found a <u>significant difference</u> between CTDR and fusion in favour of CTDR (MD -9.19; 95% CI -6.57 to -11.81; p<0.00001). Clinical relevance was low, since the pooled difference in effect size was small (< 10% of the scale).

Neck Pain - 3 months

<u>No significant difference</u> between CTDR and fusion (MD -3.67; 95% CI -9.80 to 2.46).

Neck Pain - 12-24 months

<u>Significant difference</u> between CTDR and fusion in favour of CTDR (MD -3.12; 95% CI -4.69 to -1.28). Clinical relevance was low, since the pooled difference in effect size was small (< 10% of the scale).

For this outcome and time-point it was possible to update the meta-analysis with data from one RCT by Zhang.[36] As displayed in Figure 7 the effect remained with a significant difference between CTDR and fusion in favour of CTDR (MD -2.97; 95% CI -4.62 to -1.32).

Figure 7 – Neck Pain: 12-24 months NRS/VAS mean difference (cervical total disc replacement versus fusion)
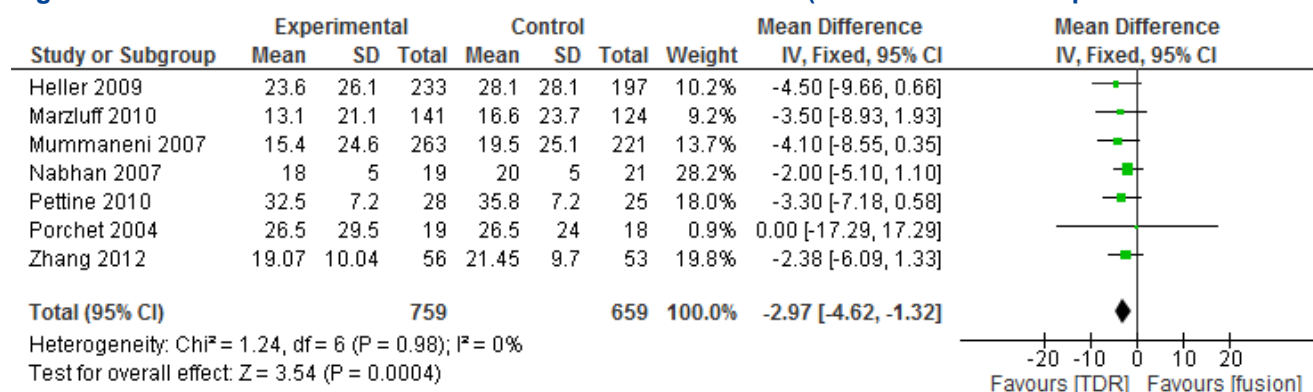
Two additional recent RCTs did not provide mean and SD values and could therefore not be incorporated in the update of the meta-analysis. One RCT by Philips with a total of 416 patients did not find significant between group differences for neck pain on the VAS scale (p=0.063).[35] The RCT by Vaccaro[34] with a total of 380 patients found a significant difference between groups according to their protocol-specified success criteria (VAS ≥ 20mm improvement). In the CTDR group 81.2% of patients experiences this level of improvement versus 72.2% in the fusion group.

Neck Pain - ≥48 months

Significant difference between CTDR and fusion in favour of CTDR (MD 5.42; 95% CI 0.21 to 10.63; p=0.04). Clinical relevance was low, since the pooled difference in effect size was small (< 10% of the scale).

Mobility

The Cochrane review reported on mobility at index level (the cervical level where the disc was removed) and at adjacent levels (upper and lower adjacent levels). The review by Ren solely reported on mobility at index level. In both reviews mobility was measured in degrees using range of motion (ROM) for flexion-extension.

3 months

Index level:

Significant difference between CTDR and fusion in favour of CTDR (MD 4.75°; 95% CI 4.45° to 5.06°).

Upper adjacent level:

Significant difference, with a slightly higher sagittal ROM in the CTDR group (MD 0.69°; 95% CI 0.16° to 1.21°).

Lower adjacent level:

No significant difference between the two groups (MD -0.37°; 95% CI -1.04° to 0.29°).

12-24 months

Index level:

Significant difference between CTDR and fusion in favour of CTDR (MD 6.90°; 95% CI 5.45° to 8.35°; p<0.00001).

Update with more recent RCTs:
For this outcome and time-point it was possible to update the meta-analysis from Cochrane with information from two more recent RCTs by Philips and Zhang.[35, 36] adding data from 451 patients. As displayed in Figure 8 the effect remained with a significant difference between CTDR and fusion in favour of CTDR (MD 7.01°; 95% CI 6.04° to 7.98°).

Figure 8 – Mobility: 12-24 months ROM mean difference (cervical total disc replacement versus fusion)
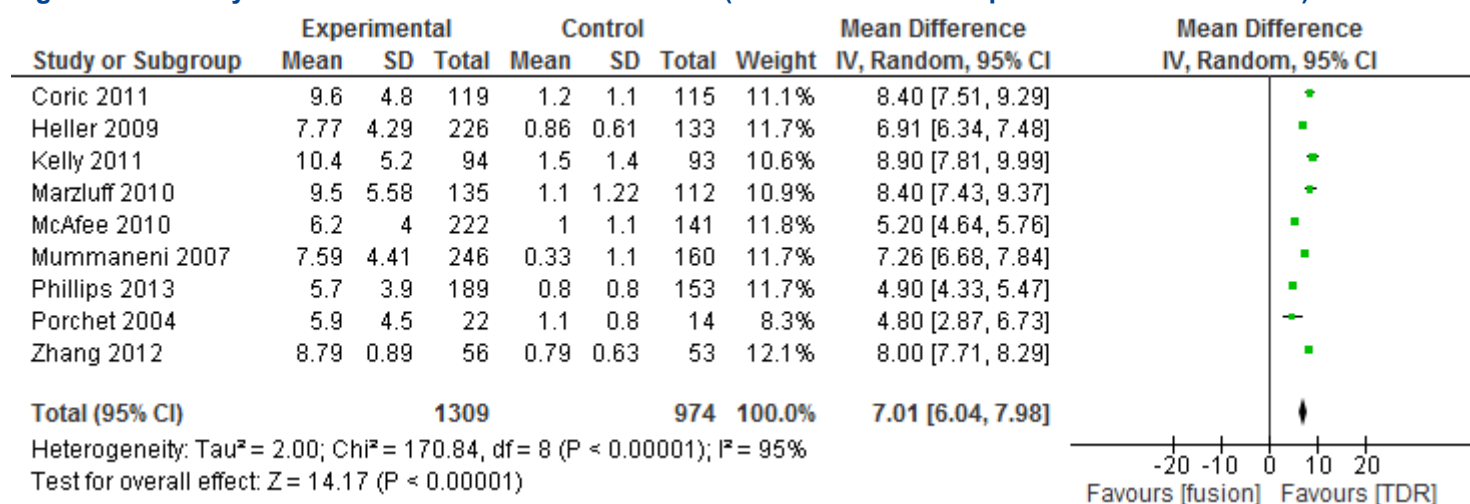
Additionally, an RCT by Vaccaro was identified.[34] This RCT did not provide SD values and could therefore not be incorporated in the meta-analysis. However, the results were in the same direction with a mean flexion-extension in the CTDR group of 9.7° at 24 months.

Upper adjacent level:

Borderline significant difference between CTDR and fusion in favour of CTDR (MD 0.53°; 95% CI 0.03° to 1.03°).

Lower adjacent level:

No significant difference between the two groups (MD -0.81°; 95% CI -1.99° to 0.36°).

≥48 months

Index level:

The review by Ren included four studies that reported the mean flexion–extension ROM at the index level, but the SD could not be calculated. In each study, the ROM was significantly higher in patients who underwent CTDR than in those who underwent fusion.

Coric et al.[43] reported a mean ROM of 8.6° in patients who underwent CTDR and 0.2° in patients who underwent fusion after 5 years.

Zigler et al.[44] reported a mean ROM of 9.42° in patients who underwent CTDR and 1.02° in patients who underwent fusion after 5 years.

Sasso et al.[45] reported a mean ROM of 8.48° in patients who underwent CTDR and a restricted (no degrees stated) ROM in patients who underwent fusion after 4 years.

Burkus et al.[46] reported a mean ROM of 9.42° in patients who underwent CTDR and a restricted (no degrees stated) ROM in patients who underwent fusion after 5 years.

Neurological outcomes

The Cochrane review defined neurological success as percentage of participants with unchanged or improved neurological status and reported this for 3 months and 12-24 months. The review by Ren reported "neurological success rates" after 48 months but did not provide further definition of this term in the review.

3 months

The difference between the two treatment groups was borderline <u>not</u> <u>significant</u> (RR 1.05; 95% CI: 0.99 to 1.12).

12-24 months

Significant difference between the two treatment groups in favour of CTDR (RR 1.05; 95% CI 1.01 to 1.09). Clinical relevance was low, since the pooled difference in effect size was small (< 10% of the scale). For this outcome and time-point it was possible to update the meta-analysis from Cochrane with information from two more recent RCTs by Philips and Vaccaro measuring neurological status in the same way.[34, 35] This added data from 632 patients. As displayed in Figure 9 the group difference remained borderline significant but with low clinical relevance (RR 1.05; 95% CI 1.02 to 1.07).

≥48 months

The neurological success rate was 93.2% in the CTDR group (3 studies with a total of 396 patients CTDR patients) and 89.9% in the fusion group (326 patients in total) but this difference was not significant (OR 1.54, 95% CI 0.91 to 2.63; p=0.11).

Figure 9 – Neurological outcomes: 12-24 months RR (cervical total disc replacement versus fusion)
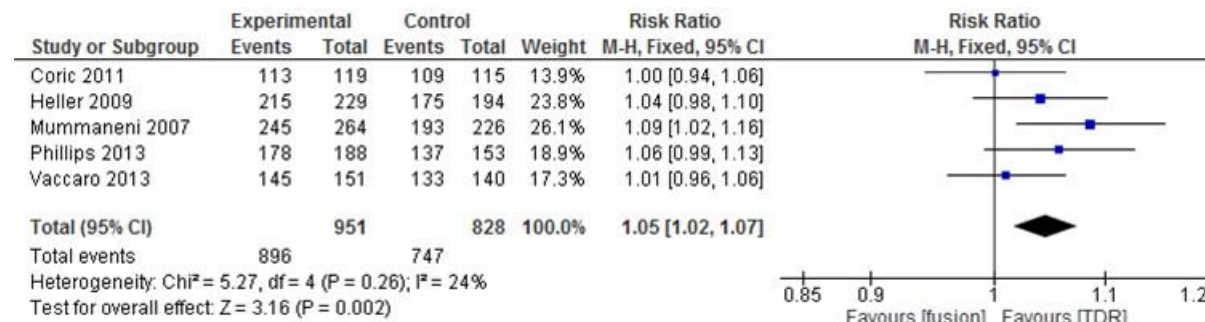
Patient Satisfaction

Only the Cochrane review[2] reported on patient satisfaction and this review only found evidence for the outcome at 12-24 months. Patient Satisfaction was measured according to Odom's outcome criteria which is a four-level assessment of success of surgery in relieving preoperative symptoms. Additionally, two RCTs reported patient satisfaction, one RCT by Philips[35] measured patient satisfaction with Odom's outcome criteria and one RCT by Vaccaro measured the percentage of patients in each group who were "definitely" or "mostly satisfied" with surgery results.[34]

3 months

No evidence found.

12-24 months

The Cochrane review found no significant difference between CTDR and fusion (RR 1.06; 95% CI 1.00 to 1.12). An RCT by Philips[35] found that 91.5% of patients in the CTDR groups reporting "excellent" or "good" outcomes according to Odom's criteria vs. 86.3% in the fusion group. Another RCT by Vaccaro[34] found superiority (significant difference) for CTDR in terms of patient satisfaction after 24 months according to their success criteria (in SECURE-C group 99.7% were "definitely" or "mostly" satisfied with the surgery results compared with 85.2% in the fusion group).

≥48 months

No evidence found.

Quality of Life

Both reviews[2, 31] reported on quality of life using SF-36 scales. The Cochrane review reported results for both SF-36 PCS and SF-36 MCS, whereas the Ren review only reported on the SF-36 PCS scale.

3 months

SF-36 PCS: Significant difference between the two treatment groups in favour of CTDR (MD 2.40; 95% CI 0.55 to 4.25).

SF-36 MCS: No significant difference between the two treatment groups (MD 1.80; 95% CI -0.10 to 3.70).

12-24 months

SF-36 PCS: Significant difference between the two treatment groups in favour of CTDR (MD 2.10; 95% CI 0.68 to 3.51).

SF-36 MCS: Significant difference between the two treatment groups in favour of CTDR (MD 1.46; 95% CI 0.10 to 2.82).

≥48 months

SF-36 PCS: Results from two studies were pooled and showed a significant difference between the two treatment groups in favour of CTDR (MD 1.91; 95 % CI 0.94 to 2.89; p=0.0001).

Note: For all 1-level quality of life results clinical relevance was low, since the pooled effect size was small (< 10% of the scale).

2.4.2.2 Results two-level disease

Functional Status

The RCT by Davis[39] and the RCT by Cheng[40] both reported functional status using the Neck Disability Index (NDI).

3 months

In the RCT by Davis no absolute values are provided but bar graphs state a significant between groups difference at 3 months (p<0.05, however, no exact p-value was provided). The Cheng study did not report statistical differences for this time-point.

12 months

In the RCT by Davis no absolute values are provided but bar graphs state a significant between groups difference at 12 months (p<0.05, however, no exact p-value was provided). Similarly, the RCT by Cheng found a significant difference in favour of CTDR (12 point average in CTDR group vs. 18 point average in the fusion group; p=0.030).

24 months

In the RCT by Davis the NDI scores at 24 months favoured CTDR with 37 (SD=20) in the CTDR group and 30 (SD=19) in the fusion group. The difference from baseline between the two treatments was significant (p<0.05, however, no exact p-value was provided). Similarly, the RCT by Cheng found a significant difference in favour of CTDR (11 point average in CTDR group vs. 19 point average in the fusion group; p=0.023).

48 months

In Davis the NDI scores at 48 months favoured CTDR with 36.5 (SD=21.3) in the CTDR group and 28.5 (SD=18.3) in the fusion group. The difference from baseline between the two treatments was <u>significant</u> (p=0.0048, using the unpaired t-test). The Cheng study did not report statistical differences for this time-point.

<u>Pain</u>

Both the RCT by Davis[39] and the RCT by Cheng[40] reported on arm pain and neck pain separately using VAS Pain Scores.

Arm Pain - 3 months

In Davis no absolute values are provided but bar graphs state a non-significant between groups difference at 3 months (p>0.05, however, no exact p-value was provided). The Cheng study did not report statistical differences for this time-point.

Arm Pain - 12 months

In Davis no absolute values are provided but bar graphs state a non-significant between groups difference at 12 months (p>0.05, however, no exact p-value was provided). The Cheng study did not report statistical differences for this time-point.

Arm Pain - 24 months

Davis reported that improvements in arm pain from baseline at 24-months were 35 (SD=29) in the CTDR group and 34 (SD=38) in the fusion group (p>0.05, however, no exact p-value was provided). On the contrary, the smaller Cheng study found a <u>significant difference</u> in favour of CTDR (1.4 point average vs. 2.7 point average; p=0.013).

Arm Pain - 48 months

The mean improvement in VAS arm pain score from baseline was similar between groups with 56 (SD=31) for CTDR patients and 53 (SD=31) for fusion patients (p>0.05, however, no exact p-value was provided). The Cheng study did not report statistical differences for this time-point.

Neck pain - 3 months

No absolute values are provided but bar graphs state a <u>significant</u> between groups difference at 3 months (p<0.05, however, no exact p-value was provided). The Cheng study did not report statistical differences for this time-point.

Neck pain - 12 months

No absolute values are provided but bar graphs state a non-significant between groups difference at 12 months (p>0.05, however, no exact p-value was provided). The Cheng study did not report statistical differences for this time-point.

Neck pain - 24 months

Improvements in neck pain from baseline at 24-months were 54 (SD=25) in the CTDR group and 53 (SD=29) in the fusion group (p>0.05, however, no exact p-value was provided). On the contrary, the small Cheng study found a significant difference in favour of CTDR (1.5 point average vs 2.6 point average; p=0.012).

Neck pain - 48 months

The mean improvement in VAS neck pain score from baseline at 48 months was 53 (SD=30) for the CTDR group and 48 (SD=29) for the fusion group (p>0.05, however, no exact p-value was provided). The Cheng study did not report statistical differences for this time-point.

<u>Mobility</u>

In the included studies range of motion, measured in degrees, for each treated level was calculated quantitatively from lateral flexion/extension radiographs and anteroposterior right/left lateral bending radiographs.

3 months

Numbers are not provided in the papers.

12 months

Numbers are not provided in the papers.

24 months

The Davis study[39] found that in the fusion group the mean ROM values at 24 months were less than 1 degree for both treated segments in both lateral flexion/extension and lateral bending.

In the CTDR group mean flexion/extension and lateral bending was maintained from baseline throughout study duration. At 24 months mean ROM was 10.1° (SD=5.9°) in flexion/extension and 5.6° (SD=3.3°) at the superior treated level. For the inferior treated level the ROM values were 8.3° (SD=5.3°) in flexion/extension and 5.4° (SD=3.3°) in lateral bending.

The Cheng study[40] found that average flexion-extension in the CTDR group was 7.9° and in the fusion group 0.5°, but no between group statistics are provided.

48 months

In the Davis study[39] on average the CTDR group maintained their baseline flexion/extension and lateral bending after 48 months without device failures observed. ROM in degrees are not provided in the paper for the fusion group but it can be assumed that they did not differ from the results at 24 months. The Cheng study did not report mobility for this time-point.

Neurological outcomes

The RCT by Davis[39] defined "neurological success" as the absence of significant neurological deterioration determined by investigator conducted evaluation that included motor assessment of muscle strength, sensory assessments and reflex assessments. This outcome was not assessed by Cheng et al.[40]

3 months

Numbers are not provided in the paper.

12 months

Numbers are not provided in the paper.

24 months

Significant difference between the two treatment groups in favour of CTDR with 5.6% of patients showing neurological deterioration in the CTDR group and 6.7% showing neurological deterioration in the fusion group (authors use the Farrington-Manning test to compare frequencies between groups

and on the basis of this test conclude there is a significant difference with p<0.0001). However, we calculated a Chi-square statistics ourselves to find that the p-value is 0.75 (not significant).[39]

48 months

At the 48-month follow-up period, 6.2% of CTDR patients experienced neurological deterioration compared with 7.6% in the fusion group (p=0.65).[39]

Patient Satisfaction

Davis measured patient satisfaction[39] with a questionnaire, which asked whether patients were very satisfied, somewhat satisfied, somewhat dissatisfied or very dissatisfied with the surgical treatment received. Cheng used the Odom's scale to measure patient satisfaction.[40]

3 months

Numbers are not provided in the papers.

12 months

Numbers are not provided in the papers.

24 months

In Davis patient satisfaction was high in both groups at 24 months with 95.8% of CTDR patients and 92.0% of fusion patients reporting being either very or somewhat satisfied with their treatment (p=0.20). Similarly, in the Cheng[40] study patient satisfaction was high in both groups. At 24 months 24 out of 30 patients in the CTDR group was rated "excellent" versus 22 out of 32 in the fusion group.

48 months

At 48 months, the percentage of patients very satisfied or somewhat satisfied with their treatment was higher in the CTDR group (96.4%) than in the fusion group (89.0%). This result was significant (p=0.03). The Cheng study did not report patient satisfaction for this time-point.

Quality of Life

The RCT by Davis[39] reported on quality of life using the SF-12 questionnaire. Patients were evaluated separately on the physical component score (PCS) and the mental component score (MCS). The RCT by Cheng reported on quality of life using the SF-36 questionnaire but only provided an evaluation for the physical component score (PCS).

3 months

In the studies quality of life appears not to be assessed for this time-point.

12 months

PCS score: In Davis no absolute values are provided but bar graphs state a significant between groups difference at 12 months (p<0.05, however, no exact p-value was provided). Similarly, in the Cheng study a significant difference on SF-36 PCS in favour of CTDR was found after 12 months (49 points vs. 46 points; p=0.033).

MCS score: No absolute values are provided but bar graphs state a non-significant between groups difference at 12 months (p>0.05, however, no exact p-value was provided).

24 months

Patients in the CTDR group had an increased mean PCS score from baseline of 13.5 points and an increased mean MSC score from baseline of 9.5 points. Patients in the fusion group had an increased mean PCS score from baseline of 10.5 points and an increased mean MSC score from baseline of 7.2 points. The authors conclude there is a <u>significant difference</u> for the PCS score (p<0.05) but not a significant difference for the MCS score (p>0.05). However no exact p-values were provided. The Cheng study found a significant difference on SF-36 PCS in favour of CTDR after 24 months (50 points vs. 45 points; p=0.013).

48 months

The CTDR group showed significantly greater improvement in SF-12 PCS scores than the fusion group, the mean improvement from baseline SF-12 PCS scores was 13 (SD=12) for the CTDR group and 10 (SD=12) for the fusion group at 48 months (p<0.05). There was not a significant difference for the MCS score (p>0.05). The Cheng study did not report on quality of life for this time-point.

Conclusions Clinical Effectiveness

1-level disease

- **The results for pain, quality of life and functional status did not reach clinical significance**
- **The statistically significancy of the results were the following:**
  o **There is a significant functional difference after 3 months, 1-2 years and after 4 years in favour of CTDR measured by the Neck Disability Index.**
  o **There is a significant difference after 3 months, 1-2 years and after 4 years in favour of CTDR on the level of arm pain.**
  o **No significant between group difference was found after 3 months on neck pain but after 1-2 years and after 4 years a significant difference in favour of CTDR emerged.**
  o **There is a significant difference after 3 months and at 1-2 years for mobility at the index level favouring CTDR.**
  o **There is a significant difference after 3 months and 1-2 years for mobility at the the upper adjacent level favouring CTDR but no significant difference after 3 months and 1-2 years at the lower adjacent level.**
  o **There is no significant difference for "neurological succes" at 3 months, at 1-2 years there is a borderline significant difference favouring CTDR but this effect does not appear to be sustained after 4 years or more.**
  o **Patient satisfaction after 1-2 years is high in both CTDR and the fusion groups but is slightly higher in the CTDR group.**
  o **There is a significant difference for quality of life on the physical component score favouring CTDR at 3 months, 1-2 years and after 4 years.**
  o **There is no significant difference for quality of life on the mental component score at 3 months but at 1-2 years a significant difference in favour of CTDR emerged.**

2-level disease

- **Results are based on two RCTs and should be interpreted with great caution**
- **The results for pain, global quality of life and functional status (NDI) did not reach clinical significance**
- **The statistically significancy of the results were the following:**
- o **There appears to be a significant difference after 2 and 4 years in favour of CTDR on the Neck Disability Index.**
- o **One relatively large RCT found a significant difference favouring CTDR for neck pain after 3 months but the effect was not sustained after 1, 2 or 4 years. Another small study only reported neck pain at 2 years and found a significant difference in favour of CTDR.**
- o **From a relatively large RCT there appears to be no difference between CTDR and fusion for arm pain at any time-point. Another small study only reported arm pain at 2 years and found a significant difference in favour of CTDR.**
- o **After 2 years the average flexion/extension range of motion in the neck was 7.9° in one study and 10.1° in another study for CTDR patients. In both studies it was less than 1° for fusion patients.**
- o **There appears to be a significant difference for quality of life on the physical component score favouring CTDR at 1, 2 and 4 years.**
- o **No significant difference was observed for quality of life on the mental component score at any time-point.**
- o **Patient satisfaction appears to be similar between groups after 2 years but after 4 years a significant effect in favour of CTDR emerged.**
- o **There appears to be no significant difference for "neurological succes" after 2 and 4 years.**
- **There is a need for more studies on 2-level disease to reliably determine clinical effectiveness**

2.4.3 Results on Safety

The two systematic reviews retained[2, 31] both reported results for safety but at different time-points. Results are presented below by outcome (domain) and subsequently by time-period. As requested by the expert group we added results from a recent meta-analysis by Verma[33] looking at the rate of ASD related surgery at 2 to 5 years to complement the information on ASD from the other SRs. Additionally, we included recent RCTs by Philips,[35] Vaccaro[34] and Zhang[36] for single-level disease, and an RCT by Davis assessing 2-level surgery patients.[39]

2.4.3.1 Results single-level disease

Revision surgery at index level/ secondary surgery at adjacent levels

The Cochrane review reported on revision surgeries at index level and at adjacent levels separately (3 months and 12-24 months), whereas the review by Ren grouped these surgery types, and reported on them as an overall rate of revision surgery:

3 months

Index level:

From one study with 290 participants there was 1 revision surgery in 139 patients in the fusion group and no revision surgeries in 151 patients in the CTDR group.

Adjacent levels:

From one study with 290 participants there was 1 revision surgery at adjacent levels in 139 patients in the fusion group and no revision surgeries in 151 patients in the CTDR group.

12- 24 months

Index level:

Significant difference between the two treatment groups in favour of CTDR (RR 0.39; 95% CI 0.23 to 0.64). This result comes from 20 events in 783 CTDR patients and 47 events in 701 fusion patients. However, it was possible to update this meta-analysis with data from the Vaccaro trial[34] adding data from 380 patients. As displayed in Figure 10 the result remained significant (RR 0.35; 95% CI 0.22 to 0.53).

Figure 10 – Revision surgery at index level: 12-24 months RR (cervical total disc replacement versus fusion)

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Coric 2011	6	119	7	115	10.0%	0.83 [0.29, 2.39]
Heller 2009	6	234	10	197	15.2%	0.51 [0.19, 1.37]
Marzluff 2010	1	141	4	127	5.9%	0.23 [0.03, 1.99]
Mummaneni 2007	5	223	23	198	34.1%	0.19 [0.07, 0.50]
Nabhan 2007	0	19	0	21		Not estimable
Pettine 2010	1	28	3	25	4.4%	0.30 [0.03, 2.68]
Porchet 2004	1	19	0	18	0.7%	2.85 [0.12, 65.74]
Vaccaro 2013	7	236	17	144	29.6%	0.25 [0.11, 0.59]
Total (95% CI)		**1019**		**845**	**100.0%**	**0.35 [0.22, 0.53]**
Total events	27		64			

Heterogeneity: Chi² = 7.05, df = 6 (P = 0.32); I² = 15%
Test for overall effect: Z = 4.79 (P < 0.00001)

Risk Ratio M-H, Fixed, 95% CI — 0.001 0.1 1 10 1000 — Favours [TDR] Favours [fusion]

Additionally, in the RCT by Philips[35] 11/211 patients (5.2%) and 10/184 patients in the fusion group (5.4%) had subsequent secondary surgical intervention. In the RCT by Zhang[36] 1 patient (out of 56) in the CTDR group and 4 patients (out of 53) in the fusion group had reoperations. However, the split between surgeries at index level and adjacent levels are unclear from both of these trials.

Adjacent levels:

The Cochrane review reported a non-significant difference between the two treatment groups (RR 0.60; 95% CI 0.35 to 1.02). This result comes from 18 events in 755 CTDR patients and 31 events in 676 fusion patients. However, it was possible to update this meta-analysis with data from the Vaccaro trial[34] adding data from 380 patients. As displayed in Figure 11 the result remains borderline non-significant (RR 0.62; 95% CI 0.37 to 1.02).

Figure 11 – Secondary surgery at adjacent levels: 12-24 months RR (cervical total disc replacement versus fusion)
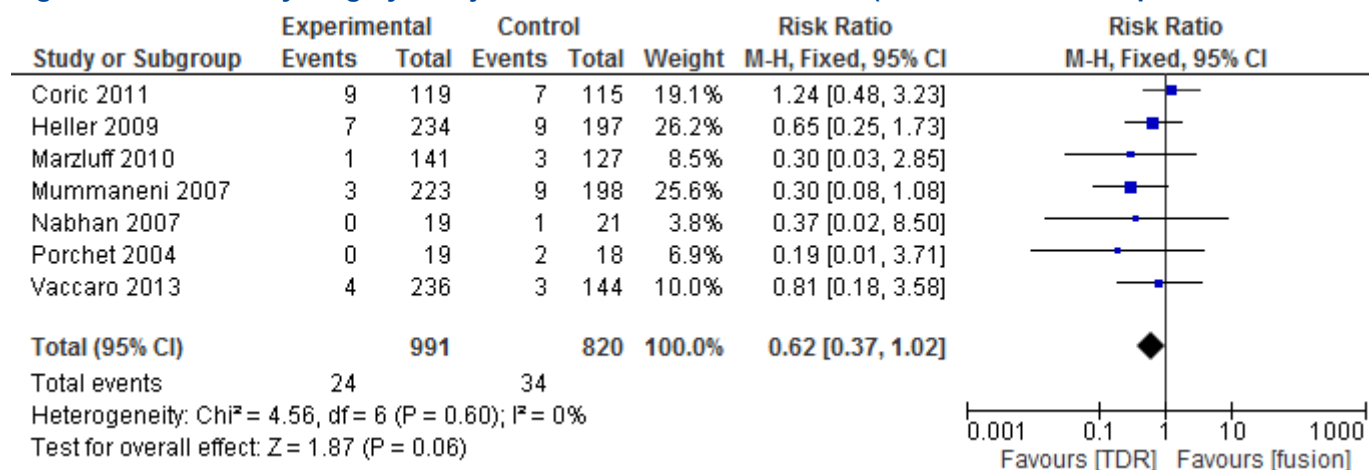
Additionally, in the RCT by Philips 11/211 patients (5.2%) and 10/184 patients in the fusion group (5.4%) had subsequent secondary surgical intervention. In the RCT by Zhang[36] 1 patient (out of 56) in the CTDR group and 4 patients (out of 53) in the fusion group had reoperations. However, the split between surgeries at index level and adjacent levels are unclear from both of these trials.

≥48 months

Overall rate (index + adjacent levels):

Significant difference between the two treatment groups in favour of CTDR (OR=0.44, 95% CI 0.22 to 0.89) were reported in the review by Ren.[31] This result comes from 26 events in 662 CTDR patients (4%) and 57 events in 624 fusion patients (9%). For consistency in outcome reporting we calculated the odds ratios to risk ratios. As displayed in Figure 12 the risk ratio was 0.43, with a 95% CI from 0.27 to 0.68.

Figure 12 – Overall revision surgery rate: ≥48 months RR (cervical total disc replacement versus fusion)
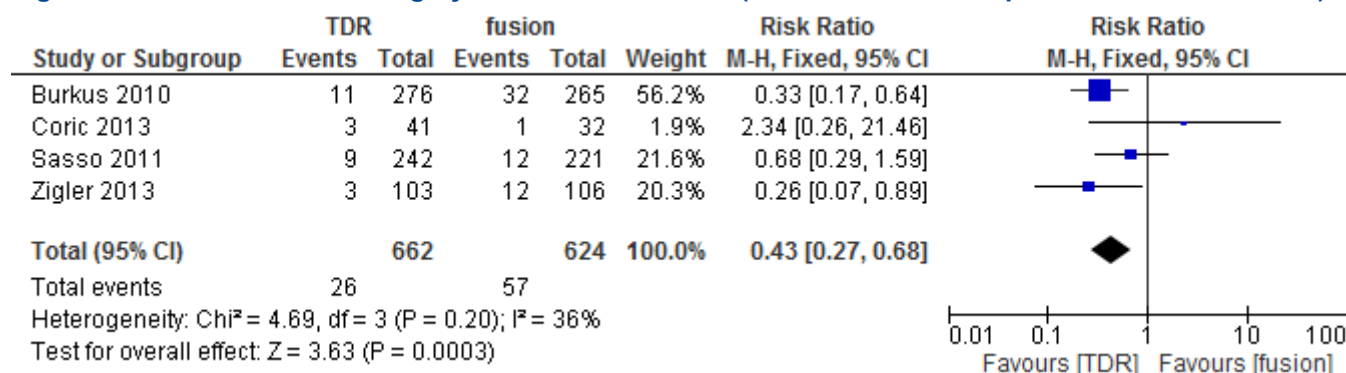
As requested by the experts we searched for additional information on the rate of ASD and identified a recent SR and meta-analysis by Verma.[33] This meta-analysis pooled studies of patients with one or two-level disease and with 2-5 years follow-up and found a significant difference between the two treatment groups in rate of revision surgery for ASD fusion versus CTDR in favour of CTDR (OR=0.74; 95% CI 0.58 to 0.93; p=0.01). However, the significant difference disappears when only patients available for follow-up are included in the analysis (patients with fusion have lower follow-up rates).

Adjacent segment degeneration

This outcome was reported in the review by Ren as an overall rate of adjacent segment disease.[31]

3 months

No evidence found.

12- 24 months

No evidence found.

≥48 months

No significant difference between the two treatment groups in rate of adjacent segment disease: CTDR 6.4% and fusion 5.7%, (OR=0.95, 95% CI 0.59 to 1.53). This result comes from 44 cases in 685 CTDR patients and 33 cases in 574 fusion patients. For consistency in outcome reporting we calculated the odds ratios to risk ratios. As displayed in Figure 13 the risk ratio was 0.95 with a 95% CI from 0.61 to 1.49.

Figure 13 – Adjacent segment degeneration: ≥48 months RR (cervical total disc replacement versus fusion)
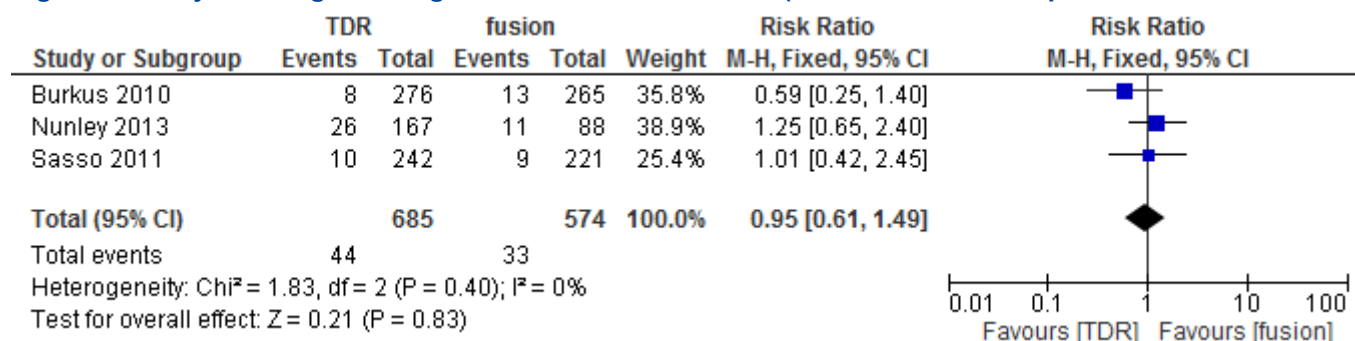
The authors of the Verma review[33] used historical data to determine the rate of ASD for each given follow-up year and predicted that 9.4% of CTDR patients and 9.2% of ACDF patients would have new symptoms of myelopathy or radiculopathy attributable to an adjacent level at the final follow-up. The authors did not evaluate this further.

Dysphagia

This outcome was only reported in the review by Ren.[31]

3 months

No evidence reported.

12- 24 months

No evidence reported.

≥48 months

Ren reported rate of dysphagia providing percentage numbers from individual studies:

One study reported a non-significant difference with 22 cases of dysphagia or dysphonia in fusion patients (8.3%) vs. 24 cases in CTDR patients (8.7%). A second study found one patient with dysphagia in the CTDR group (2.4%) and no cases in the fusion group. Finally, a third study found one patient with dysphagia in the fusion group (0.9%) and no cases in the CTDR group.

Other complications/adverse events

Below is a summary of other complications/adverse events that was reported in the systematic reviews and RCTs. In both reviews additional adverse events/complications were described narratively.

3 months

No evidence reported.

12- 24 months

The Cochrane review assesses radiological signs of fusion and writes that criteria to establish fusion vary not only between but also within studies e.g. as "bridging heterotopic ossification" in the CTDR group and as "bridging trabecular bone" in the fusion group. It was not possible to pool this outcome because of the heterogeneous definition. However, the Cochrane review concludes that the problem of "bone bridging" appears to be much more frequent in the fusion group than in the CTDR group. The RCT by Philips found a non-significant difference between groups for implant or surgery related adverse events; CTDR 5.6% (12/214) and fusion 7.4% (14/190).

≥*48 months*

The Ren review[31] found that three studies reported bridging ossification in CTDR patients (respectively in 17%, 6% and 3.2% of patients) and that none of these three studies found bridging ossification in the fusion group.

Additionally, Ren found that one study[43] reported 1 (3.1%) implant loosening in a patient who underwent fusion and no implant breakages or device failures in the CTDR patients. Finally, one study.[44] found 6 (5.7%) pseudo arthrosis in patients who underwent fusion.

2.4.3.2 *Results 2-level disease*

Subsequent Surgical Intervention

The RCT by Davis[39] reported subsequent surgical intervention classified as removal, revision, supplemental fixation, or reoperation were defined as surgical procedures to modify the study device or surgeries at the index level for new or ongoing original indications. The RCT by Cheng did not report on this outcome.

3 months

No numbers provided.

12 months

No numbers provided.

24 months

7 patients (3.1%) in the CTDR groups and 12 patients (11.4%) in the fusion group required a subsequent surgical intervention. This is a significant between groups difference result (p=0.003).

48 months

At 48 months, the cumulative percentage of patients who underwent subsequent surgeries at the index level remained significantly lower (p<0.0001) for the CTDR group at 4.0% (9 of 225 patients, with 10 surgeries) than for the fusion group at 15.2% (16 of 105 patients, 18 surgeries).

Adjacent segment degeneration

In the RCT by Davis adjacent segment degeneration was evaluated using the Kellgren-Lawrence grading scale. If the degeneration grade for a segment increased by 1 or more points from preoperative baseline this was classified as adjacent-level degeneration. The RCT by Cheng did not report on this outcome.

3 months

No numbers provided.

12 months

At 12 months 5.7% of CTDR patients and 12.5% of fusion patients had superior ASD whereas 1.5% of the CTDR patients and 9.1% of the fusion patients had inferior ASD. This was a significant difference at both levels (p<0.03).

24 months

At 24 months 13.1% of CTDR patients and 33.3% of fusion patients had superior ASD whereas 2.9% of the CTDR patients and 18.1% of the fusion patients had inferior ASD. This was a significant difference at both levels (p<0.03).

48 months

Evaluation measures were similar to the one used at 24 months. The superior levels at 48 months indicated degeneration in 64.7% of the fusion patients and 27.6% of the CTDR patients. This is a significant difference (p<0.0001). Results for the inferior levels were similar at 56.2% for the fusion group and 16.4% for the CTDR group (p<0.0001).

Dysphagia

3 months

No numbers reported.

12 months

No numbers reported.

24 months

In the RCT by Davis[39] this outcome is reported in total number of events with 9 patients (3.8%) in the CTDR groups and 8 patients (7.6%) in the fusion group developing dysphagia as an adverse event. We calculated the p-value and found this was a non-significant difference (p=0.16). Similarly, it is stated in the RCT by Cheng that one patient (out of 34) in the fusion group developed dysphagia versus no patients in the CTDR group.

48 months

No evidence was identified.

Other complications/adverse events

The study by Davis[39] reported on incidence rates of definite or possible device-related adverse events (neck and/or arm pain due to device, neurological adverse events, spinal adverse event, respiratory events etc.). The Cheng study[40] had an extremely limited reporting on complications/adverse events.

3 months

No numbers provided.

12 months

No numbers provided.

24 months

In Davis[39] the incidence rate of device-related adverse events were 16.7% (39/225) in the CTDR group and 34.3% (36/105) for the fusion patients. We calculated the p-value and found this was a significant difference (the Fisher exact test statistic value is 0.001). In Cheng it is stated that one patient in the CTDR group (out of 31) had deep vein thrombosis versus none in the fusion group.

48 months

No evidence was identified.

Conclusions Safety

1-level disease

- **At 1-2 years the CTDR group has significantly less revision surgery at index level.**

- **There is no significant between groups difference for secondary surgery at adjacent levels after 1-2 years.**

- **After 4 years there appears to be a significant difference in favour of CTDR for the overall rate of index and adjacent levels surgeries.**

- **After 4 years there is no significant difference between the two treatment groups in the rate of adjacent segment disease.**

- **There does not appear to be a between groups difference in the rate of dysphagia after 2 years.**

2-level disease

- **Results are based on two RCTs and should be interpreted with great caution.**

- **After 2 and 4 years the CTDR group appears to have significantly less subsequent surgical interventions.**

- **Patient symptoms related to a degenerative adjacent level is not well reported.**

- **There does not appear to be a between groups difference in the rate of dysphagia after 2 years.**

- **There appears to be a between groups difference for device-related adverse events after 2 years with the CTDR group having significantly less device-related adverse events.**

- **There is a high need for more studies on 2-level disease to reliably determine safety issues.**

2.4.4   Discussion

From the literature review on <u>one-level</u> cervical disc degeneration there was a clear tendency for results to be in favour of cervical total disc replacement compared with conventional fusion, and these were often statistically significant. However, although statistically significant, the differences between the two interventions did not reach the clinical relevance significance threshold, as set by the experts, for the main functional and clinical outcomes. This includes pain, functional status (NDI) and quality of life. The categorization of the clinical relevance may impede to draw clear conclusions when the outcomes mean differences are close to the chosen cut-offs (e.g. 9.99% of the scale or RR=0.80001). Nevertheless the results were never borderline following the clinical relevance criteria.

This analysis is based on a highly selected patient population. Additionally, the experts argued that the comparison between total disc replacement and prolonged non-surgical treatments does not apply in this patient group because surgery mainly is meant for the management of arm pain and that a prolonged non-surgical treatment only would be relevant in patients in which neck pain is the main complaint.

The literature review on <u>two-level</u> disease resulted in the identification of two RCTs. These RCTs both suffered from limitations with regards to quality and to outcome reporting. It is consequently difficult to make evidence based recommendations for two-level disease before more and larger studies are conducted, and before outcomes, including those relating to safety aspects, are reported.

Firstly, the prevention of adjacent segment degeneration was one of the main rationales for the introduction of disc replacement (whatever the number of levels involved). This was hoped to be a result of preserved motion at index level and consequently reduced adjacent level forces. However, because adjacent segment degeneration evolves slowly over time this needs to be evaluated based on truly long-term results. The expert group argued that at least 10 years would be appropriate for this outcome. The present review, where an insufficient number of studies evaluated this outcome, and where follow-up generally is limited to 48 months, should therefore not provide a basis for definite conclusions regarding the development of adjacent segment degeneration. That being said, it is clear that fusion provides, to a large extent, a loss in range of motion at index

level. This is, however, commonly compensated for at adjacent levels[47] thus providing no real clinical difference in terms of neck mobility for the patient.

Secondly, with the introduction of disc replacement there was a hope for a decreased rate of reinterventions. However, although the rate of reinterventions appears to be smaller in the group receiving cervical disc replacement, it was stressed by the experts that there are several limitations to the reporting of reinterventions in the literature e.g. it is a common Belgian practice experience that the complexity and severity of a reintervention after a disc replacement at times are much greater than after a fusion. This argument, however, will require much larger study populations to be fully verified.

It is important that patients eligible for cervical disc replacement are selected carefully. However, there are several limitations in the body of literature that provide a barrier for patient selection based on the evidence available e.g. a stratification based on age group was not possible to perform. Similarly, we did not undertake subgroup analysis with respect to prosthesis model because data was insufficient.

2.5   Economic evaluation

HTA CORE MODEL DOMAIN: ECO

2.5.1   Introduction

This section aims at reviewing the peer-reviewed full economic evaluations of cervical total disc replacement (CTDR) versus conventional surgery (i.e. anterior cervical discectomy and fusion, ACDF) and/or conservative treatment (CT) in adult patients with chronic cervical indications (including degenerative disc disease or cervical disc hernia) refractory to conservative treatment for at least 6 weeks.

2.5.2 Methods

2.5.2.1 Inclusion and exclusion criteria

The following criteria were developed:

- Inclusion and exclusion criteria describing the <u>Population</u>, the <u>Intervention</u> and the <u>Comparator</u> of interest are defined as in the methods chapter for the clinical effectiveness and safety assessment of the CTDR. These criteria can be found in section 2.4.1.
- <u>Design</u>: only full economic evaluations are considered, i.e. studies comparing at least two alternative treatments in terms of both costs and outcomes. Cost-minimization analyses (CMA), cost-utility analyses (CUA, with results expressed as incremental cost per quality-adjusted life year (QALY) gained), cost-effectiveness analyses (CEA, with results expressed as cost per life year (LY) gained) and cost-benefit analyses (CBA, with a monetary valuation of health outcomes) are eligible. Cost comparisons (not considering health outcomes), cost-outcome descriptions (not considering an alternative treatment) and cost-consequence analyses[b] are not relevant for inclusion. Both primary studies and reviews of full economic evaluations are relevant for inclusion; letters, news, conference proceedings and editorials were removed.
- <u>Timing</u>: a previous KCE report published in 2006 performed a systematic literature review of economic evaluations on the same interventions.[27] That search was performed up to early 2006. Only one RCT on cervical prostheses was then available, and no original full economic evaluation was found. The current search is thus limited to publications from 2006 up to April 2015.

2.5.2.2 Search strategy

Both electronic and manual searches were performed.

- Electronic search: the following databases were searched in April 2015: Medline(Ovid), Medline(Ovid) in-process and other non-indexed citations, Embase, CRD (Centre for Review and Dissemination) HTA and CRD NHS EED (National Health Service Economic Evaluation Database). A combination of MeSH, EMTREE and text word terms related to total disc replacement, cervical vertebra, arthroplasty and degenerative disc disease were combined with those related to full economic evaluations (see appendix). The websites of the HTA institutes listed on the International Network of Agencies for Health Technology Assessment (INAHTA) were also consulted.
- Manual search: the reference lists of relevant review papers and full economic evaluations were scrutinized for additional relevant articles.

2.5.2.3 Selection procedure

The selection was performed in a two-step procedure: initial assessment of the title, abstract and keywords; followed by full-text assessment if the title or the abstract or the keywords suggested relevant information, or if no abstract was available. The procedure was performed on the cervical and lumbar total disc replacement techniques combined, but the results with the final selection of the full-text articles retained are presented for each type of procedure separately. The flow chart of the selection processes can be found in appendix.

The search strategy yielded 126 unique, potentially relevant citations on the cervical and lumbar disc replacement techniques combined. After title and abstract review, 73 articles were excluded, the majority of which were not economic studies or only partial economic evaluations (i.e. mostly cost-comparisons). Of the 53 full-text articles reviewed, 41 were excluded. The primary reason for exclusion was that most studies focused on efficacy/safety without reporting any economic result. Other reasons for

---

b   In cost-consequence analyses, both costs and outcomes of different alternatives are described. In such studies however, an incremental cost-effectiveness ratio (ICER) is not calculated or the results are expressed in

disease-specific outcome (e.g. incremental cost per patient reduction of one point in the Activities of Daily Living scale). Such studies are discarded as their results cannot be compared with those of other types of economic evaluations, mainly CEA and CUA.

exclusion were that the design of the studies was limited to a cost comparison (5 studies[48-52]), or a cost-outcome description (1 study[53]).

A total of 12 articles were retained: 9 full economic evaluations among which 5 related to CTDR,[54-58] and 3 literature reviews including 2 on CTDR.[59, 60]

2.5.2.4 *Coverage of the reviews*

The review by Alvin et al.[59] covered the literature up to mid-January 2014. Their search returned no study with conservative treatment as a comparator, and 2 studies with ACDF as a comparator: the full economic evaluation by Qureshi et al.,[57] also identified in our review but only narratively described in theirs, and the study by Menzin et al.,[61] appraised here as being a cost comparison and therefore discarded from our review. Of the four additional full economic evaluations on CTDR identified in our literature search, three[54-56] had publication dates beyond the search period in Alvin et al.[59]

In the second review by Kramer et al.,[60] no methodology for the literature search/selection procedure and no search date were reported. Their review about the value (a larger concept than cost-effectiveness) of CTDR versus ACDF returned 3 studies published over the period 2008-2012;[25, 52, 62] all three articles being appraised here (as by Kramer et al.[60] themselves) as cost comparison analyses and therefore discarded from our review. Finally, we did not retain these two reviews after critical and qualitative analysis. Our analysis was based on the primary studies by Qureshi et al.,[57] Warren et al.,[58] Ament at al.,[54] Lewis et al.[55] and McAnany et al.[56]

2.5.2.5 *Data extraction and costs conversion*

Data extraction sheets, providing a comprehensive and detailed summary, were developed for each full economic evaluation (see appendix). The quality of the studies was assessed narratively.

Past costs reported in the studies were converted into 2014 costs using local Consumer Price Indices. For each country, the Purchasing Power Parities indices were used to convert the original costs into Euro values (for the whole Euro area, not for a specific country of the Euro zone). Both indices were obtained from the OECD website.[63] If no costing year was mentioned in a study, an interval of two years before the publication date was chosen. The original cost figures (i.e. before conversion) can be found in the data extraction sheets.

2.5.3 *Characteristics of the economic evaluations*

Table 6 gives an overview of the characteristics of the 5 studies assessing the cost-effectiveness of CTDR.

Table 6 – Base-case characteristics of the full economic evaluations of cervical total disc replacement

	Ament et al., 2014[54]	Lewis et al., 2014[55]	McAnany et al., 2014[56]	Qureshi et al., 2013[57]	Warren et al., 2013[58]
Country	USA	USA	USA	USA	USA
Funding source§	Industry support	No industry support	No industry support	Unknown	Unknown
Study design	CUA	CUA	CUA	CUA	CUA
Perspective	Societal (HCP*)	HCP	HCP	HCP	HCP
Time horizon	2 years	5 years	5 years	20 years	2 years
Discount rate†	3%	Not reported	3%	3%	Not reported
Costing year	2012	2014	2010	2010	Not reported
Analytic technique	Markov model	Markov model	Markov model	Decision tree	Within trial
Comparators	CTDR, ACDF	CTDR, ACDF, ACD	CTDR, ACDF	CTDR, ACDF	CTDR, ACDF
Disease level	2-level	1-level	1-level	1-level	1-level
Target patient group	45-year-old patients with radiculopathy or myelopathy, and failed CT for at least 6 weeks	Adult patients with radiculopathy secondary to symptomatic cervical disc disease, without prior operation	40-year-old patients with acute disc herniation (myelopathy/ radiculopathy) and failed CT	45-year-old patients with cervical DDD with radiculopathy, unresponsive to CT	Similar to the NYU Hospital subset of patients in the ProDisc-C trial (28 patients aged 41 years) with radiculopathy, without ASD or prior fusion
Outcome	QALY	QALY	QALY	QALY	QALY
Underlying instrument for QoL	SF-12 mapped to SF-6D	Not specified (probably different instruments combined)	SF-36 mapped to SF-6D	Not specified (probably SF-36 but unclear)	SF-36 mapped to SF-6D
Duration of QoL differential effect	2 years	5 years	5 years	Differential treatment effect up to 20 years stays the same as during the observed phase	2 years
Source for QoL differential outcome	Mobi-C trial from Davis et al., 2013[39]	Literature review	ProDisc-C trial from Zigler et al., 2013[44]	4 RCTs (Prestige,[46] Kineflex-C,[64] Bryan,[65] ProDisc-C[66]) and 1 meta-analysis[67]	ProDisc-C (1 centre) trial from Murray et al., 2009[66]
Other differential effects	Reduced post-surgical complications (supplemental fixation, revision, reoperation, device removal) with CTDR	Reduced perioperative complications and late-reoperations (index and adjacent level) with CTDR versus ACD and ACDF ‡	Reduced non-operative complications and reoperations (index and adjacent level) with CTDR	Reduced short-term (hardware failure) and long-term ASD complications with CTDR	No (complications and reoperations are not accounted for)

	Ament et al., 2014[54]	Lewis et al., 2014[55]	McAnany et al., 2014[56]	Qureshi et al., 2013[57]	Warren et al., 2013[58]
Source for other differential effects	Mobi-C trial from Davis et al., 2013[39]	Meta-regression of 156 clinical studies	Literature review	Literature review	-
Uncertainty analyses	One-way SA Scenario analyses	Probabilistic SA	One-way SA	One-way SA Scenario analyses	None

CUA: cost-utility analysis, HCP: health care payer, ACDF: anterior cervical discectomy with fusion, ACD: anterior cervical discectomy (without fusion), ASD: adjacent segment degeneration, CTDR: cervical total disc replacement, DDD: degenerative disc disease, CT: conservative treatment, QALY: quality-adjusted life year, SA: sensitivity analysis. § Funding source was classified using three categories. Industry support: any explicit acknowledgment of support from private industry (generally manufacturers of medical devices). Unknown: no information about support provided. No industry support: explicit acknowledgment of support that did not include private industry (generally from either public sources or private foundations). An explicit statement of "funding source: none" was also included in this last category.* The base-case perspective of this study is societal. To comply with Belgian guidelines, results from the HCP perspective (explored in a scenario analysis in this study) were added. † For both costs and effects. ‡ Except for late-reoperations at index level where rate for ACD is lower than that for CTDR.

2.5.3.1 Country and study design

All studies were performed in the USA. All of them were cost-utility analyses, with outcomes expressed as quality-adjusted life years. This design seems appropriate because CTDR aims at improving the quality of life of the patients, rather than extending their lives.

2.5.3.2 Perspective

Most studies adopted a health care payer perspective in their base-case, including direct medical and intervention costs only. The study by Ament et al.[54] adopted a societal perspective in their base-case, adding also indirect productivity costs due to work loss, considering that the patients belong to the working population (patients were indeed aged 45 years on average, which was confirmed by our Belgian data in section 2.3.2.2). However the recommendation from the recent Belgian guidelines on economic evaluations[68] is to use a health care payer perspective in the base-case assessment of a new intervention. As a consequence, results from Ament et al.[54] excluding the productivity costs were also added in our comparison tables.

2.5.3.3 Time horizon and discount rate

In four studies, the time horizon of the economic evaluations was limited to the time frame of the clinical trials/studies they used, i.e. 2 to 5 years.[54-56, 58] Qureshi et al.,[57] extrapolated the results observed in different studies up to a 20-year time horizon.

Short time horizons may not capture all the long-term consequences following both CTDR and ACDF, and their associated costs (e.g. ASD or revision surgery). In order to capture such long-term effects beyond the timeframe of the clinical trials, modelling is usually needed. However, as was done in Qureshi et al.,[57] this requires merging data from many different sources along with making assumptions regarding e.g. the duration and the extent of the observed treatment effect. Longer time horizons (up to 10 years instead of 2) were explored in the sensitivity analyses by Ament et al.[54]

Future costs and outcomes were discounted at a rate of 3% in three studies.[54, 56, 57] In two studies with limited time horizons of 2[58] and 5 years,[55] it was unclear whether discounting was applied.

2.5.3.4 Intervention and comparators

All studies evaluated the use of CTDR versus ACDF, and also versus ACD without fusion in one study.[55] None considered conservative treatment as a comparator.

Only one study was performed in patients with two-level disease,[54] the other studies focusing on one-level cervical disc degeneration.

2.5.3.5 Quality of life effect

The study by Ament et al.[54] on patients with 2-level disease based the quality of life (QoL) impact of CTDR versus ACDF on the SF-12 values from the Mobi-C trial by Davis et al.[39]

In Warren et al.[58] and in McAnany et al.[56] the QoL impact of CTDR versus ACDF was based on the SF-36 values of the ProDisc-C trial; the first study using the interim second-year results[66] and the latter study using the final five-year results.[44] Furthermore, in Warren et al.,[58] results from the full ProDisc-C trial population were not used but only those from a small number of patients (n=28) enrolled in one investigational centre of this trial. The resulting incremental QALYs of CTDR versus ACDF were opposite in both studies, with 0.03 QALY gained (no p-value reported)[56] with the full trial population (favouring CTDR) and -0.15 QALY gained (i.e. 0.15 QALY lost) (p=0.19)[58] with the single-centre population (favouring ACDF). As illustrated by the non-significant QoL impact, the capacity to detect a valid difference from a small subset of a trial population may be questioned.

In Lewis et al.[55] and in Qureshi et al.,[57] the QoL impact was based on several sources instead of relying on the results of an RCT for one particular implant. Qureshi et al.[57] inferred the QoL impact from four different RCTs[46, 64-66] and one meta-analysis.[67] In Lewis et al.,[55] QoL impact was derived from a literature review of the utilities associated with perioperative complications (e.g. donor site pain, dysphagia…) and late reoperations (same or adjacent level) from patients presenting diverse conditions, not all of them directly related to DDD (e.g. painful bone metastases, healthy elderly, patients with chronic subdural hematoma…). In none of these two studies, the instrument used to value the QoL impact was clearly reported.

Four studies[54-56, 58] limited the duration of the QoL effect to the timeframe of their source clinical trials/studies, such that no extrapolation of this effect in the future was done. In Qureshi et al.,[57] a 20-year QoL effect after CTDR

and ACDF was used, based on the assumption that the residual QoL effect remains the same as the effect observed during the trials. This assumption seems to be supported by the observed significant difference between CTDR and ACDF in the SF-36 physical component score at ≥48 month reported in section 2.4.2 (clinical effectiveness assessment). However there are currently no studies directly evaluating this clinical effect for such a long 20-year time horizon. Given the uncertainty about the duration and the extent of the QoL impact, the assumption of a constant long-term effect in Qureshi et al.[57] should have been tested in their uncertainty analysis, by e.g. presenting scenarios assuming a declining effect over time or a scenario with no effect after the observed trial-period.

2.5.3.6 Other differential effects

Four studies[54-57] incorporated the benefits associated with a reduced rate of adverse events (dysphagia, pain…) and/or complications (hardware failure, revisions, reoperations) with CTDR. One of those studies[57] further modelled a benefit in the rate of adjacent segment degeneration (ASD) favouring CTDR (i.e. ASD rates of 1.5% for CTDR versus 3% for ACDF), although this does not seem to be supported by the evidence at 4 years (see section 2.4.3 on safety assessment). In their study, Warren et al.[58] did not incorporate any of those effects.

Three studies reported on reoperation/revision surgeries at the index level and at adjacent levels separately. The reported values differed considerably. In McAnany et al.,[56] rates for index level revision were 1.1% for CTDR versus 2.8% for ACDF (1.7% difference). Rates in Lewis et al.[55] were higher at 2.6% for CTDR versus 3.2%-4.9% for ACDF (dependent on the ACDF technique used) (0.6%-2.3% difference); while the rate for ACD (without fusion) was valued more favourably at 1.9%. In Qureshi et al.[57] the difference in the hardware failure rate (necessitating a revision) of CTDR compared to ACDF was the highest (4% difference) with a rate of 1% assumed for CTDR versus 5% for ACDF. Similarly the rates for adjacent level revision also varied between the studies. In McAnany et al.,[56] the rates were 1.1% for CTDR versus 1.3% for ACDF (0.2% difference). Rates in Lewis et al.[55] were higher at 2.3% for CTDR (and ACD without fusion) versus 4.3%-5.4% for ACDF (dependent on the ACDF technique used) (2%-3.1% difference).

2.5.4 Results of the economic evaluations

Table 7 and Table 8 give an overview of the results of the five studies assessing the cost-effectiveness of CTDR.

2.5.4.1 Single-level disease

Reflecting the variability in their input parameters and assumptions, the results of the cost-effectiveness studies were quite different. For single-level disease, two studies reported that CTDR is a dominant strategy compared to ACDF,[56, 57] as CTDR was less costly and more clinically effective than ACDF. It should be remembered however that some assumptions by Qureshi et al.[57] were rather optimistic, as this study presented the highest difference in hardware failure rate compared to other studies, and as the beneficial effects of CTDR on QoL, ASD and hardware failure were assumed to remain constant over a 20-year period. Ignoring the potentially expensive complications and reoperations associated with both interventions, and based on a very limited number of patients on which QoL impact was measured, the cost-effectiveness results by Warren et al.[58] were more in favour of ACDF. In the study by Lewis et al.,[55] where CTDR is also compared to ACD (without fusion), ACD was found to dominate both ACDF and CTDR. However, the relevance of using ACD without fusion as a comparator to CTDR was questioned during the expert meeting on economic evaluations results. All experts expressed strong doubts about the efficacy of performing a discectomy not followed by a fusion, and further argued that such intervention (without fusion) was no standard practice in Belgium and therefore not a valid comparator. Following the experts' recommendations, we recomputed the incremental results of Lewis et al.[55] after excluding ACD (without fusion) as a comparator. We found that ACDF with allograft was dominated (more costly and less clinically effective) by ACDF with spacer and therefore discarded as a comparator; CTDR was associated with an ICER of €8535 per QALY compared with ACDF with spacer; and ACDF with autograph was dominated by CTDR. This would suggest that, in the absence of ACD (without fusion) as a comparator, CTDR would be the preferred option compared to ACDF. However, those results are only speculative as they are not consistent with the initial objectives of the Lewis et al.[55] study.

Only Lewis et al.[55] performed a probabilistic sensitivity analysis; deterministic one-way sensitivity analyses were performed in McAnany et al.[56] and in Qureshi et al.[57] No sensitivity analyses was performed in Warren et al.[58] Both McAnany et al.[56] and Qureshi et al.[57] reported that their base-case results (i.e. CTDR is dominant) were highly sensitive to the QoL impact of CTDR versus ACDF. Very small reductions in the postoperative utility values of CTDR resulted in ACDF appearing more cost-effective than CTDR. The threshold utility values for postoperative CTDR were found to be 0.713 (base-case: 0.720) in McAnany et al.,[56] and 0.796 (base-case: 0.90) in Qureshi et al.[57] This means that if the utility of CTDR is less than those threshold values, ACDF becomes more cost-effective than CTDR in the treatment of single-level cervical disc disease. The results were also sensitive to variations in the assumed index- and adjacent-level reoperation rates. In McAnany et al.[56] ACDF became more cost-effective than CTDR if the CTDR index-level revision rate was >27% (base-case 1.1%), or if the CTDR adjacent-level revision rate was >10.5% (base-case: 1.1%). In Qureshi et al.[57] ACDF became more cost-effective than CTDR if the failure rate was >29% (base-case 1%). Reducing the time horizon of the study also impacted the results; with ACDF becoming the preferred intervention with a time horizon <11 years (base-case: 20 years) in Qureshi et al.[57]

2.5.4.2 Two-level disease

In Ament et al.,[54] the ICER of CTDR compared with ACDF was found to be favourable from a societal perspective, according to the commonly accepted US threshold of $50 000 (about €40 000) per QALY gained. However, when indirect productivity costs were excluded, i.e. adopting a health care payer perspective as recommended in the Belgian guidelines,[68] the ICER rose to over €80 000 per QALY gained, such that CTDR could no longer be considered a cost-effective intervention.

The results of this study were highly sensitive to the time horizon considered (base-case: 2 years), with shorter time horizons (1 year) favouring ACDF and longer time horizons (4 and 10 years) favouring CTDR, from a societal perspective. Interestingly, the subgroup sensitivity analysis found that CTDR was more cost-effective than ACDF in patients aged <45 years (€15 900 per QALY gained), than in patients aged 46 years and older (€24 000 per QALY gained), again from a societal perspective. However one of the factors driving this difference may simply be that more retired individuals constitute the >46-year-old subgroup, thereby reducing the productivity gain obtained with CTDR.

Table 7 – Results of the full economic evaluations of cervical total disc replacement

CTDR versus ACDF	Incremental cost (Euro 2014)		Incremental outcome (QALY)	Incremental cost-effectiveness ratio	Authors' conclusion
Single-level disease					
McAnany et al., 2014[56]	- €14 800		0.03	Dominant	CTDR is dominant over ACDF
Qureshi et al., 2013[57]	- €4080		2.02	Dominant	CTDR is dominant over ACDF
Warren et al., 2013[58]	- €2445		- 0.15	€16 300 saved per QALY lost †	The ICER is in favour of ACDF
Two-level disease					
Ament et al., 2014[54]	Societal	€1714	0.087	€19 709 per QALY gained	CTDR is cost-effective compared to ACDF
	HCP	€6909	0.086	€80 342 per QALY gained	Not reported

QALY: quality-adjusted life year, ACDF: anterior cervical discectomy with fusion, CTDR: cervical total disc replacement, ICER: incremental cost-effectiveness ratio, HCP: health care payer. All costs expressed in Euro 2014.

† In the south-west quadrant of the cost-effectiveness plane, an intervention of interest is less expensive but also less clinically effective than its comparator. Thus lower costs are possible, but at the expense of lower benefits. An ICER can be calculated, although its interpretation will be the opposite of the "more frequent" ICER falling in the north-east quadrant where an intervention is more costly and more clinically effective than its comparator. ICERs in the south-west quadrant refer to a cost saving per unit of effect lost, instead of an incremental cost per QALY gained for ICERs in the north-east quadrant. In the south-west quadrant, ICERs greater than a cost-effectiveness threshold are considered acceptable.

Table 8 – Results of the incremental analysis from Lewis et al., 2014[55]

Intervention	Total cost (SD)	Total QALY (SD)	Incremental cost †	Incremental QALY †	ICER †
ACD	€12 866 (539)	4.885 (0.041)	-	-	-
ACDF with spacer	€15 182 (513)	4.787 (0.066)	€2316	-0.098	Dominated by ACD
ACDF with allograft	€15 519 (380)	4.781 (0.068)	€2653	-0.104	Dominated by ACD
CTDR	€15 660 (537)	4.843 (0.056)	€2794	-0.042	Dominated by ACD
ACDF with autograft	€15 937 (262)	4.714 (0.067)	€3072	-0.171	Dominated by ACD
Authors conclusions: ACD without fusion dominates all other surgical options					

ACDF: anterior cervical discectomy with fusion, ACD: anterior cervical discectomy (without fusion), CTDR: total disc replacement, QALY: quality-adjusted life year, ICER: incremental cost-effectiveness ratio. SD: standard deviation. All costs expressed in Euro 2014.

† In this study calculations of the ICERs follow the efficiency frontier approach. In this approach, interventions are ranked from the least to the most expensive one. Each intervention that is (extendedly) dominated by another intervention is then removed. An intervention is dominated when its effectiveness is lower and its cost higher than another intervention. ICERs are only computed for the remaining interventions, by comparing each intervention with the previous less costly and less effective intervention.

2.5.5 Discussion

The assessment of the clinical effectiveness of CTDR for <u>one-level disease</u> reported a significant statistical (but not clinical) difference for the quality of life (physical component score) favouring CTDR at different time intervals (3 months, 1-2 years and after 4 years). Despite this significant statistical difference, the current literature review of economic evaluations highlighted that, compared to ACDF, CTDR was not always the preferred option from a cost-effectiveness point of view.

CTDR was more clinically effective and less costly than ACDF in two economic evaluations.[56, 57] However the robustness of those studies' conclusions (CTDR is dominant) was weak as ACDF became the most cost-effective strategy with relatively small variations in their input data, particularly for the QoL impact after surgery. In two other economic evaluations, one of which of low quality,[58] CTDR was found to be less cost-effective than ACDF[58] or dominated by ACD with no fusion.[55] In <u>two-level disease</u>, CTDR was found to be cost-effective compared to ACDF from a societal perspective, but not from a health care payer perspective, with also a great sensitivity of the results to parameter variations. However conclusions based on a single study should be treated with caution and certainly need corroboration by other studies. There was no economic evaluation of the use of CTDR versus conservative management.

The review of the literature identified that there was a great variation between the studies in the values reported for the index- and adjacent-level revision/reoperation rates. Rates for index-level ranged from 1%[57] to 2.6%[55] for CTDR, and from 2.8%[56] to 5%[57] for ACDF; while rates for adjacent-level ranged from 1.1%[56] to 2.3%[55] for CTDR and from 1.3%[56] to 5.4%[55] for ACDF. This may be due to different methods used in the studies to collect input data, with some performing either a literature review or a meta-regression and others using data from a single trial. It can be further questioned whether the results of the studies relying on trial(s) for one particular implant may be generalized to all implants used for CTDR. Clearly more precise and robust estimates of such influential input parameters should be better documented.

The long-term costs and consequences associated with CTDR and ACDF (e.g. the evolution of the QoL effect over time and the long-term revision, reoperation and ASD rates) must be taken into account in the assessment of the cost-effectiveness of those interventions. This is particularly important as it might be expected that the longer the operated patients will live, the higher their chance to undergo reoperations. Inclusion of the long-term costs and consequences of CTDR and ACDF was only performed in one study,[57] using a rather optimistic extrapolation assumption with no scenario analysis conducted on that hypothesis. In the absence of studies documenting the long-term treatment effects and safety, economic evaluations should explore the impact of their extrapolation assumptions over a range of plausible scenarios, e.g. constant residual effect (best-case), declining residual effect after observation period, no residual effect after observation period.

None of the five economic evaluations was performed in Belgium, with costs and outcome data reflecting the Belgian health care system and organisation. The results of those economic evaluations are therefore not directly transferable to Belgium. Context-specific studies are always preferred and as such only the development of a model reflecting the Belgian situation and consequences could provide a valid assessment of the cost-effectiveness of the use of CTDR in Belgium. However the literature review highlighted the great long-term uncertainty surrounding some crucial baseline parameters determining the cost-effectiveness of CTDR. Those uncertainties should first be addressed before providing any further economic evaluation on this topic.

Therefore, given the current lack of high-quality economic evaluations and awaiting better long-term information on crucial input parameters (QoL effect and rates of revision/reoperations and ASD), it is difficult to currently draw definite conclusions regarding the cost-effectiveness of cervical total disc replacement versus ACDF in adult patients with chronic cervical indications (including degenerative disc disease and cervical disc hernia) refractory to conservative treatment for at least 6 weeks.

3 LUMBAR TOTAL DISC REPLACEMENT

3.1 Health problems

HTA CORE MODEL DOMAIN: CUR1

3.1.1 Population and condition

In the Belgian Health Interview survey 2013, one fifth of Belgians aged 15 years or more (20.8%) suffered from low back disorder or other chronic back defect in the last year (22.7% women and 18.7% men).[4] The definition of chronic low back pain differs between authors but they generally agree that beyond 12 weeks the pain can be termed as chronic.[69]

Age related degenerative processes in the intervertebral disc and facet joins are probably the most common amongst the numerous sources of low back pain.[70-72] Other sources of pain include, among others, facet arthrosis, herniated nucleus pulposus, lumbar instability, sagittal imbalance and musculoligamentous injuries.[13, 71] A significant association between low back pain and lumbar degenerative disc disease (DDD) has been reported in some studies.[6, 73] If we only look at MRI signs, the prevalence of lumbar DDD increases linearly with age and at 70 years, 80% of all lumbar discs are affected.[72] Nevertheless, most patients presenting MRI signs of DDD are asymptomatic.

The first-line treatment of low back pain is conservative. But when conservative treatment does not relieve pain, surgery may be considered. Between 10 and 20% of people suffering from lumbar DDD will not respond to conservative treatments and may be considered candidates for lumbar surgery.[7] The international literature indicates that lumbar total disc replacement (LTDR) is mainly performed after at least 6 months of failed nonoperative therapy.[74]

3.1.2 Existing treatments

Conservative treatments include physical therapy, cognitive behavioural therapy, muscle strengthening, weight control, analgesics medication including NSAIDs and epidural steroids.[7, 75]

When conservative treatment failed and surgery is considered, lumbar spinal fusion is a common procedure. There are two possible techniques: posterolateral fusion and interbody fusion, both techniques being possibly used jointly. In posterolateral fusion (PLF), a bone graft is placed between the transverse processes in the back of the spine and vertebrae are fixed together using screws or wire and metal rods. In interbody fusion, a bone graft is placed after a discectomy at the place originally occupied by the intervertebral disc. This can be done by an anterior approach called anterior lumbar interbody fusion (ALIF), by a posterior approach, either the posterior lumbar interbody fusion (PLIF) or the transforaminal lumbar interbody fusion (TLIF), or finally by a lateral approach known as the extreme lateral interbody fusion (XLIF).[76]

Amongst those requiring surgery approximately 5% are candidates for LTDR, according to the Ontario Medical Advisory secretariat (Canada). Considering the many reasons beside DDD which can be related to low back pain, many surgeons prefer to fuse the entire motion treating different possible problems at the same time (segmental instability or facet problems cannot be treated by LTDR).[13, 70] It is important to note that symptomatic lumbar DDD is the primary indication for LTDR, not radiculopathy as for CTDR.[13, 74]

3.2 Description and technical characteristics

HTA CORE MODEL DOMAIN: TEC

The LTDR restores the disc height and relieves the pain like a fusion. But contrary to the fusion, it attempts to restore motion instead of eliminating it. Like the CTDR, the aim is also to avoid stress and increased motion in adjacent segments that can lead to a degenerative process in those segments.[75] There are contraindications to LTDR including, amongst others, osteoporosis, facet problems such as ankyloses or arthrosis, scoliosis, retrolisthesis, spondylolisthesis or spondylosis.[74] The procedure is performed by a transperitoneal or retroperitoneal anterior approach.

Beside complications related to the surgery in this abdominal area (wound ruptures, vascular or neurological damages, bowel or uro-genital complications), complications that may occur after LTDR are implant subsidence (sinking of the prosthesis in the vertebral body), migration, dislocation or wear, fracture of the vertebral body, facet arthrosis and heterotopic ossifications.[77, 78] LTDR being complicated by the presence of the aorta and inferior vena cava and its branches, orthopaedic or neurosurgeons need skilled backup in vascular surgery.[76, 79]

Lumbar vertebral bodies and intervertebral discs present very different features from the cervical ones, because they serve different functions. Mainly, lumbar vertebrae are larger, more massive, they contain more proteoglycan and moisture, and they still present a nucleus pulposus while cervical ones have practically lost theirs at the age of 45 years due to fibrosis.[22, 80, 81] The lumbar spine is also able of flexion, extension, lateral flexion and rotation, but in a more restricted manner than the cervical region. Most of all, lumbar vertebral bodies need to support great compression load, ground reaction forces and muscle contraction. Furthermore, lumbar intervertebral discs that are also larger than cervical ones present a concentric arrangement of collagen fibres around their nucleus pulposus in order to resist tensile forces in nearly all directions.[22] Whereas CTDR may treat one or two levels, the number of levels involved in LTDR varies from one up to four levels.[74] Nevertheless, Belgian guidelines recommend a maximum of 2 levels (see section 3.2.1.5).

Basically, materials used for lumbar artificial discs are the same as used in cervical ones. Like cervical ones, lumbar artificial discs can be two- and three-component ball and socket models or one-piece models. Different models all present their own biomechanical advantages and disadvantages and there is no solid evidence that different artificial disc designs or materials better suit different indications.[74] Nevertheless, Buttner-Janz et al. recently proposed an overview of patient selection criteria that may be used to choose the disc design including age, psycho-social factors, medical history, pain, clinical and imaging findings and bone quality.[74]

After unpromising results by Ulf Fernström in the fifties with stainless steel balls subsiding in the vertebrae, research took up again in the eighties with the development of the AcroFlex prosthesis by Arthur Steffee, originally consisting of a rubber core between two titanium endplates. Results were not particularly successful despite the different material improvements. The story of LTDR really took off with the SB Charité designed by Karin Büttner-Janz and Kurt Schellnac in 1982 in Berlin and clinically used since 1987 (SB=Schellnac+ Büttner-Janz, Charité Universitätsmedizin Berlin). This time ultra-high molecular weight polyethylene (UHMWPE) was used for the core that was sandwiched between metal endplates. Different versions came to light and the third version, SB Charité III, became the first lumbar artificial disc approved by the Food and Drug Administration (FDA) in 2004 for use in the USA. Meanwhile, research and development gave birth to the

ProDisc, another three-piece ball and socket model developed by Thierry Marnay in 1989. Since the beginning of the century, many other models have been designed and manufactured, especially in Europe, including the Maverick disc, a two-pieces metal-on-metal model clinically used since 2002.[82]

Lumbar artificial disc elements come in different sizes, heights and angles (for the endplates) to fit the patient anatomy. Brand surgery kits are sold by the disc manufacturer, containing chisels, tamps, distraction spacers, insertors, etc. Four prostheses are currently reimbursed in Belgium (see section 3.2.1.6). Characteristics and pictures of these lumbar artificial discs can be found in Table 9 and Table 10 gives the year when FDA approved each device (PMA is the pre-market approval decision required for the prosthesis before US marketing) and when the CE mark was authorized (CE mark is required for the launching on the European market).

3.2.1.1 InMotion artificial disc

The production of Charité was stopped by Depuy Synthes Spine and the new version of the disc is now called InMotion artificial disc. Basic features of the Charité disc are nevertheless maintained. Endplates are made of cobalt chromium alloy, both fitted with teeth and coated with titanium calcium phosphate for bony on growth to reinforce the prosthesis anchorage. The sliding core is made of UHMWPE. The FDA approved the InMotion in 2007 (Charité was earlier approved by the FDA in 2004).

3.2.1.2 Prodisc-L

The current version of the Prodisc-L (Depuy Synthes), a constrained[79] disc offering 3 degrees of freedom (see section 2.2), consists of an UHMWPE fixed inlay between two cobalt chromium molybdenum endplates. These are each coated by porous titanium to promote bony on growth and equipped with a central keel and spikes for the primary fixation and rotational stability. The upper plate presents a concave bearing surface articulating with the convex polyethylene core.

3.2.1.3 *Maverick*

The Maverick disc (Medtronic) is also a ball and socket, but featuring a metal on metal articulation. It consists of two cobalt chromium molybdenum endplates, each of them plasma-sprayed with hydroxyapatite for bony on growth and fitted with a keel. The superior endplate presents a concave surface whereas the inferior one presents a convex dome. Considering that a lumbar prosthesis should last a lifetime, this design banks on the metal durability and lower risk of wear debris.

3.2.1.4 *Mobidisc (Mobi-L)*

Like the Mobi-C, the Mobi-L (LDR Medical) is a semi-constrained metal-on-polyethylene device. The metal plates are made of a cobalt chromium alloy, with a porous titanium surface that encourages bony on growth. The anchorage is also insured by a keel on the superior endplate. The motion is made possible by the mobile UHMWPE core, flat on the bottom and convex in its superior part to mould the concave surface of the superior plate. Four peripheral stops control the mobility of the core in all axes (6 degrees of freedom).

Table 9 – List of lumbar artificial discs reimbursed in Belgium

Device	Manufacturer	Belgian Distributor	Constraint	Bearing surface material	Centre of rotation	FDA Pre Market Approval	CE mark approval
In Motion Artificial disc	Depuy Synthes	Depuy Synthes	Semi-constrained	Metal (CoCr) on polyethylene	2 mm dorsal of the vertebral sagittal midline (near the centre of the disc)	2007	2007
Prodisc-L	Depuy Synthes*	Depuy Synthes*	Constrained	Metal (CoCrMo) on polyethylene	Just below lower vertebra surface	2006	2000
Maverick	Medtronic	Medtronic	Constrained	Metal (CoCrMo) on metal	In the posterior third of the device itself	No**	2002
Mobidisc (Mobi-L)	LDR Medical	LDR Medical	Semi-constrained	Metal (CoCr) on polyethylene	Lower vertebra surface	No**	2003

* Depuy and Synthes were merged in 2012, as a company by Johnson & Johnson.
** Currently not marketed in the USA.

Table 10 – Pictures of main lumbar artificial discs available in Belgium

InMotion Artificial Disc	Prodisc-L	Maverick	Mobidisc (Mobi-L)

Note: scale is not preserved between models.

There are a few other models potentially available on the Belgian market that have been notified to the INAMI – RIZIV, such as M6-L (Spinal kinetics) distributed by Cormed, Orbit (Globus Medical), active-L (Aesculap) by B. Braun Medical, Flexicore (Stryker), Cadisc-L (Ranier) by Inspine Belgium, Baguera-L (SpineArt) by Hospithera, Anatomical Lumbar Discreplacement (ALD) by Osteon, Dynardi (Zimmer), etc. Because they are not reimbursed by the INAMI – RIZIV, their sales are marginal and some are even not sold or not sold every year. Together, between 10 and 20 lumbar prostheses are implanted each year without being reimbursed, whether (mostly) they are models non-listed for reimbursement or they are listed for reimbursement but used outside the conditions for reimbursement (see conditions listed in section 3.2.1.6).

3.2.1.5 Belgian recommendations of good practice for lumbar disc replacement

The Belgian Neurosurgical Spine Society elaborated the following recommendations on the use of total lumbar disc replacements in 2007:[83]

1. The surgeon should prove his ability to perform this operation, by adequate training certificate and prove that the approach is done in collaboration with a dedicated abdominal surgeon or a vascular surgeon or that one of those is standby in the hospital during the procedure.

2. *Standard* indications limited to the following situations:
   a. Clear history of DDD, no sudden onset, preferably no radicular pain
   b. Disease affecting usually 1 level, maximum 2 levels
   c. Lumbar spine MRI showing clear signs of Degenerative Disc Disease
   d. None of the following contra indications
      1. BMI > 28
      2. Disc height under 3 mm
      3. Spinal – foraminal stenosis
      4. Fracture
      5. Spondylolysis
      6. Spondylolisthesis
      7. Extreme facet degeneration
      8. Extreme muscular degeneration or muscular disease
   e. Age 18 to 55 years
   f. Minimum of 6 months well managed conservative treatment

3. Suggested additional tests
   a. Psychological evaluation highly recommended
   b. Bone scan (to rule out posterior facet impairment)
   c. X rays full spine (assessment of sagittal balance)

d. Dynamic X rays lumbar spine

e. Positive discography should not be done routinely but only in cases of more than 2 degenerated levels (discussed)

4. Caution recommended with L5-S1 total disc replacement, as high failure rate is reported. Fusion may be preferable.

3.2.1.6 *Belgian regulation and reimbursement*

Device

There are 4 total prostheses that are currently reimbursed in Belgium:

- Maverick (Medtronic)
- ProDisc-L (Johnson & Johnson Medical – DePuy)
- InMotion Artificial Disc (Johnson & Johnson Medical – DePuy)
- Mobidisc (Mobi-L) (LDR)

SB Charité (Johnson & Johnson Medical – DePuy) was reimbursed until 1st of November 2013 and was then replaced by InMotion, the newer version of the prosthesis.

Currently, only these four discs are reimbursed up to €1800. The patient out-of-pocket expense is limited at €180. In accordance with the article 102 of the Hospitals Law, the maximum price of these discs is €1980 (including the patient out-of-pocket expense).

The sliding core of the InMotion Artificial disc and the inlay of the Prodisc-L can be reimbursed separately in case of replacement, respectively at €286.26 and €185.50 (maximum €28.62 and €18.55 out-of-pocket expenses by the patient).

The four artificial lumbar discs are reimbursed under strict conditions and after notification to the advisory physician of the Sickness fund of the patient, under the code 163015 - 163026. Details including former codes (which were used in the data analysis of section 3.3) can be found in appendix.

Inclusion criteria in order to get reimbursement are:

- Patient aged 35 year or more;
- unsuccesful conservative treatment (more than 6 months) of low back pain due to disc degeneration at 1 or 2 levels characterized by:
  - ostheophyte formation on the vertebral end plates;
  - medial disc herniation limited to level L4-L5 or L5-S1;
  - documented discopathy;
- all of the following technical examinations, performed consecutively, confirmed the degeneration diagnosis:
  - X-ray face and/or profile;
  - flexion/extension radiographs without sliding of more than 3 mm;
  - negative scintigraphy at the level of facets;
  - MRI showing signs of disc degeneration, with or without Modic[c] changes;
  - unsuccesful facet infiltrations with a local anaesthetic without the use of corticoïds;
  - discography and/or disco-CT, positive (with pain) at the level of the intervention and negative (without pain) at adjacent levels.

Exclusion criteria are:

- patient with previous lumbar disc prosthesis (at any level);
- lateral recess stenosis or neuroforamen arthrosis;
- vertebral fractures;
- metabolic diseases weakening the vertebral corpus;
- spondylolysis;
- anterior spondylolisthesis;
- lumbar and dorsolumbar scoliosis;
- primary non-medial disc herniation;
- in situ tumour;

---

[c]   Firstly described by Michaël Modic in 1988, Modic changes are MRI signal changes revealing pathological changes situated in both the body of the vertebrae and in the end plate of the neighbouring disc.

- infections;
- osteoporosis documented by DEXA-scan;
- documented radiculopathy;
- residual intervertebral height inferior to 5 mm.

Procedure

The code billed for (a lumbar fusion or) a total lumbar disc arthroplasty is 281654 - 281665 *Anterior interbody arthrodesis or screwing, including the possible graft taking*, whose reimbursement amounts to €825.45 (N650). The first code (one day hospitalisation) was used only twice between 2008 and 2011.

3.3 Current use

HTA CORE MODEL DOMAIN: CUR2

3.3.1 Methods

3.3.1.1 Data sources

In order to estimate the number of LTDRs performed in Belgium we used the same three data sources as for the CTDR (see description of the databases in section 2.3.1.1): the UNAMEC survey, the RHM – MZG database and the non-reimbursed lumbar prostheses notifications to the INAMI – RIZIV. As the lumbar artificial discs are reimbursed, we additionally used the reimbursement figures of these discs (INAMI – RIZIV Doc N).

For the selection of the LTDR hospitalisations, we extracted the hospitalisations of patients discharged between 2008 and 2011 (last available year in November 2014) presenting an ICD-9-CM procedure code 84.65 *Insertion of total spinal disc prosthesis, lumbosacral* or the following codes for lumbar fusion: 81.06 *Lumbar and lumbosacral fusion, anterior technique*, 81.07 *Lumbar and lumbosacral fusion, lateral transverse process technique* and 81.08 *Lumbar and lumbosacral fusion, posterior technique*. Selecting the fusions also allowed us to compare both populations.

3.3.1.2 Analysis

We describe the current use of LTDR using simple descriptive statistics. After checking the normality, age at admission was compared between LTDR group and fusion group using a t test, gender proportion using a χ^2 test and right skewed length of stay using a nonparametric Mann Whitney test.

3.3.2 Results

3.3.2.1 Number of lumbar prostheses implanted in Belgium

Table 11 – Number of lumbar artificial discs or lumbar total disc replacement by data source

Source	Type of data	2008	2009	2010	2011	2012	2013
UNAMEC	Lumbar artificial discs sales					221	159
RHM – MZG	Hospitalisations with LTDR coded in ICD-9-CM (84.65) and discharged in the year	401	312	317	240		
INAMI – RIZIV	Reimbursed lumbar artificial discs		100	202	184	166	
	Notified but nonreimbursed lumbar artificial discs			8 (*)	28	24	

(*) June-December.

Lumbar artificial discs sales, implantations and reimbursements are all decreasing over years.

Similarly to the figures given for the CTDR, the number of lumbar prostheses reimbursed or notified to the INAMI – RIZIV is lower than the sales of the UNAMEC, which seem to be in line with the number of LTDRs performed in the RHM – MZG. Nevertheless, as several discs can be implanted during a same procedure and as some implantations may have been recorded under a nonspecific LTDR ICD-9-CM code, the number of discs implanted each year lies most probably above the RHM – MZG figures.

The yearly number of prostheses is estimated to be below 200 prostheses implanted with a total reimbursement budget for the discs only of approximately €400 000.

Table 12 – INAMI – RIZIV reimbursed amounts for lumbar artificial discs (€)

Type of data	2009	2010	2011	2012
Lumbar artificial discs reimbursements	226 168	465 998	429 490	390 912

3.3.2.2 *Characteristics of patients undergoing lumbar total disc replacement and comparison with patients undergoing a fusion*

Sixty percent of the 1270 cases who underwent a LTDR between 2008 and 2011 were female. On average patients were 41.7 year old (SD: 9.1, median: 41, range: 18-82). Figure 14 shows the age distribution per 5-year category and per gender. Almost all patients (99.9%) were admitted at least one night in the hospital, a one day hospitalisation was recorded in 2 cases only. The mean length of stay was 5.4 days (SD: 3.3, median: 5, range 0-70). All patients were discharged alive.

Figure 14 – Lumbar total disc replacement: Age distribution per gender
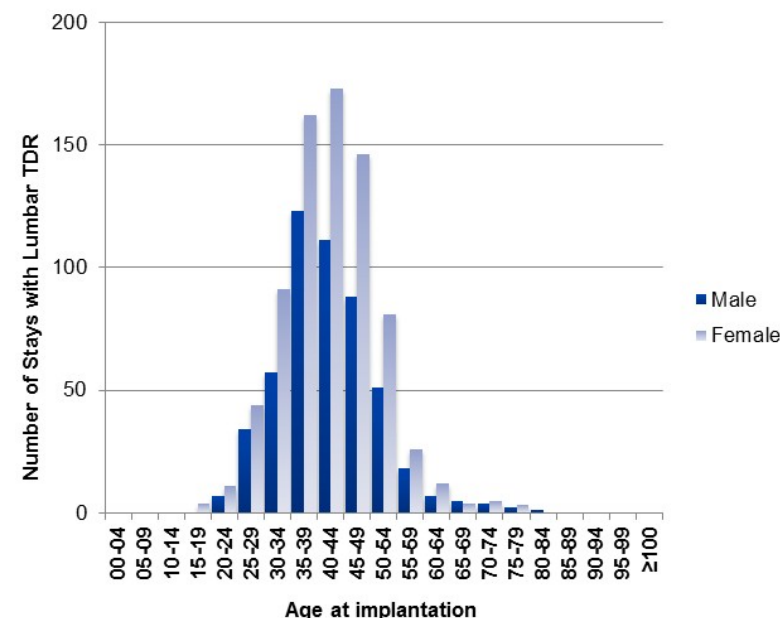
Source: RHM – MZG 2008-2011.

The most frequent principal diagnosis causing the admission was an intervertebral disc disorder (83.4% of the hospitalisations), followed by spondylosis and allied disorders (11.9%) or other and unspecified disorders of back (3.5%). The registration of a principal diagnosis of spondylosis is unexpected as this condition is a contraindication to LTDR in the literature. [74] Unfortunately, it is not possible in the scope of the present study to verify if the registration in the administrative database was correctly done. Inside the intervertebral disc disorder category, most of the cases suffered from a degeneration of lumbar or lumbosacral intervertebral disc (50.5% of all hospitalisations) or a displacement of the lumbar disc without myelopathy (16.7.% of all hospitalisations). A displacement of the lumbar disc with myelopathy was coded in 2.2% of the cases. The top 10 of the principal diagnoses in 3-digit ICD-9-CM codes, as well as the details of the first code split into 5-digit codes can be found in appendix.

We also extracted the data of the patients who underwent a lumbar fusion in order to compare their patient characteristics with those of the LTDR group. Because some hospitals also register a code of fusion in case of disc replacement, we selected the cases of fusion for whom no disc replacement was coded at all (neither partial nor total, whatever the location) and with a principal diagnosis of lumbar intervertebral disc disorder (degeneration, displacement, postlaminectomy syndrome or *Other or unspecified disorder of the lumbar intervertebral disc*: ICD-9-CM codes 722.1, 722.52, 722.73, 722.83 or 722.93). The 18 163 cases of fusion included a higher proportion of male cases than the LTDR cases (43.78% versus 40%; p<0.0046) and were significantly older (50.3 year versus 41.7 year for LTDR group; p<0.0001). Length of stay was higher than for LTDR cases (7.2 days versus 5.4 days for LTDR; p<0.0001).

3.3.2.3 Geographic variation of lumbar total disc replacement use

Thirty-three out of the 105 Belgian acute hospitals performed at least one of the 240 LTDRs recorded in 2011. As shown in Figure 15 the number of procedures performed per hospital ranged from 1 to 76, with a yearly median of 2 procedures. The mean number was 7.3 (SD: 13.7). Most hospitals (n=26, 79%) did not perform 10 procedures, and 14 of these (42%) performed only one procedure. LTDRs are not concentrated in academic hospitals and hospitals with university beds. Nevertheless one institution performed 76 LTDRs, which represents 31.7% of all procedures performed in Belgium.

Figure 15 – Number of lumbar total disc replacement per hospital performing this technique in 2011
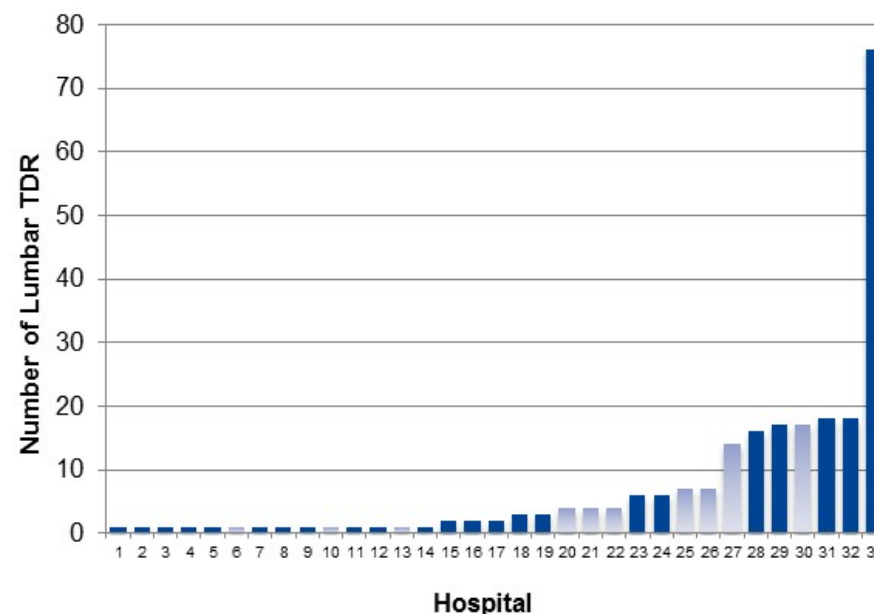
Source RHM – MZG 2011 – Academic hospitals and hospitals with university beds are in light shade

Figure 16 shows the number of procedures performed in the hospitals of a district per 100 000 inhabitants. Only 8 LTDRs were performed in the Brussels Capital Region, which represents 0.7 procedures per 100 000 inhabitants. The highest number of procedures reached 15.9 procedures per 100 000 inhabitants in the arrondissement of Liège (97 procedures in total).

Figure 16 – Number of lumbar total disc replacements performed in 2011 per hospital district ('arrondissement')

Number of lumbar TDR / 100 000 inhabitants

- 10 - 16
- 8 - <10
- 6 - <8
- 4 - <6
- 2 - <4
- >0 - <2
- 0
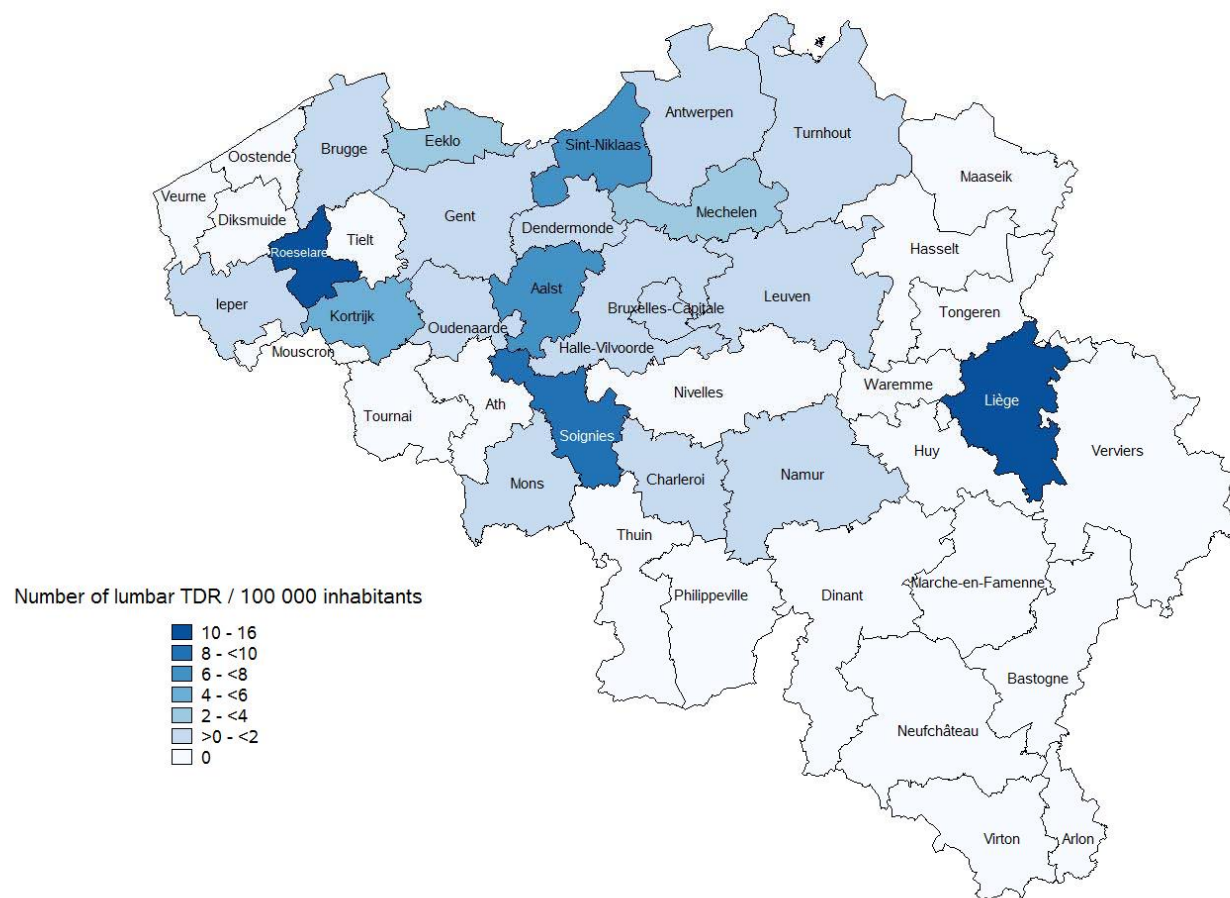
Source: RHM – MZG 2008-2011 - Population DGSIE – ADSEI 1 January 2011.

- **One out of five Belgian adults declared that he/she suffered from low back disorder or other chronic back defect in the last 12 months. Age related degenerative processes in the intervertebral disc is one of the most common aetiology of low back pain.**

- **Between 10 and 20% of people suffering from lumbar DDD do not respond to conservative treatments and may be considered candidates for lumbar surgery. Lumbar total disc replacement is an alternative to fusion of vertebrae which is the most common form of surgery.**

- **Four lumbar prostheses are currently reimbursed in Belgium and represent the vast majority of the implanted models. The reimbursement currently amounts to €1800 per prosthesis and the patient out-of-pocket payment for the device is limited to €180 maximum.**

- **The current annual number of lumbar prostheses is 200 discs at the most implanted in about 30 Belgian hospitals. Patients are on average 42 years old and 60% are female.**

3.4 Clinical effectiveness and safety

HTA CORE MODEL DOMAIN: EFF - SAF

3.4.1 Methods

3.4.1.1 Search strategy

A systematic search for HTA reports, systematic reviews and randomised controlled trials on both cervical and lumbar total disc replacement (LTDR) vs. other techniques was carried out in Medline Ovid, Embase and Cochrane databases (CDSR, DARE, HTA, Economic Evaluations and CENTRAL) on October 9th, 2014. This report is an update of a previous KCE report published in 2006,[27] which included a review on the literature for lumbar disc replacement. The current literature search was consequently limited to publications from 2006 and onwards.

Additionally, we consulted INAHTA and the EUnetHTA POP database for HTA reports recently published or ongoing.

The hits from the electronic databases were merged in a unique EndNote file and duplicates were removed. We then screened titles and abstracts to identify and exclude articles which did not fulfil the inclusion criteria (see Table 13). The remaining papers were retrieved and read in full for a final selection of studies to include in the review. No restrictions were imposed for language. While the systematic search for studies were performed on the cervical and lumbar disc replacement combined, the process of selecting potentially relevant articles was performed in two separate "rounds", one round for the cervical review and one round for the lumbar one. Bibliographies of selected studies have been hand-searched for possible relevant references. The flow chart of the selection process for both reviews is presented in the appendix, together with details on the search strategies. Systematic reviews that were included in the report was critically appraised with the AMSTAR tool[28] and the RCTs included were appraised with the Cochrane grid for risk of bias.[84] The results of these appraisals can be found in the appendix.

The PICO used for the selection of LTDR literature is presented in Table 13. For the description of the assessment instruments see Table 5 in section 0.

Table 13 – PICO table and selection criteria for lumbar total disc replacement

Selection criteria	Inclusion criteria	Exclusion criteria
Population	Adult patients (18 years of age or older) with chronic lumbar indications, including degenerative disc disease and lumbar disc hernia, refractory to conservative treatment for at least 12 weeks. Both patients with single or double-level disease will be included Literature that mixes cervical and lumbar patients will be considered only if outcomes are reported for each patient group separately	Metabolic bone disease (e.g. osteoporosis), malignancy (spinal metastasis, myeloma), haemangioma, osteonecrosis, trauma
Intervention	Lumbar total disc replacement, all disc types/brands	Partial disc replacement
Comparators	• Conventional surgery, including discectomy, fusion (with plate, cage, autograft, allograft material) alone or in combination with conservative treatment • Conservative treatment alone	
Outcomes	• Back-specific functional status (Oswestry score, Roland-Morris) • Back and leg pain (VAS, NRS or similar scale) • Mobility (ROM) • Neurological outcomes, compared to preoperative status • Patient satisfaction (Odom's outcome criteria or other patient satisfaction scale) • QoL (generic instrument e.g. SF-36, or generic utility instruments e.g. SF-6D, EQ-5D, QWB, HUI) • ASD (adjacent segment degeneration), includes radiographic adjacent segment pathology as well as to clinical adjacent segment pathology • Complications, adverse events (wound infections, heterotopic ossification, medical and surgical adverse event e.g. cardiac, respiratory, dysphagia, intraoperative blood loss) • Revision surgery at operated segment/secondary surgery at adjacent segments (reoperation rate) • Mortality	Working status Length of stay/ operating time
Study design	HTAs, systematic reviews, randomised controlled trials (including non-inferiority RCTs)	Narrative reviews, observational studies, letters, editorials, notes, abstracts

3.4.1.2 Determining clinical relevance

The threshold for determining clinical relevance follows the thresholds described by Jacobs[5] and are defined as follows:

Minimal clinically important difference of the primary outcome measurements (i.e. functional disability, perceived recovery and pain) is defined as a 30% improvement from baseline. This corresponds to a mean difference of 15 for the Visual Analogue scale (VAS) (0-100), 2 for the Numerical Rating Scale (0-10), 5 for the Roland Disability Questionnaire (0-24), 10 for the Oswestry Disability Questionnaire (0-100) and 20 for the Quebec Back Pain Disability Questionnaire (0-100). In case of absence of a clinical relevant point estimate difference, it is evaluated whether the upper bound of the 95% CI is smaller than this difference. No MCID was defined for the mobility which was deemed not clinically relevant in itself.

During the second expert meeting held at KCE on April 28th, 2015, these thresholds were agreed upon by the experts.

3.4.2 Results on Clinical Effectiveness

This chapter presents a review of the evidence on the clinical benefits of lumbar total disc replacement (LTDR) versus fusion alone or in combination with conservative treatment or versus conservative treatment alone. Our literature search identified a number of HTAs and systematic reviews relevant for this review. These included reviews by BlueCross and BlueShield (2014),[85] Jacobs (2012),[5] van den Eerenbeemt (2010),[86] Thavaneswaran (2014),[87] Rao (2014),[88] Wang (2012),[89] Wei (2013),[90] and Yajun (2010).[91]

The systematic review by Jacobs (2012),[5] a Cochrane Review, was of high methodological quality, was comprehensive and aligned with our research questions and had a relatively late search date (December 22, 2011). The other reviews were generally of lower quality, were less comprehensive and also had no additional RCTs included that were not already included in the Cochrane review.

We therefore decided to base our clinical effectiveness and safety review on this Cochrane review which included 7 RCTs: Gornet (2011),[92] Hellum (2011),[42] Berg (2009),[93] Moreno (2008),[94] Sasso (2008),[95] Zigler (2007),[96] and Blumenthal (2005).[97] It is likely that none of the RCTs were blinded although in two of the studies (Gornet, 2011 and Sasso, 2008) blinding of

participants was insufficiently described. Three studies were solely industry sponsored, [92, 95, 97] one study reported conflicting interest statements across publications,[96] one study was founded by non-commercial parties,[42] and one study did not receive financial support.[93]

Six of these RCTs[92-97] compared LTDR with fusion and only the study by Hellum (2011)[42] compared LTDR with conservative treatment (rehabilitation). In this trial rehabilitation took approximately 60 hours during three to five weeks and was conducted by a multidisciplinary team of physiotherapists and specialists in physical medicine and rehabilitation.

All the RCTs included patients with back or leg pain in the presence of degenerative disc disease unresponsive to conservative treatment for at least 12 weeks.

The follow-up of the included studies in the review was 24-months, while only the RCT by Blumenthal (2005)[97] extended to 5 years. The results after 5 years are presented separately, but should be interpreted with caution as the 5-year paper presents follow-up only on a partial cohort of patients defined prospectively before randomisation took place (only 57% of patients were contacted at 5 years). Contrary to the review on CTDR the experts involved in the accompanying group for this report agreed to include studies that mixed 1 and 2-level surgeries for the lumbar part to keep as many study data as possible.

It should be noted that the Cochrane review reported mean difference results favouring LTDR as negative values in all their meta-analyses for back and leg pain and for the Oswestry disability index whereas for mean improvement on the same three measures results favouring LTDR are reported as positive values. For comprehensibility and readability we decided to report results in a similar way (practically this means that the result signs from mean difference results for functional status (Oswestry) and for pain are reversed, whereas for mean improvements are reported in the same way as in the review).

Additionally, we searched for all randomised controlled trials that were not included in the retained systematic review with the aim of updating the results from the systematic review.

This process resulted in the selection of the following additional papers: Zigler (2012),[98] Zigler (2012),[99] Hellum (2012),[100] and Johnsen (2013).[101]

The two publications by Zigler are companion papers providing 5-year results from the RCT by Zigler (2007) that was already included in the Cochrane review for the short-term follow-up. Similarly, the two papers by Hellum and Johnsen are companion papers to the major RCT publication by Hellum (2011) that was already incorporated in the Cochrane review; thus these two papers are both comparing LTDR with rehabilitation.

Further, a recent (search up to January 17, 2014) rapid response report from the Canadian Agency for Drugs and Technologies in Health (CADTH)[102] with a comprehensive literature search and reference list was identified and was used to compare with our literature search. This review identified one systematic review by Jacobs,[5] which is the Cochrane review we decided to base this report on, and four RCTs by Hellum (2012),[100] Hellum (2011),[42] Johnsen (2013),[101] and Yanbin (2011).[103] None of these trials were missed by our search, however, one of the RCTs by Yanbin[103] was not included in the Cochrane review although it was published before the end of their search. After reviewing this study we also excluded it because it compared two different types of lumbar discs as opposed to LTDR compared with fusion. Additionally, the results from one of the RCTs identified by Hellum (2011) was already incorporated in the systematic review by Jacobs.[5] Unfortunately this Canadian report was limited to a reference list and therefore could not be used as comparison for results of the clinical review.

Quality of the evidence

Jacobs et al. provided a GRADE score by outcome: the quality of evidence was considered to be very low (HRQoL, adverse effects), low (back pain) or moderate (functional status, patient satisfaction).[5] Due either to the non-blinded nature of the supplementary RCTs, their small size and/or the incomplete outcome data for some of them we judged the quality of the body of evidence as low quality for LTDR versus fusion or rehabilitation. Consequently, we did not provide a GRADE score for each outcome separately in the scope of this rapid review. Besides, if there are very severe problems for any factor of the GRADE scoring tool, RCT evidence may fall by two levels due to that factor alone anyway.[41, 42]

3.4.2.1 Single and two-level disease – Lumbar total disc replacement vs fusion

Back-Specific Functional status

All RCTs included reported functional status using the Oswestry Disability Index. The Cochrane review reported pooled results at 6 and 24 months. The RCTs by Zigler and Blumenthal additionally reported results after 5 years.

6 months

Two studies included in the Cochrane review (Blumenthal 2005, Moreno 2008) with a total of 309 participants found a significant difference between LTDR and fusion in favour of LTDR on Oswestry scores (MD=9.90; 95% CI 5.49 to 14.31). This result does not reach the threshold for clinical relevance.

24 months

Five studies included in the Cochrane systematic review (Berg 2009, Blumenthal 2005, Gornet 2011, Moreno 2008, Zigler 2007) with a total of 1207 participants found a significant difference between LTDR and fusion in favour of LTDR (MD=6.31; 95% CI 5.49 to 9.68). This result does not reach the threshold for clinical relevance.

Additionally, the Cochrane review reported that the mean improvement in Oswestry score at 24 months in the LTDR group was 4.27 points better (95% CI 1.85 to 6.68) than in the fusion group. This result does not reach the threshold for clinical relevance.

5 years

The RCT by Zigler[98] found that both treatment groups maintained significant improvement in the Oswestry score from 2 to 5 years and that after 5 years the mean improvement in scores were similar for LTDR patients and fusion patients (p=0.4552). The study sets a success criteria of ≥ 15% improvement from baseline for the ODI score and find that after 5 years 76.5 % of fusion patients and 78.6% of LTDR patients fulfil this criteria. This is not either a statistically significant difference (p=0.84). The RCT by Blumenthal[104] similarly did not find a statistically significant difference in mean Oswestry scores between groups at the 5-year postoperative time-point (mean change from baseline was in favour of LTDR with 24 points in the LTDR group and 27.5 points in the fusion group).

Back Pain

The studies included in the Cochrane systematic review reported back pain using the VAS scale (0-100). The review reported pooled results at 6 and 24 months.

6 months

Two studies (Berg 2009, Gornet 2011) with a total of 706 participants found a significant difference between LTDR and fusion in favour of LTDR (MD=6.44; 95% CI 2.34 to -10.55). This result does not reach the threshold for clinical relevance.

24 months

Two studies (Berg 2009, Gornet 2011) with a total of 676 participants found a significant difference between LTDR and fusion in favour of LTDR (MD=5.12; 95% CI 0.64 to 9.60). This result does not reach the threshold for clinical relevance.

Additionally, the review reported that the mean improvement in back pain at 24 months in the LTDR group was 5.22 mm better on the VAS than in the fusion group (95% CI 0.18 to 10.26). This result does not reach the threshold for clinical relevance.

5 years

No evidence was identified

Leg Pain

Leg pain was measured with a numeric rating scale (NRS) to find between group differences or was measured as pre- versus postoperative leg pain scores to assess improvement.

12 months

One study (Gornet 2011)[92] with a total of 524 participants found no significant difference between groups for leg pain at 12 months (no numbers are provided in the primary study).

24 months

One study (Gornet 2011)[92] with a total of 524 participants found no significant difference between groups for leg pain at 24 months (MD=3.6; 95% CI -1.64 to 8.84).Two studies (Berg 2009, Gornet 2011),[92, 93] with a total of 676 participants reported improvement in leg pain at 24 months. The

difference between groups favoured LTDR but was not significant (MD=0.56; 95% CI -4.54 to 5.65).

5 years

No evidence was identified

Quality of life

Quality of life was measured in four studies but results were only reported in two studies. Two RCTs (Gornet 2011, Zigler 2007) included in the Cochrane systematic review reported quality of life using SF-36 (either stated as the SF-36 success rate, which was defined as any improvement from baseline in the composite score of the mental and physical components, or as the actual SF-36 PCS and MCS scores). Additionally, the RCT by Zigler reported results for the SF-36 physical component score at 5 years.

6 months

One study (Zigler 2007) with 236 single level patients found that the percentage improvements in quality of life from baseline were not significantly different (LTDR=80.4% versus fusion=75%; p=0.2333). Another study (Gornet 2011) with 554 patients found that improvements were significant in favour of LTDR (LTDR=15.6 versus fusion=12.3; p<0.001) on the SF-36 PCS scale but were not significant on the SF-36 MCS scale. No meta-analysis could be performed.

24 months

One study (Zigler 2007) with 236 single level patients[96] found that percentage improvements from baseline between groups were not significantly different (LTDR=79.2% versus fusion=70%; p=0.094).

Another study (Gornet 2011) with 524 patients[92] found quality of life improvements were significant in favour of LTDR on the SF-36 PCS scale (LTDR=17 versus fusion=14.3; p=0.009) but were not significant on the SF-36 MCS scale. No meta-analysis could be performed.

5 years

The study by Zigler 2007[98] evaluated quality of life with the SF-36 PCS scale and found that both groups maintained significant improvements in SF-36 PCS compared with baseline. The significant difference between groups in favour of LTDR that was observed at 2 years was not maintained after 5 years (p=0.168). The study by Blumenthal,[104] similarly, did not find a statistically significant between group differences in SF-36 PCS compared

with baseline. The scores for SF-36 PCS were 12.6 points in the LTDR group and 12.3 points in the fusion group.

Mobility

Mobility was measured as range of motion (ROM) in degrees. Results in the review were reported after 24 months.

24 months

Four RCTs (Blumenthal 2005, Gornet 2011, Zigler 2007, Berg 2009) measured ROM and all studies found that ROM 2 years after a LTDR was comparable to preoperative status whereas ROM for fusion was nearly zero. The clinical implications of this are not clear.

5 years

The RCT by Blumenthal[104] found that the mean ROM at index level was 6.0° for LTDR patients and 1.0° for fusion patients. The clinical implications of this are not clear.

Satisfaction

Four studies used a dichotomized measure for patient satisfaction and results could be pooled at 24 months. One study used a continuous 100 mm VAS measure for patient satisfaction and is reported separately for 24 months and 5 years.

24 months

Four studies (Berg 2009, Blumenthal 2005, Gornet 2011, Moreno 2008) with a total of 958 patients found a significant difference in favour of LTDR (OR=1.93; 95% CI 1.36 to 2.76). It should be noted that the significant result is driven by the effect from two sponsored studies whereas the two non-sponsored studies did not find a significant effect.

One study (Zigler 2007)[96] used a continuous 100 mm VAS scale and found a significant difference in favour of LTDR over fusion (9.4 mm mean difference for one-level procedures and 8.8 mm difference for two-level procedures; p=0.015). This result does not reach the threshold for clinical relevance.

5 years

The significant between group difference observed by Zigler at 2 years on the continuous VAS scale was not maintained at 5 years (LTDR mean=78.3, SD=27.1) versus fusion (mean=78.1, SD=27.7; p=0.62).[44] Similarly, the

study by Blumenthal[104] found a non-significant difference in patient satisfaction after 5 years (use of scale was not specified in the paper).

3.4.2.2 *Single and two-level disease – Lumbar total disc replacement vs rehabilitation*

The Cochrane systematic review only identified one RCT by Hellum et al. comparing LTDR with rehabilitation.[42] Our additional search for RCTs did not identify further trials but two companion papers to the major publication were identified[100, 101] and incorporated in the presentation of the results. Consequently, the results below are all based on a single RCT.[42]

Back-Specific Functional status

This outcome was measured with the Oswestry Disability Index.

12 months

There was a statistically significant mean group difference from baseline in improvements of Oswestry scores. The LTDR group scored 8.9 points higher than the rehabilitation group (172 patients, 95% CI 4.77 to 13.03). This result does not reach the threshold for clinical relevance.

24 months

The mean treatment effect (difference between groups) was 8.4 (95% CI 3.6 to 13.2) in favour of LTDR, a statistically significant mean group difference in improvements of Oswestry score. The LTDR group scored 6.9 points higher than the rehabilitation group (172 patients, 95% CI 2.23 to 11.57). These results does not reach the threshold for clinical relevance.

Back Pain

This outcome was measured with the VAS scale (0-100, with lower scores being better).

12 months

Improvement of back pain at 12 months was 14.0 mm higher in favour of the LTDR group (152 patients; 95% CI 5.0 to 23.0 mm) than in the rehabilitation group. This result is statistically significant but does not reach the threshold for clinical relevance.

24 months

Improvement of back pain at 24 months was 12.3 mm higher in favour of the LTDR group (152 patients; 95% CI 3.1 to 21.3 mm) than in the rehabilitation group. This result is statistically significant but does not reach the threshold for clinical relevance.

Leg Pain

This outcome was not reported in the study.

Quality of life

Quality of life was measured with SF-36 (physical component score and mental component score) and with EQ-5D. This outcome is not reported in the systematic review but results could be retrieved from the primary RCT.

24 months

At 24 months there was a significant difference for the SF-36 physical summary score in favour of the LTDR group (MD=5.8; 95% CI 2.5 to 9.1). For the SF-36 mental summary score and EQ-5D there were no significant between group differences at 2-year follow-up.

Mobility

This outcome was not reported in the main publication but in a separate publication.[101] In this last publication the authors determined segmental movement, disc height and sagittal alignment with the aim of assessing the correlation of biomechanical properties to clinical outcomes. Segmental movement in the sagittal plane and disc height were measured using distortion compensated röntgen analysis (DCRA) comparing radiographs in active flexion and extension. Results were reported after 24 months.

24 months

At 24 months no significant change in sagittal plane movement between treatment groups were found. Results are reported per disc level and can be found in the paper.[101] However, at no level were the group differences significant. Additionally, the authors did not find a correlation between the ROM and the patient reported outcomes in any of the two treatment groups.

Patient Satisfaction

Patient satisfaction was measured on a 7 points Likert scale which was recalculated into percentage of patients being satisfied with the outcome i.e. patients scoring a one or two on the scale ("completely recovered" or "much recovered") at 2 years.

24 months

Patients are more likely to be satisfied with the outcome at two years in the LTDR group (OR 2.65, 95% CI 1.42 to 4.96) than in the rehabilitation group.

Conclusions Clinical Effectiveness

Lumbar Total Disc Replacement (LTDR) versus fusion

- **The results for the main outcomes including pain, patient satisfaction and functional status did not at any time-point reach clinical significance.**
- **Overall, the long-term (5-years or more) clinical outcomes for lumbar disc replacement are unclear.**
- **The statistically significancy of the results were the following:**
- o **There is a statistically significant back-specific functional difference after 6 months and 2 years in favour of LTDR measured by the Oswestry Disability Index. Two studies find that this difference is not maintained at 5 years.**
- o **There is a statistically significant difference at 6 months and after 2 years in favour of LTDR for back pain measured by the VAS scale.**
- o **No significant between group difference was found after 1 year or after 2 years for leg pain measured with the NRS scale.**
- o **One study found a statistically significant difference for quality of life on the SF-36 physical component score favouring LTDR at 6 months and after 2 years. The same study did not find any statistically significant difference for quality of life on the SF-36 mental component score measured at the same time-points.**
- o **Another study did not find any significant difference for quality of life at 6 months and after 2 years measured with the SF-36 (as success rate).**

o **Two studies find that the statistically significant difference on the SF-36 physical component score is not maintained at 5 years.**

o **Studies measuring mobility consistenly find that mobility (ROM) after 2 years in the LTDR group was comparable to preoperative status whereas ROM for fusion was nearly zero. The clinical implications of this are not clear.**

o **Patient satisfaction after 2 years is high in both LTDR and the fusion groups but is statistically significantly higher in the LTDR group. However this results is driven by the industry sponsored studies. One RCT found that patient satisfaction was similar between groups after 5 years.**

Lumbar Total Disc Replacement (LTDR) versus rehabilitation

• **Results are based on one single RCT and should be interpreted with great caution**

• **The results for back pain and functional status did not reach clinical significance.**

• **The statistically significancy of the results were the following:**

o **There is a statistically significant back-specific functional difference after 1 year and 2 years in favour of LTDR measured by the Oswestry Disability Index.**

o **There is a statistically significant difference after 1 year and 2 years in favour of LTDR for back pain measured by the VAS scale.**

o **There is a statistically significant difference in quality of life after 2 years in favour of LTDR for the SF-36 physical summary score but no significant difference for the SF-36 mental summary score and the EQ-5D after 2 years.**

o **No significant group difference was found for moblity after 2 years.**

o **Patient satisfaction was statistically significantly higher after 2 years in the LTDR group**

3.4.3 Results on Safety

This chapter presents a review of the evidence on the safety of lumbar total disc replacement (LTDR) versus fusion alone, fusion in combination with conservative treatment or versus conservative treatment alone. The results are based on the same literature as for the clinical effectiveness part, namely the systematic review by Jacobs (2012),[5] the 5-year results provided by Zigler[98] and Blumenthal[104] and additionally, for the comparison between LTDR and conservative treatment (rehabilitation) an additional publications from the one RCT which was included in the Cochrane review. It was agreed by the experts not to include observational study data for the safety review due to study limitations and much selected patient groups.

3.4.3.1 Single and two-level disease - Lumbar total disc replacement vs fusion

Reoperations

This outcome was reported in the review as an overall rate of reoperations. In the majority of the studies this outcome included both revisions, removals and reoperations (all index levels), and in the largest of the studies also additional fixations. It should also be noted that there is large inconsistency amongst studies.

Additionally, the RCT by Zigler reported "device success" after 5 years defined as the absence of any reoperation to modify or remove implants and no need for additional fixation.

24 months

In five studies (Berg 2009, Blumenthal 2005, Gornet 2011, Moreno 2008, Zigler 2007) 63 of 810 patients in the LTDR group (7.8%) and 35 of 384 patients in the fusion group (9.1%) had reoperations (OR 0.8, 95% CI 0.51 to 1.24).This is not a statistically significant difference.

5 years

The study by Zigler[98] found that "device success" was achieved after 5 years in 93.2% of LTDR patients and 93.3% of fusion patients. This is not a statistically significant difference. The study also reported that secondary surgeries at the index level occurred in 9 fusion patients (12%) and 13 LTDR patients (8%), p=0.048. This is also not a statistically significant difference. The study by Blumenthal[104] reported that 7 of 90 patients in the LTDR group

(7.7%) and 7 of 43 patients in the fusion group (16.3%) had additional fixations/index level reoperation after 5 years. This was not significant measured with the Fischer Exact test (p=0.144).

Adjacent segment degeneration

This outcome was in the Cochrane review only reported in one study by Berg 2009,[93] and was in this case reported as the number of patients having "adjacent segment problems" up to 24 months. However, after 5 years the RCT by Zigler[98] and the RCT by Blumenthal (Guyer, 2009)[104] reported adjacent-level degenerative changes in patients with single-level disease. In the study by Zigler, adjacent segment degeneration (ASD) was characterized by a composite score including disc height loss, endplate sclerosis, osteophytes and spondylolisthesis. The Blumenthal study reported clinical adjacent segment pathology (ASP) as percentage of patients having clinical ASP treated with pain management. However, the results from this trial should be interpreted with caution as only a partial cohort of the randomised patients were included for follow-up. Consequently, there is a risk of reporting and selection bias for the long-term part of this study.

24 months

In the study by Berg 2009, the difference in adjacent segment degeneration was not statistically significant after 24 months. It was reported that 6 of 72 patients in the fusion group had adjacent segment problems and one out of 80 patients in the LTDR group suffered from this condition.

5 years

The RCT by Zigler found that changes in adjacent level disease at 5 years were observed in 9.2% of LTDR patients and 28.6% of fusion patients (p=0.004). Among the patients without adjacent-level disease preoperatively, new findings of ASD at 5 years post treatment were apparent in 6.7% of LTDR patients and 23.8% of fusion patients (p=0.008). The RCT by Blumenthal found that 4.4% (4/90) of LTDR patients versus 14.0% (6/43) of the fusion patients had clinical ASP after 5 years.

Thromboembolic complications

Thromboembolic complications were reported in two studies, in one study as venous thrombosis and in another study as "cardiovascular events" which was not further defined. No meta-analysis could be performed due to a lack of definition for "cardiovascular event" in one of the studies.

24 months

In a study by Blumenthal[97] with 304 patients, two cases of venous thrombosis were observed in the LTDR group and none in the fusion group. In the study by Gornet[92] with 577 patients, one cardiovascular event was observed in the LTDR group and none in the fusion group. None of these results are statistically significant.

Intraoperative blood loss

Intraoperative blood loss was reported in 5 studies (Berg 2009, Blumenthal 2005, Gornet 2011, Moreno 2008, Zigler 2007) and was measured in ml. The evidence on intraoperative blood loss is conflicting and studies are reporting results in very different directions. If all studies are pooled the results slightly favour LTDR over fusion but with a very large confidence interval (MD=-37.22, 95% CI -185.06 to 110.62), therefore no firm conclusion can be drawn for this outcome.

3.4.3.2 *Single and two-level disease - Lumbar total disc replacement vs rehabilitation*

The results below are based on one study by Hellum (2011). The original paper[42] which is reported in the Cochrane systematic review had a limited reporting regarding safety/adverse events. On adjacent level degeneration and facet joint arthroplasty we were able to update with information from an additional publications from the same RCT.[100]

Reoperations

This outcome was reported as subsequent surgeries (repeat surgeries or additional spinal surgeries within 24 months).

24 months

There was no statistical difference between groups in the rate of reoperations. Out of 77 patients in the surgery group, one patient had a repeat surgery, two patients went on to have a fusion one of these patients at level with disc prosthesis and the other at the level above. Additionally, two patients had a resection of spinous process because of possible painful contact between adjacent levels.

Adjacent segment degeneration (ASD)

This outcome was not reported in the original publication but was reported in a companion paper.[100] This paper assessed ASD by evaluating Modic changes (changes in adjacent vertebral bone marrow), posterior high intensity zone in the disc, nucleus pulposus signal, disc height, disc contour and index level facet arthropathy (FA).

24 months

ASD developed with similar frequencies (no significant difference) in patients who were (n =59) and were not (n =57) treated with surgery. Results are provided for each evaluation parameter and all results are non-significant, including the FA decrease in 1 patient (2%) both in the surgery group and the rehabilitation group.

However, in patients treated with surgery, index level FA appeared or increased in 20 patients (34%) versus 2 patients in the rehabilitation group. This difference is significant (p<0.001).

Conclusions Safety

Lumbar total disc replacement (LTDR) versus fusion

- **There is insufficient evidence to determine long-term safety outcomes for lumbar disc replacement versus fusion**
- o **There is no statistically significant group difference for intraoperative blood loss.**
- o **After 2 years and 5 years there is no statistically significant group difference for the overall rate of reoperations.**
- o **After 2 years there is no statistically significant group difference for patients with "adjacent segment problems". However, one RCT finds a between group difference in favour of LTDR for adjacent-level degenerative changes after 5 years.**
- o **After 2 years there is no statistically significant group difference for patients with "tromboembolic complications".**

Lumbar total disc replacement (LTDR) versus rehabilitation

- **There is insufficient evidence to determine long-term safety outcomes for lumbar disc replacement versus rehabilitation**

- o **After 2 years there is no statistically significant group difference for the rate of reoperations.**
- o **After 2 years there is no statistically significant group difference for the rate of adjacent segment degeneration (ASD).**

3.4.4 Discussion

From the literature review on one and two-level lumbar disc degeneration there was a tendency for clinical results to be in favour of lumbar total disc replacement compared with conventional fusion or rehabilitation in the short term whereas the long-term results are quite unclear. However, although statistically significant, in the short term, the differences between the interventions did not reach the clinically relevant thresholds, as set by the experts, for the main functional and clinical outcomes. This includes pain, functional status (Oswestry) and patient satisfaction. The categorization of the clinical relevance may impede to draw clear conclusions when the outcomes mean differences are close to the chosen cut-offs (e.g. 9.99% of the scale or RR=0.80001). Nevertheless the results were never borderline following the clinical relevance criteria.

The literature review on lumbar total disc replacement resulted in the identification of seven RCTs. These RCTs suffer from limitations with regards to quality (no blinding), to outcome reporting and presented in particular a too limited length of follow-up. Firstly, the prevention of adjacent segment degeneration was one of the main rationales for the introduction of disc replacement. This was hoped to be a result of preserved motion at index level and consequently reduced adjacent level forces. However, because the average patient receiving total disc replacement is relatively young, and because adjacent segment degeneration evolves slowly this needs to be evaluated based on truly long-term results. Although most RCTs are started long ago only few studies have published 5-year results. It is therefore difficult to make evidence based recommendations related to adjacent segment degeneration. Similar to cervical total disc replacement, it is clear that fusion provides, to a large extent, a loss in range of motion at index level compared to disc replacement. However, this is commonly compensated for at adjacent levels and through hip movements thus providing no real difference in terms of functional mobility for the patient. Secondly, with the introduction of disc replacement there was a hope for a decreased rate of

reinterventions. There does, however, not appear to be a difference in the reintervention rate between lumbar total disc replacement and fusion. Additionally, it was stressed by the experts that it is a Belgian practice experience that the reintervention of a lumbar total disc replacement is often complex and will end in a fusion surgery (conversion surgery). This argument, however, will require much larger study populations to be verified.

Additionally, this analysis is based on a highly selected patient population, and the experts argued that the question about which patients and at which stage/time in their disease development total disc replacement is preferred over conservative treatment or fusion, remains unanswered. It was also noted that in this patient population it is often difficult to determine the main origin of the pain and whether it truly originates from degenerative disc disease or other physical structures including muscles and ligaments.

Psychosocial factors might also play an important role in this patient group. One RCT showed equal improvement for main outcomes, including back pain, in patients with chronic low back pain and disc degeneration randomised to cognitive behavioural intervention and exercises, or lumbar fusion. Although not studied, it was argued by the experts, that this likely could be the case in total disc replacement patients as well.[105]

3.5 Economic evaluation

HTA CORE MODEL DOMAIN: ECO

3.5.1 Introduction

This section aims at reviewing the peer-reviewed full economic evaluations of lumbar total disc replacement (LTDR) versus conventional surgery (including discectomy, fusion) and/or conservative treatment (CT) in adult patients with chronic lumbar indications (including degenerative disc disease and lumbar disc hernia) refractory to conservative treatment for at least 12 weeks.

3.5.2 Methods

3.5.2.1 Inclusion and exclusion criteria

The following criteria were developed:

- Inclusion and exclusion criteria describing the <u>Population</u>, the <u>Intervention</u> and the <u>Comparator</u> of interest are defined as in the methods chapter for the clinical effectiveness and safety assessment of the LTDR. These criteria can be found in section 3.4.1.
- <u>Design</u>: only full economic evaluations are considered, i.e. studies comparing at least two alternative treatments in terms of both costs and outcomes. Cost-minimization analyses (CMA), cost-utility analyses (CUA, with results expressed as incremental cost per quality-adjusted life year (QALY) gained), cost-effectiveness analyses (CEA, with results expressed as cost per life year (LY) gained) and cost-benefit analyses (CBA, with a monetary valuation of health outcomes) are eligible. Cost comparisons (not considering health outcomes), cost-outcome descriptions (not considering an alternative treatment) and cost-consequence analyses[d] are not relevant for inclusion. Both primary studies and reviews of full economic evaluations are relevant for inclusion; letters, news, conference proceedings and editorials were removed.
- <u>Timing</u>: a previous KCE report published in 2006 performed a systematic literature review of economic evaluations on the same interventions.[27] That search was performed up to early 2006 but no original full economic evaluation was found. The current search is thus limited to publications from 2006 up to April 2015.

---

[d] In cost-consequence analyses, both costs and outcomes of different alternatives are described. In such studies however, an incremental cost-effectiveness ratio (ICER) is not calculated or the results are expressed in disease-specific outcome (e.g. incremental cost per patient reduction of one point in the Activities of Daily Living scale. Such studies are discarded as their results cannot be compared with those of other types of economic evaluations, mainly CEA and CUA.

3.5.2.2 Search strategy

Both electronic and manual searches were performed.

- Electronic search: the following databases were searched in April 2015: Medline(Ovid), Medline(Ovid) in-process and other non-indexed citations, Embase, CRD (Centre for Review and Dissemination) HTA and CRD NHS EED (National Health Service Economic Evaluation Database). A combination of MeSH, EMTREE and text word terms related to total disc replacement, lumbar vertebra, arthroplasty and degenerative disc disease were combined with those related to full economic evaluations (see appendix). The websites of the HTA institutes listed on the International Network of Agencies for Health Technology Assessment (INAHTA) were also consulted.
- Manual search: the reference lists of relevant review papers and full economic evaluations were scrutinized for additional relevant articles.

3.5.2.3 Selection procedure

The selection was performed in a two-step procedure: initial assessment of the title, abstract and keywords; followed by full-text assessment if the title or the abstract or the keywords suggested relevant information, or if no abstract was available. The procedure was performed on the cervical and lumbar total disc replacement techniques combined, but the results with the final selection of the full-text articles retained are presented for each type of procedure separately. The flow chart of the selection processes can be found in appendix.

The search strategy yielded 126 unique, potentially relevant citations on the cervical and lumbar disc replacement techniques combined. After title and abstract review, 73 articles were excluded, the majority of which were not economic studies or only partial economic evaluations (i.e. mostly cost-comparisons). Of the 53 full-text articles reviewed, 41 were excluded. The primary reason for exclusion was that most studies focused on efficacy/safety without reporting any economic result. Other reasons for exclusion were that the design of the studies was limited to a cost comparison (5 studies),[48-52] or a cost-outcome description (1 study).[53]

A total of 12 articles were retained: 9 full economic evaluations among which 4 related to LTDR,[76, 106-108] and 3 literature reviews including 2 on LTDR.[60, 109]

3.5.2.4 Coverage of the reviews

The review by North et al.[109] did not report any methodology or search date for the literature search/selection procedure. Their search for evidence on the cost-effectiveness of LTDR covered 2 articles published in 2011 and 2007: the full economic evaluation by Fritzell et al.,[106] also identified in our review, and the study by Levin et al.[110] appraised here (and also by the authors of the review themselves) as a cost comparison analysis not considering the effectiveness of the interventions; and therefore discarded from our review.

In the second review by Kramer et al.,[60] no methodology for the literature search/selection procedure and no search date were reported either. Their review about the value of LTDR versus fusion returned 4 studies published over the period 2007-2011; one was the full economic evaluation by Fritzell et al.,[106] also identified in our review, the remaining three articles[49, 110, 111] were cost comparisons (also reported as such by Kramer et al.[60]) that were discarded from our review according to our exclusion criteria.

The three most recent full economic evaluations[76, 107, 108] identified in our literature search were not covered by any of the reviews. Therefore, after critical analysis, none of the selected reviews was retained. Our review was thus based on the primary full economic evaluations by Johnsen et al.,[107] Parkinson et al.,[108] MSAC[76] and Fritzell et al.[106]

3.5.2.5 Data extraction and costs conversion

Data extraction sheets, providing a comprehensive and detailed summary, were developed for each full economic evaluation (see appendix). The quality of the studies was assessed narratively.

Past costs reported in the studies were converted into 2014 costs using local Consumer Price Indices. For each country, the Purchasing Power Parities indices were used to convert the original costs into Euro values (for the whole Euro area, not for a specific country of the Euro zone). Both indices were obtained from the OECD website.[63] If no costing year was mentioned in a study, an interval of two years before the publication date was chosen. The original cost figures (i.e. before conversion) can be found in the extraction sheets.

3.5.3 Characteristics of the economic evaluations

Table 14 gives an overview of the characteristics of the four publications assessing the cost-effectiveness of LTDR. Two of them were closely related; the 2013 publication by Parkinson et al.[108] being a summary of the full HTA report published in 2011 by the Australian Medical Services Advisory Committee (MSAC).[76] In the remaining of this review, both publications are treated together and are referred to as the Australian study.

Table 14 – Base-case characteristics of the full economic evaluations of lumbar total disc replacement

	Johnsen et al., 2014[107]	Parkinson et al., 2013[108] and MSAC Application, 2011[76]	Fritzell et al., 2011[106]
Country	Norway	Australia	Sweden
Funding source§	No industry support	No industry support	Industry support
Study design	CUA	CUA	CUA
Perspective	Society (HCP†)	HCP	Society and HCP
Time horizon	2 years	2 years	2 years
Discount rate‡	0% (as short time horizon)	0% (as short time horizon)	0% (not clearly reported)
Costing year	2012	2011	2006
Analytic technique	Within trial	Markov model	Within trial
Comparators	LTDR, CT	LTDR, PLIF, PLF	LTDR, fusion (PLIF or PLF)
Disease level	1- or 2-level	1- or 2-level	1- or 2-level
Target patient group	41-year-old patients with chronic low back pain (> 1 year) who have failed conservative treatment	40-year-old patients suffering from significant axial back pain and/or radicular (nerve root) pain, secondary to disc degeneration or prolapse, who have failed conservative treatment	40-year-old patients suffering from chronic low back pain (> 1 year) who have failed conservative treatment
Outcome	QALY	QALY	QALY
Underlying instrument for QoL	EQ-5D and SF-36 mapped to SF-6D	EQ-5D	EQ-5D

	Johnsen et al., 2014[107]	Parkinson et al., 2013[108] and MSAC Application, 2011[76]	Fritzell et al., 2011[106]
Duration of QoL differential effect	2 years	2 years	2 years
Source for QoL differential outcome	ProDisc-L trial by Hellum et al., 2011[42]	CHARITE, ProDisc-L or Maverick trial by Berg et al., 2009[93]	CHARITE, ProDisc-L or Maverick trial by Berg et al., 2009[93]
Other differential effects	Overall reoperation rate with LTDR	Different index level reoperation rates (replacement, removal, supplementation, revision, other reoperation) with LTDR, PLIF and PLF	Reduced overall reoperation rate (mostly removal of implant) with LTDR
Source for other differential effects	ProDisc-L trial by Hellum et al., 2011[42]	Review of 4 trials: CHARITE by Blumenthal., 2005,[97] ProDisc-L by Zigler., 2007,[96] CHARITE, ProDisc-L or Maverick by Berg et al., 2009[93] and FlexiCore by Sasso et al., 2008[95]	CHARITE, ProDisc-L or Maverick trial by Berg et al., 2009[93]
Uncertainty analyses	Probabilistic sensitivity analysis Scenario analyses	One-way sensitivity analyses Scenario analysis	Probabilistic sensitivity analysis Scenario analyses

CUA: cost-utility analysis, HCP: health care payer, PLF: posterolateral fusion, PLIF: posterior lumbar interbody fusion, ASD: adjacent segment degeneration, LTDR: lumbar total disc replacement, DDD: degenerative disc disease, CT: conservative treatment, QALY: quality-adjusted life year.
§ Funding source was classified using three categories. Industry support: any explicit acknowledgment of support from private industry (generally manufacturers of medical devices). Unknown: no information about support provided. No industry support: explicit acknowledgment of support that did not include private industry (generally from either public sources or private foundations). An explicit statement of "funding source: none" was also included in this last category. † The base-case perspective of this study is societal. To comply with Belgian guidelines, results from a HCP perspective were added based on personal computations from the costs reported in the original study. ‡ For both costs and effects.

3.5.3.1 Country and study design

The studies were performed in Norway,[107] Sweden[106] and Australia.[76, 108] All of them were cost-utility analyses, with outcomes expressed as QALYs. This design seems appropriate because LTDR aims at improving the quality of life of the patients, rather than extending their lives.

3.5.3.2 Perspective

The Australian study[76, 108] and Fritzell et al.[106] adopted a HCP perspective in their base-case, including direct medical and intervention costs only. Fritzell et al.[106] further adopted a societal perspective (with indirect productivity costs for work absenteeism and direct non-medical costs such as travel costs, etc.) as a second base-case.

In Johnsen et al.,[107] a societal perspective is used, adding also the productivity costs due to work absenteeism from the patients and the informal care costs due to relatives caring for a sick patient. This does not

correspond to the recommendation by the recent Belgian guidelines on economic evaluations[68] to use a health care payer perspective in the base-case assessment of a new intervention. As a consequence, results from Johnsen et al.,[107] were re-computed excluding the productivity and informal care costs.

3.5.3.3 Time horizon and discount rate

The time horizon of the studies was limited to 2 years, as in their source RCTs. This is too short to capture the long-term costs and consequences following LTDR and fusion (e.g. adjacent segment degeneration, revision surgery). None of the studies explored the impact of longer time horizons.

Modelling could have been used to extend the 2-year time horizons of the studies. However, this would require merging data from many different sources along with making assumptions on e.g. the duration and the extent of the treatment effect. The clinical assessment section (see section 3.4.2.1) identified two RCTs[97, 98] of LTDR versus fusion with a 5-year follow-up in which no statistically significant difference in quality of life was found. Longer-term data are currently unavailable.

Due to the short time horizons, costs and outcomes were not discounted in none of the studies.

3.5.3.4 Intervention and comparators

In Johnsen et al.,[107] LTDR was compared to conservative treatment, described as an outpatient programme focusing on exercises and cognitive behavioural interventions, and lasting for approximately 60 hours during 3 to 5 weeks. In the Australian study[76, 108] and in Fritzell et al.,[106] LTDR was compared to fusion. Either specific fusion procedures were considered separately (i.e. PLIF and PLF)[76, 108] or the concept of fusion was considered,[106] irrespective of the procedure used. All studies mixed patients with one- and two-level lumbar disc degeneration surgeries.

3.5.3.5 Quality of life effect

Johnsen et al.[107] based the QoL impact of LTDR versus conservative treatment on EQ-5D and SF-36 (converted to SF-6D) values, both collected during the ProDisc-L trial.[42] Surprisingly, the EQ-5D QoL input values used in Johnsen et al.[107] differed from the 1- and 2-year values reported in the original trial by Hellum et al.,[42] with no explanation provided for this

difference. Further Johnsen et al.[107] reported that a significant difference was found at 1- and 2-year of follow-up in favour of LTDR, although this was only the case for the 1-year results in Hellum et al.[42] The SF-6D QoL input data used in the cost-utility analysis by Johnsen et al.[107] were not reported. Note that, in order to increase the comparability of the results between the studies, the Belgian guidelines on economic evaluations[68] recommend the use of EQ-5D as the outcome measure for the cost-utility analyses.

In the Australian study[76, 108] and in Fritzell et al.,[93, 106] the QoL impact of LTDR versus fusion was obtained from the EQ-5D valued collected from the CHARITE, ProDisc-L or Maverick trial by Berg et al.[93] It is surprising though that the QoL input values used in Fritzell et al.[106] differed from the values reported in the original trial by Berg et al.,[93] with no explanation provided for this difference.

3.5.3.6 Other differential effects

Being within-trial economic evaluations, Johnsen et al.[107] and Fritzell et al.[106] incorporated the overall reoperation rates observed in their companion trials.[42, 93] No distinction was made in the rates between reoperations performed at the index and at the adjacent levels. In the study by Johnsen et al.,[107] comparing LTDR with conservative treatment, a reoperation rate of 7.3% with LTDR was used. In Fritzell et al.,[106] the reoperation rates were 10% with LTDR versus 36% with lumbar fusion. This does not seem to coincide with the values stated in their source RCT,[93] where similar reoperation rates for LTDR and fusion (10%) were reported. No explanation for this potential discrepancy was provided.

In the Australian study,[76, 108] index-level reoperation rates (for replacement, removal without replacement, supplementation with further fixation, revision not involving replacement or supplementation or removal, and other reoperations not involving the device itself, separately) were obtained from a review of 4 RCTs,[93, 95-97] from which average values were derived. The between groups differences (expressed as relative risks with PLF or PLIF versus LTDR) were not always in favour of LTDR (e.g. replacement and other reoperation risk were lower for fusion versus LTDR). However, except for the rate of removal without replacement (0% with LTDR and PLIF, 47.7% with PLF), none of the relative risks were found to be statistically significant. Adjacent-level reoperations were not considered in these studies, due to a lack of evidence and as these were estimated as rare by the authors.

3.5.4   *Results of the economic evaluations*

Table 15 gives an overview of the results of the three studies assessing the cost-effectiveness of LTDR.

3.5.4.1   *LTDR versus conservative treatment*

In Johnsen et al.,[107] a significant overall QoL differential effect in favour of LTDR versus CT was found, although the QoL gain measured with the SF-6D (0.11, 95% CI 0.05 to 0.17) was much lower than the EQ-5D QoL gain (0.34, 95% CI 0.18 to 0.50). From their base-case societal perspective, the ICER of LTDR compared with conservative treatment was found to be favourable with the EQ-5D QoL data (€41 424 per QALY gained), but unfavourable with the SF-6D QoL data (€133 738 per QALY gained), at the Norwegian threshold of €74 600 (Norwegian KR 500 000) per QALY gained. Excluding indirect productivity costs (lost work days) and informal care costs (care provided by relatives), i.e. adopting a health care payer perspective as recommended in the Belgian guidelines,[68] the ICER decreased to €20 355 (with EQ-5D) and €62 917 (with SF-6D) (own computations), which could be considered as cost-effective according to the Norwegian threshold. Results from the HCP perspective were more favourable as patients undergoing surgery required on average more care by their relatives (65 days, SD 157) compared to conservative treatment (35 days, SD 93); such that excluding these costs favoured LTDR. The number of days spent out of work on average by the patients was similar in both groups; i.e. 225 (SD 195) for LTDR and 219 (SD 210) for conservative treatment.

3.5.4.2   *Lumbar total disc replacement versus fusion (PLIF and/or PLF)*

The incremental QALYs for LTDR versus fusion were opposite with a small gain of 0.01 QALY in favour of LTDR in Fritzell et al.,[106] and a loss of 0.01 QALY in favour of fusion in the Australian study.[76, 108] The small QALY gain in Fritzell et al.[106] is further reported to be non-significant.

This minimal and non-significant QALY gain prompted Fritzell et al.[106] to conclude that it was impossible to determine whether LTDR or fusion was more cost-effective within the 2-year time frame, even though there was a non-significant (societal) or significant (HCP) cost difference in favour of LTDR. From the HCP perspective, fusion was significantly more costly than LTDR mainly due to a high reoperation rate assumed in this group, that we could not confirm from the source RCT.[93]

Fritzell et al.[106] presumed that patients receiving a disc prosthesis would return to work earlier than fusion patients, and that this could be used as an additional statement in favour of LTDR. The number of sick leave days for the patients in this study was indeed found to be less in the LTDR group than in the fusion group (67 days less with LTDR; 185 (SD 146) vs. 252 (SD 189)). However this difference was not found to be significant.

In the Australian study,[76, 108] the cost-effectiveness of LTDR was found to depend on the comparator. Results were in favour of LTDR when compared with PLIF. However, LTDR was found to be less clinically effective and more costly (thus dominated) than PLF. Although not reported it is most likely that none of those results would be statistically significant. Deterministic univariate sensitivity analyses were performed on many uncertain parameters, of which the QoL differential effect was the most influential. When the upper value of the 95% CI for the QoL effect was used (value not reported by the authors), the results were more favourable to LTDR, with a situation of dominance for LTDR against PLIF and an incremental cost per QALY gained of €811 versus PLF. When the lower value of the 95% CI was used, this was detrimental to the cost-effectiveness of LTDR, with either PLIF or PLF becoming the preferred options compared to LTDR.

Table 15 – Results of the full economic evaluations of lumbar total disc replacement

1- or 2-level disease	Incremental cost (Euro 2014)		Incremental outcome (QALY)		Incremental cost-effectiveness ratio	Authors' conclusion
LTDR versus conservative treatment (rehabilitation)						
Johnsen et al., 2014[107]	Society	€14 074 (-€4627 to €32 778)	EQ-5D	0.34 (0.18–0.50)	€41 424 per QALY gained (€16 664–€68 412)	LTDR is cost-effective compared to CT
			SF-6D	0.11 (0.05–0.17)	€133 738 per QALY gained (€53 493–€229 178)	LTDR is not cost-effective compared to CT
	HCP †	€6921	EQ-5D	0.34 (0.18–0.50)	€20 355 per QALY gained †	Not reported
			SF-6D	0.11 (0.05–0.17)	€62 917 per QALY gained †	Not reported
LTDR versus fusion (PLIF and/or PLF)						
Parkinson et al., 2013[108] and MSAC Application, 2011[76]	PLF	€447	-0.01		Dominated	LTDR is dominated by PLF
	PLIF	-€2571	-0.01		€331 769 saved per QALY lost ‡	The ICER is in favour of LTDR
Fritzell et al.,[106]	Society	-€10 278 (-€25 510 to €5429)	0.01 (NS)		Dominant (NS)	Inconclusive as NS cost and QALY results
	HCP	-€2737 (-€5125 to -€143)	0.01 (NS)		Dominant	Inconclusive as NS QALY results

Note: mean values are reported, with 95% confidence intervals in ( ) when available. HCP: health care payer, QALY: quality-adjusted life year, LTDR: lumbar total disc replacement, ICER: incremental cost-effectiveness ratio, NS: non-significant. PLIF: posterior lumbar interbody fusion, PLF: posterolateral lumbar fusion. All costs expressed in Euro 2014.

† Personal computations excluding the productivity costs for patients work absenteeism and the caregivers' costs reported in the original study. ‡ In the south-west quadrant of the cost-effectiveness plane, an intervention of interest is less expensive but also less clinically effective than its comparator. Thus lower costs are possible, but at the expense of lower benefits. An ICER can be calculated, although its interpretation will be the opposite of the "more frequent" ICER falling in the north-east quadrant where an intervention is more costly and more clinically effective than its comparator. ICERs in the south-west quadrant refer to a cost saving per unit of effect lost, instead of an incremental cost per QALY gained for ICERs in the north-east quadrant. In the south-west quadrant, ICERs greater than a cost-effectiveness threshold are considered acceptable.

3.5.5 Discussion

This literature review highlighted the diversity of the results of the economic evaluations of LTDR. One study found that, compared to conservative treatment, LTDR could potentially be cost-effective when EQ-5D QoL differential effects are used.[107] None of the studies comparing LTDR with fusion (as a whole or PLF/PLIF separately) could demonstrate a significant improvement in QoL with LTDR.[76, 106] In those studies LTDR was found either cost-effective compared to PLIF or not cost-effective (dominated) compared to PLF,[76, 106, 108] while no conclusions could be driven when compared to fusion in general.[106]

Early return to working activities after surgery is often mentioned as an important complementary statement for the use of LTDR. In two studies however, the number of sick leave days up to 2 years after surgery was indeed lower in the LTDR group (185 versus 252 for fusion in Fritzell et al.[106] and 225 versus 219 for conservative treatment in Johnsen et al.[107]) but the differences were never found to be significant.

Our review also highlighted that the quality of some studies could be questioned, as some input data used in the two trial-based studies[106, 107] did not seem to correspond to the data reported in their source trials.[42, 93]

Several limitations were further identified that precludes drawing conclusions on the cost-effectiveness of LTDR so far.

1. **Not confirmed QoL benefit associated with LTDR**. The clinical assessment section (see section 3.4.2) concluded that there was no clear evidence at 2 years of a QoL improvement after LTDR compared to fusion (i.e. significant difference in SF-36 Physical score but not on SF-36 Mental score, nor on EQ-5D). Two studies could not confirm that the advantage with SF-36 Physical score for LTDR was maintained at 5 years.[98, 104] Evidence on QoL improvement after LTDR versus conservative treatment was too scant (only 1 study[42]) to draw any firm conclusion. Although this parameter was found to be highly influential on the results of the cost-effectiveness studies, so far it is not possible to say with certainty if a QoL benefit with LTDR exists. The duration and the extent of this effect over time are also unknown. New studies providing evidence on this effect would be beneficial.

2. **Short time horizon**. A key limitation of the economic evaluations of LTDR is their very short time horizon, limited to 2 years. However device failure and complications will certainly occur after 2 years and the associated costs and impact on health outcome should be considered. Modelling should be used to extend the time horizon of the studies, and different extrapolation assumptions should be assessed in the scenario analyses. Studies documenting the long-term safety of those interventions would also be beneficial.

None of the four economic evaluations was performed in Belgium, with costs and outcome data reflecting the Belgian health care system and organisation. The results of those economic evaluations are therefore not directly transferable to Belgium. Context-specific studies are always preferred and as such only the development of a model reflecting the Belgian situation and consequences could provide a valid assessment of the cost-effectiveness of the use of LTDR in Belgium. However the literature review highlighted the high long-term uncertainty surrounding some crucial baseline parameters determining the cost-effectiveness of LTDR. Those uncertainties should first be addressed before providing any further economic evaluation on this topic.

Therefore, given the current lack of high-quality economic evaluations and awaiting better long-term information on crucial input parameters (QoL effect and safety), it is difficult to currently draw definite conclusions regarding the cost-effectiveness of lumbar total disc replacement versus fusion or versus conservative treatment in adult patients with chronic lumbar indications refractory to conservative treatment.

■ REFERENCES

1. Delgado-Lo Pez PD, Rodri Guez-Salazar A, Castilla-Di Ez JM, Marti NVV, Ferna Ndez-Arconada O. [Role of surgery in spinal degenerative disease. Analysis of systematic reviews on surgical and conservative treatments from an evidence-based approach]. Neurocirugia (Astur). 2005;16(2):142-57.

2. Boselie TF, Willems PC, van Mameren H, de Bie R, Benzel EC, van Santbrink H. Arthroplasty versus fusion in single-level cervical degenerative disc disease. Cochrane Database of Systematic Reviews. 2012;9.

3. Goertz M, Thorson D, Bonsell J, Bonte B, Campbell R, Haake B, et al. Adult Acute and Subacute Low Back Pain. 2012. Health Care Guideline Available from: https://www.icsi.org/_asset/bjvqrj/LBP.pdf

4. Belgian Health Interview Survey - Interactive Analysis (HISIA) [Web page].2013 [cited 24 November 2014]. Available from: https://hisia.wiv-isp.be/SitePages/Home.aspx

5. Jacobs W, Van der Gaag NA, Tuschel A, de Kleuver M, Peul W, Verbout AJ, et al. Total disc replacement for chronic back pain in the presence of disc degeneration. Cochrane Database of Systematic Reviews. 2012;9.

6. Teraguchi M, Yoshimura N, Hashizume H, Muraki S, Yamada H, Minamide A, et al. Prevalence and distribution of intervertebral disc degeneration over the entire spine in a population-based cohort: the Wakayama Spine Study. Osteoarthritis Cartilage. 2014;22(1):104-10.

7. Medical Advisory Secretariat (MAS). Artificial disc replacement for lumbar and cervical degenerative disc disease- update: an evidence-based analysis. Report. Toronto: Ontario Ministry of Health and Long-Term Care; 2006.

8. Grob D. Surgery in the degenerative cervical spine. Spine (Phila Pa 1976). 1998;23(24):2674-83.

9. Spine Society of Australia. Use of artificial disc replacement in patients with cervical degenerative disc disease. Medical Services Advisory Committee; 2010, January. Available from:

http://www.msac.gov.au/internet/msac/publishing.nsf/Content/1145-public

10. BlueCross BlueShield Association. Artificial intervertebral disc arthroplasty for treatment of degenerative disc disease of the cervical spine. Report. 2011 November 2011. Chicago IL: Blue Cross Blue Shield Association (BCBS) PUB: BlueCross BlueShield Association (BCBS) 265 Available from: http://www.google.be/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&cad=rja&uact=8&ved=0CCEQFjAAahUKEwjV1ofyo9jlAhXh_HIKHeMqCzA&url=http%3A%2F%2Fwww.bcbs.com%2Fblueresources%2Ftec%2Fvols%2F28%2F28_13.pdf&usg=AFQjCNFQzx77s-hAWxT14o3USl9NhFkErQ

11. Nandyala SV, Marquez-Lara A, Fineberg SJ, Singh K. Comparison between cervical total disc replacement and anterior cervical discectomy and fusion of 1 to 2 levels from 2002 to 2009. Spine (Phila Pa 1976). 2014;39(1):53-7.

12. Fejer R, Kyvik KO, Hartvigsen J. The prevalence of neck pain in the world population: a systematic critical review of the literature. Eur Spine J. 2006;15(6):834-48.

13. Auerbach JD, Jones KJ, Fras CI, Balderston JR, Rushton SA, Chin KR. The prevalence of indications and contraindications to cervical total disc replacement. Spine J. 2008;8(5):711-6.

14. Walker DH, Mummaneni PV. Current concepts in Spinal Fusion versus Nonfusion. In: Kim DH, Cammisa FPJ, Fessler RG, editors. Dynamic reconstruction of the spine. New York: Thieme; 2006. p. 16-23.

15. Todd AG. Cervical spine: degenerative conditions. Curr Rev Musculoskelet Med. 2011;4(4):168-74.

16. Sekhon LH, Ball JR. Artificial cervical disc replacement: principles, types and techniques. Neurol India. 2005;53(4):445-50.

17. Baaj AA, Uribe JS, Vale FL, Preul MC, Crawford NR. History of cervical disc arthroplasty. Neurosurg Focus. 2009;27(3):E10.

18. Ramani PS, Shoda M, Zileli M, Dohrmann GJ. Surgical Management of Cervical Disc Herniation. Jaypee Brothers Medical Pub; 2012.

19. Porchet F, Metcalf NH. Clinical outcomes with the Prestige II cervical disc: preliminary results from a prospective randomized clinical trial. Neurosurg Focus. 2004;17(3):E6.

20. Goffin J, Casey A, Kehr P, Liebig K, Lind B, Logroscino C, et al. Preliminary clinical experience with the Bryan Cervical Disc Prosthesis. Neurosurgery. 2002;51(3):840-5; discussion 5-7.

21. Vital J-M, Guérin P, Gille O, Pointillart V. Prothèses discales cervicales. In: Techniques chirurgicales - Orthopédie-Traumatologie. Paris: EMC (Elsevier Masson SAS, Paris); 2011. p. 44-162.

22. Levangie PK, Norkin CC. Joint structure and function: a comprehensive analysis. Fourth ed. Philadelphia, PA: F.A. Davis; 2005.

23. Bogduk N, Mercer S. Biomechanics of the cervical spine. I: Normal kinematics. Clin Biomech (Bristol, Avon). 2000;15(9):633-48.

24. Depreitere B, Fransen P, Goffin J, Lubansu A, Put E, Scordidis V, et al. Recommendations of good practice for cervical disc replacement [Web page]. Belgian Neurosurgical Spine Society 2008. Available from: http://www.bnss.be/link/sources/2008-guideline-cervical.pdf

25. Nesterenko SO, Riley LH, 3rd, Skolasky RL. Anterior cervical discectomy and fusion versus cervical disc arthroplasty: current state and trends in treatment for cervical disc pathology. Spine (Phila Pa 1976). 2012;37(17):1470-4.

26. Qureshi SA, Koehler SM, Lu Y, Cho S, Hecht AC. Utilization trends of cervical artificial disc replacement during the FDA investigational device exemption clinical trials compared to anterior cervical fusion. J Clin Neurosci. 2013;20(12):1723-6.

27. Vlayen J, Camberlin C, Paulus D, Ramaekers D. Rapid assessment of emerging spine technologies: intervertebral disc replacement and vertebro/balloon kyphoplasty. Health Technology Assessment (HTA). Brussels: Belgian Health Care Knowledge Centre (KCE); 2006 16/10/2006. KCE Reports 39 Available from: https://kce.fgov.be/publication/report/rapid-assessment-of-emerging-spine-technologies-intervertebral-disc-replacement-a

28. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.

29. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions: Wiley; 2008. p. 187-241.

30. Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion ; Clinical article. Journal of Neurosurgery: Spine. 2013;18(2):154-60.

31. Ren C, Song Y, Xue Y, Yang X. Mid- to long-term outcomes after cervical disc arthroplasty compared with anterior discectomy and fusion: a systematic review and meta-analysis of randomized controlled trials. Eur Spine J. 2014;23(5):1115-23.

32. Luo J, Huang S, Gong M, Dai X, Gao M, Yu T, et al. Comparison of artificial cervical arthroplasty versus anterior cervical discectomy and fusion for one-level cervical degenerative disc disease: a meta-analysis of randomized controlled trials. Eur. J. Orthop. Surg. Traumatol. 2014.

33. Verma K, Gandhi SD, Maltenfort M, Albert TJ, Hilibrand AS, Vaccaro AR, et al. Rate of adjacent segment disease in cervical disc arthroplasty versus single-level fusion: meta-analysis of prospective studies. Spine. 2013;38(26):2253-7.

34. Vaccaro A, Beutler W, Peppelman W, Marzluff JM, Highsmith J, Mugglin A, et al. Clinical outcomes with selectively constrained SECURE-C cervical disc arthroplasty: Two-year results from a prospective, randomized, controlled, multicenter investigational device exemption study. Spine. 2013;38(26):2227-39.

35. Phillips FM, Lee JYB, Geisler FH, Cappuccino A, Chaput CD, Devine JG. A prospective, randomized, controlled clinical investigation comparing PCM cervical disc arthroplasty with anterior cervical discectomy and fusion: 2-year results from the US FDA IDE clinical trial. Spine. 2013;38(15):E907-E18.

36. Zhang X, Zhang X, Chen C, Zhang Y, Wang Z, Wang B, et al. Randomized, controlled, multicenter, clinical trial comparing BRYAN cervical disc arthroplasty with anterior cervical decompression and fusion in China. Spine. 2012;37(6):433-8.

37. Cheng L, Nie L, Li M, Huo Y, Pan X. Superiority of the Bryan(®) disc prosthesis for cervical myelopathy: a randomized study with 3-year followup. Clinical orthopaedics and related research. 2011;469(12):3408-14.

38. Skeppholm M, Olerud C. Comparison of dysphagia between cervical artificial disc replacement and fusion: data from a randomized controlled study with two years of follow-up. Spine. 2013;38(24):E1507-10.

39. Davis RJ, Kim KD, Hisey MS, Hoffman GA, Bae HW, Gaede SE, et al. Cervical total disc replacement with the Mobi-C cervical artificial disc compared with anterior discectomy and fusion for treatment of 2-level symptomatic degenerative disc disease: a prospective, randomized, controlled multicenter clinical trial: clinical article. Journal of neurosurgery. Spine. 2013;19(5):532-45.

40. Cheng L, Nie L, Zhang L, Hou Y. Fusion versus Bryan Cervical Disc in two-level cervical disc disease: a prospective, randomised study. International orthopaedics. 2009;33(5):1347-51.

41. Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64(4):380-2.

42. Hellum C, Johnsen LG, Storheim K, Nygaard OP, Brox JI, Rossvoll I, et al. Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. BMJ (Clinical research ed.). 2011;342:d2786.

43. Coric D, Kim PK, Clemente JD, Boltes MO, Nussbaum M, James S. Prospective randomized study of cervical arthroplasty and anterior cervical discectomy and fusion with long-term follow-up: Results in 74 patients from a single site - Presented at the 2012 Joint Spine Section Meeting. Clinical article. Journal of Neurosurgery: Spine. 2013;18(1):36-42.

44. Zigler JE, Delamarter R, Murrey D, Spivak J, Janssen M. ProDisc-C and anterior cervical discectomy and fusion as surgical treatment for single-level cervical symptomatic degenerative disc disease: five-year results of a Food and Drug Administration study. Spine. 2013;38(3):203-9.

45. Sasso RC, Anderson PA, Riew KD, Heller JG. Results of cervical arthroplasty compared with anterior discectomy and fusion: four-year clinical outcomes in a prospective, randomized controlled trial. Orthopedics. 2011;34(11):889.

46. Burkus JK, Haid Jr RW, Traynelis VC, Mummaneni PV. Long-term clinical and radiographic outcomes of cervical disc replacement with the prestige disc: Results from a prospective randomized controlled clinical trial - Presented at the 2009 joint spine section meeting. Journal of Neurosurgery: Spine. 2010;13(3):308-18.

47. Schwab JS, DiAngelo DJ, Foley KT. Motion compensation associated with single-level cervical fusion: Where does the lost motion go? Spine. 2006;31(21):2439-48.

48. Bhadra AK, Raman AS, Casey AT, Crawford RJ. Single-level cervical radiculopathy: clinical outcome and cost-effectiveness of four techniques of anterior cervical discectomy and fusion and disc arthroplasty. Eur Spine J. 2009;18(2):232-7.

49. Guyer RD, Tromanhauser SG, Regan JJ. An economic model of one-level lumbar arthroplasty versus fusion. Spine J. 2007;7(5):558-62.

50. Health Quality Ontario. Artificial discs for lumbar and cervical degenerative disc disease -update: an evidence-based analysis. Ontario Health Technology Assessment Series. 2006;6(10):1-98.

51. Wiedenhöfer B, Nacke J, Stephan M, Richter W, Carstens C, Eichler M. Is Total Disc Replacement a Cost Effective Treatment for Cervical Degenerative Disc Disease? . J Spinal Disord Tech. . 2014 (Oct 10. [Epub ahead of print]).

52. Wohns R. Safety and cost-effectiveness of outpatient cervical disc arthroplasty. Surgical neurology international. 2010;1(77).

53. Bronsard N, Litrico S, Hovorka I, Paquis P, Gastaud B, Daideri G, et al. Medicoeconomic evaluation of total disc replacement based on French National Health Care System data. Orthop Traumatol Surg Res. 2011;97(5):533-40.

54. Ament JD, Yang Z, Nunley P, Stone MB, Kim KD. Cost-effectiveness of cervical total disc replacement vs fusion for the treatment of 2-level symptomatic degenerative disc disease.[Erratum appears in JAMA Surg. 2014 Dec;149(12):1295]. JAMA Surg. 2014;149(12):1231-9.

55. Lewis DJ, Attiah MA, Malhotra NR, Burnett MG, Stein SC. Anterior surgical management of single-level cervical disc disease: a cost-effectiveness analysis. Spine. 2014;39(25):2084-92.

56. McAnany SJ, Overley S, Baird EO, Cho SK, Hecht AC, Zigler JE, et al. The 5-year cost-effectiveness of anterior cervical discectomy and fusion and cervical disc replacement: a Markov analysis. Spine. 2014;39(23):1924-33.

57. Qureshi SA, McAnany S, Goz V, Koehler SM, Hecht AC. Cost-effectiveness analysis: comparing single-level cervical disc replacement and single-level anterior cervical discectomy and fusion: clinical article. J Neurosurg Spine. 2013;19(5):546-54.

58. Warren D, Andres T, Hoelscher C, Ricart-Hoffiz P, Bendo J, Goldstein J. Cost-utility analysis modeling at 2-year follow-up for cervical disc arthroplasty versus anterior cervical discectomy and fusion: A single-center contribution to the randomized controlled trial. International Journal of Spine Surgery. 2013;7(1):e58-e66.

59. Alvin MD, Qureshi S, Klineberg E, Riew KD, Fischer DJ, Norvell DC, et al. Cervical degenerative disease: systematic review of economic analyses. Spine. 2014;39(22 Suppl 1):S53-64.

60. Kramer J, Javidan Y, Smith J. The value of disk replacement? Semin. Spine Surg. 2014;26(1):45-51.

61. Menzin J, Zhang B, Neumann P.J, Lines L.M, Polly D.W, Barnett-Myers S, et al. A health-economic assessment of cervical disc arthroplasty compared with allograft fusion. Tech. Orthop. 2010;25(2):133-7.

62. Steinmetz MP, Patel R, Traynelis V, Resnick DK, Anderson PA. Cervical disc arthroplasty compared with fusion in a workers'

compensation population. Neurosurgery. 2008;63(4):741-7; discussion 7.

63. OECD Statistics Directorate. Prices and purchasing power parities (PPP) [Web page].2015 [cited 12/06/2015]. Available from: http://www.oecd.org/std/prices-ppp/

64. Coric D, Nunley PD, Guyer RD, Musante D, Carmody CN, Gordon CR, et al. Prospective, randomized, multicenter study of cervical arthroplasty: 269 patients from the Kineflex|C artificial disc investigational device exemption study with a minimum 2-year follow-up: clinical article. J Neurosurg Spine. 2011;15(4):348-58.

65. Heller JG, Sasso RC, Papadopoulos SM, Anderson PA, Fessler RG, Hacker RJ, et al. Comparison of BRYAN cervical disc arthroplasty with anterior cervical decompression and fusion: clinical and radiographic results of a randomized, controlled, clinical trial. Spine (Phila Pa 1976). 2009;34(2):101-7.

66. Murrey D, Janssen M, Delamarter R, Goldstein J, Zigler J, Tay B, et al. Results of the prospective, randomized, controlled multicenter Food and Drug Administration investigational device exemption study of the ProDisc-C total disc replacement versus anterior discectomy and fusion for the treatment of 1-level symptomatic cervical disc disease. Spine J. 2009;9(4):275-86.

67. McAfee PC, Reah C, Gilder K, Eisermann L, Cunningham B. A meta-analysis of comparative outcomes following cervical arthroplasty or anterior cervical fusion: results from 4 prospective multicenter randomized clinical trials and up to 1226 patients. Spine (Phila Pa 1976). 2012;37(11):943-52.

68. Cleemput I, Neyt M, Van de Sande S, Thiry N. Belgian guidelines for economic evaluations and budget impact analyses : second edition. Health Technology Assessment (HTA). Brussels: Belgian Health Care Knowledge Centre (KCE); 2012. KCE Reports 183C (D/2012/10.273/54) Available from: https://kce.fgov.be/sites/default/files/page_documents/KCE_183C_economic_evaluations_second_edition_0.pdf

69. Nielens H, Van Zundert J, Mairiaux P, Gailly J, Van Den Hecke N, Mazina D, et al. Chronic low back pain. Good Clinical practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE);

2006 17/01/2007. KCE Reports 48C (D/2006/10.273/71) Available from: https://kce.fgov.be/sites/default/files/page_documents/d20061027371.pdf

70. de Kleuver M, Oner FC, Jacobs WC. Total disc replacement for chronic low back pain: background and a systematic review of the literature. Eur Spine J. 2003;12(2):108-16.

71. Deyo RA, Weinstein JN. Low back pain. N Engl J Med. 2001;344(5):363-70.

72. Kalichman L, Hunter DJ. The genetics of intervertebral disc degeneration. Familial predisposition and heritability estimation. Joint Bone Spine. 2008;75(4):383-7.

73. Livshits G, Popham M, Malkin I, Sambrook PN, Macgregor AJ, Spector T, et al. Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine Study. Ann Rheum Dis. 2011;70(10):1740-5.

74. Buttner-Janz K, Guyer RD, Ohnmeiss DD. Indications for lumbar total disc replacement: selecting the right patient with the right indication for the right total disc. Int J Spine Surg. 2014;8.

75. Geisler FH. CHARITE Artificial Disc. In: Kim DH, Cammisa FPJ, Fessler RG, editors. Dynamic reconstruction of the spine. New York: Thieme; 2006. p. 160-78.

76. Medical Services Advisory Committee. Review of interim funded service: Artificial intervertebral disc replacement - lumbar. Report. 2011. Canberra: Medical Services Advisory Committee (MSAC) PUB: Medical Services Advisory Committee (MSAC) Available from: http://www.msac.gov.au/internet/msac/publishing.nsf/Content/1090.1/$File/1090.1_Final_Report.pdf

77. Lee S-H, Shim CS. Complications of Lumbar Disk Arthroplasty. In: Kim DH, Cammisa FPJ, Fessler RG, editors. Dynamic reconstruction of the spine. New York: Thieme; 2006. p. 227-33.

78. Garreau de Loubresse C. [Total disc arthroplasty of the lumbar spine: actual concepts]. e-mémoires de l'Académie Nationale de Chirurgie. 2009;8(1):6.

79. Vital JM, Boissiere L. Total disc replacement. Orthop Traumatol Surg Res. 2014;100(1 Suppl):S1-14.

80. Bland JH, Boushey DR. Anatomy and physiology of the cervical spine. Semin Arthritis Rheum. 1990;20(1):1-20.

81. Mercer S, Bogduk N. The ligaments and annulus fibrosus of human adult cervical intervertebral discs. Spine (Phila Pa 1976). 1999;24(7):619-26; discussion 27-8.

82. Shibata KM, Kim DH. Historical review of Spinal Athroplasty and Dynamic Stabilizations. In: Kim DH, Cammisa FPJ, Fessler RG, editors. Dynamic reconstruction of the spine. New York: Thieme; 2006. p. 3-15.

83. Depreitere B, Dewaele L, Dewitte O, Fransen P, Lubansu A, Martens F, et al. Recommendations of good practice for lumbar disk replacement. Surgical Neurology. 2009;71(1):138-9.

84. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

85. BlueCross BlueShield Association. Artificial Lumbar Disc Arthroplasty Assessment Program. 2014;Volume 28, No. 7.

86. van den Eerenbeemt KD, Ostelo RW, van Royen BJ, Peul WC, van Tulder MW. Total disc replacement surgery for symptomatic degenerative lumbar disc disease: a systematic review of the literature. Eur Spine J. 2010;19(8):1262-80.

87. Thavaneswaran P, Vandepeer M. Lumbar artificial intervertebral disc replacement: a systematic review. ANZ J Surg. 2014;84(3):121-7.

88. Rao MJ, Cao SS. Artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials. Arch Orthop Trauma Surg. 2014;134(2):149-58.

89. Wang JC, Arnold PM, Hermsmeyer JT, Norvell DC. Do lumbar motion preserving devices reduce the risk of adjacent segment pathology compared with fusion surgery? A systematic review. Spine. 2012;37(22 Suppl):S133-43.

90. Wei J, Song Y, Sun L, Lv C. Comparison of artificial total disc replacement versus fusion for lumbar degenerative disc disease: a meta-analysis of randomized controlled trials. Int Orthop. 2013;37(7):1315-25.

91. Yajun W, Yue Z, Xiuxin H, Cui C. A meta-analysis of artificial total disc replacement versus fusion for lumbar degenerative disc disease. Eur Spine J. 2010;19(8):1250-61.

92. Gornet MF, Burkus JK, Dryer RF, Peloza JH. Lumbar disc arthroplasty with Maverick disc versus stand-alone interbody fusion: a prospective, randomized, controlled, multicenter investigational device exemption trial. Spine. 2011;36(25):E1600-11.

93. Berg S, Tullberg T, Branth B, Olerud C, Tropp H. Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. European spine journal. 2009;18(10):1512-9.

94. Moreno P, Boulot J. Comparative study of short-term results between total artificial disc prosthesis and anterior lumbar interbody fusion. Revue de chirurgie orthopédique et réparatrice de l'appareil moteur. 2008;94(3):282-8.

95. Sasso RC, Foulk DM, Hahn M. Prospective, randomized trial of metal-on-metal artificial lumbar disc replacement: Initial results for treatment of discogenic pain. Spine. 2008;33(2):123-31.

96. Zigler J, Delamarter R, Spivak JM, Linovitz RJ, Danielson GO, Haider TT, et al. Results of the prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of the ProDisc-L total disc replacement versus circumferential fusion for the treatment of 1-level degenerative disc disease. Spine. 2007;32(11):1155-62; discussion 63.

97. Blumenthal S, McAfee PC, Guyer RD, Hochschuler SH, Geisler FH, Holt RT, et al. A prospective, randomized, multicenter Food and Drug Administration Investigational Device Exemptions study of lumbar total disc replacement with the CHARITE(trademark) artificial disc versus lumbar fusion - Part I: Evaluation of clinical outcomes. Spine. 2005;30(14):1565-75.

98. Zigler JE, Delamarter RB. Five-year results of the prospective, randomized, multicenter, Food and Drug Administration investigational device exemption study of the ProDisc-L total disc

replacement versus circumferential arthrodesis for the treatment of single-level degenerative disc disease. Journal of neurosurgery. Spine. 2012;17(6):493-501.

99. Zigler JE, Glenn J, Delamarter RB. Five-year adjacent-level degenerative changes in patients with single-level disease treated using lumbar total disc replacement with ProDisc-L versus circumferential fusion. Journal of neurosurgery. Spine. 2012;17(6):504-11.

100. Hellum C, Berg L, Gjertsen Ø, Johnsen LG, Neckelmann G, Storheim K, et al. Adjacent level degeneration and facet arthropathy after disc prosthesis surgery or rehabilitation in patients with chronic low back pain and degenerative disc: second report of a randomized study. Spine. 2012;37(25):2063-73.

101. Johnsen LG, Brinckmann P, Hellum C, Rossvoll I, Leivseth G. Segmental mobility, disc height and patient-reported outcomes after surgery for degenerative disc disease: a prospective randomised trial comparing disc replacement and multidisciplinary rehabilitation. The bone & joint journal. 2013;95-B(1):81-9.

102. CADTH. Compressible non-articulating disc prostheses in adult patients with degenerative disc disease: clinical effectiveness, safety, cost-effectiveness, and guidelines (Structured abstract). Health Technology Assessment Database. 2014(3).

103. Yanbin Z, Yu S, Zhongqiang C, Zhongjun L. Sagittal alignment comparison of Bryan disc arthroplasty with ProDisc-C arthroplasty: a prospective, randomized controlled clinical trial. Journal of spinal disorders & techniques. 2011;24(6):381-5.

104. Guyer RD, McAfee PC, Banco RJ, Bitan FD, Cappuccino A, Geisler FH, et al. Prospective, randomized, multicenter Food and Drug Administration investigational device exemption study of lumbar total disc replacement with the CHARITE artificial disc versus lumbar fusion: Five-year follow-up. Spine Journal. 2009;9(5):374-86.

105. Brox JI, Sorensen R, Friis A, Nygaard O, Indahl A, Keller A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. Spine (Phila Pa 1976). 2003;28(17):1913-21.

106. Fritzell P, Berg S, Borgstrom F, Tullberg T, Tropp H. Cost effectiveness of disc prosthesis versus lumbar fusion in patients with chronic low back pain: randomized controlled trial with 2-year follow-up. Eur Spine J. 2011;20(7):1001-11.

107. Johnsen LG, Hellum C, Storheim K, Nygaard OP, Brox JI, Rossvoll I, et al. Cost-effectiveness of total disc replacement versus multidisciplinary rehabilitation in patients with chronic low back pain: a Norwegian multicenter RCT. Spine. 2014;39(1):23-32.

108. Parkinson B, Goodall S, Thavaneswaran P. Cost-effectiveness of lumbar artificial intervertebral disc replacement: driven by the choice of comparator. ANZ J Surg. 2013;83(9):669-75.

109. North RB, Shipley J, Wang H, Mekhail N. A review of economic factors related to the delivery of health care for chronic low back pain. Neuromodulation. 2014;Suppl 2:69-76.

110. Levin DA, Bendo JA, Quirno M, Errico T, Goldstein J, Spivak J. Comparative charge analysis of one- and two-level lumbar total disc arthroplasty versus circumferential lumbar fusion. Spine. 2007;32(25):2905-9.

111. Patel VV, Estes S, Lindley EM, Burger E. Lumbar spinal fusion versus anterior lumbar disc replacement: the financial implications. J Spinal Disord Tech. 2008;21(7):473-6.