Autologous chondrocyte implantation with Spherox for treating articular cartilage defects in the knee. (NICE ID 851)

Warwick Evidence ERG report October 19th 2017

Authors

Ewen Cummins

Xavier Armoiry

Andrew Metcalfe

Pamela Royle

Hema Mistry

Rhona Johnston

Jeremy Rodrigues

Norman Waugh

Address for correspondence;

Prof Norman Waugh, Division of Health Sciences, Warwick Medical School

University of Warwick, Coventry CV4 7AL norman.waugh@warwick.ac.uk

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 15/69/10.

Competing interests of the authors: none.

Rider on responsibility for report. The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be cited as;

Cummins E, Armoiry X, Metcalfe A, Royle P, Mistry H, Johnston R, Rodrigues J, Waugh N. Spherox for treatment of articular cartilage defects in the knee. A Single Technology Appraisal. Warwick Evidence, 2017.

Contribution of authors;

Norman Waugh and Xavier Armoiry wrote the clinical effectiveness section, assisted by Jeremy Rodrigues. Norman Waugh co-ordinated the project and wrote Chapters 1 and 5. Ewen Cummins wrote the cost-effectiveness section, assisted by Rhona Johnston and Hema Mistry. Pamela Royle conducted literature searches. Andrew Metcalfe provided expert advice and commented on drafts. Norman Waugh and Pamela Royle edited the final document.

Text highlighted in yellow is academic in confidence.

Table of Contents

List of Tables

Table of Figures

LIST OF ABBREVIATIONS

ACI	Autologous chondrocyte implantation
AE	Adverse event
AEs	Adverse events
ANCOVA	Analysis of covariance
BMI	Body mass index
BSC	Best supportive care
CC	ChondroCelect
CEAC	Cost-effectiveness acceptability curve
CHMP	Committee for Medicinal Products for Human Use
EMA	European Medicines Agency
EPAR	European public assessment report
GAGs	Glycosaminoglycans
ICER	Incremental cost-effectiveness ratio
ICRS	International Cartilage Repair Society
IKDC	International Knee Documentation Committee
ITT	Intention-to-treat
K-L	Kellgren-Lawrence
KOOS	Knee Injury and Osteoarthritis Outcome Score
KR	Knee replacement
MACI	Matrix-Applied Characterized Autologous Cultured Chondrocytes
MF	Microfracture
MOCART	Magnetic resonance observation of cartilage repair tissue
MRI	Magnetic resonance imaging

NFR	No further repair
NMA	Network meta-analysis
NMB	Net monetary benefit
OA	Osteoarthritis
OCD	Osteochondritis dissecans
OP	Outpatient
PKR	Partial knee replacement
PP	Per protocol
QALY	Quality-adjusted life year
QoL	Quality of life
RCT	Randomized clinical trial
ROB	Risk of Bias
RR	Relative risk
SD	Standard Deviation
SE	Standard Error
SPHX	Spherox
TKR	Total knee replacement
WTP	Willingness-to-pay

1 Summary

1.1 Critique of the decision problem in the Co-Don submission

The Co-Don summary of the decision problem is similar to the NICE scope, except that Co-Don consider that some of the comparators were inappropriate, including osteotomy and mosaicplasty. The ERG agrees with the Co-Don position. Mosaicplasty is little used in the UK and we think it would be used only for small lesions.

The FAD from the recent MTA of autologous chondrocyte implantation (ACI) recommended that it should be used, subject to some restrictions, one of which was size of the articular cartilage (chondral) defect. This means that the NICE scope issued for this STA is out of date, and microfracture is no longer a comparator for defects over 2 cm^2. The key comparators are now other forms of ACI, though there are problems with the availability of these, as reported later. So the decision problem as defined by Co-Don is also out of date – the timing was unfortunate.

The outcomes in the Co-Don decision problem are as in the NICE scope.

1.2 Summary of the clinical effectiveness evidence submitted by Co-Don

Trials

The Co-Don submission presents the results from two RCTs (one phase II and one phase III).

The phase II RCT was conducted prior to the phase III and aimed to identify the optimal strength of Spherox by comparing three arms with different doses. This study included people with large defects (4-10cm^2). The KOOS score improved from a baseline mean of [redacted], median [redacted], to a mean of [redacted], median [redacted] at four years. So the key result is that improvements seen at 24 months are sustained.

The phase III study, called COWISI, was the pivotal trial to support the approval of Spherox.

COWISI is a prospective, randomised, open label, multicentre phase III clinical trial that compared Spherox to microfracture (MF) in 102 patients with defect sizes between 1 and 4cm^2. The outcomes in the trial match the NICE scope. The primary outcome was the change of overall KOOS from day 0 to assessment at 24 months after treatment completion. The KOOS score, together with other outcomes

such as MOCART (MRI score) will be evaluated with longer term follow-up durations (36, 48, 60 months) that are not yet available.

COWISI was a good quality trial though blinding of intervention was impractical because the Spherox group had two procedures. The sample size was calculated to show non-inferiority of Spherox against MF whereas other trials of ACI (SUMMIT, TIG/ACT, and ACTIVE) were designed to show if ACI was superior to MF.

The KOOS scores improved from baseline to 24 months with both Spherox (improvement of [redacted]) and MF ([redacted]). The repeated-measures ANCOVA testing for non-inferiority of Spherox against MF showed a [redacted] difference with a lower bound [redacted].

Safety was assessed through the incidence of adverse events (AE) that were probably or possibly related to treatments. In the phase II trial only [redacted] of 73 patients treated had severe adverse events, [redacted].

Network meta-analysis:

The Co-Don submission presented indirect evidence for comparisons of Spherox with two other forms of ACI, ChondroCelect and Vericel MACI, via a network meta-analysis (NMA). The network included three RCTs and used microfracture as a common comparator. Two outcomes were assessed, responders and failures. The studies varied in how response was reported, with response was defined in two trials as a gain of 10 or more points in the overall KOOS scale, and in the third as gains in several KOOS subscales. Failure was a need for revision surgery. [redacted]. The proportions of [redacted] ChondroCelect. [redacted]. At clarification stage, the Company provided the dataset used to run the NMA on the two outcomes. Using a frequentist framework, the ERG was able to replicate the findings of the Company's NMA on responders but found different results for the NMA on failure. However, this was not considered as a major issue given that failure inputs in the cost-effectiveness model used data from RCTs and not those from the NMA.

1.3 Summary of the ERG critique of the clinical effectiveness evidence submitted by Co-Don

COWISI was a good quality trial though blinding of intervention was impossible, and protocol deviations were seen in [redacted]. The largest number of deviations was because of failure to attend visits, with taking of prohibited pain medications next (mainly in the MF group). The main problem with the trial at present is that results are only available to 24 months. Longer-term follow-up is planned, to five years.

Although the Phase II dosage study was of reasonable quality, apart from the [redacted] drop-out rates due to protocol violations, it is of limited interest because it did not include any comparator listed in the NICE scope and therefore was not used to inform the cost-effectiveness model.

The ERG has identified several methodological flaws in the NMA, in particular focusing on the assumptions of homogeneity and similarity.

- The transitivity assumption does not hold, since the distribution of population characteristics that are effect modifiers differ across the treatment comparisons of the network. Three such treatment effect modifiers in the Co-Don NMA are the baseline KOOS score, the lesion size at baseline, and previous repair attempts. The uneven distribution of these effect modifiers across the network comparisons violates the transitivity assumption.
- The networks compared interventions for two outcomes, namely the proportion of responders and failure rate. However there was some variation in the definition of both outcomes which means that the outcomes were not assessed consistently across studies. Furthermore, failure rates were not evaluated over the same time periods across studies. Outcomes using time-varying events should be assessed consistently to enable a valid comparison.

The ERG doubts whether it was appropriate to do an NMA, and considers the validity of the estimate for the indirect comparisons to be very questionable.

Given the paucity of RCT data, the ERG looked to see if anything could be gleaned from case series. However these are mainly small, three are just pilot studies, two are available only as conference abstracts, and others have duration of only for around a year. Without control groups, their value is limited. They do report before and after improvements, showing that Spherox is clinically effective, and also that Spherox can be implanted arthroscopically.

1.4 Summary of cost effectiveness evidence submitted by Co-Don

The company model structure is meant to mirror that of the model of the ACI MTA. It is a markov model with an annual cycle and a lifetime horizon. All patients receive the 1st repair of the sequence during the 1st cycle of the model. These patients can move into one of three health states.

- Success
- No further repair (NFR)
- 2nd repair, if necessary

Subsequent to the 1st cycle those who were a success either remain a success or move to 2nd repair. All those in NFR remain in NFR.

The patients who receive the 2nd repair of the sequence can move into one of two health states.

- Success
- No further repair (NFR)

Those who were a success either remain a success or move to NFR. All those in NFR remain in NFR. No further repairs are possible after a 2nd repair.

From age 55, a common probability of patients receiving knee replacements is applied.

Four main sets of comparators are included.

- A 1st microfracture repair with the possibility of a 2nd microfracture repair
- A 1st microfracture repair with the possibility of a 2nd ACI repair
- A 1st ACI repair with the possibility of a 2nd microfracture repair
- A 1st ACI repair with the possibility of a 2nd ACI repair

When a 1st ACI repair is followed by a 2nd ACI repair the same ACI is given. ERG expert opinion suggests that this it reasonable because centres are likely to specialise in a single type of repair.

The company derives the clinical effectiveness estimates from its NMA on success rates and its NMA on failure rates. Quality of life values are aligned with those of the ACI MTA. Unit costs are largely aligned with those of the 1st AG model of the ACI MTA. Cell costs are £10,000 for Spherox and £16,000 for MACI and ChondroCelect.

Following clarifications the company has revised some of its clinical effectiveness estimates. The company revised cost effectiveness estimates are that the cost effectiveness of Spherox relative to microfracture is around £4k per QALY and the cost effectiveness of MACI compared to Spherox is around £18k per QALY.

There are no sensitivity analyses around the revised company estimates. The original modelling was most sensitive to the assumption that all microfracture repair successes fail at year 5.

1.5 Summary of the ERG critique of the cost effectiveness evidence submitted by Co-Don

The company model differs from that of the model of the ACI MTA in one crucial respect. 1st repair successes cannot lose response and move into the no further repair health state. This is likely to bias the analysis in favour of the ACIs. It may also further bias the analysis in favour of MACI and ChondroCelect if their loss of response is similar to that of Spherox, because their initial success proportion is a bit higher.

The response estimates for 2nd repairs are only applied once within the modelling and as a consequence the company method used to derive these is incorrect.

The company accepts that the probabilities of 2nd repair successes losing success and moving to no further repair are incorrect. It suggests revising these to be based upon the annualised 1st repair non-response probabilities at 2 years. These estimates are applied every year of the model, do not really relate to a loss of response, and are probably too high.

The company clinical effectiveness estimates are incorrect and biased in favour of Spherox.

The company quality of life estimates are aligned with those of the ACI MTA.

The company does not apply the preferred set of unit costs of the ACI MTA FAD.

1.6 ERG commentary on the robustness of evidence submitted by Co-Don

The ERG has attempted to revise the company model to have inputs similar to those of the 1st AG report of the ACI MTA. This is imperfect but appears to suggest that the company model estimates roughly double the patient gains compared to the model of the 1st AG report of the ACI MTA. The cost effectiveness estimates of the ACI MTA also tended to worsen as the assessment progressed and publicly available time to event data for loss of response was incorporated. The company model structure may be too optimistic for the comparison with microfracture.

The company accepts that all the clinical effectiveness estimates for the model of its original submission are wrong and biased in favour of Spherox. It has provided a revised set of estimates for a subset of these. These still appear to be incorrect and biased in favour of Spherox.

The clinical effectiveness estimates for Spherox are little different from those of microfracture. The model estimates quite large QALY gains from Spherox compared to microfracture. These are almost entirely due to the assumption that all microfracture successes fail at year 5.

1.7 Summary of additional work undertaken by the ERG.

The ERG is limited to the company model structure and while it has made some revisions to this it cannot revise the model to reflect the full model structure of the model of the ACI MTA.

In the light of the ACI MTA FAD the ERG also limits the number of comparators. A 1st microfracture repair cannot be followed by a 2nd microfracture repair or by a 2nd ACI repair.

The ERG has revised all the clinical effectiveness estimates of the model, and has aligned the unit costs with the preferred set of unit costs of the ACI MTA FAD.

If all microfracture successes fail at year 5 the company model estimates the cost effectiveness of Spherox compared to microfracture to be £4-5k per QALY. It also estimates the cost effectiveness of MACI compared to Spherox to be £12-18k per QALY.

If microfracture successes are as durable as ACI successes the company model estimates that Spherox results in few patient gains relative to microfracture and its cost effectiveness is very poor. The cost effectiveness of MACI compared to microfracture also typically rises above £30k per QALY.

1.8 ERG conclusions

Spherox is clinically effective in the treatment of chondral defects, and the Phase II trial shows benefit maintained for up to [redacted] years. [redacted].

2 Introduction

Members of the Appraisal Committee who are familiar with ACI may wish to go straight to Section 2.4

2.1 Cartilage injuries

The ends of the bones and the inner surface of the patella in the knee are covered with articular cartilage. Articular cartilage should not be confused with the meniscal cartilages that are cushions of cartilage between the bones – when people talk of "cartilage problems" in the knee, they often mean the meniscal cartilage.

Normal "hyaline" cartilage is a rubber-like substance that is normally very smooth, promoting smooth frictionless movements of the joints and also acting as a shock absorber. It is formed mainly of a protein called type 2 collagen. Under the articular cartilage are the bones of the knee – femur in thigh, tibia below the knee, and the patella or knee-cap.

Cartilage has no blood vessels and has very limited ability to repair itself. Epidemiological studies show a relationship between knee injury and later development of osteoarthritis. In some people, this will lead in the long-term to a need for a knee replacement with an artificial joint, usually total knee replacement (TKR), though there can be partial knee replacement of just one side.

Loss of articular cartilage is referred to as a chondral defect, and loss of cartilage and bone as an osteochondral defect. Cartilage damage can be caused by injury, by various types of arthritis, or spontaneously in a condition called osteochondritis dissecans (OCD) in which a bit of bone and attached cartilage breaks off. Cartilage damage may also arise because of knee instability or abnormal loading, for example secondary to a ligament injury[1] or damaged meniscal cartilages.[2] Serious obesity may also affect knee cartilage.[3] Conversely, physical activity without injury may be protective.[4]

In young people the most common cause of hyaline cartilage damage is sporting injuries. Aroen and colleagues[5] reported that in patients having knee arthroscopy in Norway, injuries occurred in sport in 55%, in the home in 15%, at work in 12% and in road traffic accidents in 5%. In 13% the cause was unknown.

It should be noted that cartilage defects without any underlying bone involvement may not cause pain – there are no nerves in cartilage. The source of pain in knees with damaged cartilage is poorly understood but may come from many sources including ligaments, the joint capsule and the underlying bone.[6] So results from series of symptomatic patients may not be entirely representative of all people with cartilage damage. The commonest symptom is pain, with others being temporary locking of the knee in one position, and swelling. Pain and disability from symptomatic cartilage lesions have been shown to be as significant in magnitude as that from severe arthritis of the knee.[7]

The longer-term consequence of chondral injury is osteoarthritis (OA), which develops over time and often leads to a need for knee replacement. Knee replacement has been of great benefit to many people, by relieving the pain of OA, but it does not restore the full range of function in the knee, and replacements do not last forever. Failure is common after 10-15 years, and while a replacement can be replaced, second knee replacements are more difficult, about double the cost, and are accompanied by a greater risk of complications. Orthopaedic surgeons try to avoid doing knee replacements done before the age of 55 in OA. (In other forms of arthritis such as rheumatoid arthritis (RA), they may be done at younger ages but may last longer because people with RA are limited in other ways and put less stress on the new joint.) So a treatment for chondral defects that removes symptoms could be very useful even if it did not give a permanent repair, by acting as an interim solution till patients were able to have knee replacements.

The International Cartilage Repair Society (ICRS) has a scoring system for grading the severity of cartilage damage [8];
Grade 1: soft indentation and/or superficial cracks
Grade 2: small cracks or lesion extending down to under half of cartilage depth
Grade 3: deep cracks or gaps of over 50% of cartilage depth
Grade 4: cracks through the total thickness of cartilage down to the underlying bone
Grade 5: defects of the full thickness of cartilage involving the sub-chondral bone

Grading is done by arthroscopic examination. An arthroscope is a fibreoptic telescope inserted into the knee joint so that the surgeon can look at the injury.

2.2 Autologous chondrocyte implantation

The cells that produce cartilage are called chondrocytes. In autologous chondrocyte implantation (ACI), a small piece of cartilage is removed from the knee, and the chondrocytes are cultured the laboratory until they number millions. They are then put into the damaged area of articular cartilage

as a patch. The hope is that this patch will repair the damaged area and form a new layer of natural articular cartilage, called hyaline cartilage. Autologous means that the cells implanted in ACI come from the patient's own cartilage.

Chondrosphere or Spherox is the latest form of autologous chondrocyte implantation (ACI) to be appraised by NICE, and the fourth appraisal of ACI. The FAD (Box 1) from the third appraisal was issued on 4th October 2017. It does not specify any particular ACI product, but gives a general approval to ACI.

Box 1. FAD on ACI

FAD for ACI

Autologous chondrocyte implantation (ACI) is recommended as an option for treating symptomatic articular cartilage defects of the knee, only if:

☐ the person has not had previous knee repair surgery

☐ there is minimal osteoarthritic damage to the knee (as assessed by clinicians experienced in investigating knee cartilage damage using a validated measure for knee osteoarthritis)

☐ the defect is over 2 cm^2

☐ the procedure is done at a tertiary referral centre.

One point to note is that the restriction to people who have not had previous attempts at repair such as microfracture (debridement does not count as a cartilage repair procedure) is based on ICERS which were higher after previous repairs because ACI is less successful if the subchondral bone has been damaged. However those ICERS assume cell cost of £16,000 (the list price). The cell costs are one of the key drivers in the cost-effectiveness analysis, and a significantly lower price might produce acceptable ICERs.

One issue which will need to be clarified is the tertiary referral process. Referral could be based on MRI in the first centre, with both harvesting and implantation both done in the tertiary centre.

2.3 Treatments for chondral injury

There are several possible interventions after chondral injury

Conservative management

One option is no surgical treatment, but to use symptomatic relief, with or without physiotherapy. Three case series [9-11] reported high levels of return to activities after cartilage injuries after 14 years, 9 years and 9 years respectively. Messner and Maletius reported a case series of young athletes (mean age 25, range 14-38) who had no treatment. 14 years later, most (21 out of 28) had returned to activity and 22 had excellent or good function.[9] However despite lack of symptoms, most showed radiological changes suggestive of early osteoarthritis. The NICE guidance specifies "symptomatic articular cartilage defects of the knee", but in some people, symptoms resolve. However the cartilage defect will not, and they are likely to develop OA, and some will need knee replacement in later years.

The UK knee surgeons' consensus recommends that all patients being considered for ACI should have had physical therapy first, since that may relieve symptoms.[12]

Lavage and debridement.

In lavage, an arthroscope is inserted into the knee and saline is poured in through a cannula. This is usually done under general anaesthesia on a day case basis. The saline washes out loose debris which comes out through the cannula or is sucked out using a suction/shaving device. It is also thought to wash out compounds that cause inflammation.

Debridement is done under arthroscopic vision and is the removal of damaged cartilage or bone. It is not a repair procedure. Debridement and lavage are often done at the same time.

The evidence for effectiveness of debridement is sparse and mixed. One three-armed RCT of lavage alone, lavage plus debridement and a sham arm reported no difference at 2 years.[13] Another by Hubbard had methodological weaknesses, but reported that debridement and lavage was better than lavage alone.[14] The NICE intervention procedures guidance on lavage with or without debridement (IPG230) noted uncertainty about the efficacy of the procedure.[15]

ACI

ACI has been used since at least 1987.[16] The procedure has evolved over time, with different ways of implanting the chondrocytes into the chondral defect.

In the first generation of ACI, the cultured chondrocytes were placed in the defect, in liquid form, and then covered with a cap made from a patch of periosteum, the tough fibrous tissue that covers bones such as the tibia – ACI-P. This led to problems with pain at the periosteal harvest site in the immediate post-operative period, and a need for further procedures to remove overgrowth in the graft. It is now obsolete but comes up in some of the older Chondrosphere studies.

The second generation of ACI used a collagen cap (ACI-C) instead of the periosteal one, but still used cells in a liquid suspension

In the third generation of ACI, the chondrocyte cells are loaded or embedded, or "seeded", on to a porcine collagen membrane ACT-C or matrix (MACI – matrix induced chondrocyte implantation), with a patch cut to fit. These patches can be implanted by a less invasive form of surgery, by mini-arthrotomy, requiring less surgical time than ACI-C.[17] (Arthrotomy = opening of a joint). This has become the main method used.

The membrane used in MACI is composed of type I/III collagen, with a rough side wherein the chondrocytes are seeded and a smooth side which faces into the joint cavity.[17] The membrane is tough enough to be cut to shape or stitched in place, though it is more often glued in place.[17] The membrane is bio-degradable. The term "scaffold" is often used instead of membrane. However the membrane needs careful handling to minimize chondrocyte death during implantation.[18]

Box 2. The evolution of ACI

First generation	ACI-P. Liquid suspension of cultured chondrocyte cells placed in the defect covered with a cap made from periosteum.
Second generation	ACI-C. Liquid suspension of cells placed in the defect and covered with a collagen cap.
Third generation	The cultured cells are seeded on to a membrane or "scaffold" as in MACI (matrix associated chondrocyte implantation).
Characterized chondrocytes	Not all chondrocytes are equally good at producing cartilage. Some are more "chondrogenic" (cartilage-producing) than others. The most useful can be selected and are known as "characterized".
Fourth generation	Newer developments include the implantation not of cells that will form cartilage, but of tissue-engineered cartilage grown from autologous chondrocytes in the laboratory. Some of the chondrocytes used may come from cartilage from the nose or ear.

Spherox (formerly known as Chondrosphere and ACT3D) is a form of fourth generation ACI in which the cells are not only multiplied in the laboratory, but are persuaded to generate cartilage. Chondrocytes are harvested from healthy articular cartilage, cultivated for 8-10 weeks in the laboratory, and condensed into spheroids (chondrospheres) of cells plus cartilage. The 3-dimensional spheroids are then implanted into the defect. The Co-Don submission says that the spheroids adhere to

the defect (presumably to the subchondral bone) and that no cap or fibrin glue is required to keep the in place.
Spheroids of human autologous matrix-associated chondrocytes are licensed in Germany for the treatment of articular cartilage defects of the knee, hip, shoulder, elbow and ankle. Unlike MACI, the procedure does not require any non-human collagen scaffold.

Microfracture

The main alternative method of repair has been microfracture, in which small holes are drilled through the surface of the bone in the area of damaged cartilage. This allows bleeding from the bone marrow, and the blood carries stem cells into the area where the damaged cartilage has been debrided. These cells form scar cartilage called fibrocartilage, composed of type 1 collagen. This is regarded as being inferior to hyaline cartilage, being less hard-wearing and it is not expected to last as long.

Microfracture may be combined with the insertion of a collagen membrane to cover the microfracture clot, known as augmented microfracture.

Microfracture can be done arthroscopically (i.e. without opening the knee joint) and can be done at the same time as debridement and lavage.

Mosaicplasty

Mosaicplasty (sometimes called OATS – osteochondral autograft transfer system) involves transplanting small sections of cartilage and underlying bone from a less weight-bearing part of the knee into the damaged area. The pieces are in little cylinder shapes and once transplanted, have an appearance not unlike a mosaic – hence the name. Mosaicplasty can only be used for small areas of damage because the transplanted sections have to come from elsewhere in the knee, usually the trochlea. (In some countries, allograft cadaver donor tissue is used, but this appears to be rare in the UK because of issues around local funding and arrangements for the sourcing of the allografts.)

Mosaicplasty appears to be little used now. In the ACTIVE trial [19] of ACI versus standard methods such as microfracture and mosaicplasty, few surgeons chose mosaicplasty.

Comparator ACIs.

In the last appraisal of ACI by NICE, three forms of ACI were appraised.

- The ChondroCelect ACI system from TiGenix, a form of ACI-C in which the cultured cells are combined with a biodegradable collagen I/III patch, with characterised chondrocytes.

ChondroCelect received European marketing authorisation in October 2009.[20] It was being marketed by Swedish Orphan Biovitrum, but following the initial negative NICE decision, production ceased, and ChondroCelect is no longer on the market

- The Matrix ACI system (MACI® – short for "matrix applied characterised autologous cultured chondrocyte implant") originally developed by Sanofi. The matrix refers to a collagen membrane into which the chondrocytes are loaded at operation. The Sanofi MACI was approved in Europe in June 2013.[21] This product was taken over by Aastrom Biosciences who changed their name to Vericel. They recently received FDA approval for their MACI product now being marketed in the USA. They do not at present have any manufacturing facility in Europe, so the EMA has suspended their European licence. However we have heard that the EMA will be inspecting the US production facility and that cells may be provided to Europe from there. (Note that MACI is used both to refer to third generation ACI, and as a trade name.)
- ACI using cells cultured in the John Charnley Laboratory, an NHS laboratory at the Robert Jones and Agnes Hunt (RJAH) Orthopaedic Hospital in Oswestry, England. The facility has cultured and provided autologous chondrocytes (OsCells) for use in ACI since 1997. The facility has a Hospital Exemption Licence under the advanced therapy medicinal products regulations that enables OsCell to supply chondrocytes for use in ACI. This is the only NHS facility that currently cultures cells for use in ACI. NICE refers to OsCells as "traditional ACI".

2.4 Some decision issues

As noted in the ACI FAD, ACI is less successful in patients who have had previous attempts at repair, usually by microfracture, which damages the bone immediately under the cartilage (subchondral bone). When comparing results of Spherox and other forms of ACI, the proportions with previous repair attempts needs to be considered.

There may be a question about how soon cartilage defects should be treated. In the TIG/ACT trial of ACI versus microfracture, outcomes were better in those treated within three years of symptom onset compared to those with longer duration.[22] However the 3-year division is somewhat misleading, because the under 3-year group had an average duration of injury of just under one year, and the over 3 years group had average duration of almost 8 7.8 years. The groups also differed in other ways. Mithoefer and colleagues have also reported better results with ACI sooner after injury, in football players.[23] Harris and colleagues also concluded that results were better in patients with shorter duration of symptoms and fewer prior procedures.[24] So duration of injury should also be considered when comparing results.

Patient factors.

The patient group, as stated in the scope from NICE, is "people with an articular cartilage defect". The EMA approval mentions adults and symptoms. The NICE FAD states that ACI should not be used in advanced OA.

There are three issues here: adults, symptomatic, and defining advanced OA.

Adults. In most past trials, patients had a mean age of 32, range 16 to 49, with about 60% men. In most cases, the cartilage damage was due to injury, usually from sport. However there are now several trials in teenagers (ages 15-17). Some studies of Spherox included patients as young as 15.

Symptoms. Some people with chondral injuries have symptoms which resolve. The UK consensus summarised in Box 3 below, would restrict ACI to people with symptoms and with higher grade lesions. As the statement recognises, some people may have symptoms relieved by physiotherapy. However physiotherapy cannot repair chondral defects, so this group will still be at risk of progression to osteoarthritis.

Box 3. UK Cartilage Consensus [12]

The surgical management of symptomatic articular cartilage defects of the knee: consensus statement from UK knee surgeons.

The statement notes variations in provision of repair of articular cartilage in the knee, and financial constraints on the more expensive treatment options.

The consensus relates to management of an isolated chondral lesion in a knee that is free of other defects, or in which these have been corrected. Key points include;

- Surgical treatment should be considered for symptomatic lesions of ICRS grade 3 or worse.
- Microfracture leads to fibrocartilagenous scar tissue that has poorer biomechanical properties that normal hyaline cartilage, and this repair tissue degenerates. Short-term improvement in symptoms does not persist.
- Mosaicplasty can give good short-term results in small lesions but longer-term results are poorer. It is not suitable for larger lesions, or for patellar defects.
- In small defects, less than $2cm^2$, microfracture, mosaicplasty and ACI may all be considered.
- For lesions > $2cm^2$, cell therapy (ACI) is the most effective treatment based on current evidence

- Outcomes are poorer in smokers, patients with BMI>30, and those with a long duration of symptoms
- When ACI is considered appropriate, it should be first-line treatment because results are poorer if it is used after failure of other procedures
- Physical therapy may be effective in controlling symptoms and should be provided before surgery is considered.

Osteoarthritis.

NICE considered the OA issue and chose a form of words in the FAD which may lead to debate:

"there is minimal osteoarthritic damage to the knee (as assessed by clinicians experienced in investigating knee cartilage damage using a validated measure for knee osteoarthritis)"

The most common method for assessing structural changes in knee osteoarthritis is plain radiography, graded using the Kellgren-Lawrence (K-L) classification.[25] Care has to be taken in interpreting plain radiographic findings, as K-L grades have moderate but not strong correlations with other measures of structural change such as MRI measures of osteoarthritis or operative findings.[26-31]

The K-L classification is a widely accepted tool in osteoarthritis research and good reliability has been quoted in series in which the assessors were experienced in its use.[27, 29] However, it is based on a subjective assessment of structural changes and different authors often apply different criteria to define the boundaries between the grades, making comparisons across studies difficult.[32]

The boundary between K-L grade 2 and 3 is often difficult to define as the interpretation of 'possible' and 'definite' joint space narrowing can be very subjective.[33] The distinction between lower KL grades is also difficult is dependent on the interpretation of small osteophytes which can variably give a score of 0, 1 or 2 depending on the exact definitions used and the radiological technique.[32] Patients with an isolated chondral lesion and no OA may, simply from the result of loss of joint space due to the chondral lesion, be mistaken for having OA based on the K-L grade. The ERG therefore feels that the recommendation made for defining OA in the NICE ACI FAD is a good and pragmatic solution.

3 The Co-Don submission.

Co-Don have been unfortunate in the timing of the Spherox appraisal. They have based their submission largely on their single RCT which compared Spherox with MF. However, NICE has now approved ACI in place of most MF. So the key comparators are the other forms of ACI, and in particular Vericel's MACI, because it has the only licence in Europe, albeit temporarily suspended. MACI is used by Vericel as a trade name, but it is also used as a general term to describe third generation ACI. When referring to the Vericel product, we will use VerMACI.

3.1 Manufacturer's description of health problem.

Co-Don provide a concise but accurate description of chondral injuries, making the key points;

- Articular cartilage has very limited self-repair capacity
- Chondral injuries are common, especially after sporting or occupational injuries
- Because the chondral lesions don't heal, they lead to osteoarthritis
- The people who sustain such injuries are often in their 20s and 30s
- So they are much too young for knee replacements
- We need interventions to repair the chondral injuries to relieve symptoms and to prevent, or at least delay, progression to OA.

3.2 Manufacturer's description of current services

The Co-Don submission correctly notes that in the current clinical pathway in the UK, people only progress to ACI once conservative treatment, such as physiotherapy and analgesia has failed. This is in line with the UK Knee Surgeons consensus statement in Chapter 1. The submission also noted the then draft FAD on ACI, which recommended ACI as first line surgical treatment following conservative care, with the restrictions reported in Section 2.2 above.

So Co-Don provided a correct overview of an evolving situation, since it had to be written before the final FAD was released.

However, the submission does not give an account of current provision of ACI in the UK. The NICE guidance of 2004 recommended ACI only in research, and the 2015 ACD repeated that recommendation. So very little ACI has been done.

The approval by NICE of ACI, subject to certain restrictions, is likely to be welcomed by orthopaedic surgeons in the UK. Commissioners of care will be expected to fund ACI. Patients with chondral defects will look forward to an effective treatment.

Unfortunately, provision of cells may be a problem. TiGenix has discontinued production of ChondroCelect. They may resume but that would take time, and marketing authorisation was discontinued. The licence for Vericel MACI is currently suspended because they have no European production facility, but may be reinstated after the EMA has inspected the production site in the USA. We do not know if Vericel will open a new facility in Europe.

OsCells is authorised to produce cells only for use in the RJAH Hospital in Oswestry. They can and do accept referrals from elsewhere but their capacity is limited.

Other NHS units may seek to develop cell production facilities but would have to obtain MHRA approval and developing the facilities would be a lengthy and difficult process.
So in the short term, there may be a mismatch between supply and demand.

3.3 Co-Don definition of decision problem

The Co-Don summary of the decision problem is similar to the NICE scope, except that Co-Don consider that some of the comparators were inappropriate, including osteotomy and mosaicplasty. The ERG agrees with the Co-Don position. Mosaicplasty is little used in the UK and we think it would be used only for small lesions.

However following the recent MTA, the NICE scope is out of date, and microfracture is no longer a comparator. So the decision problem as defined by Co-Don is also out of date – the timing was unfortunate.

The outcomes in the Co-Don decision problem match the NICE scope.

The NICE scope mentions “people” with no age restriction. The EMA SPC states that “safety and efficacy of Spherox in children aged 15-18 are not established”.[34] However the Co-Don submissions notes that two studies, cod 16 HS 16 (2012) and cod 16 HS 17 paed (2016) (with some overlap of patients), have shown that Spherox was considered safe and effective in adolescents of 14 to 17 years of age. The EMA approved a paediatric investigation plan in November 2012. It appears that only an interim analysis of these studies has yet been carried out, so presumably data will be provided to EMA in due course.

The SPC from the EMA says "Application of Spherox in obese patients is not recommended." No reason is given, but both trials of Spherox excluded patients with BMI >30. This may cause problems because the commonest cause of chondral injury is sport, and in sports such as rugby, many players have BMIs over 30, especially the forwards. However they are muscular rather than obese. A blanket restriction by BMI would be inappropriate.

The SPC recommends a few other restrictions;

- Primary (generalized) osteoarthritis
- Advanced OA of the affected joint, defined as exceeding Kellgren Lawrence grade 2
- Other joints. The SPC states that safety and efficacy are not established beyond the patella and femoral condyles.

3.4 Intervention: Manufacturer's description of Spherox

As noted above, Spherox consists of implants of both chondrocytes and the cartilage they have produced in the laboratory. The Co-Don submission reports that Spherox received a marketing authorisation from the European Medicines Agency (EMA)[35] in July 2017, but also that it has been used, with a marketing authorisation, in Germany since 2004, in around 11,000 patients in 120 orthopaedic departments. It has also been used in five other European countries.

However, the regulatory situation changed, and in order to comply with the regulations on tissue-engineered products (Article 2 (1) (b) of Regulation 1394/2007/EC) Chondrosphere became subject to a centralised authorisation procedure, which required a clinical trial.

The approved indication is for the repair of symptomatic articular cartilage International Cartilage Repair Society (ICRS) grade III or IV defects on the femoral condyle and on the patella, for defects of up to 10 cm^2 in adults.

The EMA verdict was not unanimous, and 16 members expressed dissent (EPAR report)[35], and argued that Spherox was "not approvable due to a negative benefit/risk ratio". Reasons for dissent included;

- Only clinical non-inferiority to MF has been shown
- Pain medication could have been a confounding factor
- Efficacy based on MOCART structural endpoints was not proven, and most of the seven biopsies after Spherox showed mixed fibrous tissue, not hyaline cartilage

- The number of non-responders in both the trials was >30%, and since Spherox required two operations compared to one for MF, benefit for patients was not demonstrated.
- The dissenters was also concerned about production processes and whether problems therein were related to non-responder rates.

Note that at the time Spherox was being considered, only 12 month data from the COWISI trial were available, and the dissenters stated that the 24 month data were required before the benefit/risk assessment could be completed. So some may not now dissent.

The price of the spheroids is given as £10,000, and this is not flagged as confidential. It includes transportation costs. Harvesting and implantation costs are added and Co-Don have used the costs from the recent MTA, adjusted for inflation. This is despite an assertion (page 19) that Spherox requires less invasive surgery for implantation, arthroscopically or by mini-arthrotomy, which may result in less theatre time.

However MACI can also be done by mini-arthrotomy. (And arthroscopically, but cell viability and speed are better when ACI is done by mini-arthrotomy than arthroscopically.[36]
Several of the case series from Germany report that Spherox can be implanted arthroscopically, so we can accept that a slightly shorter operation is required, perhaps saving 10 minutes of theatre time. This will have little effect on overall costs.

3.5 Clinical effectiveness - trials

The Co-Don submission presents the results from two trials, one Phase II and the other phase III, but mentions some earlier case series in an appendix. They carried out systematic searches for studies, using what we consider to be reliable search strategies. No systematic reviews of Spherox were found.

The Phase II trial, called HS14, was conducted prior to the Phase III trial and aimed to identify the optimal strength of Spherox by comparing three arms with different doses. There was no non-Spherox arm.

The Phase III compares Spherox with MF. This trial, which provides evidence for the modelling, is NCT01222559, now known as COWISI, but formerly called HS13. It is described in the submission as:
Phase III clinical trial designed to compare the efficacy and safety of the treatment with the autologous chondrocyte transplantation product Spherox with microfracture in subjects with cartilage defects of the knee with a defect size between 1 and 4 cm^2

The COWISI trial

This is summarised in Table 1, adapted from Table 3 of the Co-Don submission

Table 1 Summary of the COWISI trial

Study	NCT01222559 (COWISI)				
Study design	Prospective, randomised, open label, multicentre Phase III clinical trial				
Population	The analysis population comprised 102 patients (41 women, 61 men) aged 37 ± 9 years, with ICRS grade 3 or 4 chondral defects on femoral condyles.				
Intervention(s)	Implantation of Spherox into the cartilage defect. There are two study operations: harvesting of chondrocytes at arthroscopy and, after approximately 2 months, implantation of Spherox.				
Comparator	Microfracture				
Indicate if trial supports application for marketing authorisation	Yes	X	Indicate if trial used in the economic model	Yes	X
	No			No	
Reported outcomes specified in the decision problem	Change of overall KOOS (Knee Injury and Osteoarthritis Outcome Score) from Day 0 (baseline for both treatment groups = pre arthroscopy assessment) to assessment at 24 months, compared between Spherox and microfracture. Overall KOOS including 5 subscores (pain, knee function including long-term function, activities of daily living, other symptoms and quality of life). Activity levels, avoidance of osteoarthritis including knee replacement, adverse effects of treatment, health-related quality of life				
All other reported outcomes	MOCART (MRI Score), ICRS and ICRS II Visual Histological Assessment Score, Bern Score, Change of ICRS/IKDC, Change of modified Lysholm Score. Days of absence from work				

As in other trials, microfracture was performed by the method developed by Steadman et al.[37]

The entry criteria excluded people with BMI over 30, but Table 7 reports a range of BMIs up to 31.2. Further follow-up visits are planned at 36, 48 and 60 months. The current results were from visits at 3, 12, 18 and 24 months, but we focus on the 24 month results. The exclusion criteria in Table 5 also list radiological signs of OA as an exclusion but according to Table 8, four people with OA were included.

The KOOS assesses pain, symptoms, activities of daily living, sport and recreational activities, on a scale of 0 to 100, where 100 is best.

The MOCART score (magnetic resonance observation of cartilage repair tissue) is based on imaging by MRI (magnetic resonance imaging). It was recorded at 12 months and 24 months, but our focus is

on the 24 month data because that gives more time for the implanted cartilage to mature. MOCART has subscores that look at issues such as whether the chondral defect (the gap of missing articular cartilage) has filled completely, and at the smoothness of the surface, which could be an indication of whether the gap has been filled with hyaline cartilage or less durable fibrocartilage.

The ICRS scores are based on inspection of the repair by arthroscopy, and on the histology of biopsies of the repair. Only a minority of patients had arthroscopy – 10 from the Spherox arm and 7 from the microfracture arm. The Bern score also examines the composition of transplanted cartilage.
The Lyshom score is based on patient reports on 8 aspects: pain, limping, locking, stair-climbing, need for supports, instability, swelling and squatting. It has a range 0 to 100 (best),
Days of absence from work is useful, but another option, not used in this trial, is time to resumption of previous activities, which is particularly relevant to sportspeople, who may be able to work but may not be able to play sports again. Some recent studies have used return to sport as an outcome.
IKDC (International Knee Documentation Committee) is another symptom score with range from 0 (worst) to 100 (best), based on function, symptoms, and range of motion.

Quality
As assessed by the Cochrane risk of bias score (Appendix 1), COWISI was a good quality trial though blinding of intervention was impossible. The submission notes that MRI and follow-up biopsies were assessed centrally by blinded independent radiologists and pathologists, respectively. However the key outcomes are neither radiological nor pathological, but symptoms. One source of bias may have been avoided because (pages 28-20)

"*Patient-Reported Outcomes data were entered directly by the patients into an ePRO (electronic Patient-Reported Outcome) system specifically designed for the trial."*

That removes the chance for non-blinded clinical staff to influence patient responses.
There were 102 patients randomised, not far short of the 118 in the TIG-ACT trial[38], but less than the 144 in SUMMIT.[39]

Baseline matching was good, [redacted]. The table (Table 7) of baseline characteristics does not provide details of duration of injury and proportions having previous attempts at repair. The defect sizes after debridement were similar: [redacted].

There were what were described as major protocol violations in [redacted] in the Spherox group and [redacted] in the MF group. These included some violations that may not seem major. They included (CSR pages 82-83):

[redacted]

The trialists seem to have been quite strict.

The sample size was based on showing non-inferiority which seems odd. We would have expected the trial to be aimed at showing that Spherox was better than MF, since that is what other trials of ACI aimed to do. Non-inferiority was taken to be shown if the KOOS score with Spherox was not 8.5 points lower than with MF. A clinically meaningful difference in KOOS is usually taken to be 10 points or more, but some researchers accept 8 as a meaningful difference.

In a non-inferiority trial, one should justify the choice of the non-inferiority margin, which corresponds to some loss of efficacy that might be accepted, with regards to other benefits, like safety ones, that the new intervention might have over the compared intervention. There is no such justification in the Co-Don submission.

Pages 42-43 of the Co-don submission outlines the testing for non-inferiority, and elsewhere there are references to power for non-inferiority. [redacted]

It may be that the aim was to show similarity with other trials of ACI versus MF, which do not usually show differences in the early years, but COWISI will be collecting data at 5 years, by which time an effective form of ACI may be giving better results than MF. So we might have expected the longer-

term aim of COWISI being to show superiority over MF. This is mentioned later, just after Table 14, where it is stated;

"The study was designed to test the non-inferiority and possible superiority of Spherox"

Results

Table 2 Results of COWISI trial

	Spherox	MF
Baseline KOOS	Mean [redacted] Median [redacted]	Mean [redacted] Median [redacted]
24 month KOOS	Mean [redacted] Median [redacted]	Mean [redacted] Median [redacted]
Change baseline to 24 months	Mean [redacted] Median [redacted]	Mean [redacted] Median [redacted]
Baseline MOCART	Not reported	Not reported
24-month MOCART	[redacted]	[redacted]

In the text below Table 16, we are told that the ANCOVA difference in change in KOOS is [redacted], which does not fit with the 24-month figures of [redacted]. Shortly below, we are told that ANOVA analysis gives figures of [redacted] for Spherox and [redacted] for MF, a difference of [redacted]. [redacted]

KOOS subscore results are given in Co-Don Table 17, [redacted], but with p values not given. [redacted]. Co-Don Table 18 gives changes from baseline in KOOS subscores, without p values, but reporting in the text that the improvement in one subscore, function in daily living, [redacted] and median changes [redacted] for Spherox and MF respectively. Since the [redacted], we do not think the subscore analysis adds anything of note.

The MOCART scores (Co-Don Table 20) at 24 months show [redacted]. MF gave slightly better results but the difference had confidence interval (presumably 95% CI, but not stated) of [redacted]. The submission notes (page 112 and table 29) that there was "at most - a

very weak correlation" between MOCART and KOOS scores. Some figures in Table 29 appear to have been misplaced.

ICRS results at 24 months were available from only 10 Spherox and 7 MF patients. Arthroscopic assessment showed no significant differences between arms. Histological assessment is reported in Co-Don Table 22, reproduced in Table 3.

Table 3 Cartilage repair assessment: numbers of patients and biopsy results

	Spherox	MF
Hyaline	[redacted]	[redacted]
Mixed hyaline and fibrocartilage	[redacted]	[redacted]
Fibrocartilage	[redacted]	[redacted]
Fibrous tissue	[redacted]	[redacted]

The Bern score results showed no difference.

The IKDC examination has four grades. The baseline and 24 month results are shown in Table 4. The text states that the [redacted] but no statistical test is provided. The SE of the [redacted] for Spherox at 24 months is 4.7%, and for the [redacted] for MF is 6.3% so the CIs overlap. (ERG calculations).

Table 4 IKDC Knee Examination results

	Spherox		Microfracture	
Grade	Baseline (47)	24 months (48)	Baseline (48)	24 onths (49)
A. Normal	[redacted]	[redacted]	[redacted]	[redacted]
B. Nearly normal	[redacted]	[redacted]	[redacted]	[redacted]
C. Abnormal	[redacted]	[redacted]	[redacted]	[redacted]
D. Severely abnormal	[redacted]	[redacted]	[redacted]	[redacted]

There are 10 IKDC Current Health Assessment subscores, [redacted] versus [redacted] on a scale of 0 to 100. Given the number of tests this may be a chance finding.

The Lysholm scores [redacted].

The proportions of recruits improving by 10 or more points on the KOOS score ("responders") at 24 months were [redacted].

Overall, in the planned analysis, there was [redacted].

Once the results were available, an alternative analysis was carried out, using a one-sided confidence level of alpha = 0.05. [redacted]. The ERG is doubtful as to whether this post-hoc analysis with a changed alpha represents good practice.

In the alternative analysis, superiority was also reported for change in the physical functioning score of the IKDC current health assessment subscore, but no figures or p value were provided.

Additional analyses

The results for two age groups, 18-34 and 35-50 years, were compared. Both age groups are reported to have had significant improvements, but neither baseline KOOS scores or changes from baseline are not given, only 24 month scores.

[redacted].

The Clinical.Trials.gov registration includes the outcome of days of absence from work (employment) and/or days of inability to follow usual activities during the last year or since the last visit, respectively, and time point when patient was back to work and/or to follow usual activities, but this is not reported in the submission.

Defect sizes

The COWISI trial included patients with (page 23 of Co-Don submission) defect sizes after debridement of >1 cm^2 to <4cm^2. The NICE ACI FAD recommends that ACI should be used only for lesions greater than 2cm^2. We therefore asked Co-Don as part of the clarification process, to split the COWISI results by defect size. We requested this breakdown because it is known that the effectiveness of microfracture declines as lesion size increases, and in our clarification request we hypothesised that the microfracture results in the smaller defects (<2 cm^2) might be better relative to Spherox, than in larger lesions. So the overall results of COWISI might have been missing a greater effect in the group to which the NICE FAD on ACI restricts it.

The results are in Table 5 – see row in bold. Figure 1 shows the flowchart for participants with lesion size >2cm^2. [redacted].

Table 5 Changes in KOOS score by defect size

KOOS score	*24 months*			
	Strata Defect size and treatment group			
	ACT:1-≤2cm²	*MF:1-≤2cm²*	*ACT:>2-4cm²*	*MF:>2-4cm²*
KOOS (overall)	[redacted]	[redacted]	[redacted]	[redacted]
Changes from baseline	[redacted]	[redacted]	[redacted]	[redacted]
Missing	[redacted]	[redacted]	[redacted]	[redacted]
Mean	[redacted]	[redacted]	[redacted]	[redacted]
SD	[redacted]	[redacted]	[redacted]	[redacted]
Minimum	[redacted]	[redacted]	[redacted]	[redacted]
Lower quartile	[redacted]	[redacted]	[redacted]	[redacted]
Median	[redacted]	[redacted]	[redacted]	[redacted]
Upper quartile	[redacted]	[redacted]	[redacted]	[redacted]
Maximum	[redacted]	[redacted]	[redacted]	[redacted]

These figures are based on the ITT populations. Numbers are quite small (see figure 1), and fall even further if those with protocol violations are removed. In their response to clarification questions, Co-Don reported that non-inferiority was shown between Spherox and MF in both defect size groups.

Figure 1 Flowchart by size of defect.

Other clarification responses.

Co-Don explained how they had calculated failure rates in the NMA, when there were no failures in COWIS and SUMMIT;

The median RR of 0.9894 was calculated assuming that in each arm 0.5 patients experienced the event. This approach was used per the NICE DSU document (Dias et al. 2016; reference provided with this submission) which recommends this in the case that no events are observed in one arm of the trials. Due to the larger sample size of the SUMMIT trial, a RR in favour of MACI was obtained. However, for the purpose of the economic model, 0 events were assumed for both interventions.

The Phase II trial (NCT01225575)

The aim of this trial was to compare three doses of Spherox. There was no control group. It recruited people with defects of 4-10 cm^2 in area, and about two-thirds had patellar defects. So the group studied is different from those in the COWISI trial, which had no recruits with lesions that large and was almost entirely of condylar defects. The restriction to large defect sizes was stated (Becher et al 2017[40]) to be because ACI was already regarded as the standard of care for medium (3-4 cm^2) defects.

The trial is summarised in Table 6, adapted from Table 4 of the Co-Don submission.

Table 6 Summary of dosages trial

Study	Prospective, randomised, open-label, multicentre Phase II clinical trial to investigate the efficacy and safety of the treatment of large defects (4–10 cm^2) with 3 different doses of Spherox in subjects with cartilage defects of the knee (Trial no. cod 16 HS 14)		
Study design	Dose-response study.		
Population	Males and females between ages of 18 and 50 years with an isolated single cartilage defect of the knee joint		
Intervention(s)	Spherox Group A:patients receiving 3-7 spheroids/cm^2 Group B:patients receiving 10-30 spheroids/cm^2 Group C: patients receiving 40-70 spheroids/cm^2		
Comparator(s)	Not applicable		
Indicate if trial supports application for marketing authorisation	Yes	**Indicate if trial used in the economic model**	No
Rationale for use/non-use in the model	Not used in the model as not comparison with microfracture that could be included in the network meta-analysis.		
Reported outcomes specified in the decision problem	Change of overall KOOS from baseline to final assessment at 12 months after implantation. Follow-up visits are planned at 24, 36-, 48- and 60-months.		
All other reported outcomes	• Changes in KOOS • MOCART (magnetic resonance observation of cartilage repair tissue) • Modified Lysholm score • IKDC (International Knee Documentation Committee) knee examination form • IKDC current health assessment form • IKDC subjective knee evaluation form • Bern score • International Cartilage Repair Society rating		

An unusual feature of this study, which has been published in part (Niemeyer et al 2016 [41]with the 12-month follow-up, Becher et al 2017 [40] with safety data) in that 63% of chondral defects were on the patella and only 37% on the femoral condyle. Patellar lesions tend to do less well than femoral condyle ones. Results are not provided separately for patella and condyle.

The trial appears to be well-designed, but for our purposes the lack of a control group reduces its value, and 30% withdrew prematurely. [redacted]. One entry criterion was defect size 4-10 cm^2 but the mean defect size was 5.6 cm^2 and only 10 of the 75 patients had 7-10 cm^2 defects.[40] The table of baseline characteristics gives no details of duration of injury or of previous attempts at repair. The groups were well-matched at baseline.

There were [redacted], especially in the high dose group, sometimes due to inadequate cell proliferation in culture. The rest include failure to attend visits or to complete data collection.

The final results showed no important difference amongst the three groups, so we only report the whole group results here.

The KOOS score [redacted].

No analysis by duration of defect, or by history of previous repair attempts, was reported, but at clarification stage, Co-Don provided data showing no difference by duration of injury;

"Due to the study design of this Phase II dose confirmation study, the results are based on Spherox data only. For a total of [redacted] treated patients, the 4 year follow-up analysis yielded the following results:

- *< 1 year: n=[redacted]; mean ± SD,* [redacted]
- *>1 year: n=[redacted]; mean ± SD,* [redacted]"

The ERG identified the article by Becher et al[40] presenting the safety outcomes. We used this reference, that was not included in the Co-Don submission (it was published on-line on 12th May, perhaps too late), to check the results from the phase II RCT in the submissions against those in this manuscript. The aim of this paper was to report the safety outcomes at 36 months post treatment so no effectiveness outcomes were presented. The occurrence of severe adverse events (AE) over time were described consistently with the Co-Don submission together the baseline characteristics of included patients. In the Co-Don submission, adverse events in the trial were reported at 12 and 48 months, meaning that the ERG could not compare table 44 of the submission against the Becher et al. paper. Treatment-related AEs were infrequent – [redacted] arthralgia and [redacted] of chondropathy (cartilage disease).

Meta-analysis

Section B.2.8 provides a meta-analysis of the phase II and COWISI trials, but since these recruited mutually exclusive groups, the meta-analysis does not seem to add much.

ERG comments on Summary by Co-Don

Section B.2.12 states that "*Spherox is a fourth generation ACI and represents a marked improvement over microfracture*". This is not what the evidence summarised above shows. A number of statements are made by Co-Don about the comparison with MF. These are in italics below with our comments added

Spherox demonstrates the following improvements over MF:

- *Spherox aims to produce hyaline-like cartilage whereas MF is associated with the production of fibrocartilage which is inferior cartilage.* ERG comment: this was not shown in the COWISI trial, as reported in Table 3.

- *Spherox is shown to be more effective than MF across age categories studied.* ERG comment: Spherox was not shown to be more effective than MF.

- *Spherox can be used for large defects (up to 10 cm^2) whereas MF is generally used on smaller defects (1-4cm^2)* ERG comment: This comment is fair, because the larger the defect, the poorer the result with MF. However Co-Don did not provide any comparison with MF in defects larger than 4cm^2.

- *Spherox is associated with fewer serious adverse effects than MF*. ERG comment: There is a little support for this statement. In the Spherox arm of the COWISI trial there were no serious AEs related to the procedure. In the MF arm there were three AEs possibly related to the procedure, one deep vein thrombosis, one arthralgia and one adhesions.

- *Spherox may reduce the following complications because of the autologous cells used in the procedure:*
    - *Rejection and incompatibilities – where patients may require further procedures*
    - *Viral contaminations*
    - *Overcomes any objections to the procedure on religious grounds - no porcine derived collagen membrane*

ERG comment: none of these comments are relevant to a comparison to traditional MF, though the last might be if MF is used with a cap, or when Spherox is being compared with older forms of ACI. (Allografts were not included amongst the comparators.)

- *Using Spherox as first line surgical treatment before MF could be more effective than using MF 1st line before Spherox.* ERG comment: no evidence has been produced to support this statement because both the Cowisi and the Phase II trial excluded patients who had had previous MF. Based on research on other forms of ACI, we expect it to be true. However the FAD on ACI recommends ACI as first line in defects greater than 2 cm^2 so this comment is now superseded.

3.6 Clinical effectiveness - network meta-analysis

The ERG has appraised the methodology of the NMA, in particular focusing on the assumptions of homogeneity, similarity, and consistency. The NMA used only two outcomes, proportion of responders and failures (defined as requiring further surgery). KOOS is not used, despite being the primary outcome in the COWISI trial.

Baseline characteristics of included studies

In Table 7, we show the baseline characteristics of the three trials included in the Co-Don NMA. Co-Don provide the results of the NMA, and then report the assessment of heterogeneity based on the key studies characteristics in section B.2.9.3. The ERG believes it would have been more appropriate to do the heterogeneity assessment prior to running the NMA, because we think this should have led to a decision not to undertake the NMA. Note that the Co-Don review of heterogeneity does not consider one of the most important factors, namely whether patients had had previous attempts at repair.

Table 7 Baseline characteristics of the three trials included in the Co-Don NMA

Variable	COWISI		SUMMIT		TIG/ACT	
Study sponsor	Co-Don		Sanofi (Vericel)		TiGenix	
Region/Country	EU: Germany and Poland		EU: Czech Republic, France, Netherlands, Norway, Poland, Sweden, UK		EU: Belgium, Croatia, Germany, Netherlands	
Number of centres	11		16		13	
Study period	Dec 2010-February 2017		Began May 2008		February 2002-January 2008	
Compared interventions	Spherox	MF	MACI	MF	ChondroCelect	MF
Sample size	52	50	72	72	57	61
Age ±SD	36 ±10	37±9	34.8 ±9.2	32.9 ±8.8	33.9±8.5	33.9±8.6
Male sex (%)	33 (63.5)	28 (56.0)	45 (62.5)	48 (66.7)	35 (61)	41 (67)
BMI (kg/cm2) ±SD	25.7 ± 3.3	25.8 ± 3.0	26.2 ± 4.3	26.4 ± 4.0	28 (49%) and 26 (46%) with a BMI≤25 and>25 to ≤30 respectively	31 (51%) and 24 (39%) with a BMI≤25 and >25 to ≤30 respectively
Lesion size cm^2	2.2 ± 0.7	2.0 ± 0.8	4.9 ± 2.8	4.7 ± 1.8	2.6 ± 1.0	2.4 ± 1.2
Previous repair procedures affecting subchondral bone n (%)	[redacted]		marrow stimulation techniques (34.6%),		14% (MF 5, drilling 3, abrasion 1)	7% (MF 1, drilling 2, abrasion 1)
Duration of symptoms (years)	[redacted]	[redacted]	5.8 (0.05-28.0)	3.7 (0.1-15.4)	1.97	1.57
Type of lesions	Isolated ICRS grade III or IV single-defect chondral lesion on femoral condyle		Cartilage defects of the medial femoral condyle (MFC), lateral femoral condyle (LFC) and/ or trochlea		single grade III to IV symptomatic cartilage defects of the femoral condyles	
Outerbridge grade n (%)						
III	[redacted]	[redacted]	21 (29.2)	15 (20.8)	10 (18)	16 (26)
IV	[redacted]	[redacted]	51 (70.8)	57 (79.2)	47 (82)	45 (74)
Location n (%)						
Medial femoral condyle	52 (100)	49 (98)	54 (75.0)	53 (73.6)	57 (100)	61 (100)

Lateral femoral condyle			13 (18.1)	15 (20.8)		
Trochlea	0	0	5 (6.9)	4 (5.6)	0	0
Origin n (%)						
Acute trauma	19 (36.5)	24 (48)	33 (45.8)	45 (62.5)	NA	NA
Chronic degeneration	1	[redacted]	18 (25.0)	9 (12.5)	NA	NA
Osteochondritis dissecans	none	[redacted]	8 (11.1)	12 (16.7)	NA	NA
Unknown	none	[redacted]	9 (12.5)	6 (8.3)	NA	NA
Other	32	[redacted]	4 (5.6)	0	NA	NA
Baseline KOOS score						
Overall	[redacted]	[redacted]	NA	NA	56.3 ± 13.6	59.5 ± 14.9
Pain	[redacted]	[redacted]	37.0±13.5	35.5±12.1	62.1 ±18.73	65.5 ±17.1
Function	[redacted]	[redacted]	14.9 ± 14.7	12.6 ± 16.7	NA	NA
Concomitant surgery	0	0	36%	31%	7%	11%

Table 8 Results of MF in the three trials.

	Response				Failure	
Trial	Definition	KOOS score		Responders, n/N (%)	Definition	Failure, n/N (%)
		Baseline	24 months			
COWISI	At least 10-point improvement on overall KOOS score	[redacted]	[redacted]	[redacted]	Objective clinical findings by the investigator, which are directly correlated with subjective patient complaints resulting in a deterioration of the subjective clinical outcome as assessed by the total KOOS and the 5 KOOS	[redacted] at 24 months

					subscores. Or need for revision surgery.	
SUMMIT	At least 10-point improvement in both the KOOS pain and function subscales	Pain: 35.5 ± 12.1 Function: 12.6 ± 16.7	Pain: 70.9 ± 24.2 Function: 48.7 ± 30.3	49/72 (68.1)	After week 24, a patient and physician global assessment result that was the same or worse than at baseline, a <10% improvement in the KOOS pain subscale, physician diagnosed failure ruling out all other potential causes, and the physician deciding that surgical retreatment was needed	2/72 (2.8%) at 24 months
TIG/ACT	Overall KOOS of at least 10 and/or an increase from baseline of at least 10 in at least 3 of the 4 KOOS subdomains and/or an improvement from baseline in the degree of knee disorder severity of at least one category or a decrease from baseline of at least 20 points in VAS pain score and/or an improvement in the degree of knee disorder severity of at least one category.	Overall: 59.5 ± 14.9 Pain: 65.5 ±17.1	NA	31/51 (61%) at 36 months. (Note error in Table 3 of Saris 2009 – correct denominator is 51)	If the surgeon decided that reintervention in the index lesion was necessary because of the persistence or recurrence of symptoms	7/61(11.5%) at 36 months About 10% at 24 months (from graph)

The Co-Don critique of their NMA (pages 81-82 of their submission) is quite rigorous, gives several reasons why the NMA was inappropriate, and does cast doubt (page 83, last paragraph) on their comparability.

The ERG would phrase this more strongly. There is considerable heterogeneity in the baseline characteristics across studies that were included in the NMA as shown in Table 7.

The studies were conducted over different time periods and settings. There could be variations in techniques for both MF and AC depending on the practice and experience of centres, especially given the long experience with Spherox in Germany.

There were differences in inclusion criteria across studies particularly with regards to the baseline KOOS score, much lower in the SUMMIT trial, and the lesion size, much larger in SUMMIT. This led to differences in baseline characteristics of patients across studies for these two variables.

Because the SUMMIT trial included patients with moderate to severe KOOS pain scores (<55), this resulted in a major imbalance in KOOS between SUMMIT, and COWISI and TIG/ACT. The KOOS score at baseline appears to be an effect modifier for one of the outcomes used in the Co-Don NMA, namely the proportion of responders with responders being defined as having at least a 10-point improvement in one or several KOOS subscales. It is likely that the achievement of response was easier with a lower KOOS score at baseline, as in the SUMMIT trial, compared to higher KOOS scores at baseline, as in COWISI and TIG/ACT.

The SUMMIT trial included patients with a minimum lesion size of $3cm^2$, which also results in a considerable imbalance in the mean lesion sizes at baseline (between [redacted] in COWISI and TIG/ACT vs 4.7-4.9 cm^2 in SUMMIT). The lesion size is an effect modifier because there is evidence suggesting that ACI has a better outcome compared to MF in people with larger lesions, in which MF is less successful (for review see Mistry et al 2017[42]). So one might expect the MF group in SUMMIT to do less well than the MF group in COWISI.

However the most important difference is the absence of previous attempts at repair in the COWISI patients, whereas 35% and 14% of the ACI groups in SUMMIT and TIG/ACT had had previous repair attempts, mainly MF.

One way of assessing heterogeneity is to compare the results of MF in the three trials, as in Table 8. The proportion of responders was [redacted] in COWISI ([redacted]) than in the other two trials: SUMMIT 68% and TIG/ACT 62%. The proportions of failures also varied. This provides more evidence that the patient groups were different, and that an NMA might have been inappropriate.

The OA criteria in the three trials varied, as shown in Table 9.

Table 9 Osteoarthritis criteria in the three trials.

	Criteria regarding osteoarthritis in the three trials
COWISI	Exclusion criteria: Radiological signs of osteoarthritis, taking specific osteoarthritis drugs such as chondroitin sulphate, diacerein, N-glucosamine, piascledine, capsaicin within two weeks of baseline.
SUMMIT	Exclusion criteria: Kellgren-Lawrence grade 3 or 4 osteoarthritis
TIG/ACT	Exclusion criteria: Advanced osteoarthritis (as defined by Radiographic Atlas of Osteoarthritis, grade 2-3), taking specific osteoarthritis drugs, such as chondroitin sulfate, diacerein, n-glucosamine, piascledine, and capsaicin, within 2 weeks of the baseline visit

The effects of the heterogeneity are mixed;

- Comparing Spherox and VerMACI using COWISI and SUMMIT should disadvantage Spherox because of the baseline KOOS scores and defect sizes
- Comparing Spherox and VerMACI might disadvantage the latter because of the longer duration, if we extrapolate from TIG/ACT 5-year data which showed that ACI was less successful in defects with longer duration
- Comparing Spherox with both the other trials should disadvantage VerMACI and ChondoCelect because of the previous repair attempts

Transitivity assumption

The Co-Don submission does not discuss whether or not they assessed the transitivity assumption and whether it was violated. If the transitivity assumption is compromised or does not hold, the consistency assumption is also violated, leading to biased estimates in the network meta-analysis. The ERG examined the transitivity assumptions applicable to the NMA included in the CS.

The transitivity assumption does not hold if the distribution of population characteristics that are effect modifiers differ across the treatment comparisons of a network. Three such treatment effect modifiers in the Company's NMA are the KOOS score, the lesion size at baseline and previous repair attempts. The networks for the proportion of responders and failure rate include three RCTs of clinically diverse populations based on the KOOS score and lesion size at baseline, rendering the compared treatments in the networks not jointly randomizable. The uneven distribution of these effect modifiers across the network comparisons violates the transitivity assumption.

Another threat to transitivity assumption is the potential difference within the microfracture interventions as previously described, which means that these interventions may not be exactly considered as one node of microfracture.

Lastly, there was some variation on the definition of responders, one of the NMA outcomes, which means that the number of responders was not assessed consistently across studies. Failure rates were reported over different timescales (2 years for SUMMIT and COWISI, 3 years for TIG/ACT, though 2-year data for TIG/ACT were available).

Overall, owing to the violations on transitivity assumption, the validity of the estimate for the indirect comparisons is very questionable.

ERG comments

On page 78, there is a statement: "*The median RRs suggest that Spherox is associated with a higher number of responders when compared to MF"*. This is not what was reported from the trial in Co-Don Table 30 – there were [redacted] responders for Spherox and MF respectively, and in the forest plot the RR is 0.9684. The text does note that the results are not statistically significant.

On page 80, there is a comment that TIG/ACT only published outcomes at three years. This is not entirely correct. Saris and colleagues[38] provide 2-year data (in the figures) for KOOS scores and treatment failures, showing a clear separation of KOOS scores and failure rates between ACI and MF arms by 24 months.

On page 83, table 38 shows differences in the MF KOOS results for the MF arms in COWISI and SUMMIT, with the statement that:

At the end of the 24th month, patients receiving Spherox and MACI report [redacted].

A more likely explanation is lesion size, which is much smaller in COWISI than SUMMIT (means of about 2.1 and 4.8 respectively). So we would expect much better MF results in COWISI.

Table 82 has a few unimportant errors. The studies by Clave et al[43] and Jones et al[44] were probably meant to be listed as exclusions. We note that Knutsen et al 2016[45] is listed as an inclusion, but is not in the NMA. Knutsen 2004[46] is listed as an exclusion but it could be argued that it is relevant, as an RCT of ACI versus MF.

3.7 Evidence from case series

The Co-Don submission reports that Chondrosphere has been used in Germany since 2004, and that more than 10,000 patients have been treated. Unfortunately this widespread use does not seem to have been accompanied by an equivalent amount of data collection. In the submission to the EMA, 11 studies were included, but all but one were "case studies, conference posters and study reports without detailed information of the conducted study". (CHMP 2017). Some were about ACI in the hip joint. Spherox has been used in knee, ankle, shoulder and hip.[47]

The ERG has identified some case series. Co-Don did not use these in their submission, except in a list in an appendix. Given that we have evidence from only one (as yet) short-duration RCT with an active comparator, we have looked at some case series to see what can be gleaned.

Quality assessments are provided in Appendix 1. Note that studies can be assessed as poor quality for two reasons;

- The study was of poor quality
- The study might have been good quality but insufficient details are provided to assess quality

Fickert et al 2012 [48]

This case series was assessed as fair quality. Fickert et al from Mannheim in Germany recruited 37 patients with isolated chondral defects in the knee, roughly half patellar and half femoral condyle. 13% had had previous attempts at repair. Duration of defects ranged from 2 months to 11 years, but analysis of results by duration under one year or over showed no difference in most outcomes, Tegner being the exception.

Implantation was by medial mini-arthrotomy with mean operation time 60 minutes. The authors noted the possibility of arthroscopic implantation.

Seven of the 37 had AEs, mainly local such as effusion and locking, but with one deep vein thrombosis (DVT) and pulmonary embolism. The patient who had the DVT and embolism was aged 46, and had a longer than usual (148 minutes) operation that included ACL reconstruction.

There were no important differences by defect site, leading Fickert et al to suggest that Spherox may be more effective in patellar defects than other forms of ACI.

Improvements in SF-36 are reported but no p values are given and the improvements, while definite, do not appear from the graph to be statistically significant.

One weakness of the study is that follow-up was only for 12 months, but longer follow-up was planned. However, we have found no further publications from Fickert and colleagues.

The lack of a control group is the main weakness. One strength is that the recruits may be more typical of routine practice than RCT recruits. Six had BMIs over 30 and several were over 50 years of age.

Siebold 2015 [49]

Siebold and colleagues from Heidelberg performed "second-look" arthroscopy on 57 cartilage lesions in 41 patients at a median of 10 months, mean 13 months (range 6 to 72 months) after arthroscopic spheroid implantation. No information is given on what proportion of all patients treated with spheroids had second-look arthroscopy, but all who did had another reason for arthroscopy (table 3 of paper) – none of the arthroscopies were done just to evaluate the cartilage repair. So this case series may not reflect the outcomes for the generality of Spherox patients. It is noted that 27 patients (66%) had ACI combined with other procedures, which is common, and understandable in the interests of patient care, but which does make interpretation of the benefits of ACI more difficult.

The ICRS Cartilage Repair Assessment grading, based on visual inspection and probing, was reported to be normal in 12 lesions (21%), nearly normal in 40 (70%) and abnormal in 5. Clinical follow-up data (KOOS etc) was not available in 24%, but in any case, baseline pre-operation data were not provided. None of the patients reported by Siebold et al had had previous repair attempts such as microfracture.

Maiotti 2012 [50]

This study was available only as an abstract with sparse detail making quality assessment difficult. It reports on only 23 patients, of whom only three had follow-up biopsies. One useful item was that the spheroids were all implanted arthroscopically.

Roessing 2010 [51]

This is available only as an abstract from an ICRS meeting, so details are sparse, and we have not attempted quality assessment. 42 patients had spheroids implanted arthroscopically. The aim of the study was to show that spheroids could be implanted arthroscopically, in which it succeeded. Follow-up was for 2 years, during which time no failures requiring further surgery occurred, and symptoms improved (no figures given). The patients in Roessing may include some from the unpublished Co-Don document cod RS1 SR 2015, which had 19 patients.

Schreyer 2010 [52]

Schreyer and colleagues from Darmstadt in Germany compared three ways of implanting Co-Don chondrocytes: by ACI-P (40 patients 1998 to 2004), and as spheroids (2005-2009) by arthrotomy (15) or arthroscopically (16). They concluded that uncapped implantation was as good as with ACI-P.

Other studies

Some other data were supplied to the EMA and reported in the CHMP assessment report 2017. These included an unpublished case series by Zinser in Dinslaken, Germany, in which before and after improvement in the IKDC from 39 to 61 points was reported. However only 36 of 90 patients treated agreed to the analysis, raising questions of selection bias. Data from three pilot studies, including six patients treated by Dr Schreyer in Darmstadt, 26 from Dr Ruhnua in Buer and 10 from Dr Baum in Gundelfinger, are summarized in a book chapter by Libero and colleagues [47] which also provides a good account of the pre-clinical research on spheroids. These three pilots all report useful improvements in clinical scores, but their usefulness is limited, because of lack of control groups or even natural history studies. The chapter states that implantation was by "mini-arthroscopy", but we assume this means mini-arthrotomy.

In summary, these case series provide evidence of before and after improvement, and that Spherox can be implanted arthroscopically. Without comparators, their usefulness is limited.

3.8 Conclusion and discussion

The ERG's main conclusions at this stage are;

1. COWISI was a good quality trial, though blinding of intervention was impractical, duration of patient follow-up is as yet only two years, and it included patients with defects smaller (<2cm^2) than NICE currently approves for ACI

2. Spherox is clinically effective in treatment of chondral defects, and the improvement lasts for at least four years.

3. However, the comparative effectiveness is an issue. The evidence presented does not show [redacted]. This is perhaps not surprising, since MF is effective in the short-term, in smaller defects. So with longer follow-up, we would expect the benefits of MF to wane. We note that in the comparator trials, TIG/ACT and SUMMIT, ACI was showing an advantage over MF by 2 years, but these differed in some ways from COWISI.
4. We doubt whether it was appropriate to do the NMA given the heterogeneity. We do not regard the results of the NMA as robust, and insufficient to support the cost-effectiveness analysis.

If these conclusions are accepted, no positive results on clinical effectiveness are available to feed into the modelling, and we might stop here. However, the Appraisal Committee may take a more sympathetic view, so in the next section we provide a critique of the Co-Don cost-effectiveness analysis.

4 Cost-effectiveness

4.1 ERG comment on manufacturer's review of cost-effectiveness evidence

In the main, the company literature review provides a good summary of the available papers and their central cost effectiveness estimates. More could have been made of the scenario analyses of Mistry et al[42] particularly the scenario analyses around the effects of previous interventions and the effects of severity. It would also have been much improved if the company had summarised the evolving debate, cost effectiveness estimates and conclusions of the ACI MTA [TA477].[53]

4.2 Summary and critique of manufacturer's submitted economic evaluation

4.2.1 NICE reference case checklist

Attribute	**Reference case and TA Methods guidance**	**Does the de novo economic evaluation match the reference case**
Comparator(s)	The scope specifies: • Microfracture • ACI • Debridement • Mosaicplasty • BSC	The submission considers: • Microfracture (MF) • ACI: Spherox • ACI: ChondroCelect • ACI: MACI These are only considered in sequences where a 2nd repair is possible: • MF->MF • MF->ACI • ACI->MF • ACI->ACI Where the 1st ACI is followed by a 2nd ACI, the 2nd ACI is assumed to be the same as the 1st ACI.
Patient group	As per NICE scope. "*People with articular cartilage defects*"	The submission only considers knee repair. This is in line with the SmPC and the recent ACI assessment [TA477].

		The pivotal trial was limited to defects of between 1cm^2 and 4cm^2. The SmPC permits treatment of defects up to 10cm^2. This complicates the NMA, which is further complicated by the proportions of patients having had a previous repair differing between the trials. The recent ACI assessment [TA477] has approved ACI for defects of more than 2cm^2, in part due to a consensus statement by a group of experts.
Perspective costs	NHS & Personal Social Services	Yes.
Perspective benefits	All health effects on individuals	Yes.
Form of economic evaluation	Cost-effectiveness analysis	Yes. Cost utility.
Time horizon	Sufficient to capture differences in costs and outcomes	Yes. Lifetime.
Synthesis of evidence on outcomes	Systematic review	Yes. A systematic review and NMA are undertaken.
Outcome measure	Quality adjusted life years	Yes.
Health states for QALY	Described using a standardised and validated instrument	Yes. The quality of life values for 1st and 2nd repairs are taken from and are in line with those of TA477. TA477 derived values from Gerlier et al [54] who[1] analyse the TIG/ACT trial 5 year follow-up SF-36 data mapped to the QoL using the Brazier et al [55] algorithm. The quality of life values for knee repairs are also taken from

[1] Sponsored by TiGenix NV, provider of ChondroCelect

		TA477, but are not entirely aligned with it.
Benefit valuation	Time-trade off or standard gamble	Yes. Standard gamble.
Source of preference data for valuation of changes in HRQL	Representative sample of the public	Yes. 611 members of the UK general public.
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes.
Equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes.
Probabilistic modelling	Probabilistic modelling	Yes. But the clinical effectiveness estimates are varied independently.
Sensitivity analysis		A reasonable range of sensitivity analyses are conducted.

4.2.2 Model structure

A markov model with an annual cycle is developed based on the recent model of Mistry et al [42]. While the model structure is similar to that of Mistry et al, the transition probabilities differ quite considerably from it. In the opinion of the ERG the presentation of the model and the transition probabilities of tables 47 and 48 of the submission does not accurately or transparently present the implementation of the model. Section 3.3.3 of the submission should be read alongside the detail of section 5.2.6 on treatment effectiveness and extrapolation below.

The model compares 10 sequences, each sequence having two treatments or repairs. Up to the age of 55 only the two repairs of the sequence may be received. Thereafter patients may receive knee replacements.

Model structure to the age of 55

All patients receive the 1st repair of the sequence during the 1st cycle of the model. These patients can move into one of three health states[2].

- Success
- No further repair (NFR)
- 2nd repair

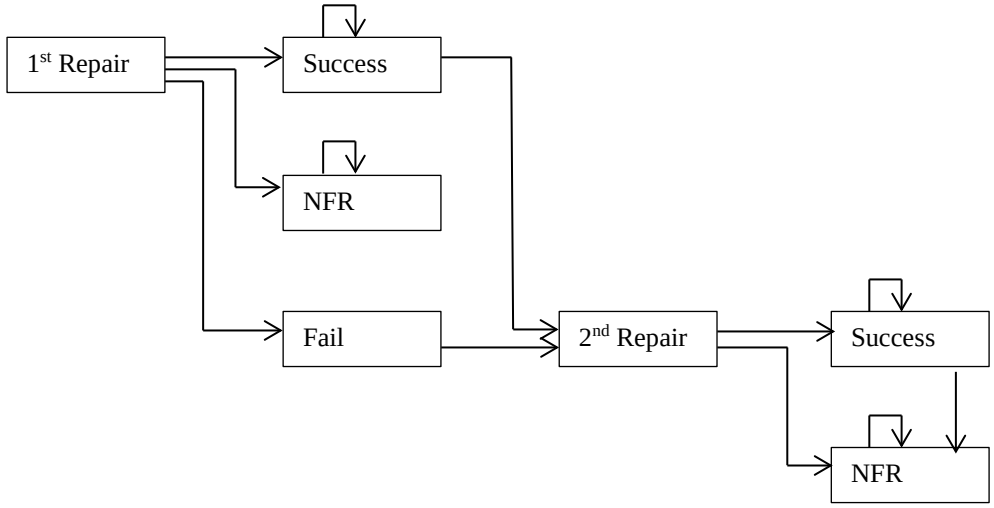
Subsequent to the 1st cycle those who were a success either remain a success or move to 2nd repair. All those in NFR remain in NFR.

The patients who receive the 2nd repair of the sequence can move into one of two health states.

- Success
- No further repair (NFR)

Those who were a success either remain a success or move to NFR. All those in NFR remain in NFR.

Figure 2 Model structure to age 55



The above is a slight simplification. Both the 1st repair and 2nd repair successes are divided into 5 health states: four annual tunnel health states of success for years 1 to 4 after the repair and a fifth health state of success in years 5+ after the repair. This is in line with Mistry et al. It enables quality of life values specific to the duration of success, years 1, 2, 3, 4 and 5+, to be applied. These differ between:

- ACI repairs and microfracture repairs
- 2nd ACI repairs after a 1st microfracture repair and 2nd ACI repairs after a 1st ACI repair

[2] Death from all-cause mortality is possible from all health states but is largely ignored in this description for sake of simplicity. The 1st knee replacement increases the probability of death in the year of operation by 0.35%, and subsequent knee replacements by1.10%.

A key difference between the company model structure and that of Mistry et al is that there is no possibility of 1st repair successes losing the benefits of success and transitioning into the NFR health state. In the company model 1st repair successes can only transition to 2nd repairs.

Model structure subsequent to age 55

From the age of 55 the model structure is augmented by a knee replacement (KR) module. There is a common annual 1.01% probability of receiving a 1st KR for patients who are in the 1st repair success, the 1st repair NFR, the 2nd repair success and the 2nd repair NFR health states. Those receiving a 1st KR can move into one of three health states:

- Success
- No further repair (NFR)
- Subsequent KR

Subsequent to the 1st KR those who were a success either remain a success, move to NFR or receive a subsequent knee replacement. Note that this NFR health state differs from the NFR health state of those moving directly from their 1st KR to NFR without success and is associated with a different quality of life. Those in NFR remain in NFR.

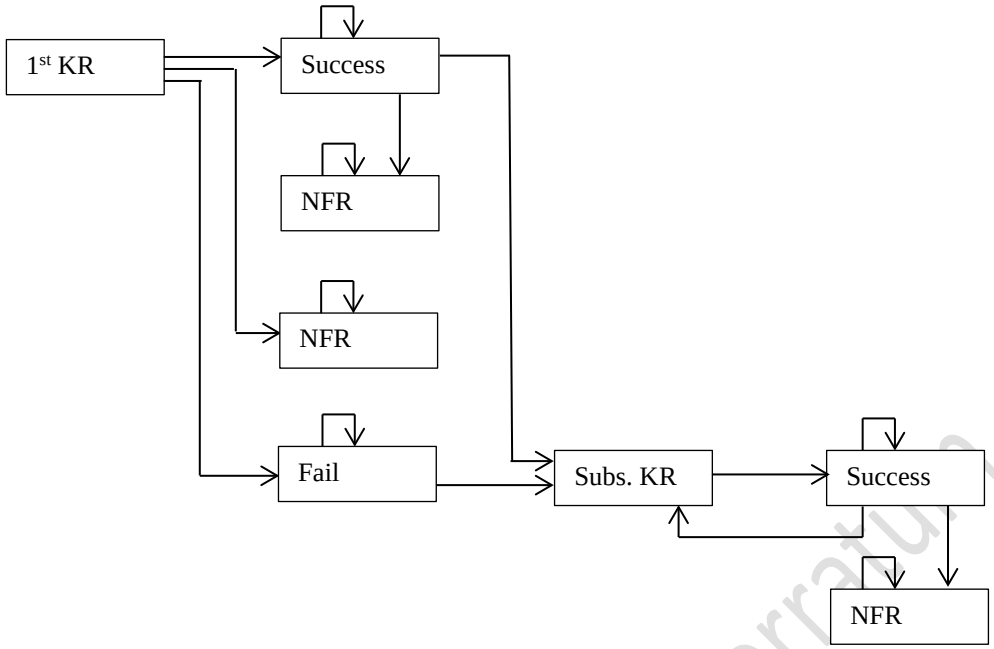
Those receiving a subsequent knee replacement can move into one of two health states.

- Success
- No further repair (NFR)

Those who were a success can either remain a success, move into NFR or receive a subsequent KR. The feedback loop between success and subsequent KR means that there is no limit on how many KRs a patient may receive. Those in NFR remain in NFR.

For the 1st knee replacement there is a 50:50 balance between total knee replacement (TKR) and partial knee replacement (PKR). All subsequent knee replacements are TKRs. This complicates the implementation of the KR module within the company electronic model

Figure 3 Knee replacement module from age 55



4.2.3 Population

The patient population reflects the baseline characteristics of the Phase III trial with a mean age at baseline of 34 years and 60% male.

4.2.4 Interventions and comparators

There are four main interventions:

- Microfracture (MF)
- Spherox (SPHX)
- MACI (MACI)
- ChondroCelect (CC)

All of these interventions are modelled as being part of a possible sequence of two repairs. The 1st treatment is applied to all patients. The 2nd treatment is applied to those requiring repairs after having received the 1st treatment. The 10 sequences that are compared are:

- Microfracture followed by another treatment:
    - MF->MF
    - MF->SPHX
    - MF->MACI
    - MF->CC

- Each ACI followed by microfracture:
    - SPHX->MF
    - MACI->MF
    - CC->MF
- Each ACI followed by itself:
    - SPHX->SPHX
    - MACI->MACI
    - CC->CC

4.2.5 Perspective, time horizon and discounting

The time horizon is 67 years, i.e. to 100 years of age, which is effectively a lifetime horizon. The perspective and discounting are as per the NICE reference case.

4.2.6 Treatment effectiveness and extrapolation

The company submission contains many errors. As outlined in section 5.4 the ERG has revised all the clinical inputs to the model derived from the trials and the company NMA. The company accepts that many revisions are required. But it has not documented its new method, suggests revised values that differ from those of the ERG and has not provided a coherent set of responses and additional analyses. In the light of this much of the original submission is irrelevant. The detail of the submission is presented below for completeness and to explain the ERG critique and the ERG changes to the company model. Most readers may wish to move forward to section 5.2.7 on quality of life.

Treatment effectiveness: Response rates and probabilities of remaining a success

During the 1st annual cycle of the model the two year probabilities of response, P_2, are applied to the 1st repairs. In effect the 1st cycle of the model is two years rather than one year, though the QALY and cost calculations do not particularly take this into account.

For instance, the 1st cycle applies the 81% probability of response for Spherox 1st repairs. The probabilities of response for the other comparator treatments are derived by applying the relative risks of the NMA to the Spherox response rates.

The company submission and the clinical effectiveness section raise serious issues around the NMA and its validity. If the NMA is invalid it may still be possible to consider a head to head of Spherox with microfracture. This is complicated by the estimated benefits of Spherox over microfracture not

arising from the clinical effectiveness estimates of the trial but stemming almost entirely from the assumption that all microfracture fails at 5 years.

The response rates are calculated as rate = $-\ln(1-P_2)*RR$, with these rates being back transformed along the lines of 1-exp(-1*rate) to yield the two year probability of response for the comparator. This is equivalent to estimating the two year probability of response for the comparator as $1-(1-P_2)^{RR}$, or defining the two year non-responder or failure rate as $F_2 = (1-P_2)$ is more simply $1-F_2^{RR}$.

Table 10 Two year probabilities of response for 1st repairs

	SPHX	CC	MACI	MF
Spherox 2yr probability of response $Sp(P_2)$	[redacted]			
Relative risk (RR)	1.000	1.209	1.223	0.968
1st repair 2yr probability of response $P_2 = 1-(1-Sp(P_2))^{RR}$	[redacted]	86.57%	86.88%	79.98%

To be able to outline what the company has applied for the treatment effectiveness of 2nd repairs and the ongoing probabilities of failure rates requires a small digression on the conversion of two year probabilities to annual probabilities[3].

If the two year probability of response is P_2 then for modelling purposes it is possible to take the square root of this to yield an annual probability of $P_1 = (P_2)^{½}$. While slightly curious in the current context, this annual probability of P_1 could then be applied during two cycles and the cumulative probability would be $P_1 * P_{1=} P_2$ and the correct proportion of responders would be modelled as occurring at the end of the 2nd year.

As an example the 2 year probability of response for Spherox is [redacted]. The square root of [redacted] is [redacted] which is the company estimate of the annual probability of response for Spherox as a 2nd repair. But the company model only applies this [redacted] annual probability once for 2nd repairs and does not compound it over two years to arrive at the [redacted] two year response rate. The response rates for 2nd repairs are consequently modelled as being much higher than the response rates for 1st repairs.

A similar logic can be applied if the probabilities of failure are to be modelled rather than probabilities of response, where by definition $F_1 = (1-P_1)$. By substitution $F_1 = 1-(P_2)^{½}$, and since $F_2 = (1-P_2)$ implies that $P_2 = (1-F_2)$ this in turn implies that $F_1 = 1-(1-F_2)^{½}$.

[3] This is simplified to only consider probabilities. The company implementation converts 2 year probabilities to annual rates along the lines of $r=-\ln(1-P_2)/2$ and from there to annual probabilities by $P_1=1-e^{-r}$. Substitution causes the exponentiation of the logarithm to disappear resulting in $P_1=1-(1-P_2)^{½}$.

The company applies the formula for failures to the two year probability of response P_2 rather than the two year failure rate F_2. For instance, applying the formula for failures to the [redacted] two year probability of response for Spherox yields a probability of [redacted] = 56.41%. The company takes this to be the annual probability of a successful Spherox repair failing and the patient moving into the No Further Repair (NFR) health state. This annual probability is not limited to being applied once, but is applied every cycle of the model to those with a successful Spherox 2nd repair.

This is most easily seen in table 48 on page 114 of the submission in the Spherox followed by Spherox transition probability matrix entry of 0.5641 for the probability of failing and moving from a successful 2nd repair to NFR. The residual of 0.4359 is the annual probability of remaining a successful 2nd repair.

Table 48 of the Co-Don submission suggests that a similar probability of 1st repair successes failing and moving into the NFR health state is applied. The 0.5500 entry of table 48 is based upon the 0.5641 probability, adjusted for the 0.0063 probability of 1st repair successes failing and receiving a 2nd repair. In the opinion of the ERG there is no probability of 1st repair successes moving into the NFR health state. The only means of exiting the 1st repair successes health state[4] is via a 2nd repair and for those age 55+ via a 1st KR.

This results in the following estimates for the probabilities of response from 2nd repairs and the annual probabilities successes from 2nd repairs failing and moving into the NFR health state[5]. These are independent of the type of 1st repair.

Table 11 Probabilities of 2nd repair responses and 2nd response successes to NFR

	SPHX	CC	MACI	MF
Spherox 2yr probability of response $Sp(P_2)$	[redacted]			
Relative risk (RR)	1.000	1.209	1.223	0.968
1st repair 2yr probability of response $P_2 = 1-(1-Sp(P_2))^{RR}$	[redacted]	86.57%	86.88%	79.98%
2nd repair probability of response = $P_2^{½}$	[redacted]	93.04%	93.21%	89.43%
Annual probability of 2nd success failing to NFR = $1-(1- P_2)^{½}$	56.41%	63.36%	63.78%	55.25%

Probabilities of a 2nd repair

The probabilities of a 2nd repair differ between the 1st cycle of the model and subsequent model cycles. For the 1st cycle of the model for both Spherox and MACI these are assumed to be zero. For

[4] Ignoring death.

[5] Though as already outlined the annual probabilities of success from 1st repairs moving into the NFR health state

the 1st cycle of the model for ChondroCelect and microfracture these are derived from the NMA input of a probability of [redacted] for Spherox coupled with the NMA relative risks of 2.032 for ChondroCelect and 6.979 for microfracture. As for the probabilities of response the relative risks are applied to the rate for Spherox and then back transformed to probabilities, resulting in a 2 year probability of failure of 1.01% for ChondroCelect and 3.44% for microfracture.
For subsequent cycles of the model, those who are an ACI success are assumed to have a 2 yearly probability of a 2nd repair of 1.25% as taken from Mistry et al. This is converted to an annual probability of 0.63%.

For subsequent cycles of the model, those who are a microfracture success are assumed to have a 2 yearly probability of a 2nd repair of 3.44% / 2 = 1.72%. This is converted to an annual probability of 0.86%. In effect the 2 year 3.44% probability derived from the NMA is quartered.

Knee replacement module
From the age of 55 all those remaining with a successful repair or having failed and fallen into the NFR health state have a common annual probability of being given a 1st knee replacement of 1.01%. There is a 50:50 balance between total knee replacements and partial knee replacements. Those who have had a knee replacement also have a common probability of 1.01% of having another knee replacement, all of which are total knee replacements.

Based upon Mistry et al 1st total knee replacements are associated with an increased probability of death of 0.7%, and 2nd total knee replacements 1.1%.

1st knee replacements have a 0.20% annual probability of failing and receiving no further treatment and a 0.58% annual probability of failing and receiving a subsequent knee replacement. The remainder have a successful 1st knee replacement.
Subsequent knee replacements have a 2.09% annual probability of failing and receiving no further treatment. The remainder have a successful subsequent knee replacement.

Those with a successful knee replacement, whether a 1st or a subsequent knee replacement, have a 1.62% annual probability of failing and receiving no further treatment and a 1.08% annual probability of failing and receiving a subsequent knee replacement.

4.2.7 Health related quality of life

Most of the quality of life values within the submission are taken from Mistry et al. Mistry et al derived their quality of life values for repair health states from Gerlier et al [54], and their quality of life values for knee replacement health states from Dong and Buxton[56], Gerlier et al[54] and Jansson and

Granath[57]. Gerlier et al mapped to the SF-36 data collected during the 5 year follow-up of the TIG/ACT trial to quality of life values using the Brazier et al[55]SF-36 to SF-6D to quality of life mapping function.

There are two key assumptions.

- In common with Mistry et al the company assumes that for microfracture all successes fail completely at year 5. This causes them to fall back to the baseline quality of life value of 0.654. The AC of the ACI MTA [TA477] requested a scenario analysis that assumes that the quality of life is maintained at 0.817. The company also provides this scenario analysis.
- For those receiving an ACI as a 2nd repair after a 1st repair of microfracture their 2nd repair deteriorates at year 4. This causes their quality of life to be the midpoint between the 1st year quality of life value of 0.760 and the quality of life of success of 0.817: 0.789. This assumption makes it less likely that reserving ACI to be only an option as a 2nd repair will be cost effective.

Table 12 Quality of life values for successful repairs

	1st repair		2nd repair		
	ACI	MF	ACI post ACI	ACI post MF	MF
Year 1	0.760	0.760	0.760	0.760	0.760
Year 2	0.817	0.817	0.817	0.817	0.817
Year 3	0.817	0.817	0.817	0.817	0.817
Year 4	0.817	0.817	0.817	0.789	0.817
Years 5+	0.817	0.654	0.817	0.789	0.654

In addition to the above quality of life values those moving into the failure and NFR health state after their repair have a quality of life value of 0.691. Those requiring a 2nd repair receive a quality of life value of 0.654 for that cycle.

For knee replacements the quality of life values are as follows.

Table 13 Quality of life values for knee replacements

Health state	QoL
1st KR	0.615
Subs KR	0.557
Success	0.780
NFR from 1st KR	0.691
NFR from 1st KR success	0.557
NFR from 2nd KR	0.557
NFR from 2nd KR success	0.557

4.2.8 Resources and costs

The resource use and many of the unit costs within the submission are taken from Mistry et al. With the exception of cell costs, the unit costs taken from Mistry et al are in 2012/13 prices and so are inflated by 3.4% to be in 2015-16 prices. These costs in Mistry et al are sourced from Clar et al 2005[58] and inflated from 2013-12 prices.

A company assumption is that Spherox implantation is done arthroscopically so requires a less invasive and shorter implantation procedure than other ACIs and so only incurs costs of £734 for both harvesting and implantation. The balance between total knee replacements and partial knee replacements is assumed to be 50:50 for 1st knee replacements, with all subsequent knee replacements being total knee replacements.

Unit costs of visits are taken from NHS reference costs. Unit costs of knee replacements are taken from the 2016-17 National Prices and Tariff.

Table 14 Unit costs

	Cost	Source
Harvesting	£734	Mistry et al, Arthroscopy, Table 22, inflated
Implanting SPHX	£734	Mistry et al, Arthroscopy, Table 22, inflated
Implanting CC and MACI	£1,065	Mistry et al, Arthrotomy, Table 22, inflated
Procedure MF	£3,122	Mistry et al, Procedure, Table 22, inflated
1st knee replacement	£5,556	2015-16 National Tariff
2nd knee replacement	£13,396	Mistry et al, 2nd TKR, Table 22, inflated
Outpatient visit	£121	Ref Cost: WF01A: OP: NA: FF: CL
Rehabilitation visit	£345	Ref Cost: REHBL2: rehabilitation for joint replacement

This, coupled with the cell costs and the visit and rehabilitation schedule of Mistry et al, results in the following total costs.

Table 15 Total costs of procedures

	SPHR	CHON	MACI	MFRC	1st KR	Subs KR
Cost of cells	£10,000	£16,000	£16,000	..	..	..
Harvesting	£734	£734	£734	..	..	..
Implantation	£734	£1,065	£1,065	..	..	..
Procedure	..	..	..	£3,122	£5,566	£13,397
Procedure cost	£11,468	£17,799	£17,799	£3,122	£5,566	£13,397
OP	6	6	6	3	2	2
Rehabilitation	3	3	3	3	0	0

Total Cost	£13,226	£19,556	£19,556	£4,518	£5,807	£13,638

4.2.9 Cost effectiveness results

As already outlined, the company accepts that the methods of its submission are incorrect. Following further clarification the company has submitted a deterministic set of results. In brief among the non-dominated sequences these are as follows. (Note that second procedures are only if required.)

Table 16 Revised company cost effectiveness results

	Cost	QALYs	Δ Cost	ΔQALYs	ICER
MF->MF	£5,762	15.878			
SPHX->MF	£14,174	17.955	8,412	2.077	£4,051
SPHX->SPHX	£14,993	18.000	819	0.045	£18,137
MACI->MACI	£22,312	18.395	7,319	0.395	£18,523

A key point to note in the above is that the cost effectiveness estimate for SPHX->SPHX compared to SPHX->MF of £18,137 per QALY is only slightly below the implied cost effectiveness estimate of MACI->MACI compared to SPHX->MF of £18,483 per QALY. It will only take a small increase in the effectiveness of MACI for SPHX->SPHX to be extendedly dominated by MACI->MACI. The ERG revised estimates suggest such an increase compared to the company revised estimates, as outlined in greater detail in section 5.3.4 below.

The revised company deterministic results are not accompanied by a revised electronic model, probabilistic modelling or sensitivity analyses. In the light of this the results of the original submission are presented below for completeness. But other than to inform the examination of the original company sensitivity analyses they are largely irrelevant.

Original submission results

The company base case deterministic results are as follows.

Table 17 Company deterministic base case results

	Cost	QALYs	Δ Cost	Δ QALY	ICER
MF->MF	£5,763	15.851	..	..	..
MF->SPHX	£7,156	15.851	..	..	Ext. Dom.
MF->CC	£8,168	15.849	..	..	Dominated
MF->MACI	£8,168	15.849	..	..	Dominated
SPHX->MF	£14,182	17.971	£8,419	2.120	£3,971
SPHX->SPHX	£15,017	17.972	..	..	Ext. Dom.
MACI->MF	£20,544	18.117	£6,362	0.146	£43,676

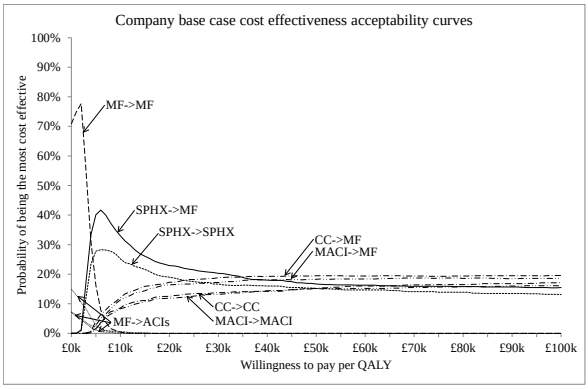
CC->MF	£20,588	18.110	..	..	Dominated
MACI->MACI	£22,091	18.116	..	..	Dominated
CC->CC	£22,283	18.109	..	..	Dominated

The central estimates of the company probabilistic modelling over 1,000 iterations are broadly similar to the deterministic. The main change is that the cost effectiveness of MACI->MF compared to SPHX->MF falls to £33,206 per QALY. MACI->MF also no longer dominates MACI-MACI, though the cost effectiveness of MACI->MACI remains very poor compared to MACI->MF.

Table 18 Company probabilistic base case results

						Pr(c/e) with WTP	
	Cost	QALYs	Δ Cost	Δ QALY	ICER	@£20k	@£30k
MF->MF	£5,601	15.827	..	..	..	0%	0%
MF->SPHX	£6,827	15.833	..	..	Ext. Dom.	0%	0%
MF->CC	£7,727	15.831	..	..	Dominated	0%	0%
MF->MACI	£7,793	15.828	..	..	Dominated	0%	0%
SPHX->MF	£14,029	18.001	£8,469	2.134	£3,959	18%	20%
SPHX->SPHX	£14,783	17.994	..	..	Ext. Dom.	19%	17%
MACI->MF	£20,392	18.109	£6,348	0.191	£33,206	19%	17%
CC->MF	£20,444	18.155	..	..	Dominated	15%	19%
MACI->MACI	£21,687	18.110	£1,266	0.003	£477k	14%	13%
CC->CC	£21,996	18.148	..	..	Dominated	14%	14%

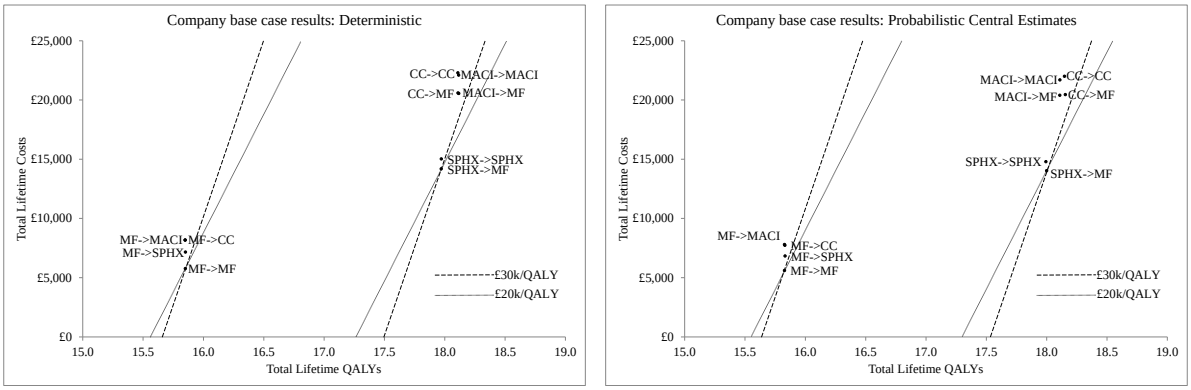
Figure 4 Company base case CEAC



There is an argument for considering the set of sequences with microfracture as the 1st repair separately from the set of sequences with a form of ACI as the 1st repair. This is most simply achieved graphically on the cost effectiveness plane. Since the results suggest two possible sequences as bases, MF->MF and SPHX->MF, total amounts rather net amounts are presented in what follows. This does

not affect the relative position of the sequences, and cost effectiveness lines for £20k/QALY and £30k/QALY can be drawn using the two possible bases as the "origin".

Figure 5 ***Company base case results in the cost effectiveness plane***



There is little probability of a microfracture 1st repair followed by any of the ACIs as 2nd repairs being cost effective at any willingness to pay values.

The likelihood of an ACI as 1st repair followed by itself as 2nd repair being the most cost effective is always less than that of the same ACI as 1st repair followed by microfracture as 2nd repair.

4.2.10 Sensitivity analyses

As outlined at the start of section 5.2.9 the company accepts that there are major errors in its submission. It has provided a revised deterministic base case as summarised at the start of section 5.2.9. It has not provided probabilistic estimates or sensitivity analyses around this. The sensitivity analyses of the original submission are presented below. These are still of some use in showing the structural uncertainty around the model.

Original submission sensitivity analyses

The company presents a range of sensitivity analyses for SPHX->MF compared to MF->MF. This is presented as the effect upon the net monetary benefits (NMB) valued at a willingness to pay of £20k per QALY[6]. This appears to vary a number of inputs by an arbitrary ±20% and concludes that the NMB at £20k per QALY for SPHX->MF compared to MF->MF remains positive throughout. Results are most sensitive to varying the quality of life values for years 5+ that are applied to Spherox and to microfracture.

A number of deterministic scenario analyses are also presented by the company. These broadly preserve the ordering of the sequences and the patterns of dominance and extended dominance. Their

[6] For full details see Table 64 page 144 and Figure 18 page 145 of the company submission.

main effect is to alter the deterministic base case cost effectiveness estimate for SPHX->MF compared to MF->MF and for MACI->MF compared to SPHX->MF.

Table 19 Company scenario analyses

Sequence	SPHX->MF	MACI->MF
Comparator	MF->MF	SPHX->MF
Base case ICER	£3,971	£43,676
SA01: 5 year time horizon	£75,395	£206k
SA02: 15 year time horizon	£8,497	£78,218
SA03: All 1st KR are TKR	£3,971	£43,676
SA04: All 1st KR are PKR	£3,971	£43,676
SA05: QoL NFR = QoL Failure = 0.654	£4,008	£32,838
SA06: QoL Failure = QoL Success = 0.817	£3,991	£43,333
SA07: QoL Failure = QoL Success = 0.746 (midpoint)	£3,982	£43,481
SA08: QoL MF Yr5+ = QoL ACI Yr5+ = 0.817	Ext. Dominated	£62,927*
SA09: QoL prior to 2nd KR = QoL prior to 1st KR = 0.615	£3,971	£43,676
SA10: Spherox implantation same as other ACIs = £1,065	£4,127	£41,405
SA11: Spherox same responder rate as microfracture	£3,936	£43,680
SA12: Spherox same failure rate as microfracture	£4,061	£43,676
* ICER for MACI->MF vs MF->MF due to SPHX->MF being Ext. Dominated		

The main result of interest is that if success with microfracture persists SPHX->MF is extended dominated and the cost effectiveness of MACI->MF compared to MF->MF is poor.

No subgroup analyses are presented.

4.2.11 Model validation and face validity check

Original model face validation

The company model estimates that MF->SPHX results in more QALYs than MF->MACI and MF->CC. 2nd repairs with MACI and ChondroCelect are estimated to have higher probabilities of response than 2nd repairs with Spherox. This raises questions about the face validity of the model and in particular the modelling of 2nd repairs.

The company model also estimates that MACI->MF results in more QALYs than MACI->MACI, and that CC-> MF results in more QALYs than CC->CC. 2nd repairs with MACI and ChondroCelect are estimated to have higher probabilities of response than 2nd repairs with microfracture. This again raises questions about the face validity of the model and in particular the modelling of 2nd repairs.

The AC for TA477 expressed some scepticism that all microfracture repair successes would fail after 5 years. This is the main source of the QALY gain for Spherox over microfracture. The company has addressed this with a scenario analysis which applies the same quality of life value for microfracture repair successes for years 5+ as for year 1-4. This has a large impact upon results. In the opinion of the ERG it is such an important assumption that it warrants full exploration. The company model structure only permits this assumption to be turned on or turned off. As a consequence, the ERG will present a full set of analyses with the assumption that all microfracture repairs fail after 5 years and without it.

ERG revised model face validation

The main validation work that can be conducted is to revise the model inputs to be broadly in line with those of the various reports that underlie the ACI MTA [TA477] and check if the model outputs are broadly in line with those of the ACI MTA reports.

The 1st AG report clinical effectiveness estimates of response for ACI of 83% and microfracture of 62% are aligned with those of the TIG/ACT trial and the company model can be revised to apply these estimates.

The 1st AG report clinical effectiveness estimates for 1st repair successes failing and requiring a 2nd repair are 0.63% for ACI and 1.61% for microfracture. These compare to 0.63% for ACI and 3.44% for microfracture in the company model.

The quality of life values and cost inputs of the 1st AG report and ACI monograph are broadly in line with those of the company model. Only a value in the knee replacement module has to be revised and the ACI cost set to £16,000 to largely align the company model inputs with those of the 1st AG report.

Table 20 Model validation against AG reports ACI MTA

	Company model		1st AG report (table 16)		3rd AG report (table 2)	
	Costs	QALYs	Costs	QALYs	Costs	QALYs
MF->ACI	£7,608	15.966	£6,607	17.028	£6,248	17.135
ACI->ACI	£21,636	18.098	£20,921	18.023	£22,461	17.995
Net	£14,028	2.131	£14,314	0.994	£16,213	0.860
ICER	£6,186		£14,395		£18,844	

The company model cost estimates are reasonable aligned with those of the 1st AG report. The total QALYs for ACI->ACI are also broadly aligned. But the total QALYs for MF->ACI are considerably less and result in the net QALY gain more than doubling to 2.131 QALYs. As a consequence the cost effectiveness estimate of the company model is around half that of the 1st AG report.

The ERG has not managed to identify why there is this discrepancy. A possible source is the different model structure with 1st repair successes only being able to transition to a 2nd repair and not to lose the benefits and move into the NFR health state. But the ERG would anticipate this further reducing the total QALY estimates in both arms, albeit by more in the ACI arm than in the microfracture arm.

During the course of the ACI MTA the AC requested longer term time to event data on loss of success be incorporated into the AG modelling, with the 3rd AG report reflecting this. This also revised the costs of harvesting to £870 and the costs of implantation to £2,396. The company model does not reflect the publicly available time to event data. But the AG incorporation of this data appears to reduce the net QALY gain from ACI over microfracture.

4.3 ERG cross check and critique

4.3.1 Base case results

The company model is constructed in an extremely convoluted manner, with a number of odd constructs and a number of dead ends in terms of inputs and TPMs not feeding through to the actual model. This is in part the reason for table 48 of the submission having limited relevance to the actual model inputs. In section 4.4 the ERG makes extensive changes to the company model and the base case results change markedly.

The ERG has rebuilt the company deterministic model using the same assumptions as the company and gets near complete agreement with the company model results.

Table 21 ERG rebuild vs company model: Company base case

	ERG Rebuild		Company model	
	QALY	Cost	QALY	Cost
MF->MF	15.851	£5,765	15.851	£5,763
MF->SPHX	15.851	£7,157	15.851	£7,156
MF->CC	15.849	£8,170	15.849	£8,168

MF->MACI	15.849	£8,170	15.849	£8,168
SPHX->MF	17.971	£14,184	17.971	£14,182
SPHX->SPHX	17.972	£15,018	17.972	£15,017
MACI->MF	18.117	£20,546	18.117	£20,544
CC->MF	18.110	£20,590	18.110	£20,588
MACI->MACI	18.116	£22,092	18.116	£22,091
CC->CC	18.109	£22,283	18.109	£22,283

4.3.2 Data Inputs: Correspondence between written submission and sources cited

Clinical effectiveness

A variety of clinical inputs are derived from Mistry et al. The following elements cross check:

- The 1.25% 2 yearly ongoing probability of moving from a successful ACI 1st repair to a 2nd repair.
- All the probabilities associated with knee replacement.

There is slight divergence between:

- The 3.44% 2 year probability of moving from a successful microfracture 1st repair to a 2nd repair of the model which implies an annual probability of 1.73%, and the 1.61% estimate Mistry et al derive from Saris et al.[39]

Quality of life

The quality of life values applied by the company for repairs cross check with those of Mistry et al, including the assumptions that:

- quality of life among microfracture 1st repair and 2nd repair successes for years 5+ after the repair declines to 0.654, and
- quality of life among ACI 2nd repair successes after a microfracture 1st repair for year 4 and years 5+ after repair declines to 0.789.

The quality of life values applied by the company for knee replacements do not entirely cross check with those of Mistry et al. In Mistry et al those with no further repair (NFR) had a common quality of life value of 0.691. The company revises these for most of the NFR health states to 0.557. This worsens the cost effectiveness of sequences that result in more knee replacements.

Table 22 Knee replacement quality of life values cross check

	QoL	
Health state	Company	Mistry et al
1st KR	0.615	0.615
Subs KR	0.557	0.557
Success	0.780	0.780
NFR from 1st KR	0.691	0.691
NFR from 1st KR success	0.557	0.691
NFR from 2nd KR	0.557	0.691
NFR from 2nd KR success	0.557	0.691

The ERG will apply the quality of life values of Mistry et al.

Resource use and unit costs

The resource use in terms of outpatient visits and rehabilitation visits cross check with Mistry et al. The unit costs sourced from Mistry et al table 22 cross check when a 3.4% inflation uplift is taken into account

The HRG codes for OP visits and rehabilitation visits cross check with those of Mistry et al.

- The unit cost of £121 for OP paediatric trauma and orthopaedics has been applied, incorrectly. The unit cost of OP trauma and orthopaedics of £110 should be applied.
- The unit cost of rehabilitation cross checks.

The 2012-13 HRG code of HB21C major knee procedure: non-trauma, cat 2, no CC appears to have been superseded in the 2015-16 reference cost HRG codes. The 2015-16 reference cost HRG codes with the closest description to these are HN23A to HN23D for Major Knee Procedures for Non-Trauma, 19 years and over with different CC scores. These are as below for elective inpatients.

Table 23 Major Knee Procedures: 2015-16 Reference costs

HRG	CC Score	FCEs	Mean cost	Mean LoS
For Non-Trauma				
HN23A	4+	330	£5,746	6.0
HN23B	2-3	1025	£4,118	2.7
HN23C	0-1	7318	£3,587	1.4

The company does not use NHS reference costs, but unusually chooses to use the 2016-17 National Prices and Tariff of £5,566. This cross checks with Annex A: HRG code HB21C: Major Knee Procedures for Non-Trauma, Category 2, without CC. It is also broadly in line with the uninflated cost of knee replacement of £5,676 of Mistry et al. The model is not sensitive to the cost of knee replacement.

The cell costs of £16,000 for MACI and ChondroCelect cross check with Mistry et al table 22. However, the TA477 AG report noted that CIC discounts were available to these costs and over the course of the assessment undertook a range of scenario analyses that varied the cell costs to £16,000, £12,000, £8,000 and £6,000.

For the ACI MTA [TA477] OsCell initially reported cell costs of around £4,100 but the AC was concerned that this did not account for overheads. OsCell supplied another costing of £6,000 inclusive of overheads, and £9266 including both procedures.

Many of these costs have been superseded by the FAD of the MTA of ACI [TA477] which preferred:

- Harvesting costs of £870 (HRG HB25F)
- Implantation costs of £2,396 (HRG HB22C)
- OsCell cell costs of £6,000 inclusive of overheads, though the FAD suggests that this may still be an underestimate due to not fully accounting for start-up costs

4.3.3 Data Inputs: Correspondence between written submission and electronic model

Transition probabilities

As already outlined, table 48 of the submission has only limited relevance to the electronic model. The transition probabilities that are applied in the original model are summarised in section 5.2.6 above. These have subsequently been heavily revised by the company as outlined in section 5.3.4 below.

Knee replacement quality of life values

Table 51 suggests a common quality of life value of 0.691 for all NFR subsequent to knee replacement health states. As already outlined above, this is incorrect. This value is only applied for those moving immediately from a 1st KR to NFR. Those moving to NFR from a 1st KR success, immediately from a subsequent KR, and from a subsequent KR success have a quality of life of 0.557 applied. This increases the cost effectiveness of a treatment which avoids knee replacements.

4.3.4 ERG commentary on model structure, assumptions and data inputs

Comparators

The AC of TA477 noted that those failing after a 1st microfracture repair would not receive a 2nd microfracture repair. The FAD of TA477 approved ACI with various restrictions, among them that "*the person has not had previous knee repair surgery*". This suggests that a comparator of only a 1st microfracture repair should be considered, and that ACI subsequent to microfracture should not be considered. This limits the relevant comparators of the company analyses to:

- MF
- SPHX->MF
- MACI->MF
- CC->MF
- SPHX->SPHX
- MACI->MACI
- CC->CC

It can be further argued that the FAD of TA477 may not permit 2nd repairs with ACI, it limiting ACI to patients who have "*not had previous knee repair surgery*". However this should refer only to previous procedures that damage the sub-chondral bone, and the ERG interpretation is that a 1st ACI repair can be followed by a 2nd ACI repair, but the FAD is ambiguous.

Model structure

Successes from a 1st repair cannot lose response and move into the NFR health state. To the age of 55 they can only exit to a 2nd repair. This is a fundamental difference from the model structure of Mistry et al. To put this more clearly into context, if the model is used to explore there only being 1st repairs all the successful repairs remain successes to the age of 55 after which a small proportion each year receive knee replacements. This will overstate the benefits of treatment successes compared to the model structure of Mistry et al.

Probabilities of 1st repair success failing and requiring a 2nd repair

The likelihood of failure and requiring a 2nd repair is based upon the 2 year trials' data and the NMA. These probabilities are applied through the model time horizon.

For the modelling of the MTA of ACI the AC requested that this applied publicly available time to event data. This appears to worsen the cost effectiveness estimates for ACI compared to microfracture.

Modelling microfracture success duration

The AC of TA477 was critical of microfracture failures being modelled by microfracture successes having a lower quality of life applied for years 5+ after the successful repair, and suggested that this might be better handled through the transition probabilities.

In the ERG reduced set of comparators, a microfracture repair is never followed by another repair (in line with NICE guidance, and a decision during the MTA that ruled out second MF). Up to the age of 55 microfracture repair patients cannot exit to another intervention. It is consequently reasonable to apply a reduced quality of life among these patients after the average duration of repair. The model applies this for years 5+ after the repair. The limitation of this is that the model structure does not permit the average duration of microfracture success to be explored, other than assuming that it is indefinite[7]. The company supplies a scenario analyses that applies the indefinite duration of microfracture success assumption, and the ERG will do likewise.

Application of the NMA relative risks of response

The relative risks of response are applied to the 2 year response rate of Spherox. This seems peculiar. The resulting 2 year probabilities or risks imply relative risks that are very different from those of the NMA. When these are based upon the Spherox probability of response of 81% they imply the probabilities of the 2nd to last row of the table below. The ERG will apply these values for its revised base case.

But there may be an argument that the resulting probabilities of response for ChondroCelect and MACI are infeasibly high. This is due to the Spherox trial probability of response for microfracture being much higher than those of the other trials. There may be an argument for applying the relative risks to the mean microfracture rate of the trials of 69.59%. This would imply the response probabilities of the last row of the table below. The ERG will apply these as a sensitivity analysis.

Table 24 Alternative ERG application of the NMA relative risks of response

	SPHX	CC	MACI	MF
Spherox 2yr probability of response Sphx(P_2)		[redacted]		
Relative risk (RR)	[redacted]	1.209	1.223	0.968
Company 1st repair 2yr probability of response P_2 = 1-(1-Sphx(P_2))^{RR}	[redacted]	86.57%	86.88%	79.98%
Relative risks implied by 2yr probabilities of response P_2/Sphx(P_2)	[redacted]	1.069	1.073	0.987
ERG 1st repair 2yr probability of response P_2 = Sphx(P_2)*RR	[redacted]	97.93%	99.06%	78.41%
ERG 1st repair 2yr probability of response P_2 with MF = 69.59%	[redacted]	86.88%	87.89%	69.59%

[7] This is a slight simplification for the company full set of comparators since there is the possibility of exiting to a 2nd repair. But this does not apply to the reduced set of comparators.

In response to a 2nd clarification the company states that: "*The original approach was an incorrect application of the NMA data*". The company supplies an alternative set of estimates and sources, but does not outline the arithmetic of these estimates. The last row of the table below contains the relative risks that appear to be implied by these estimates as calculated by the ERG.

Table 25 Alternative company estimates of response probabilities

	SPHX	CC	MACI	MF
1st repair 2yr probability of response P_2	[redacted]	91.59%	92.28%	78.44%
Source	NMA + OR from trial			NMA
Relative risks implied by 2yr probabilities of response P_2/Sphx(P_2)	1.000	1.152	1.161	0.987

The relative risks implied by the company revised estimates still appear to be different from the central estimates of figure 12 of the company submission and biased in favour of Spherox relative to MACI and ChondroCelect. The stated sources are also peculiar with the trials' odds ratios apparently being applied to the NMA. In the absence of further information (requested 10th October) about the revised company calculations the ERG will only apply these in a sensitivity analysis.

Application of the NMA relative risks of failure

The same considerations around the application of the NMA relative risks of failure as outlined above for the NMA relative risks of response apply. The company has applied these to rates rather than to probabilities. This is relatively minor due to the low probabilities of failure.

The ERG revises the model to apply the NMA relative risks of failure to the failure probability for Spherox as inputted to the NMA.

2 year probabilities of response for 2nd repairs

For 2nd repairs the probability of response is calculated as the square root of the 2 year probability of response. The intention here appears to have been that this should be compounded over 2 model cycles and so after 2 annual cycles result in the 2 year probability of response. But in the model every incident patient that gets a 2nd repair has this 2nd repair probability of response applied only once. This causes the model to overestimate the initial proportion of patients achieving successes and seems

likely to result in some bias in favour of ChondroCelect and MACI and some bias against microfracture.

Table 26 Probabilities of 2nd repair responses

	SPHX	CC	MACI	MF
Spherox 2yr probability of response $Sp(P_2)$	[redacted]			
Relative risk (RR)	1.000	1.209	1.223	0.968
1st repair 2yr probability of response $P_2 = 1-(1-Sp(P_2))^{RR}$	[redacted]	86.57%	86.88%	79.98%
2nd repair probability of response = $P_2^{½}$	[redacted]	93.04%	93.21%	89.43%

In the opinion of the ERG given the model structure the best mean of addressing this is to treat 2nd repairs as 1st repairs; i.e. to apply the 2 year probability of response to the incident patients during a single annual model cycle. While not correct, this is probably more correct than the original model implementation.

Derivation of the probabilities of 2nd successes becoming failures

The derivation of the annual probabilities of 2nd repair successes becoming failures is invalid. As previously outlined in section 4.2.6 above there are peculiar calculations based upon the initial 2 year probabilities of success. This results in typically fewer than half of successes being estimates to remain as such each year.

The company suggests revising this to apply the same probability of moving from a 2nd repair success to NFR as that of the 1st repair success to NFR. But, for example, this means that for SPHX->MF, the 2nd repair of MF has the probability of the 1st repair of Spherox applied to it. This seems peculiar and the ERG will revise this so that the probability a 2nd repair success becoming NFR is equal to the corresponding probability of a 1st repair of the same type, MF in this example.

But there is a more general problem with the company method. The 2 year probability of response or success, P_2, is treated as implying a probability of failure of $(1-P_2)$. For 1st repairs these probabilities are only applied during the first cycle. But for 2nd repairs the probability of failure, or the success going to NFR, is applied not just to the year of repair but every year. While this is correct at year 2, it is not obviously correct to extrapolate an ongoing failure rate using an annualised $(1-P_2)$. In the light of this the annualisation of the 2 year probability for 2nd repairs is retained by the ERG. There is no simple means of correcting this in the company model. But provided that the modelling does not consider MF->MF, an exploration that sets these probabilities to zero does not particularly affect the cost effectiveness estimates.

4.4 Exploratory and sensitivity analyses undertaken by the ERG

For the main analyses, as outlined at the start of section 5.3.4 above in the light of TA477 and its FAD the ERG reduces the list of comparator sequences to:

- MF with no 2nd repair
- SPHX->MF
- MACI->MF
- CC->MF
- SPHX->SPHX
- MACI->MACI
- CC->CC

It can be argued that TA477 does not formally bar Spherox as a 2nd repair after microfracture and that this should be considered. But the ERG thinks it unlikely to apply (the ICERS were higher than usually considered acceptable, though assuming cell cost of £16,000) and considering it within the set of comparator sequences adds relatively little to the analysis.

The company base case assumption that all microfracture repair successes lose all their quality of life gains at 5 years is central to the comparisons with microfracture. The company model structure does not permit this assumption to be relaxed such that the gains are lost gradually after 5 years. The TA477 AC expressed concerns around this assumption. It is sufficiently central for two full sets of analyses to be presented, one that assumes that all microfracture repair successes lose all quality of life gains at 5 years and one that does not[8].

The ERG has revised the company model to:

- Multiply the Spherox 2 year probability of response by the 2 year relative risks of response to derive the comparator 2 year probabilities of response.
- Apply the above 2 year probabilities of response to 2nd repairs, albeit within an annual cycle.
- Multiply the Spherox probability of failure and 2nd repair by the relative risks of failure and 2nd repair to derive the comparator probabilities of failure and 2nd repair.
- Remove the double halving of the 2 year probability of failure and repair for microfracture.
- Revise the probabilities of moving from a 2nd repair success to NFR to be based upon those of 1st repairs.
- Apply the quality of life values of Mistry et al for knee replacement.
- Apply the costs of the FAD of the MTA of ACI [TA477].

[8] This would also seem to require that the quality of life for success from an ACI 2nd repair after a microfracture 1st repair does not deteriorate after 5 years. But this is not considered in the ERG set of possible sequences.

All the clinical inputs derived from the trials and the company NMA have been heavily revised. The ERG has not been in this situation before. As already noted, the model is quite convoluted in its construction with a number of dead ends. It is desirable that the company spend some time checking the ERG model revisions before the 1st AC.

The ERG also undertakes the following sensitivity analyses:

- SA01: Pooling the MF response data across the three trials to yield an estimate of 70% and using the company NMA to provide estimates of 72% for Spherox, 88% for MACI and 87% for ChondroCelect.
- SA02: Applying the company revised estimates of the probability of response.
- SA03: No 2nd repairs.
- SA04: A 2nd MF repair after 1st MF repair being possible.

Given the concerns around the NMA, a head to head comparison of Spherox with microfracture would seem possible. For this the ERG applies the response probabilities of the COWISI trial.

4.4.1 ERG revised results: Microfracture successes lose all gains at 5 years

The ERG revised base case is as below.

Table 27 ERG base case CEAC: MF success lost at year 5

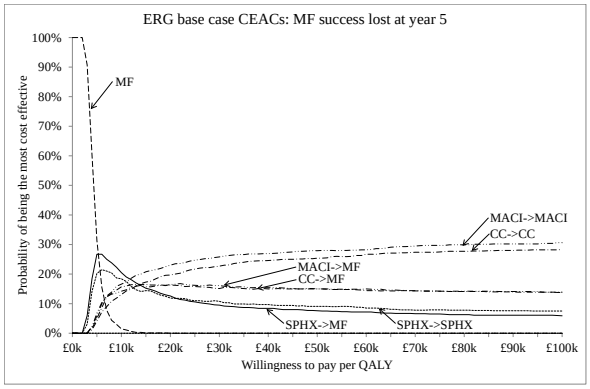
	Costs	QALYs	ICER
MF	£5,043	15.779	..
SPHX->MF	£15,980	17.989	£4,949
SPHX->SPHX	£16,987	18.035	Ext. Dom.
MACI->MF	£22,076	18.437	Ext. Dom.
CC->MF	£22,116	18.410	Dominated
MACI->MACI	£24,011	18.640	£12,336
CC->CC	£24,198	18.629	Dominated

The model suggests that ACI increases costs by roughly the extent of the cell costs. This is much as would be expected given that the harvesting and implantation costs are roughly the same as the costs of microfracture. The model estimates quite large QALY gains and SPHX->MF is estimated to be more costly than MF at £4,949 per QALY.

But MACI is more effective than Spherox, and the cost effectiveness of MACI->MACI relative to SPHX->MF is also good at £12,336 per QALY. At conventional willingness to pay thresholds MACI is estimated to be more cost effective than Spherox.

ChondroCelect is the same price as MACI but slightly less effective. This causes MACI to be estimated to dominate it. But this is better read as MACI and ChondroCelect being of much the same clinical effect and cost effectiveness.

Figure 6 ERG base case CEAC: MF success lost at year 5



The probabilistic modelling suggests that at low willingness to pay values microfracture has the highest probability of being cost effective. Spherox may be the most likely to be cost effective if the willingness to pay lies between £5k and £10k per QALY. At conventional willingness to pay thresholds MACI followed by MACI and ChondroCelect followed by ChondroCelect are more likely to be the most cost effective.

If the VerMACI and ChondroCelect ACIs, being of the same intervention cost and of similar effectiveness, were to be grouped the likelihood of these being the most cost effective would lie somewhat above that of the grouped Spherox likelihood.

None of the ERG scenario analyses change the cost ordering of the various strategies which eases their presentation. In what follows, SA03 does not permit a 2nd repair and as a consequence the label SHPH->SPHX is really just SPHX for this scenario, and likewise for MACI->MACI and CC->CC. Similarly, SA05 permits a 2nd MF repair after a 1st MF repair and so the label MF is really MF->MF for this scenario.

Table 28 ERG scenario analyses: MF success lost at year 5

	Base	SA01	SA02	SA03	SA04

MF	..	..	..	..	..
SPHX->MF	£4,949	£5,554	£5,030	n.a.	£4,791
SPHX->SPHX	Ext. Dom.	Ext. Dom.	Ext. Dom.	£4,360	Ext. Dom.
MACI->MF	Ext. Dom.	£15,310	Ext. Dom.	n.a.	Ext. Dom.
CC->MF	Dominated	Dominated	Dominated	n.a.	Dominated
MACI->MACI	£12,336	£15,177	£18,284	£12,180	£12,336
CC->CC	Dominated	Dominated	Dominated	Dominated	Dominated

SA01 reduces the microfracture response rate to the average across the three main trials, with the response rates for the ACIs being based upon this coupled with the relative risks of the company NMA. This worsens the cost effectiveness of ACI in general compared to microfracture. This in turn causes MACI-MF to no longer be dominated.

SA02 applies the company revised response estimates. This has little effect upon Spherox and microfracture but it worsens the effectiveness of MACI and ChondroCelect. As a consequence, the cost effectiveness of MACI->MACI relative to SPHX->MF worsens to £18,248 per QALY.

SA03 only compares 1st repairs with no 2nd repairs being possible. This slightly improves the cost effectiveness of Spherox relative to microfracture due to SPHX being estimated to result in slightly greater total QALYs than SPHX->SPHX. This would appear to raise some concerns around the modelling of 2nd repairs, but it may rather be a reflection of the modelling of 1st repairs not permitting patients to move from a successful repair into the NFR health state. If there are no 2nd repairs patients remain trapped in the 1st repair success health state.

SA04 permits a 2nd microfracture repair after a 1st microfracture repair. This slightly worsens the cost effectiveness of MF->MF and as a consequence the cost effectiveness of SPHX->MF relative to MF->MF improves slightly.

4.4.2 ERG revised results: Microfracture successes do not lose all gains at 5 years

The ERG revised base case is as below.

Table 29 ERG base case CEAC: MF success not lost at year 5

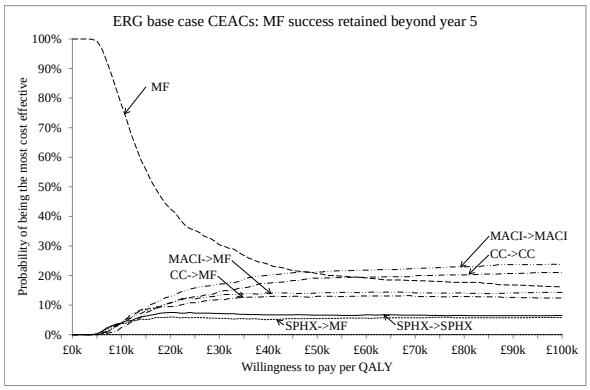
	Costs	QALYs	ICER
MF	£5,043	18.119	..
SPHX->MF	£15,980	18.036	Dominated
SPHX->SPHX	£16,987	18.035	Dominated
MACI->MF	£22,076	18.494	Ext. Dom.

CC->MF	£22,116	18.472	Dominated
MACI->MACI	£24,011	18.640	£36,425
CC->CC	£24,198	18.629	Dominated

The costs are as per the previous analyses. But the total QALYs for microfracture increase quite markedly to be roughly the same as those for Spherox. This is much as would be expected given the small difference in response rates between microfracture and Spherox.

The model formally estimates a higher total QALY for microfracture than for Spherox followed by microfracture, despite microfracture having a very slightly lower response rate than Spherox. In the opinion of the ERG this is due to the model in the absence of 2nd repairs causing 1st repair successes to remain successes indefinitely and none to lose response and move into the NFR health state. This view is given some support by SA03 below which only models 1st repairs for all comparators and causes the total QALYs for Spherox to rise very slightly above those of microfracture.

Figure 7 ERG base case CEAC: MF success not lost at year 5



The probabilistic modelling suggests that at a willingness to pay of £30k per QALY microfracture remains likely to be the most cost effective by a reasonable margin. But this is to compare microfracture with the individual ACI sequences. The probability of microfracture being the most cost effective falls below 50% at willingness to pay values above around £15,000 per QALY. But even at £30k per QALY its probability of being the most cost effective is still about 30%.

If the conventional ACIs, being of the same intervention cost and of similar effectiveness, were to be grouped the absolute separation between these and the grouped Spherox would increase.

Table 30 ERG scenario analyses: MF success not lost at year 5

	Base	SA01	SA02	SA03	SA04
MF	..	..	..	..	..

SPHX->MF	Dominated	Dominated	Dominated	n.a.	Ext. Dom.
SPHX->SPHX	Dominated	Dominated	Dominated	Ext. Dom.	Dominated
MACI->MF	Ext. Dom.	Ext. Dom.	Ext. Dom.	n.a.	Ext. Dom.
CC->MF	Dominated	Dominated	Dominated	n.a.	Dominated
MACI->MACI	£36,425	£51,698	£71,489	£29,349	£20,601
CC->CC	Dominated	Dominated	Dominated	Dominated	Dominated

The pattern of dominance is retained throughout the scenario analyses. Microfracture is estimated to result in slightly higher total QALYs than SPHX->MF and SPHX->SPHX. But the differences are small and as for the base case this is probably more accurately seen as Spherox resulting in similar total QALYs as microfracture but at somewhat greater cost.

SA01 reduces the response probability for MF. Applying the relative risks from the NMA, reduces the response for MACI. This in turn worsens the cost-effectiveness of MACI>MACI relative to MF. SA02 applies the company revised response probabilities for MACI which are worse than those of the company's base case. This again worsens the cost-effectiveness of MACI.MACI relative to MF.

SA03 causes the total gains for Spherox to rise slightly to 18.189 QALYs and so be greater than the 18.119 QALYs of microfracture. But its cost effectiveness compared to microfracture is poor at £150k per QALY.

More surprising for SA03 is the extent of the improvement in the cost effectiveness of MACI relative to MF if there are no 2nd repairs. The net QALY gains of the base case are much reduced if microfracture success is not assumed to be lost at 5 years. The small increase in total QALYs changing from MACI->MACI to MACI has a larger proportionate effect. SA04 also provides similar cause for concern in terms of the modelling of 1st repairs compared to 2nd repairs as outlined in section 5.4.1 above.

The results of SA04 that models a 2nd MF repair after a 1st MF repair are also surprising. The cost effectiveness estimate drops quite markedly. But this appears to be due to the handling of the probabilities of 2nd repair successes moving into NFR. If these probabilities are set to zero the cost effectiveness of MACI->MACI relative to MF->MF still falls, but only to £29,062 per QALY.

4.4.3 ERG revised results: Spherox head to head with microfracture

In the light of the problems with the modelling of 1st repairs compared to 2nd repairs this section considers both MF against SPHX->SPHX and MF against SPHX.

Table 31 MF against SPHX->SPHX

	MF success lost yr5		MF success not lost yr5	
	Costs	QALYs	Costs	QALYs
MF	£5,043	15.779	£5,043	18.119
SPHX->SPHX	£16,987	18.035	£16,987	18.035
net	£11,944	2.256	£11,944	-0.084
ICER	£5,294		Dominated	

Table 32 MF against SPHX

	MF success lost yr5		MF success not lost yr5	
	Costs	QALYs	Costs	QALYs
MF	£5,043	15.779	£5,043	18.119
SPHX	£15,549	18.189	£15,549	18.189
net	£10,506	2.410	£10,506	0.070
ICER	£4,360		£150,506	

The broad conclusion from the above is that if microfracture success persists, there is little clinical difference between microfracture and Spherox. If so, the cost effectiveness of Spherox relative to microfracture is estimated to be poor. But if microfracture success is lost at 5 years the cost effectiveness of Spherox relative to microfracture is estimated to be good.

The results of this section should be viewed with caution and alongside the validation data presented in section 5.2.11 above.

Rehabilitation costs

There is a comment on page 92 of the Co-Don submission that rehabilitation needs are reduced compared to other forms of ACI, as a result of the less invasive surgical procedure.

We doubt that. The rehabilitation after the surgery (in effect, wound healing) is unimportant compared to the rehabilitation associated with maturation of the cartilage in the defect. And MACI can also be done by mini-arthrotomy.

We note comments in various places in the CoDon submission that patients were required to adhere to a strict rehabilitation programme. (And that they should not receive ACI unless they agreed, which is sensible. One of our clinical advisors in the MTA had problems with keen sportsmen returning to sport too soon and damaging the repair.) We also note that in Table 53, the same number of rehabilitation visits is assumed for Spherox and MACI. And on pages 117 and Table 2, rehabilitation is envisaged to take up to a year, as with MACI.

Most rehabilitation is actually patients doing exercises at home at no cost to NHS.

The bottom line is that we do not think there is any significant difference in rehabilitation duration or costs between Spherox and MACI.

4.5 Conclusions on cost effectiveness

The model structure differs from the model structure of the ACI MTA [TA477] in one fundamental aspect. It is not possible for 1st repairs successes to subsequently lose response and patients to move into the no further repair health state. This seems likely to bias the model in favour of the more effective treatment. When coupled with the assumption that all microfracture successes fail at year 5[9] it is also likely to bias the analysis in favour of Spherox compared to microfracture.

The ERG has attempted to revise the company model to have inputs similar to those of the 1st AG report of the ACI MTA. This is imperfect but appears to suggest that the company model estimates roughly double the patient gains compared to the model of the 1st AG report of the ACI MTA. The company model structure may be too optimistic for the comparison with microfracture.

Over the course of the ACI MTA the model inputs evolved. The AC requested that publicly available time to event data be used to estimate the probability of loss of response. The cost effectiveness estimates also appear to worsen over the course of the ACI MTA. Not reflecting the publicly available time to event data may mean that the company model is again too optimistic for the comparison with microfracture.

For 2nd repairs the probability of response is only applied once and as a consequence the company method to derive the estimates appear to be too high. This biases the analysis in favour of MACI and ChondroCelect. The ERG changes this in its revised base case.

For 2nd repairs the possibility of a success losing response and moving into the NFR health state is allowed for in each cycle. The original company estimates for this were incorrect. The revised company estimates are more reasonable. But they are still based upon data that does not particularly relate to this aspect of the model and may be too high. If the model is revised to permit 1st repair successes to lose response with the probability of this being derived from publicly available time to event data, this would probably be the best source for 2nd repairs as well.

The clinical effectiveness of MACI and ChondroCelect is similar. They are assumed to have the same costs. For the probabilistic modelling it may be clearer to consider these as a single treatment.

[9] While this sounds like microfracture successes are failing and so moving into the NFR health state, the model implementation is that they remain successes but have a lower quality of life value applied to them from year 5.

The application of the relative risks of the company NMA is wrong. The resulting company estimates imply relative risks that differ from those of the NMA and that are biased in favour of Spherox. The company has supplied a revised set of response estimates but does not explain their calculation. They still appear to imply relative risks that differ from those of the NMA and that are biased in favour of Spherox.

The ERG revised base case applies clinical effect estimates for both 1st repair and 2nd repair that differ quite markedly from those of the original model and that differ from the company revised response estimates. The ERG has also revised the unit costs to reflect those preferred during the ACI MTA.

In the COWISI trial, [redacted]. Virtually all the modelled gains are a consequence of the assumption that all microfracture successes fail at 5 years. If this is applies the company model estimates that the cost effectiveness of Spherox compared to microfracture is very good. But the company model estimates that MACI yields additional patient gains albeit at a higher cost, and the cost effectiveness of MACI relative to Spherox is also good. If microfracture repairs are as durable as ACI repairs the company model estimates the cost effectiveness of Spherox to be poor.

5 Discussion

5.1 Principal findings

The principal findings in this report are;

- Spherox is clinically effective in the treatment of chondral defects
- However, the phase III COWISI trial has not yet, at 24 months, [redacted]
- The Phase II dosage study shows that the benefit of Spherox implantation varies little by dose, and that the benefit is sustained for up to 4 years
- Around [redacted] of the defects treated in the Phase 2 study were patellar.
- We think the network meta-analysis was inappropriate due to heterogeneity of the included trials
- Taking the above into account, we are doubtful that there is sufficient evidence of benefit to support the economics modelling
- The Appraisal Committee may take a more sympathetic view, so we have critiqued the Co-Don modelling
- The company model structure differs from that of the ACI MTA in that it does not permit 1st repair successes to lose response and move into the no further repair health state. This seems likely to bias the model in favour of the more effective treatments. When coupled with the assumption that all microfracture fails at 5 years it seems likely to bias the analysis in favour of Spherox compared to microfracture.
- ERG validation work suggests that the company model may overestimate the patient gains from ACI relative to microfracture compared to the model of the 1st AG report to the ACI MTA. The modelling of the ACI MTA also evolved to incorporate time to event data. The cost effectiveness estimates appear to have worsened over the course of the ACI MTA. The company model may consequently be too optimistic.
- The company application of relative risks is incorrect and biased in favour of Spherox. The company has supplied revised estimates for the probabilities of response. These still appear to be biased in favour of Spherox.
- The modelled patient gains from Spherox over microfracture are almost entirely due to the assumption that all microfracture successes fail at year 5. These gains cause the company model to estimate Spherox to be cost effective relative to microfracture. But MACI results in greater gains albeit at a higher cost, and the company model estimates that its cost effectiveness relative to Spherox is good.

- If microfracture repairs are as durable as ACI repairs the cost effectiveness of Spherox compared to microfracture is poor.

5.2 Differences in results with microfracture.

The COWISI trial found [redacted] in outcomes between ACI and MF at 24 month. For example, the median changes in KOOS scores from baseline were [redacted] points. This may not be surprising since MF is usually effective in the short term. The 5-year ACTIVE trial results, presented at the 11th Oswestry Cartilage Symposium on 5th October 2017 (Samir Mehta, personal communication), reported no significant differences between ACI and control groups (mainly MF) at 5 years.

However, the trials of ChondroCelect and VeriMACI examined in the recent MTA of ACI, did show some differences at 2 years. In the SUMMIT trial, the 24 month results included;

- Responders 87.5% with MACI, 68.1% with MF
- KOOS subscales all statistically significantly better with ACI
- No failure with MACI, two with MF
- Cincinnati scores 1.05 points better with MACI (p =0.002)
- IKDC 5.9 point better with MACI (p = 0.069)
- But no difference in EQ5D

In the TIG/ACT trial at 24 months, KOOS scores had improved by about 20 points after ACI and by about 13 points with MF, with no overlap of 95% Cis. There were two failures with ACI and 8 with MF.

The COWISI, TIG/ACT and SUMMIT trials differed in the characteristics of participants as reported earlier. For example, a possible explanation for the poorer results of MF in SUMMIT compared to COWISI is the defect sizes, with [redacted] defects in COWISI (mean [redacted]) than in SUMMIT (mean 4.9cm^2, and all over 3cm^2).

The ERG view is that the benefits of ACI compared to MF are seen mainly in later years. Longer-term data from the ACTIVE trial are not yet published. Evidence from observational studies was reported in the assessment report for the MTA, but in brief;

- Solheim and colleagues [59] reported results 10-14 years after microfracture in a prospective cohort of 110 patients. 46% had a poor outcome, defined as needing knee replacement or a Lysholm score under 64. Symptom scores did improve from baseline but few had normal knee function

- Many people with chondral defects are sportsmen or women and return to sport is a useful outcome. Two good quality systematic reviews by Campbell and colleagues[60] and Krych and colleagues[61] reported that proportions returning to sport were higher with ACI than MF – 84% versus 75% (Campbell) and 82% versus 58% (Krych) . In professional athletes, clinical outcome scores were similar at 2 years follow-up but were significantly (p = 0.005) better in the ACI group at 7.4 years, because they were stable in the ACI group but declined over time in the MF group.
- A systematic review by DiBartola and colleagues[62] reported poorer histological outcomes after microfracture compared to ACI. However, there were only six studies of MF compared to 30 of ACI.
- A very large follow-up study by Layton et al[63] of over 3000 patients in routine care who had MF, reported failure rates (defined as requiring further surgery) of 9% within one year, 18% by 3 years, and 32% by 5 years. Others did not have further surgery, but required powerful analgesics.
- A recent study by Volz and colleagues[64] reported that most of the benefits of MF were lost by 5 years.

Figure 8 Modified Cincinatti Score (from Volz et al 2017)

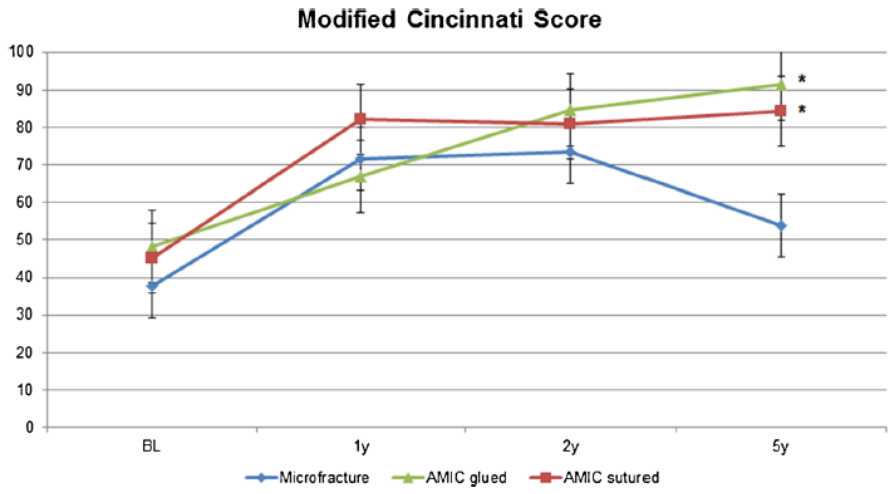


Fig. 3 Modified Cincinnati score. Number of patients available at five years follow-up (n): Microfracture (9), AMIC glued (14, p = 0.002), and AMIC sutured (16, p = 0.01); * significance versus Microfracture at five years

Conversely the 15-year results from the Knutsen trial[45] reported that long-term results with MF were as good as with ACI, though only 40 patients were randomised to each arm.

The assessment group in the MTA took a "middle view" on duration of benefit of MF, being more optimistic than the SUMMIT and TIG/ACT trials, less optimistic than the Knutsen 2016 study, and close to the Volz et al 2017 study.

5.3 Is Spherox more effective in patellar lesions than other forms of ACI?

ACI has been regarded as less successful for patellar lesions than condylar ones though results have been improving.[65]

The patello-femoral joint has features that would potentially make good results more difficult to achieve, including a less congruous joint surface which is made even more difficult in the (common) setting of a mis-shapen trochlea or patella (trochlear dysplasia or patellar dysplasia). The joint also undergoes high contact loads and shearing forces, explaining why the cartilage in a healthy knee is thickest under the patella. It is possible that ACI with caps (ACI-C) or matrices (MACI) more be more likely to be sheared off than spheroids, and this may be a plausible explanation for a difference in results in this region.

We cannot compare results in patellar and condylar lesions from the available trials. The SUMMIT trial recruited mainly medial femoral condylar defects, with small numbers of lateral femoral condylar and trochlear defects. In TIG/ACT all recruits has femoral condylar lesions. In ACTIVE, only 12% had patellar or trochlear problems. In the COWISI trial primary defects were all condylar, but there is no comparison of results by condyle.

The data on [redacted] comes from the phase II dosages trial. The CSR mentions, but does not include, some tables of results by site. The text includes KOOS results to visit 8, which are as good for patellar as for condylar defects. [redacted] of recruits in the Spherox Phase II dosage trial had patellar lesions. The Co-Don submission does not provide separate data for the two condyles.
Nawaz et al[66] reported that best results came from lateral femoral condylar defects, with little difference in outcomes after treatment of MFC and patellar defects.
In conclusion, Spherox may be better in patellar defects than older forms of ACI, but we do not currently have clinical effectiveness evidence that would support any cost-effectiveness comparison.

5.4 Extrapolation from older forms of ACI

In our assessment report for the ACI MTA, we reviewed long-term results of previous generations of ACI. We assumed that the long-term results of third generation ACI would be at least as good as first generation ACI-P. All the first three generation were based on implanting chondrocytes which then produced cartilage in vivo. Can we extrapolate from cell implants to spheroid implants? The question here is whether the spheroid cartilage integrates as well with the cartilage surrounding the defect, as the cartilage produced after MACI. There is evidence from basic science studies that provides reassurance that this is the case, so it appears that we do not need to worry about possible weaknesses around the "join".

5.5 Could a pragmatic case be made for Spherox?

To recap.

1. We know that Spherox works, in the sense that it improves patient symptoms as reflected in scores such as KOOS. This has been shown in the two trials and in the before and after case series.
2. The benefit is sustained for at least 4 years as in the Phase 2 trial of different doses, in patients with large defects.
3. The evidence from trials and case series suggests that there are no serious safety concerns.
4. The main problem is comparative effectiveness. [redacted]. This is perhaps not surprising since MF works in the short-term, and the smaller the defect, the more competitive MF is with ACI. The results from the ACTIVE trial, first released on 5th October 2017 at the UK meeting of the International Cartilage Repair Society (ICRS) and *Arthritis Research UK Tissue Engineering Centre* (ARUK TEC) in the RJAH Orthopaedic Hospital in Oswestry, show no advantage over MF at 5 years.
5. The assessment group report for the ACI MTA concluded that in the longer-term, the benefits of MF were not sustained, but that the benefits of ACI were, albeit varying amongst patients (hence the restricted approval). The Appraisal Committee had concerns about the quality of the evidence base for comparing long-term outcomes of ACI and MF, but did recommend that ACI be used.

Given the above, it could be argued that the lack of evidence of cost-effectiveness of Spherox is due to lack of long-term results, and that with longer follow-up, it would achieve acceptable ICERs.

5.6 OsCells

The Co-Don submission does not mention OsCells or the ACTIVE trial, which is fair enough given that the trial has not yet been published, and that data used in the recent MTA was academic in

confidence and not available to Co-Don. One issue is whether OsCells should have been included as a comparator. The ERG view is that OsCells are not a comparator because they are not available outwith the RJAH Orthopaedic Hospital. (And in the ACTIVE trial, most cells were provided by Genzyme.) The data on OsCells could be used to illustrate the potential for production of cells in other NHS facilities but those would take time to set up because of the regulatory burden.

5.7 Age limits

The upper age limit in COWISI was 50. In our last report on ACI for NICE, we noted that Filardo and colleagues have suggested that the consensus against ACI in older patients should be challenged.[67] Filardo and colleagues analysed results in their series of 157 patients treated with MACI, after excluding any with OA (defined as Kellgren-Lawrence grades 3-4). They divided the patients into those aged under 40, mean age 26, and those over 40, mean age 46. After adjustment for other prognostic variables, Filardo and colleagues concluded that although results in the under 40s were better, the over-40s also benefitted from ACI. When function scores were compared against people in each age group with healthy knees, there was no difference in relative benefits. This is in contrast to comparing functional results in younger and older ACI recipients. Failure rates at 10 years were similar; 11% for under-40s at ACI and 14% for over-40s. Filardo and colleagues therefore argue that age alone should not be a contra-indication to MACI. They note that some previous studies may have included subjects who were not just older, but had osteoarthritis (OA). Secondly, older people receiving ACI may be less active and so put less strain on the repair.

This appraisal specifies use of ACI in adults, 18 years and over. As noted in the recent MTA, there is some evidence of benefit from older forms of ACI in teenagers, and we have noted that the studies of Spherox in people aged under 18.

5.8 Research needs

The COWISI trial. The most important research need is for longer-term follow-up of the COWISI trial, but this is planned. We note that the ICRS has set up a registry for long-term follow-up of ACI and other knee procedures, and if Spherox is approved (now or later), it could be under condition that patients are registered with ICRS so that long-term data will accrue.
The aims of the ICRS registry[68] are:

Our mission is to create the best source of unbiased outcomes data for treatments of painful articular cartilage lesions in the world, which is paramount for improvement of existing and discovery of new cartilage repair strategies, ultimately beneficial for millions of patients around the world.

The ICRS Registry is a mechanism of allowing you and your doctor to track your individual progress following diagnosis and/or treatment of your knee problem. It is suitable for anyone with cartilage damage, whether or not the cartilage damage itself is treated. The response of patients to cartilage damage and treatments can be variable, treatments can also be forefront of medical advances, many are expensive. It is vital to you and your doctor that your progress is monitored. With your permission, the ICRS Registry makes your data anonymous so you cannot be personally identified, and pool together large numbers of patients results so that doctors around the world have the most accurate picture of which techniques are working best in which patients.

This helps patients of the future with similar injuries or cartilage problems, and rapidly identifies treatments that are showing great benefit, those that may not be performing as well as hoped, and also what happens naturally if nothing is done.

The cost-effectiveness of ACI is driven by the duration of repair success for MF, Spherox, VerMACI and ChondroCelect. The ACTIVE trial will provide 10-year data on outcomes after MF and ACI in a few years. Any economic modelling of Spherox based on the CTIVE results would probably have to assume the same duration of repair success for Spherox as for the ACI in ACTIVE.

Defects smaller than 2cm^2.

The British Knee Surgeons consensus considered that interventions such as MF and mosaicplasty should be considered in defects < 2cm^2, stating

"In the absence of comparative trials in small lesions showing superiority of cell therapy, the cost of cell therapy would need special circumstances to justify use."

The SUMMIT trial included only people with defects of 3 cm^2 or greater. The TIG/ACT trial included defects in the range 1-5 cm^2 with mean area 2.6 cm^2 in the ACI arm, but did not give a breakdown of results by defect size. The COWISI trial includes defects between1-2cm^2 but numbers are small.

There is therefore a case for a trial of ACI versus microfracture in small lesions, with follow-up for at least 5 years, and with a cost-effectiveness analysis.

Can results of ACI be improved?

Another issue is whether results of cartilage repair can be improved. In the third assessment report for the recent MTA of ACI, we noted that return to sporting activity was a useful indicator of success. Campbell and colleagues[60] provide a high quality systematic review (admittedly of mostly low-level studies with only one RCT) of return to sport by both amateur and professional athletes. The proportion returning was higher with ACI than MF – 84% versus 75% ($p<0.01$). In professional athletes, clinical outcome scores were similar at 2 years follow-up but were significantly ($p = 0.005$)

better in the ACI group at 7.4 years, because they were stable in the ACI group but declined over time in the MF group. However, return was much faster after MF (return to athletics by 3-6 months) than after ACI (10 to 18 months).

In another good quality review, Krych and colleagues[61, 69] came to similar conclusions, probably because they used most of the studies used by Campbell et al, though they added as many more. Campbell et al included 20 studies whereas Krych et al included 44. The Campbell review was rather more focused on high level athletes including professionals, where the Krych review was mainly in recreational sports people, and for more recent years (1998-2016). Krych et al concluded that 82% returned to sport at some level after ACI compared to 58% after MF.

However return to sport may not be at the level reached before injury. In a good quality review, Schmitt and colleagues[70] reviewed a number of indicators of performance, including muscle (mainly quadriceps) strength and performance achieved, after cartilage repair procedures, both MF and ACI. They found that significant quadriceps strength deficits and functional shortfalls were common 5-7 years after repair procedures. This does not necessarily mean that the repairs were the problem – it could be that the injury and resulting inactivity were the reasons. However they conclude that research is needed into why previous function was not restored, and into different rehabilitation regimens. They do note that one possible reason is impatience, with some sports-people starting weight-bearing and then activity too early.

Can success be predicted before implantation?

The chondrocyte cells from each patient are cultured in the laboratory and encouraged to grow spheroids of cartilage, which are then implanted. Some patients have better results than others. If we could predict failure before implantation, then the implantation cost could be avoided and the spheroids discarded.

Co-Don have done work on this, looking at biomarkers for chondrogenesis in the spheroids and the culture serum. Some of this has been promising but inconclusive. Glycosaminoglycans (GAGs) are one component of the extracellular matrix in hyaline cartilage, and contribute to the shock-absorbing function of cartilage. Bartz et al[71] measured the GAG content of spheroids and of the surrounding culture medium, and found variations, in cultures from different donors, in the proportions of GAG retained in the spheroids or released into the culture serum. A low bound to retained ratio was associated with poorer regeneration after implantation, but showed a trend rather than being sufficiently predictive of failure.

However another component of hyaline cartilage, aggrecan (which is the main proteoglycan in extracellular matrix of articular cartilage – Fox et al[72]) did show a stronger correlation with successful regeneration. A high level of aggrecan in spheroids before implantation was associated with a better repair.

Measurement of biomarkers may have potential for development as a method for triage of spheroids before implantation, which might improve cost-effectiveness.

Source of chondrocytes

In the assessment report for the recent MTA of ACI, we noted the work of Mizuno and colleagues [73], using chondrocytes from the ear, so far only in dogs. We also note the work of Mumme and colleagues[74] in humans, using nasal chondrocytes to produce cartilage grafts for ACI. At the 11th Oswestry Cartilage Symposium in October 2017, Ivkovic from Zagreb (personal communication) presented further work on "nose to knee" ACI.

This is another example of the problems in the evaluation of evolving technologies such as ACI. Lilford et al[75] outlined the problems;

"When should researchers start a randomised controlled trial in a clinical area where there is rapid technological change? Start too early and the resultant comparisons may seem likely to turn out to be irrelevant, but start too late and the chance of collecting much good quality data will have been lost, perhaps forever if clinical opinion has "gelled" despite the absence of randomised controlled trial data. The problem is compounded by the consid-erable time it takes to design, commission, and establish a full scale clinical trial."

They concluded that there was a need for "tracker trials" that allowed for evolution of the technology under study, without prefixed sample size or duration, and with interim analyses. However getting such trials funded may be difficult.

New forms of microfracture

Research into the reasons for differing results of MF from past studies may not be a high priority, since new methods of microfracture are being trialed, such as AMIC (autologous matrix-induced chondrogenesis). Volz et al[64] compared microfracture alone (13), or MF with a collagen cap (ChondroGide) either glued (17) or sutured (17) in place, in 39 patients. Mean defect size was 3.6cm^2, range 2.1 to 6.6cm^2, quite large for MF. In symptoms and function, all groups improved by 2 years, but improvement was sustained at 5 years in the capped group, but not in the standard MF group whose Cincinnati scores had declined by 5 years with over half the benefit at 2 years lost. Defect filling assessed by MRI at 5 years showed better filling in the capped group.

A trial by Shive and colleagues[76] also reported 5 years results of capped MF, using the BST-Cargel scaffold, reported improved MRI filling compared to MF alone, but there was no difference in symptoms: Western Ontario and McMaster Universities Osteoarthritis (WOMAC) or Short-Form 36 (SF-36).

Given the high cost of ACI, further research into enhanced MF may be worthwhile.

5.9 Conclusions

There seems to be no doubt that Spherox implantation is beneficial in chondral defects, but its comparative efficacy is as yet uncertain.

6 References

1. Indelicato PA, Bittar ES. A perspective of lesions associated with ACL insufficiency of the knee. A review of 100 cases. *Clin Orthop Relat Res* 1985:77-80.

2. Lewandrowski KU, Muller J, Schollmeier G. Concomitant meniscal and articular cartilage lesions in the femorotibial joint. *Am J Sports Med* 1997;**25**:486-94.

3. Mezhov V, Ciccutini FM, Hanna FS, Brennan SL, Wang YY, Urquhart DM, *et al.* Does obesity affect knee cartilage? A systematic review of magnetic resonance imaging data. *Obes Rev* 2014;**15**:143-57. http://dx.doi.org/10.1111/obr.12110

4. Urquhart DM, Tobing JF, Hanna FS, Berry P, Wluka AE, Ding C, *et al.* What is the effect of physical activity on the knee joint? A systematic review. *Med Sci Sports Exerc* 2011;**43**:432-42. http://dx.doi.org/10.1249/MSS.0b013e3181ef5bf8

5. Aroen A, Loken S, Heir S, Alvik E, Ekeland A, Granlund OG, *et al.* Articular cartilage lesions in 993 consecutive knee arthroscopies. *Am J Sports Med* 2004;**32**:211-5.

6. Felson DT. The sources of pain in knee osteoarthritis. *Curr Opin Rheumatol* 2005;**17**:624-8.

7. Heir S, Nerhus TK, Rotterud JH, Loken S, Ekeland A, Engebretsen L, *et al.* Focal cartilage defects in the knee impair quality of life as much as severe osteoarthritis: a comparison of knee injury and osteoarthritis outcome score in 4 patient categories scheduled for knee surgery. *Am J Sports Med* 2010;**38**:231-7. http://dx.doi.org/10.1177/0363546509352157

8. International Cartilage Repair Society. *ICRS SCORE/GRADE*. 2000. URL: https://cartilage.org/society/publications/icrs-score/ (Accessed 25/07/2016).

9. Messner K, Gillquist J. Cartilage repair. A critical review. *Acta Orthop Scand* 1996;**67**:523-9.

10. Prakash D, Learmonth D. Natural progression of osteo-chondral defect in the femoral condyle. *Knee* 2002;**9**:7-10.

11. Shelbourne KD, Jari S, Gray T. Outcome of untreated traumatic articular cartilage defects of the knee: a natural history study. *J Bone Joint Surg Am* 2003;**85-A Suppl 2**:8-16.

12. Biant LC, McNicholas MJ, Sprowson AP, Spalding T. The surgical management of symptomatic articular cartilage defects of the knee: Consensus statements from United Kingdom knee surgeons. *Knee* 2015;**22**:446-9. http://dx.doi.org/10.1016/j.knee.2015.06.001

13. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, *et al.* A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;**347**:81-8. http://dx.doi.org/10.1056/NEJMoa013259

14. Hubbard MJ. Articular debridement versus washout for degeneration of the medial femoral condyle. A five-year study. *J Bone Joint Surg Br* 1996;**78**:217-9.

15. National Institute for Health and Care Excellence. *Arthroscopic knee washout, with or without debridement, for the treatment of osteoarthritis: NICE interventional procedures guidance [IPG230]*. 2007. URL: http://www.nice.org.uk/guidance/IPG230 (Accessed 19/08/2017).

16. Brittberg M, Lindahl A, Nilsson A, Ohlsson C, Isaksson O, Peterson L. Treatment of deep cartilage defects in the knee with autologous chondrocyte transplantation. *N Engl J Med* 1994;**331**:889-95. http://dx.doi.org/10.1056/nejm199410063311401

17. Brittberg M. Cell carriers as the next generation of cell therapy for cartilage repair: a review of the matrix-induced autologous chondrocyte implantation procedure. *American Journal of Sports Medicine* 2010;**38**:1259-71.

18. Hindle P, Hall AC, Biant LC. Viability of chondrocytes seeded onto a collagen I/III membrane for matrix-induced autologous chondrocyte implantation. *J Orthop Res* 2014;**32**:1495-502. http://dx.doi.org/10.1002/jor.22701

19. Keele University. *ACTIVE trial web site*. 2011. URL: http://www.active-trial.org.uk/ (Accessed 25/07/2016).

20. European Medicines Agency. *ChondroCelect*. 2014. URL: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/000878/human_med_000698.jsp&mid=WC0b01ac058001d124 (Accessed 25/07/2016).

21. European Medicines Agency. *MACI: matrix-applied characterised autologous cultured chondrocytes*. 2014. URL: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002522/human_med_001660.jsp&mid=WC0b01ac058001d124 (Accessed 25/07/2016).

22. Vanlauwe J, Saris DB, Victor J, Almqvist KF, Bellemans J, Luyten FP, *et al.* Five-year outcome of characterized chondrocyte implantation versus microfracture for symptomatic cartilage defects of the knee: early treatment matters. *American Journal of Sports Medicine* 2011;**39**:2566-74.

23. Mithoefer K, Peterson L, Saris DBF, Mandelbaum BR. Evolution and Current Role of Autologous Chondrocyte Implantation for Treatment of Articular Cartilage Defects in the Football (Soccer) Player. *Cartilage* 2012;**3**:31s-6s.

24. Harris JD, Siston RA, Pan XL, Flanigan DC. Autologous Chondrocyte Implantation A Systematic Review. *Journal of Bone and Joint Surgery-American Volume* 2010;**92A**:2220-33. http://dx.doi.org/10.2106/jbjs.j.00049

25. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;**16**:494-502.

26. Kijowski R, Blankenbaker D, Stanton P, Fine J, De Smet A. Arthroscopic validation of radiographic grading scales of osteoarthritis of the tibiofemoral joint. *AJR Am J Roentgenol* 2006;**187**:794-9. http://dx.doi.org/10.2214/ajr.05.1123

27. Link TM, Steinbach LS, Ghosh S, Ries M, Lu Y, Lane N, *et al.* Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. *Radiology* 2003;**226**:373-81. http://dx.doi.org/10.1148/radiol.2262012190

28. Sheehy L, Culham E, McLean L, Niu J, Lynch J, Segal NA, *et al.* Validity and sensitivity to change of three scales for the radiographic assessment of knee osteoarthritis using images from the Multicenter Osteoarthritis Study (MOST). *Osteoarthritis Cartilage* 2015;**23**:1491-8. http://dx.doi.org/10.1016/j.joca.2015.05.003

29. Szebenyi B, Hollander AP, Dieppe P, Quilty B, Duddy J, Clarke S, *et al.* Associations between pain, function, and radiographic features in osteoarthritis of the knee. *Arthritis Rheum* 2006;**54**:230-5. http://dx.doi.org/10.1002/art.21534

30. Williams A, Sharma L, McKenzie CA, Prasad PV, Burstein D. Delayed gadolinium-enhanced magnetic resonance imaging of cartilage in knee osteoarthritis: findings at different radiographic stages of disease and relationship to malalignment. *Arthritis Rheum* 2005;**52**:3528-35. http://dx.doi.org/10.1002/art.21388

31. Wright RW. Osteoarthritis Classification Scales: Interobserver Reliability and Arthroscopic Correlation. *J Bone Joint Surg Am* 2014;**96**:1145-51. http://dx.doi.org/10.2106/jbjs.m.00929

32. Schiphof D, Boers M, Bierma-Zeinstra SM. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. *Ann Rheum Dis* 2008;**67**:1034-6. http://dx.doi.org/10.1136/ard.2007.079020

33. Felson DT, Niu J, Guermazi A, Sack B, Aliabadi P. Defining radiographic incidence and progression of knee osteoarthritis: suggested modifications of the Kellgren and Lawrence scale. *Ann Rheum Dis* 2011;**70**:1884-6. http://dx.doi.org/10.1136/ard.2011.155119

34. European Medicines Agency. *Summary of Product Characteristics: Spherox 10-70 spheroids/cm2 implantation suspension*. 2017. URL: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002736/WC500231919.pdf (Accessed 15/08/2017).

35. European Medicines Agency. *Spherox*. 2017. URL: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002736/human_med_002138.jsp&mid=WC0b01ac058001d124 (Accessed).

36. Biant LC, Simons M, Gillespie T, McNicholas MJ. Cell Viability in Arthroscopic Versus Open Autologous Chondrocyte Implantation. *Am J Sports Med* 2017;**45**:77-81. http://dx.doi.org/10.1177/0363546516664338

37. Steadman JR, Briggs KK, Rodrigo JJ, Kocher MS, Gill TJ, Rodkey WG. Outcomes of microfracture for traumatic chondral defects of the knee: average 11-year follow-up. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2003;**19**:477-84.

38. Saris DB, Vanlauwe J, Victor J, Almqvist KF, Verdonk R, Bellemans J, *et al.* Treatment of symptomatic cartilage defects of the knee: characterized chondrocyte implantation results in better

clinical outcome at 36 months in a randomized trial compared to microfracture. *Am J Sports Med* 2009;**37 Suppl 1**:10s-9s. http://dx.doi.org/10.1177/0363546509350694

39. Saris D, Price A, Widuchowski W, Bertrand-Marchand M, Caron J, Drogset JO, *et al.* Matrix-Applied Characterized Autologous Cultured Chondrocytes Versus Microfracture: Two-Year Follow-up of a Prospective Randomized Trial. *Am J Sports Med* 2014;**42**:1384-94. http://dx.doi.org/10.1177/0363546514528093

40. Becher C, Laute V, Fickert S, Zinser W, Niemeyer P, John T, *et al.* Safety of three different product doses in autologous chondrocyte implantation: results of a prospective, randomised, controlled trial. *J Orthop Surg Res* 2017;**12**:71. http://dx.doi.org/10.1186/s13018-017-0570-7

41. Niemeyer P, Laute V, John T, Becher C, Diehl P, Kolombe T, *et al.* The Effect of Cell Dose on the Early Magnetic Resonance Morphological Outcomes of Autologous Cell Implantation for Articular Cartilage Defects in the Knee: A Randomized Clinical Trial. *Am J Sports Med* 2016;**44**:2005-14. http://dx.doi.org/10.1177/0363546516646092

42. Mistry H, Connock M, Pink J, Shyangdan D, Clar C, Royle P, *et al.* Autologous chondrocyte implantation in the knee: systematic review and economic evaluation. *Health Technol Assess* 2017;**21**:1-294. http://dx.doi.org/10.3310/hta21060

43. Clave A, Potel JF, Servien E, Neyret P, Dubrana F, Stindel E. Third-generation autologous chondrocyte implantation versus mosaicplasty for knee cartilage injury: 2-year randomized trial. *J Orthop Res* 2016;**34**:658-65. http://dx.doi.org/10.1002/jor.23152

44. Jones HR, Crawford DC. An autologous tissue implant, neocart, for treatment of hyaline cartilage injury in the knee. *Operative Techniques in Orthopaedics* 2014;**24**:264-70.

45. Knutsen G, Drogset JO, Engebretsen L, Grontvedt T, Ludvigsen TC, Loken S, *et al.* A Randomized Multicenter Trial Comparing Autologous Chondrocyte Implantation with Microfracture: Long-Term Follow-up at 14 to 15 Years. *J Bone Joint Surg Am* 2016;**98**:1332-9. http://dx.doi.org/10.2106/JBJS.15.01208

46. Knutsen G, Engebretsen L, Ludvigsen TC, Drogset JO, Grontvedt T, Solheim E, *et al.* Autologous chondrocyte implantation compared with microfracture in the knee. A randomized trial. *J Bone Joint Surg Am* 2004;**86-A**:455-64.

47. Libera J, Ruhnau K, Baum P, Lüthi U, Schreyer T, Meyer U, *et al.* Cartilage engineering. In: Meyer U, Meyer, Th., Handschel, J., Wiesmann, H.P. , editor. *Fundamentals of Tissue Engineering and Regenerative Medicine*: Springer; 2009:233-42.

48. Fickert S, Gerwien P, Helmert B, Schattenberg T, Weckbach S, Kaszkin-Bettag M, *et al.* One-Year Clinical and Radiological Results of a Prospective, Investigator-Initiated Trial Examining a Novel, Purely Autologous 3-Dimensional Autologous Chondrocyte Transplantation Product in the Knee. *Cartilage* 2012;**3**:27-42. http://dx.doi.org/10.1177/1947603511417616

49. Siebold R, Karidakis G, Feil S, Fernandez F. Second-look assessment after all-arthroscopic autologous chondrocyte implantation with spheroides at the knee joint. *Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA* 2016;**24**:1678-85.

50. Maiotti M, Massoni C, Bondì R, Siodla V. Transplantation of Tissue-engineered Cartilage for Repair of Osteochondral Defects in the Knee (SS-27). *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2012;**28**:e15-e6.

51. Roessing S, Schreyer T, Baum P, Thermann H, Paessler H. The full arthroscopic technique of autologous chondrocyte transplantation using chondrospheres-safety study with 2 years follow up. *Cartilage* 2010;**1 (2 Supplement 1)**:55S.

52. Schreyer T. Comparison in clinical outcome of frst generation ACT with ACT of third generation with spheroids. *Cartilage* 2010;**1 (2 Supplement 1)**:58S.

53. National Institute for Health and Care Excellence. *Autologous chondrocyte implantation for treating symptomatic articular cartilage defects of the knee: Technology appraisal guidance [TA477]*. 2017. URL: https://www.nice.org.uk/guidance/ta477 (Accessed 15/10/2017).

54. Gerlier L, Lamotte M, Wille M, Kreuz PC, Vanlauwe J, Dubois D, *et al.* The Cost Utility of Autologous Chondrocytes Implantation Using ChondroCelect (R) in Symptomatic Knee Cartilage Lesions in Belgium. *Pharmacoeconomics* 2010;**28**:1129-46. http://dx.doi.org/10.2165/11584920-000000000-00000

55. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;**21**:271-92.

56. Dong H, Buxton M. Early assessment of the likely cost-effectiveness of a new technology: A Markov model with probabilistic sensitivity analysis of computer-assisted total knee replacement. *Int J Technol Assess Health Care* 2006;**22**:191-202. http://dx.doi.org/10.1017/S0266462306051014

57. Jansson KA, Granath F. Health-related quality of life (EQ-5D) before and after orthopedic surgery. *Acta Orthop* 2011;**82**:82-9. http://dx.doi.org/10.3109/17453674.2010.548026

58. Clar C, Cummins E, McIntyre L, Thomas S, Lamb J, Bain L, *et al.* Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: Systematic review and economic evaluation. *Health Technology Assessment* 2005;**9**:iii-48.

59. Solheim E, Hegna J, Inderhaug E, Oyen J, Harlem T, Strand T. Results at 10-14 years after microfracture treatment of articular cartilage defects in the knee. *Knee Surg Sports Traumatol Arthrosc* 2016;**24**:1587-93. http://dx.doi.org/10.1007/s00167-014-3443-1

60. Campbell AB, Pineda M, Harris JD, Flanigan DC. Return to Sport After Articular Cartilage Repair in Athletes' Knees: A Systematic Review. *Arthroscopy* 2016;**32**:651-68.e1.

61. Krych AJ, Pareek A, King AH, Johnson NR, Stuart MJ, Williams RJ, 3rd. Return to sport after the surgical management of articular cartilage lesions in the knee: a meta-analysis. *Knee Surgery, Sports Traumatology, Arthroscopy* 2016;**25**:3186-96.

62. DiBartola AC, Everhart JS, Magnussen RA, Carey JL, Brophy RH, Schmitt LC, *et al.* Correlation between histological outcome and surgical cartilage repair technique in the knee: A meta-analysis. *Knee* 2016;**23**:344-9.

63. Layton A, Arnold RJ, Graham J, Frasco MA, Cote E, Lynch NM. *Long-term failure rates associated with knee microfracture surgery [poster]*. 2015. URL: http://quorumconsulting.com/docs/MicrofractureFailureRates_ISPOR_May2015.pdf (Accessed 22/05/2017).

64. Volz M, Schaumburger J, Frick H, Grifka J, Anders S. A randomized controlled trial demonstrating sustained benefit of Autologous Matrix-Induced Chondrogenesis over microfracture at five years. *International Orthopaedics* 2017;**41**:797-804. http://dx.doi.org/10.1007/s00264-016-3391-0

65. Gomoll AH, Gillogly SD, Cole BJ, Farr J, Arnold R, Hussey K, *et al.* Autologous chondrocyte implantation in the patella: a multicenter experience. *American Journal of Sports Medicine* 2014;**42**:1074-81.

66. Nawaz SZ, Bentley G, Briggs TWR, Carrington RWJ, Skinner JA, Gallagher KR, *et al.* Autologous chondrocyte implantation in the knee: Mid-term to long-term results. *Journal of Bone and Joint Surgery - American Volume* 2014;**96**:824-30.

67. Filardo G, Andriolo L, Sessa A, Vannini F, Ferruzzi A, Marcacci M, *et al.* Age Is Not a Contraindication for Cartilage Surgery: A Critical Analysis of Standardized Outcomes at Long-term Follow-up. *Am J Sports Med* 2017;**45**:1822-8. http://dx.doi.org/10.1177/0363546517695088

68. International Cartilage Repair Society. *ICRS Patient Registry*. 2017. URL: https://cartilage.org/society/icrs-patient-registry/ (Accessed 18/10/2017).

69. Krych AJ, Pareek A, King AH, Johnson NR, Stuart MJ, Williams RJ, 3rd. Return to sport after the surgical management of articular cartilage lesions in the knee: a meta-analysis. *Knee Surgery, Sports Traumatology, Arthroscopy* 2016:18.

70. Schmitt LC, Quatman CE, Paterno MV, Best TM, Flanigan DC. Functional outcomes after surgical management of articular cartilage lesions in the knee: a systematic literature review to guide postoperative rehabilitation. *J Orthop Sports Phys Ther* 2014;**44**:565-a10. http://dx.doi.org/10.2519/jospt.2014.4844

71. Bartz C, Meixner M, Giesemann P, Roel G, Bulwin GC, Smink JJ. An ex vivo human cartilage repair model to evaluate the potency of a cartilage cell transplant. *J Transl Med* 2016;**14**:317. http://dx.doi.org/10.1186/s12967-016-1065-8

72. Fox AJS, Bedi A, Rodeo SA. The basic science of articular cartilage: structure, composition, and function. *Sports Health* 2009;**1**:461-8. http://dx.doi.org/10.1177/1941738109350438

73. Mizuno M, Kobayashi S, Takebe T, Kan H, Yabuki Y, Matsuzaki T, *et al.* Brief report: reconstruction of joint hyaline cartilage by autologous progenitor cells derived from ear elastic cartilage. *Stem Cells* 2014;**32**:816-21. http://dx.doi.org/10.1002/stem.1529

74. Mumme M, Barbero A, Miot S, Wixmerten A, Feliciano S, Wolf F, *et al.* Nasal chondrocyte-based engineered autologous cartilage tissue for repair of articular cartilage defects: an observational first-in-human trial. *The Lancet* 2016;**388**:1985-94. http://dx.doi.org/10.1016/S0140-6736(16)31658-0

75. Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast changing technologies: the case for tracker studies. *Bmj* 2000;**320**:43-6.

76. Shive MS, Stanish WD, McCormack R, Forriol F, Mohtadi N, Pelet S, *et al.* BST-CarGel Treatment Maintains Cartilage Repair Superiority over Microfracture at 5 Years in a Multicenter Randomized Controlled Trial. *Cartilage* 2015;**6**:62-72.

77. 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. http://dx.doi.org/10.1136/bmj.d5928

7 Appendices

Appendix 1.

Quality assessment of the COWISI trial

The ERG's quality assessment of the COWISI trial used the Cochrane ROB tool [77] (Table 33). The quality assessment focused on the primary outcome, namely the change of overall KOOS from baseline to final 24 months after the end of the treatment.

Table 33 Quality assessment of the COWISI trial

Domain	Description	Assessment of risk of bias
Random sequence generation	"A randomisation list was prepared and retained by Statconsult GmbH". Little detail given.	Probably low
Allocation concealment	Central telephone randomisation was used to assign patients to one of the two groups	Low
Blinding of participants and personnel	Blinding was not practical for the primary outcome because MF requires only one procedure and Spherox has two.	High
Blinding of outcome assessment	The primary outcome was the change of overall KOOS which was rated by patients that were aware of the treatment allocation. Those allocated to MF might have been disappointed?	High
Selective reporting	The primary outcome was pre-specified and reported consistently. All outcomes were reported in 2.6.1	Low
Incomplete outcome data	No imbalance in study discontinuations between the two arms.	Low
Other bias	The study was funded by Co-Don	Uncertain

Appendix 2. Quality assessment of case series

Before-after (pre-post) studies with no control group. Fickert 2012.

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the study question or objective clearly stated?	X		
2. Were eligibility/selection criteria for the study population prespecified and clearly described?	X		
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?	X		Note 1
4. Were all eligible participants that met the prespecified entry criteria enrolled?			CD
5. Was the sample size sufficiently large to provide confidence in the findings?		X	Note 2
6. Was the test/service/intervention clearly described and delivered consistently across the study population?	X		
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	X		
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?		X	Note 3
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis?		X	Note 4
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	X		
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?	X		Note 5
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?			N/A

Quality Rating: Fair
Additional Comments: Notes

1. Not entirely clear, but the patients recruited were a wider range than seen in trials, for example BMI and age range and on the second page, foot of RH column, they do say “patients from daily practice “ etc.
2. The sample size was large enough for some results to be statistically significant, but with only 37 patients, extrapolation to larger use may be unsafe.
3. There is mention of “independent readers” but since they knew that all patients had had Spherox, we think blinding was impossible. The “independent” appears to refer ti duplicate assessment. And the patients could not be blinded. So outcome blinding unclear, but unlikely. If unblinded, then MRI findings likely to be at high risk of bias
4. Loss to FU 19% but no account given of why lost.
5. Multiple times after intervention but not before. But methods comparable with most ACI studies.

*CD, cannot determine; NA, not applicable; NR, not reported

Before-after (pre-post) studies with no control group. Maotti 2012

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the study question or objective clearly stated?		X	
2. Were eligibility/selection criteria for the study population prespecified and clearly described?		X	
3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest?			CD
4. Were all eligible participants that met the prespecified entry criteria enrolled?			N/A
5. Was the sample size sufficiently large to provide confidence in the findings?		X	
6. Was the test/service/intervention clearly described and delivered consistently across the study population?			CD
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	X		Note 1
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?		x	CD note 2
9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis?			NR
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?	X		

11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?			CD
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?			N/A

Quality Rating: Poor
Additional Comments: Abstract only, therefore unable to determine many domains. Notes. 1. Yes overall. Yes for pre-specified, clearly defined and reliable, no for consistent assessment 2. Both a no and CD, because some outcomes assessed by patients who knew what they had had, and others assessed by radiologists or histology with no details given.

*CD, cannot determine; NA, not applicable; NR, not reported

Before-after (pre-post) studies with no control group. Siebold 2015

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the study question or objective clearly stated?	X		
2. Were eligibility/selection criteria for the study population pre-specified and clearly described?	X		
3. Were the participants in the study representative of those who would be eligible for the intervention in the general or clinical population of interest?			CD Note 1
4. Were all eligible participants that met the prespecified entry criteria enrolled?	X		
5. Was the sample size sufficiently large to provide confidence in the findings?			CD
6. Was the intervention clearly described and delivered consistently across the study population?	X		
7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants?	X		
8. Were the people assessing the outcomes blinded to the participants' exposures/interventions?		X	

9. Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis?		X	Note 2
10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes?		X	Note 3.
11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)?		X	
12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?			N/A

Quality Rating: Fair
Additional Comments: Reasonable design and conduct Notes. 1. The series was of all patients who had arthroscopic assessment after ACI. No information is given as to whether this was done in all patients receiving spheroids. 2. Only 76% had clinical follow-up 3. No pre-op data provided.

*CD, cannot determine; NA, not applicable; NR, not reported