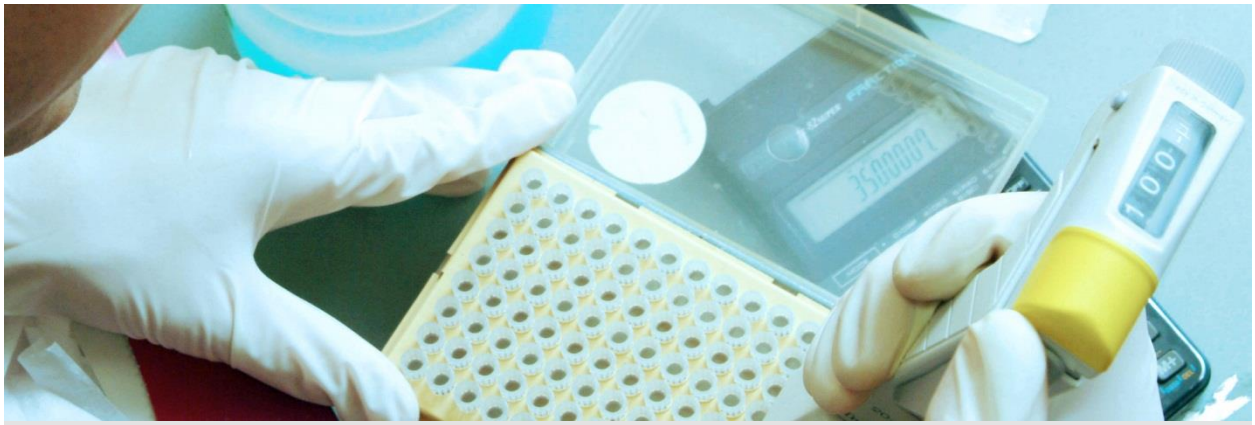
swiss medical board

Health Technology Assessment of knee arthroscopy for the treatment of degenerative changes



Assessment

Final assessment report
October 10th 2018

Impressum

Swiss Medical Board
Stampfenbachstrasse 30
Postfach, 8090 Zürich

Geschäftsstelle
Susanna Marti Calmell

Telefon +41 43 529 52 11
info@medical-board.ch
www.medical-board.ch

Assessment

The assessment team included the following persons (in alphabetical order)
Heiner C. Bucher[1], Dominik Glinz[1], Viktoria Gloy[1], Chandni Patel[1], Heike Raatz[1], Matthias Schwenkglenks[2,3], Yuki Tomonaga[3]

Clinical Expert:

Prof. Dr. med. Niklaus F. Friederich, Senior Consultant, Orthopaedics and Traumatology, University Hospital Basel

Reviewer Group:

Prof. Dr. rer. pol. Stefan Felder, Department of Business and Economics, University of Basel

Prof. Dr. Thomas Hügle, Department of Rheumatology, University Hospital Lausanne (CHUV) and Osteoarthritis Research Center Basel

PD Dr. med. Sandro Kohl, Chief Physician, Orthopaedics and Traumatology, University Hospital Bern and Hirslanden AG, 8152 Glattpark

Prof. Dr. med. Martin Tramèr, Head Division of Anaesthesiology and Director Department of Anaesthesiology, Pharmacology and Intensive Care, Geneva University Hospitals

Dr. phil. nat. Klazien Matter-Walstra, Leiterin Sektion HTA und Dr. med. Dorota A. Hofer-Zglinski, MPH, Bundesamt für Gesundheit (BAG)

The assessment team thanks the clinical expert and the members of the review group for their support. The assessment team is responsible for the contents of this report.

Contributions:
Coordination: HR and MS; **Scoping:** DG, VG, CP, HR, MS and YT; **Systematic review:** DG, VG, HR wrote the protocol; VG, CP and HR developed and CP conducted the literature search; DG and CP screened the literature; DG and CP were responsible for data extraction; DG, CP and HR graded the risk of bias and the quality of evidence; DG and CP conducted data analysis; DG, CP and HR are responsible for data interpretation and wrote this systematic review: **Health economic analyses:** MS and YT wrote the protocol; YT conducted the literature search; YT screened the identified literature; YT and MS were responsible for the data extraction and for the quality assessment of the literature; YT adapted the costs for Switzerland; MS and YT are responsible for data interpretation and wrote the health economic section; **final report:** HCB, DG, VG, CP, HR, MS and YT agreed upon the final content of this report.

[1] Basel Institute for Clinical Epidemiology and Biostatistics (CEB), Universitätsspital Basel
[2] Institute of Pharmaceutical Medicine, University of Basel
[3] Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich

Contents

Abbreviations

ACR	American College of Rheumatism
AIMS	Arthritis Impact Measurement Scale
APDRG	All Patients Diagnosis-Related Group
APM	Arthroscopic partial meniscectomy
AE	Adverse event
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CHOP	Schweizerische Operationsklassifikation (Swiss Classification of Surgical Interventions)
CI	Confidence Interval
CRD	Centre for Review and Dissemination
DMK	Degenerative meniscus of the knee
DRG	Diagnosis-Related Group
EQ-5D VAS	EuroQol 5 dimensions Visual Analogue Scale
FDA	Food and Drug Administration
FOPH	Federal Office of Public Health
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HSSKRS	Hospital for Special Surgery Knee Rating Score
HTA	Health Technology Assessment
ICD	International Classification of Diseases
ICER	Incremental Cost-Effectiveness Ratio
ID	Identification
IQR	Interquartile range
ISPOR	International Society For Pharmacoeconomics and Outcomes Research
KK	Komorbiditäten und Komplikationen (Comorbidities and Complications)
KL	Kellgren-Lawrence
KSRS	Knee Society rating system
KOOS	Knee injury and Osteoarthritis Outcome Score
LI	Lequesne Index
MCID	Minimal clinically important difference
MD	Mean difference
MOS SF-36	Medical Outcomes Study Short Form 36
N.A.	Not applicable
N.R.	Not reported
OA	Osteoarthritis
Obsan	Swiss Health Observatory
OIS	Optimal information size
OKS	Oxford Knee Score
OR	Odds ratio
QALY	Quality-adjusted life year

RCT	Randomised controlled trial
RoB	Risk of bias
RR	Relative risk ratio
SAE	Serious adverse event
SD	Standard deviation
SFOPH	Swiss Federal Office of Public Health
SMD	Standardised mean difference
SMB	Swiss Medical Board
TARMED	Tarif Médical
VAS	Visual Analogue Scale
WHO	World Health Organization
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
WOMET	Western Ontario Meniscal Evaluation Tool

Tables

Figures

1 Executive summary

1.1 Summary

1.1.1 Background

In the context of the Health Technology Assessment (HTA) program of the Swiss federation, services reimbursed by the compulsory health insurance are being re-evaluated. One of the topics selected in 2015 is arthroscopy of the knee, based on the Swiss Health Observatory (Obsan) report No. 42 "Variations géographiques dans les soins de santé. La situation en Suisse", which reported an increase of the rate of inpatient knee arthroscopies by 20% between 2005 and 2011. According to the report, the rates of inpatient arthroscopies differed markedly between cantons, i.e. cantons below the 10th percentile showed a standardized rate of 121 arthroscopies per 100'000 inhabitants, while cantons above the 90th percentile showed a standardized rate of 415.6 per 100'000 inhabitants. In addition, the proportion of inpatient and outpatient arthroscopies varied strongly from 16% up to 75% between cantons. These observations raise questions regarding the appropriate use and the benefit of knee arthroscopy.

1.1.2 Aim

This HTA report aims at assessing the clinical effectiveness and safety of

- therapeutic knee arthroscopy compared to any other treatment in patients with degenerative changes of the knee – irrespective of whether they are primarily due to meniscal damage, osteoarthritis of the knee or a mix of both;
- inpatient compared to outpatient therapeutic knee arthroscopy.

Also, this HTA report aims at assessing

- the cost-effectiveness of therapeutic knee arthroscopy compared to any other treatment in patients with degenerative changes of the knee – irrespective of whether they are primarily due to meniscal damage, osteoarthritis of the knee or a mix of both;
- the cost-effectiveness of inpatient compared to outpatient therapeutic knee arthroscopy;
- the budget impact of knee arthroscopy in patients with degenerative changes of the knee primarily due to meniscal damage.

1.1.3 Clinical effectiveness and safety

For this HTA report, clinical effectiveness and safety of knee arthroscopy in patients with symptomatic knee degeneration were addressed. Clinical effectiveness and safety of the setting (inpatient versus outpatient) in which therapeutic arthroscopy was performed were also studied. The literature search was conducted in July 2017 and filters for randomised controlled trials (RCTs) were used. The search was not restricted by time period or by language. RCT characteristics and results of the included RCTs were presented in tables and summarised

descriptively. The main focus of the analysis was the results at short-term follow-up (≤6 months), intermediate follow-up (>6 months and <7 years) and long-term follow-up (longest follow-up time including and longer than seven years). Risk of bias was assessed according to the Cochrane Handbook, and the quality of evidence was assessed according to GRADE for short-term and intermediate follow-up. When possible, clinical outcome results were summarised quantitatively in a meta-analysis by using inverse variance models assuming random effects. Effect estimates (overall and for each RCT) with corresponding 95% confidence intervals (CIs) were presented as forest plots. Relative risks (RRs) were calculated for binary outcomes. Continuous outcomes were presented as mean differences (MDs). If continuous outcomes were measured with different instrument (i.e. different scales) by the individual RCTs, the results of the individual RCTs were standardized, and the standardized mean differences (SMDs) were reported. In case of considerable heterogeneity, methodological and clinical factors that might explain the heterogeneity were explored in subgroup and sensitivity analyses where possible. Pre-specified subgroup/sensitivity analyses addressed

- patients with symptoms mainly due to meniscal degeneration (degenerative meniscus of the knee, DMK) versus patients with symptoms mainly due to osteoarthritis (OA) of the knee versus mixed populations with concurrent DMK and OA;
- RCTs in which a non-active comparator was used versus those RCTs in which an active comparator was used.

There were 21 RCTs (including more than 2,000 patients) identified for the assessment of clinical effectiveness and safety of arthroscopy in patients with degenerative knee symptoms. Data were extracted for short-term follow-up and intermediate follow-up, for the outcomes of pain, function, global assessment (combined pain, function and/or stiffness), joint stiffness, occurrence of total knee replacement, quality of life, adverse events and serious adverse event. There were no statistically significant differences between arthroscopy and comparator found for the outcomes of function, global assessment, joint stiffness, total knee replacement and quality of life at short-term (≤6 months) or intermediate follow-up (>6 months and <7 years). There was a small statistically significant effect in favour of arthroscopy in the outcome of pain at short-term follow-up (SMD -0.16, 95% CI [-0.31, -0.01]), while no statistically significant difference was found at intermediate follow-up. None of the included RCTs reported outcomes at long-term follow-up. There was limited evidence of harms reported by the RCTs; therefore, it was difficult to assess the overall clinical effectiveness with regards to benefits and harms of arthroscopy.

The overall quality of evidence was judged to be very low at short-term follow-up because of the very low quality of evidence for the critical outcome of function. The overall quality of evidence was judged to be low at intermediate follow-up because of the low quality of evidence for the critical outcomes of pain, function and global assessment.

One RCT (100 patients) was identified for the assessment of the clinical effectiveness and safety of arthroscopy in the inpatient and outpatient settings. This RCT reported only on pain within a week of discharge and found no difference. The overall quality of evidence was judged to be very low.

1.1.4 Cost-effectiveness and budget impact

To gain understanding of the cost-effectiveness of therapeutic knee arthroscopy compared to any other treatment in patients with degenerative changes of the knee, the available published literature was analysed. The analysis was based on a literature search using the same search terms as in the clinical effectiveness and safety section of this report, in combination with economic terms. After screening of the search results to identify eligible studies, extraction of relevant information, assessment of quality of reporting according to the CHEERS checklist, and assessment of transferability to Switzerland (for international studies) were performed. For the studies judged to be qualitatively transferable, cost estimates were adapted to Switzerland and cost-effectiveness results re-calculated.

Four cost-effectiveness studies were identified. Two compared knee arthroscopy to non-operative treatment and showed discordant results: one suggested that knee arthroscopy was more expensive and less effective than non-operative interventions alone from both the societal and healthcare payer perspective; the other suggested that knee arthroscopy was more costly and more effective (incremental cost-effectiveness ratio [ICER] better than CHF 30,000 per quality adjusted life year [QALY] gained) from a healthcare payer perspective or even cost-saving from a societal perspective, if compared to physical therapy alone. To understand the reasons behind this discrepancy, both publications as well as the sources used in their calculations were analysed in detail. In both cases, relevant methodological flaws were found. Two other studies comparing preoperative status with postoperative status suggested that knee arthroscopy may be cost-effective with ICERs of CHF 7,200 to 7,300 per QALY gained. However, pre-post clinical studies do not represent a reliable basis for deciding whether an intervention is cost-effective or not. Given the very limited health economic evidence, it is difficult to draw firm conclusions on the cost-effectiveness of arthroscopic surgery in patients with degenerative changes of the knee. Although three of four available studies reported arthroscopy to be cost-effective, this cannot be regarded as convincing.

The budget impact analysis focused on the costs of knee arthroscopy in patients with degenerative changes of the knee primarily due to meniscal damage, from a health insurance system perspective in Switzerland. Two steps were performed: first, the annual frequency of knee arthroscopies in Switzerland was investigated; second, based on the annual frequency, the total annual costs were estimated. The frequency in the inpatient setting was investigated using diagnosis-related (DRG) codes, ICD-10 codes, and Swiss classification of surgical interventions (CHOP) codes provided in the Swiss Hospital Statistics for the years 2010 to 2014. Two analytical strategies were used: in the first, only patients who reported at the same time a relevant DRG code, at least one relevant diagnosis (ICD-10 code), and at least one relevant treatment (CHOP code) were included. In the second strategy, all patients who had at the same time at least one relevant diagnosis (ICD-10 code) and one relevant treatment (CHOP code), irrespective of the reported DRG codes, were included. To assess the frequency of arthroscopies in the outpatient setting, an analysis of medical tariff (TARMED) codes in 2013 and 2014 provided by the Swiss Federal Office of Public Health was used. The unit costs for the budget impact calculation were based on data from the diagnosis-related case costs statistics provided by the Swiss Federal Office of Statistics ("Statistik diagnosebezogener Fallkosten") and on published estimates.

The results of the budget impact analysis suggested that the total expenditure for knee/meniscus derangement in Switzerland ranged from CHF 53.52 Mio. to CHF 71.93 Mio. in 2013 and from CHF 52.30 Mio. to CHF 67.73 Mio. in 2014. Outpatient costs accounted for 20-28% of the total costs.

The results of the analysis according to our second patient selection strategy, based on ICD-10 codes and CHOP codes, are considered more realistic than those according to the first strategy also considering DRG codes. This second strategy suggested total inpatient costs of CHF 58.10 Mio. in 2010, CHF 55.87 Mio. in 2011, CHF 58.44 Mio. in 2012, CHF 57.20 Mio. in 2013 and CHF 54.47 Mio. in 2014. Total outpatient costs were estimated to be CHF 14.73 Mio. in 2013 and CHF 13.26 Mio. in 2014. The results of the budget impact analysis according to the first patient selection strategy seemed to be quite uncertain due to the changes (from All Patients Diagnosis-Related Group [APDRG] until 2011 to Swiss DRG from 2012 onwards) and inconsistencies in the DRG coding system.

The results of the budget impact analysis are in line with those results reported in a very recent publication of the Swiss Health Observatory (Obsan). Assuming approximately 14,000 meniscectomies per year and inpatient costs of CHF 4,889 per case, the authors of the Obsan report estimated total inpatient costs of CHF 55.6 Mio. in 2016.

1.1.5 Conclusion

In conclusion, there is no evidence that knee arthroscopic interventions in patients with degenerative changes of the knee have any benefit on outcomes measured at short or intermediate follow-up, with the exception of a small effect on the reduction of pain at short follow-up. Therefore, it remains unclear whether knee arthroscopy has an effect on the assessed outcomes. Long-term follow-up data are not available. Reporting on harm was scarce and no conclusions can be drawn regarding the benefit-harm balance. The overall quality of evidence at short and intermediate follow-up was judged to be very low and low, respectively. In addition, there is no evidence that the subgroup of patients with solely DMK has a benefit from arthroscopic treatment. The clinical effectiveness findings of this assessment may be generalizable to a broader population experiencing knee pain due to a degenerative knee disorder. The findings are consistent with recently published systematic reviews.

Given very limited health economic evidence, it is difficult to draw firm conclusions on the cost-effectiveness of arthroscopic surgery in patients with degenerative changes of the knee. Although three of four available studies reported arthroscopy to be cost-effective, this cannot be regarded as convincing, given the methodological issues described and also in light of the results of the assessment of effectiveness.

1.2 Zusammenfassung

1.2.1 Hintergrund

Im Rahmen des Health Technology Assessment (HTA)-Programms des Bundesamts für Gesundheit (BAG) werden zu erstattende Leistungen der obligatorischen Krankenpflegeversicherung der Schweiz neu bewertet. Eines der 2015 gewählten Themen war Kniearthroskopie, basierend auf dem Bericht des Schweizerischen Gesundheitsobservatoriums (Obsan) Nr. 42 „*Variations géographiques dans les soins de santé. La situation en Suisse*“, welcher zwischen 2005 bis 2011 einen Anstieg um 20% von stationären Kniearthroskopien in Schweizer Spitälern fand. Gemäss dem Obsan-Bericht variiert die Häufigkeit der stationären Eingriffe markant zwischen den Kantonen, d.h. in Kantonen unter der zehnten Perzentile wurde eine standardisierte Häufigkeit von 121 Arthroskopien pro 100‘000 Einwohnern beobachtet, während in Kantonen über der neunzigsten Perzentile 415.6 Arthroskopien pro 100‘000 Einwohner durchgeführt wurden. Des Weiteren variiert das Verhältnis von stationären und ambulanten Arthroskopien stark zwischen den Kantonen. Diese Beobachtungen warfen die Frage der angemessenen Anwendung sowie des Nutzens von Kniearthroskopie auf.

1.2.2 Ziel

Das Ziel dieses HTA Berichts ist, die klinische Wirksamkeit und Sicherheit zu untersuchen der

- therapeutischen Kniearthroskopie, verglichen mit jeder anderen Behandlungsform in Patienten mit degenerativen Veränderungen des Knies, ungeachtet ob diese primär durch Meniskusschaden, Osteoarthritis oder beides verursacht wurden; und
- therapeutischen Kniearthroskopie mit stationärer im Vergleich zu ambulanter Durchführung.

Ausserdem untersucht dieser HTA Bericht

- die Kosteneffektivität der therapeutischen Kniearthroskopie im Vergleich zu jeder anderen Behandlungsform bei Patienten mit degenerativen Veränderungen des Knies, ungeachtet ob diese primär durch Meniskusschaden, Osteoarthritis oder beides verursacht wurden;
- Die Kosteneffektivität der stationären im Vergleich zur ambulanten durchgeführten therapeutischen Kniearthroskopie;

Der *Budget Impact* von Kniearthroskopie bei Patienten mit degenerativen Veränderungen des Knies, welche primär durch Meniskusschäden verursacht wurden, wird ebenfalls untersucht.

1.2.3 Klinische Wirksamkeit und Sicherheit

Für diesen HTA Bericht wurde die klinische Wirksamkeit und Sicherheit von Kniearthroskopie bei Patienten mit symptomatischen, degenerativen Kniebeschwerden untersucht. Zusätzlich zur klinischen Wirksamkeit und Sicherheit wurde die Behandlungssituation (ambulant oder stationär) untersucht. Die Literatursuche wurde im Juli 2017 durchgeführt und Filter für randomisiert-kontrollierte Studien (RCTs) wurden verwendet. Die Literatursuche wurde nicht

zeitlich oder sprachlich eingeschränkt. Studiencharakteristika und Resultate der eingeschlossenen Studien wurden in Tabellen aufgeführt und deskriptiv zusammengefasst. Der Hauptfokus der Analyse lag auf den Resultaten kurzer (bis zu sechs Monate), intermediärer (über sechs Monate und unter 7 Jahre) und langer Nachbeobachtungszeiten (einschliesslich oder länger als sieben Jahre). Das Verzerrungspotential der Studienergebnisse wurde entsprechend dem Cochrane Handbuch bestimmt und die Qualität der Evidenz nach GRADE bewertet. Wenn möglich, wurden die Resultate mittels Meta-Analysen mit inversen Varianzmodellen unter der Annahme von zufälligen Effekten quantitativ zusammengefasst. Effektschätzer (einzeln und gepoolt) mit dem dazugehörigen 95% Konfidenzintervall (CI) wurden in Forest Plots dargestellt. Relative Risiken (RRs) wurden für binäre Endpunkte berechnet. Kontinuierliche Endpunkte wurden als Mittelwertdifferenzen (MDs) dargestellt. Wenn kontinuierliche Endpunkte der einzelnen RCTs mit unterschiedlichen Instrumenten (d.h. mit unterschiedlichen Skalen) gemessen wurden, dann wurden die Resultate standardisiert und als standardisierte Mittelwertdifferenzen (SMDs) dargestellt. Im Falle von beträchtlicher Heterogenität wurden methodische und klinische Faktoren, die diese erklären könnten, soweit möglich durch Subgruppen- und Sensitivitätsanalysen untersucht. Vorab festgelegte Subgruppen-/Sensitivitätsanalysen betrafen unter anderem

- Patienten mit Symptomen hauptsächlich verursacht durch degenerative Meniskusschäden (DMK) versus Patienten mit Symptomen hauptsächlich verursacht durch Osteoarthritis (OA) versus Patienten mit DMK und OA.
- RCTs mit nicht-aktiven Vergleichsinterventionen versus RCTs mit aktiven Vergleichsinterventionen.

Es wurden 21 RCTs (mit mehr als 2000 Patienten) für die Untersuchung der klinischen Wirksamkeit und Sicherheit der Arthroskopie bei Patienten mit degenerativem Kniesymptomen eingeschlossen. Für die kurze und intermediäre Nachbeobachtungszeit wurden Informationen für folgende Endpunkte extrahiert: Schmerz, Funktion, globale Bewertung (Kombination aus Schmerz, Funktion und/oder Gelenksteifigkeit), Gelenksteifigkeit, Häufigkeit von komplettem Knieersatz, Lebensqualität, Nebenwirkungen und schwerwiegende Nebenwirkungen. Für die Endpunkte Funktion, globale Bewertung, Gelenksteifigkeit, Häufigkeit von komplettem Knieersatz, Lebensqualität wurden sowohl für die kurze (≤6 Monate) als auch intermediäre Nachbeobachtungszeit (>6 Monate und <7 Jahre) keine statistisch signifikanten Unterschiede zwischen Arthroskopie und der Vergleichsintervention gefunden. Ein geringer, statistisch signifikanter Effekt zugunsten der Arthroskopie wurde für den Endpunkt Schmerz bei kurzer Nachbeobachtungszeit gefunden (SMD -0.16, 95% CI [-0.31, -0.01]), während für die intermediären Nachbeobachtungszeit kein statistisch signifikanter Unterschied gefunden wurde. Die RCTs berichteten selten Nebenwirkung, daher ist die Evidenz eingeschränkt. Folglich ist die Abschätzung der klinischen Wirksamkeit betreffend Nutzen und Schaden der Arthroskopie schwierig.

Die Gesamtqualität der Evidenz wurde aufgrund der sehr niedrigen Qualität der Evidenz für den kritischen Endpunkt Funktion bei kurzer Nachbeobachtungszeit als sehr niedrig beurteilt. Die Gesamtqualität der Evidenz wurde aufgrund der niedrigen Qualität der Evidenz für die kritischen Endpunkte Schmerz, Funktion und globale Bewertung bei intermediärer Nachbeobachtungszeit als niedrig beurteilt.

Ein RCT (100 Patienten) wurde für die Untersuchung der klinischen Wirksamkeit und Sicherheit von Arthroskopie als stationäre versus ambulante Behandlungsform eingeschlossen. Dieser RCT

untersuchte nur Schmerz innerhalb der ersten Woche nach der Entlassung als Endpunkt und fand keine Unterschiede. Die Gesamtqualität der Evidenz wurde als sehr niedrig beurteilt.

1.2.4 Kosten-Effektivität und Budget Impact

Die vorhandene Literatur zur Kosten-Effektivität der therapeutischen Kniearthroskopie, im Vergleich zu jeglichen anderen Therapien, bei Patienten mit degenerativen Knieveränderungen wurde untersucht. Die Literatursuche basierte auf der Suchstrategie des Teils «klinische Wirksamkeit und Sicherheit» dieses Berichts, in Kombination mit ökonomischen Suchbegriffen. Nach Screening der Suchresultate und der Identifizierung von relevanten Artikeln wurden relevante Informationen extrahiert. Die Qualität des Reporting wurde mit Hilfe der Consolidated Health Economic Evaluation Reporting Standards (CHEERS)-Checkliste evaluiert. Für internationale Studien wurde die qualitative Übertragbarkeit auf die Schweiz ermittelt. Bei Studien, die als qualitativ auf die Schweiz übertragbar eingeschätzt wurden, wurden die Kostenschätzungen adaptiert, um die Kosten-Effektivitäts-Resultate neu zu berechnen.

Vier publizierte Kosten-Effektivitätsstudien wurden identifiziert. Zwei Studien verglichen die Kniearthroskopie mit nicht-operativen Behandlungen und zeigten diskordante Resultate: eine Studie fand, dass Kniearthroskopie in Vergleich zu physikalischer Therapie teurer und weniger effektiv war. Die andere Studie fand, dass Kniearthroskopie in Vergleich zu physikalischer Therapie sowohl aus der Perspektive eines Zahlers von Gesundheitsleistungen als aus der gesellschaftlichen Perspektive teurer, aber auch effektiver war (inkrementelles Kosten-Effektivitäts-Verhältnis [ICER] <CHF 30’000 pro gewonnenem qualitätsadjustiertem Lebensjahr [QALY]). Um diese Diskrepanz zu verstehen, wurden beide Publikationen sowie die Quellen, die den Berechnungen zugrunde lagen, genauer untersucht. Bei beiden Studien wurden signifikante Mängel identifiziert. Zwei weitere Studien verglichen den präoperativen Status mit dem postoperativen Status von Patienten und befanden, dass Kniearthroskopie mit ICERs von CHF 7’200-7’300 pro QALY kosteneffektiv sein könnte. Klinische Vorher-Nachher-Vergleiche stellen jedoch keine verlässliche Basis für Kosten-Effektivitätsanalysen dar. Da die gesundheitsökonomische Evidenz sehr limitiert ist, ist es schwierig, klare Schlussfolgerungen betreffend Kosten-Effektivität der Kniearthroskopie im Vergleich zu anderen Behandlungen bei Patienten mit degenerativen Knieveränderungen zu ziehen. Obgleich drei von vier vorhandenen Studien berichteten, dass die Kniearthroskopie kosteneffektiv sei, sind diese Resultate nicht überzeugend.

Die Budget Impact-Analyse konzentrierte sich auf die Kosten von Kniearthroskopien bei Patienten mit degenerativen Knieveränderungen durch Meniskusschäden. Die Analyse wurde aus der Schweizer Krankenversicherungsgesetz (KVG)-Perspektive durchgeführt. Sie erfolgte in zwei Schritten. Erstens wurde die jährliche Häufigkeit von Kniearthroskopien in der Schweiz untersucht. Zweitens wurden auf dieser Basis die gesamten jährlichen Arthroskopiekosten geschätzt. Die Häufigkeit von Arthroskopien im Rahmen stationärer Aufenthalte wurde anhand der *Diagnosis-related group* (DRG)-Codes, der ICD-10-Codes und der Codes der schweizerischen Operationsklassifikation (CHOP) untersucht, welche von der Krankenhausstatistik der Jahre 2010 bis 2014 des Bundesamts für Statistik zur Verfügung gestellt wurden. Zwei analytische Strategien wurden angewendet: in der ersten Strategie wurden Patienten eingeschlossen, die gleichzeitig einen relevanten DRG-Code, mindestens eine relevante Diagnose (ICD-10-Code) und mindestens eine relevante Behandlung (CHOP-Code) hatten. In der zweiten Strategie wurden alle Patienten eingeschlossen, die gleichzeitig mindestens eine relevante Diagnose (ICD-10-Code) sowie

mindestens eine relevante Behandlung (CHOP-Code) hatten, unabhängig vom angegebenen DRG-Code. Um die Häufigkeit von Kniearthroskopien im ambulanten Bereich zu schätzen, wurde eine vom Bundesamt für Gesundheit bereitgestellte Analyse auf der Basis von ambulanten Abrechnungscodes (TARMED-Codes) der Jahre 2013 und 2014 verwendet. Die verwendeten Kosten pro durchgeführter Arthroskopie basierten auf den Daten der Statistik diagnosebezogener Fallkosten des Bundesamtes für Statistik sowie auf publizierten Schätzungen.

Die Budget Impact-Analyse ergab für das Jahr 2013 in der Schweiz Gesamtkosten für arthroskopische Knie-/Meniskusbehandlungen zwischen CHF 53.52 Mio. und CHF 71.93 Mio. und im Jahr 2014 zwischen CHF 52.30 Mio. und CHF 67.73 Mio. Davon entfielen 20-28% auf ambulante Kosten. Die Analyseresultate gemäss der zweiten Patientenselektionsstrategie (basierend auf ICD-10-Codes und CHOP-Codes) wurde für realistischer gehalten als die der ersten Strategie, die auch DRG-Codes berücksichtigte. Ausgehend von der zweiten Strategie wurden die jährlichen Spitalkosten auf CHF 58.10 Mio. im Jahr 2010, CHF 55.87 Mio. im Jahr 2011, CHF 58.44 Mio. im Jahr 2012, CHF 57.20 Mio. im Jahr 2013 und CHF 54.47 Mio. im Jahr 2014 geschätzt. Die gesamten ambulanten Kosten wurden auf CHF 14.73 Mio. im Jahr 2013 und CHF 13.26 Mio. im Jahr 2014 geschätzt. Die Resultate der Budget Impact-Analyse gemäss der ersten Patientenselektions-Strategie erschienen sehr unsicher wegen der Änderungen (von *All Patients Diagnosis-Related Group* [APDRG] bis 2011 hin zu Swiss DRG ab 2012) und Inkonsistenzen des DRG-Codierungssystems.

Die Resultate der Budget Impact-Analyse decken sich gut mit vor kurzem publizierten Schätzungen des Obsans. Unter der Annahme, dass pro Jahr 14'000 Meniskektomien durchgeführt werden, und die Spitalkosten bei CHF 4'889 pro Patient liegen, schätzte Obsan, dass die gesamten Spitalkosten im Jahr 2016 bei CHF 55.6 Mio. lagen.

1.2.5 Schlussfolgerung

Es gibt keine Evidenz eines Nutzens der Kniearthroskopie für Patienten mit degenerativen Kniebeschwerden nach kurzer oder intermediärer Nachbeobachtungszeit. Lediglich eine leichte, klinisch nicht relevante, Schmerzreduktion zugunsten der Arthroskopie war bei kurzer Nachbeobachtungszeit sichtbar. Daher bleibt unklar ob Kniearthroskopie einen Effekt auf die untersuchten Endpunkte hat. Zur langfristiger Nachbeobachtungszeit lagen keine Daten vor. Nebenwirkungen wurden selten von den RCTs berichtet, folglich ist eine Abwägung von Nutzen und Schaden nicht möglich. Die Gesamtqualität der Evidenz wurde für die kurze und intermediäre Nachbeobachtungszeit als sehr niedrig und niedrig bewertet. Des Weiteren wurde keine Evidenz gefunden, dass die Subgruppe von Patienten mit alleinigem Meniskusschaden einen Nutzen aus der Kniearthroskopie zieht. Die Ergebnisse der Untersuchung der klinischen Wirksamkeit können auf eine breite Population von Patienten mit Knieschmerzen durch degenerative Knieveränderungen verallgemeinert werden. Diese Ergebnisse sind konsistent mit kürzlich publizierten systematischen Übersichtsarbeiten.

Die Evidenz für die gesundheitsökonomische Untersuchung ist sehr eingeschränkt und deshalb ist es schwierig, eine klare Schlussfolgerung hinsichtlich der Kosten-Effektivität der Arthroskopie bei Patienten mit degenerativen Veränderung des Knies zu formulieren. Obwohl drei der vier verfügbaren Studien berichteten, dass Arthroskopie kosteneffektiv ist, kann dies nicht als überzeugend beurteilt werden, wegen der beschriebenen methodischen Mängel dieser Studien sowie in Anbetracht der Resultate der Untersuchung der Wirksamkeit.

1.3 Résumé

1.3.1 Contexte

Dans le cadre du programme d'évaluation des technologies de la santé (*Health Technology Assessment*, HTA) de la Confédération helvétique, les soins remboursés par l'assurance maladie obligatoire font actuellement l'objet d'une réévaluation. L'arthroscopie du genou est l'un des thèmes sélectionnés en 2015, sur la base du rapport n° 42 de l'Observatoire suisse de la santé (Obsan), intitulé « Variations géographiques dans les soins de santé. La situation en Suisse », qui fait état d'une augmentation de 20 %, entre 2005 et 2011, du taux d'arthroscopies du genou des patients hospitalisés. Selon ce rapport, le taux d'arthroscopies des patients hospitalisés différait sensiblement d'un canton à l'autre : les cantons se situant au-dessous du 10e percentile présentaient un taux standardisé de 121 arthroscopies pour 100 000 habitants, tandis que les cantons situés au-dessus du 90e percentile présentaient un taux standardisé de 415,6 pour 100.000 habitants. En outre, la proportion d'arthroscopies des patients hospitalisés et ambulatoires variait fortement entre cantons, soit de 16 % à 75 %. Ces constats soulèvent des questions sur la pertinence de l'utilisation et l'intérêt de l'arthroscopie du genou.

1.3.2 Objectif

Le présent rapport HTA vise à évaluer l'efficacité et la sécurité de

- l'arthroscopie thérapeutique du genou par rapport aux autres traitements des patients présentant des altérations dégénératives du genou, que ces altérations soient principalement dues à une lésion méniscale, à une ostéoarthrite du genou ou aux deux à la fois ;
- l'arthroscopie thérapeutique du genou des patients hospitalisés par rapport à celle des patients ambulatoires.

Ce rapport HTA a également pour objectif d'évaluer

- le rapport coût-efficacité de l'arthroscopie thérapeutique du genou par rapport aux autres traitements des patients présentant des altérations dégénératives du genou, que ces altérations soient principalement dues à une lésion méniscale, à une ostéoarthrite du genou ou aux deux à la fois ;
- le rapport coût-efficacité de l'arthroscopie thérapeutique du genou des patients hospitalisés par rapport à celle des patients ambulatoires ;
- l'impact budgétaire de l'arthroscopie du genou des patients présentant des altérations dégénératives du genou principalement dues à une lésion méniscale.

1.3.3 Efficacité et sécurité

Aux fins du présent rapport HTA, on s'est penché sur l'efficacité et la sécurité de l'arthroscopie du genou des patients présentant une dégénérescence symptomatique du genou. On a également étudié l'efficacité et la sécurité de l'environnement (hospitalier par rapport à ambulatoire) dans

lequel ont été réalisées les arthroscopies thérapeutiques. La recherche documentaire a été effectuée en juillet 2017 et des filtres ont été utilisés pour les études contrôlées randomisées (*randomised controlled trial*, RCT). La recherche n'était pas limitée dans le temps ou par la langue. Les caractéristiques et les résultats des RCT incluses ont été présentés sous forme de tableaux et de résumés descriptifs. L'analyse portait principalement sur les résultats des suivis à court terme (≤6 mois), à moyen terme (>6 mois et <7 ans) et à long terme (période de suivi la plus longue, ≥7 ans). Le risque de biais a été évalué selon le manuel Cochrane, et la qualité des preuves l'a été selon l'approche GRADE pour les suivis à court terme et à moyen terme. Lorsque cela a été possible, les résultats cliniques ont été résumés quantitativement dans une méta-analyse à l'aide de modèles de variance inverse à effets aléatoires. Les estimations des effets (globalement et pour chaque étude), avec des intervalles de confiance (IC) 95 %, ont été présentées sous forme de *forest plots*. Des risques relatifs ont été calculés pour les résultats binaires. Les résultats continus ont été présentés en termes de différences moyennes. Lorsque les résultats continus ont été mesurés avec des instruments différents (c'est-à-dire des échelles différentes) par les différentes RCT, les résultats des différentes RCT ont été standardisés, et on a relevé les différences moyennes standardisées (DMS). Dans les cas de grande hétérogénéité, les facteurs méthodologiques et cliniques pouvant expliquer cette hétérogénéité ont été explorés par des analyses de sous-groupes et de sensibilité chaque fois que cela a été possible. Les analyses de sous-groupes pré-spécifiés/de sensibilité portaient sur :

- des patients dont les symptômes étaient principalement dus à une dégénérescence méniscale (ménisque dégénératif du genou) par rapport à des patients dont les symptômes étaient principalement dus à une ostéoarthrite (OA) du genou et par rapport à des populations mixtes présentant à la fois un ménisque dégénératif du genou et une OA ;
- des RCT dans lesquelles un comparatif non actif a été utilisé par rapport aux RCT dans lesquelles un comparatif actif a été utilisé.

Vingt-et-une RCT (regroupant plus de 2 000 patients) ont été identifiées pour l'évaluation de l'efficacité et de la sécurité de l'arthroscopie chez des patients présentant des symptômes dégénératifs du genou. Les données ont été extraites pour le suivi à court terme et le suivi à moyen terme, pour les résultats sur la douleur, la fonction, l'évaluation globale (douleur, fonction et/ou raideurs en même temps), les raideurs articulaires, l'implantation d'une prothèse totale du genou, la qualité de vie, les effets indésirables et les effets indésirables graves. Aucune différence statistiquement significative n'a été trouvée entre l'arthroscopie et le comparatif pour les résultats sur la fonction, l'évaluation globale, les raideurs articulaires, l'implantation d'une prothèse totale du genou et la qualité de vie dans les suivis à court terme (≤ 6 mois) ou à moyen terme (> 6 mois et < 7 ans). Un léger effet statistiquement significatif en faveur de l'arthroscopie a été trouvé dans le résultat sur la douleur dans le suivi à court terme (DMS – 0,16, IC 95 % [-0,31 ; – 0,01], tandis qu'aucune différence statistiquement significative n'a été trouvée dans le suivi à moyen terme. Aucune des études incluses n'a révélé de résultats dans le suivi à long terme. Les études n'ont fourni que des preuves limitées sur les effets secondaires; il a donc été difficile d'évaluer l'efficacité clinique globale au niveau des avantages et des inconvénients de l'arthroscopie.

La qualité générale des preuves a été jugée très faible dans le suivi à court terme en raison de la très faible qualité des preuves pour le résultat critique sur la fonction. La qualité générale des

preuves a été jugée faible dans le suivi à moyen terme en raison de la faible qualité des preuves pour les résultats critiques sur la douleur, la fonction et l'évaluation globale.

Une étude randomisée (100 patients) a été identifiée pour l'évaluation de l'efficacité et de la sécurité cliniques de l'arthroscopie dans les environnements hospitalier et ambulatoire. Cette étude ne portait que sur la douleur ressentie dans la semaine qui suivait la sortie de l'établissement et n'a révélé aucune différence. La qualité générale des preuves a été jugée très faible.

1.3.4 Rapport coût-efficacité et impact budgétaire

Pour mieux comprendre le rapport coût-efficacité de l'arthroscopie thérapeutique du genou par rapport aux autres traitements des patients présentant des altérations dégénératives du genou, on a analysé la documentation publiée disponible. L'analyse s'est basée sur une recherche systématique utilisant les mêmes termes de recherche que ceux utilisés pour la section de ce rapport sur l'efficacité et la sécurité, en association avec des termes économiques. Après avoir passé au crible les résultats de recherche pour identifier les études éligibles, on a procédé à l'extraction des informations pertinentes, à l'évaluation de la qualité des rapports en se référant à la liste de vérification CHEERS, ainsi qu'à l'évaluation de la transférabilité à la Suisse (pour les études internationales). Pour les études jugées qualitativement transférables, les estimations de coûts ont été adaptées à la Suisse et les résultats du rapport coût-efficacité recalculés.

Quatre études coût-efficacité ont été identifiées. Deux comparaient l'arthroscopie du genou à un traitement non chirurgical et présentaient des résultats discordants : une étude suggérait que l'arthroscopie du genou était plus coûteuse et moins efficace que les seules interventions non chirurgicales, du point de vue sociétal comme de celui du payeur des soins de santé ; l'autre étude suggérait que l'arthroscopie du genou était plus onéreuse et plus efficace (rapport coût-efficacité différentiel [*incremental cost-effectiveness ratio*, ICER] supérieur à 30 000 CHF par année de vie pondérée par la qualité [QALY] gagnée) du point de vue du payeur des soins de santé, voire économique du point de vue sociétal, si on la comparait à la seule physiothérapie. Pour comprendre les raisons de cette divergence, les deux publications et les sources utilisées pour leurs calculs ont été analysées en détail. Dans les deux cas, des failles méthodologiques à l'origine de cette divergence ont été trouvées. Deux autres études comparant l'état pré-opératoire à l'état post-opératoire ont suggéré que l'arthroscopie du genou pouvait présenter un rapport coût-efficacité positif avec des ICER compris entre 7 200 CHF et 7 300 CHF par QALY gagnée. Toutefois, des études cliniques avant-après ne constituent pas une base fiable pour décider si une intervention est rentable ou non. Étant donné la quantité très limitée de données probantes médico-économiques, il est difficile de tirer des conclusions fermes sur le rapport coût-efficacité de la chirurgie arthroscopique chez les patients présentant des altérations dégénératives du genou. Bien que trois des quatre études disponibles aient révélé que l'arthroscopie était rentable, on ne peut considérer cela comme convaincant.

L'analyse d'impact budgétaire s'est concentrée sur les coûts de l'arthroscopie du genou des patients présentant des altérations dégénératives du genou principalement dues à une lésion méniscale, du point de vue du système d'assurance maladie en Suisse. Deux étapes ont été réalisées : en premier lieu, la fréquence annuelle des arthroscopies du genou en Suisse a été examinée ; en deuxième lieu, sur la base de cette fréquence annuelle, les coûts totaux annuels

ont été estimés. La fréquence chez les patients en milieu hospitalier a été étudiée à l'aide des codes DRG (*diagnosis-related groups*), des codes CIM-10 et des codes CHOP (Classification suisse des interventions chirurgicales) fournis par les statistiques hospitalières suisses pour les années 2010 à 2014. Deux stratégies analytiques ont été utilisées : dans la première, seuls les patients présentant à la fois un code DRG pertinent, au moins un diagnostic pertinent (code CIM-10) et au moins un traitement pertinent (code CHOP) ont été inclus. Dans la seconde stratégie, tous les patients présentant à la fois au moins un diagnostic pertinent (code CIM-10) et un traitement pertinent (code CHOP), indépendamment des codes DRG relevés, ont été inclus. Pour évaluer la fréquence des arthroscopies en ambulatoire, une analyse des codes TARMED (tarifs médicaux) pour 2013 et 2014 fournie par l'Office fédéral de la santé publique a été utilisée. Les coûts unitaires utilisés pour calculer l'impact budgétaire étaient basés sur les données issues de la statistique des données économiques par cas (*Statistik diagnosebezogener Fallkosten*) fournie par l'Office fédéral de la statistique et sur des estimations déjà publiées.

Les résultats de l'analyse de l'impact budgétaire ont suggéré que les dépenses totales liées aux lésions du genou/ménisque en Suisse étaient comprises entre 53,52 millions CHF et 71,93 millions CHF en 2013 et entre 52,30 millions CHF et 67,73 millions CHF en 2014. Les coûts liés aux patients ambulatoires représentaient entre 20 % et 28 % des coûts totaux. Les résultats de l'analyse découlant de notre seconde stratégie de sélection des patients, basée sur les codes CIM-10 et les codes CHOP, sont considérés comme plus réalistes que ceux découlant de la première stratégie qui prenait également en compte les codes DRG. Cette seconde stratégie suggérait un total des coûts liés aux patients hospitalisés de 58,10 millions CHF en 2010, de 55,87 millions CHF en 2011, de 58,44 millions CHF en 2012, de 57,20 millions CHF en 2013 et de 54,47 millions CHF en 2014. Le total des coûts liés aux patients ambulatoires était estimé à 14,73 millions CHF en 2013 et à 13,26 millions CHF en 2014. Les résultats de l'analyse de l'impact budgétaire découlant de la première stratégie de sélection des patients semblent assez incertains en raison des changements (utilisation d'un système « DRG tous patients » [*All Patients Diagnosis-Related Group*, APDRG] jusqu'en 2011 et du système SwissDRG à partir de 2012) et d'incohérences dans le système de codification des DRG.

Les résultats de l'analyse de l'impact budgétaire s'alignent sur ceux présentés dans une très récente publication de l'Observatoire suisse de la santé (Obsan). En supposant environ 14 000 méniscectomies par an et des coûts liés au patient hospitalisé de 4 889 CHF par cas, les auteurs du rapport Obsan ont estimé le total des coûts liés aux patients hospitalisés à 55,6 millions CHF en 2016.

1.3.5 Conclusion

En conclusion, rien ne prouve que les interventions arthroscopiques des genoux des patients présentant des altérations dégénératives du genou apportent un avantage sur les résultats mesurés dans les suivis à court ou moyen terme, à l'exception d'un léger effet sur la réduction de la douleur dans le suivi à court terme. Par conséquent, il demeure difficile de savoir si l'arthroscopie du genou a un effet sur les résultats évalués. On ne dispose pas de données sur le suivi à long terme. Les données sur les effets secondaires sont peu abondantes et aucune conclusion ne peut être tirée en ce qui concerne la balance avantages/inconvénients. La qualité générale des preuves dans les suivis à court et à moyen terme a été jugée très faible et faible respectivement. En outre, aucune preuve n'existe que le sous-groupe des patients présentant

uniquement un ménisque dégénératif du genou tire un bénéfice du traitement arthroscopique. Les résultats de cette évaluation sur l'efficacité clinique pourraient être applicables à une population plus large de patients souffrant de douleurs au genou en raison de troubles dégénératifs du genou. Ces résultats concordent avec des revues systématiques récemment publiées.

Étant donné la quantité très limitée de données probantes médico-économiques, il est difficile de tirer des conclusions fermes sur le rapport coût-efficacité de la chirurgie arthroscopique chez les patients présentant des altérations dégénératives du genou. Bien que trois des quatre études disponibles aient révélé que l'arthroscopie était rentable, on ne peut considérer cela comme convaincant, compte tenu des problèmes méthodologiques décrits, mais aussi des résultats de l'évaluation de l'efficacité.

2 Introduction

2.1 Background

In the context of the Health Technology Assessment (HTA) program of the Swiss federation, services reimbursed by the compulsory health insurance are being re-evaluated. One of the topics selected in 2015 is arthroscopy of the knee, based on the Swiss Health Observatory (Obsan) report No. 42 "*Variations géographiques dans les soins de santé. La situation en Suisse*", which reported an increase of inpatient knee arthroscopies and an increase of the rate of knee arthroscopies by 20% between 2005 and 2011.[1] According to the report, the rates of inpatient arthroscopies differed markedly between cantons, i.e. cantons at the lower 10th percentile showed a standardized rate of 121 arthroscopies per 100,000 inhabitants, while cantons at the upper 90th percentile showed a standardized rate of 415.6 per 100,000 inhabitants.[1] In addition, the proportion of inpatient and outpatient arthroscopies varied strongly between cantons (from 16% in the lowest decile to 75% in the highest), raising questions regarding the appropriate use and the benefit of knee arthroscopies.[1] A recent observational study by Muheim et al. reported that around 68% of Swiss patients who underwent arthroscopic partial meniscectomy or debridement and lavage were inpatients.[2] This might be due to the fact the strong financial incentives exist in Switzerland to provide knee arthroscopy as in hospital service.

During the scoping process conducted on behalf of the Swiss Federal Office of Public Health (SFOPH), first data on existing evidence syntheses regarding the clinical effectiveness and costs of knee arthroscopies were identified. In addition, Swiss data on the use of knee arthroscopy, across all indications, were analysed. Table 40 in Appendix 1 illustrates the main results, concerning the most frequent diagnoses according to the Swiss Hospital Statistics 2014.[3]

Overall, it was estimated that about 25,000 patients were hospitalized in 2014 with a diagnosis of knee or meniscus derangement or meniscus tear.[3] Depending on the costs and effects of the treatments, such a large number of patients may imply a high budget impact for the Swiss healthcare system.

2.2 Scoping process

Based on the discussion of the results with clinical experts and the review group (composed by SFOPH), SFOPH commissioned the Swiss Medical Board (SMB) with the assessment of the clinical effectiveness and cost-effectiveness (*Wirksamkeit und Wirtschaftlichkeit*) of arthroscopic interventions in patients with degenerative changes of the menisci and to update the March 2014 report of the German Institute for Quality and Efficiency in Health Care (IQWiG) on the arthroscopic treatment of osteoarthritis (OA) of the knee.[4]

Clinical effectiveness

Based on the discussion in the IQWiG report[4] and discussions with clinical experts it was apparent that symptoms that are due to degenerative changes of the meniscus and those that are due to OA of the knee cannot be clearly distinguished. The clinical effectiveness of arthroscopic interventions in the knee for OA has been recently assessed by the IQWiG report and hence the main interest of SFOPH regarded the question whether the evidence for these interventions in the

case of degenerative meniscal changes is any different. In order to take into account the difficulty of separating the two pathologies and the main interest of SFOPH, it was decided to first perform a pooled analysis for both populations. This had the additional advantage of allowing us to consider RCTs conducted in mixed populations.

During the scoping process, which involved a review process with clinical experts and a stakeholder consultation, the original questions by SFOPH were transformed into specific PICO research questions, specifying populations (P), interventions (I), comparators (C) and outcomes (O) of interest.

Given the above considerations, it was decided first to assess the clinical effectiveness and safety of knee arthroscopy in patients with degenerative changes of the knee in general (i.e. including patients with mainly meniscal symptoms, those with symptoms mainly due to OA of the knee, and those with mixed degenerative changes or covered in RCTs with mixed populations). Potential differences between populations were to be assessed in subgroup analyses. Only if relevant differences between the populations were to be found in the subgroup analyses, further analyses would be performed separately, with the addition of a separate second PICO.

During the scoping process, a third PICO was defined to assess the differences in effects depending on whether knee arthroscopy was performed in an inpatient or outpatient setting. For the purposes of this report, the third PICO addressing inpatient or outpatient setting was changed to 'PICO 2' as the original second PICO was dropped since we found no relevant subgroup effects.

The evaluation of the quality of the evidence regarding clinical effectiveness was done according to GRADE (Grading of Recommendations Assessment, Development and Evaluation).

Health economic characteristics

For the health economic analyses, a similar process was undertaken. It was decided to systematically review the available economic literature in order to investigate the impact of arthroscopic knee interventions in terms of cost-effectiveness, from a Swiss perspective. Moreover, it was decided to perform a budget impact analysis using Swiss epidemiological data and cost data.

2.3 Aim of the assessment report

According to SFOPH and SMB, the aim of this HTA is twofold: First, to provide an evidence base for discussion with different stakeholders, regarding the benefit, harm, and economic implications, of arthroscopic meniscectomy and other small arthroscopic interventions in patients with degenerative changes in the knee joint. Second, to provide an evidence base regarding the appropriate setting (i.e. inpatient versus outpatient) for therapeutic knee arthroscopy.

Specifically, this HTA report aims at assessing the clinical effectiveness and safety of

- therapeutic knee arthroscopy compared to any other treatments in patients with degenerative changes of the knee – irrespective of whether they are primarily due to meniscal damage, OA of the knee or a mix of both;
- inpatient compared to outpatient therapeutic knee arthroscopy.

Also, this HTA report aims at assessing

- the cost-effectiveness of therapeutic knee arthroscopy compared to any other treatment in patients with degenerative changes of the knee – irrespective of whether they are primarily due to meniscal damage, osteoarthritis of the knee or a mix of both;
- the cost-effectiveness of inpatient compared to outpatient therapeutic knee arthroscopy;
- the budget impact of knee arthroscopy in patients with degenerative changes of the knee primarily due to meniscal damage.

The focus on any degenerative changes of the knee in the clinical and cost-effectiveness parts *versus* on changes primarily due to meniscal damage in the budget impact part follows the priorities defined in the final scoping document.

3 Clinical effectiveness and safety

3.1 Methods

3.1.1 Literature search

This assessment aims at updating the report[4] completed by the IQWiG in 2014 on patients with OA of the knee. Therefore, the same search strategies as in the IQWiG report were used. Additional search terms were added to cover patients with degenerative changes of the menisci.

Relevant RCTs were identified by searching the following electronic databases:

- Medline via Ovid SP
- EMBASE via Ovid SP
- Cochrane Database of Systematic Reviews via Wiley Online Library
- Cochrane Central Register of Controlled Trials via Wiley Online Library
- Database of Abstracts of Reviews of Effects via Wiley Online Library
- Health Technology Assessment Database via Wiley Online Library

The databases were searched from inception until July 2017. The search strategy combined search terms for knee arthroscopy, osteoarthritis and degenerative disease with a search filter for RCTs, literature reviews and meta-analyses, as used in the 2014 IQWiG report.[4] Search terms related to 'meniscus' were added to the originally conducted IQWiG searches. See Appendix 2 for details of search strategies used.

Additionally, the reference lists of relevant systematic and narrative reviews, HTAs and guidelines were screened to identify further relevant RCTs.

Searches for on-going RCTs were conducted in the following clinical trial registries:

- U.S. National Institutes of Health. ClinicalTrials.gov [online]. URL: http://www.clinicaltrials.gov
- World Health Organization. International Clinical Trials Registry Platform Search Portal [online]. URL: http://apps.who.int/trialsearch

3.1.2 Screening of the literature

After removing duplicates, two reviewers independently screened the titles and abstracts of the identified records for potentially eligible RCTs, reviews, HTAs and guidelines. If one of the reviewers deemed a study to be potentially relevant, the study was included, and the corresponding full text article was screened. Subsequently, two reviewers independently screened the full text articles of the potentially eligible studies in order to identify eligible RCTs. Reference lists of relevant reviews, HTAs and guidelines were screened independently by two reviewers in order to identify potentially eligible RCTs. Discrepant screening results were discussed and resolved by consensus or by third party arbitration.

3.1.3 Inclusion and exclusion criteria - PICO 1

3.1.3.1 Population

Patients with symptoms due to degenerative changes of the knee – irrespective of whether they are primarily due to meniscal damage, OA of the knee or a mix of both – were included.

RCTs on patients with symptoms mainly due to OA of the knee were identified based on the criteria used in the IQWiG report, namely: The diagnosis of OA of the knee should have been made based on the diagnostic criteria of the American College of Rheumatology (ACR).[5] Alternatively, similar definitions of OA of the knee were used as long as they were based on the classic diagnostic criteria for arthrosis (knee pain, morning stiffness less than 30 minutes, crepitus on active motion, and/or osteophytes). In order to be included, at least 80% of the study population had to fit these criteria.

The definition of patients with symptoms due to meniscal degenerative changes was based on the definition given by the authors of eligible RCTs.

RCTs in patients where pre-operative symptoms or intraoperative findings indicative of serious, primarily non-arthrotic changes in the knee (e.g. (acute) traumatic meniscal tears, free joint bodies) dominated were excluded.[4]

RCTs of patients with traumatic knee injuries to other structures of the knee that tend to be associated with significant additional trauma to the meniscus or cartilage (e.g. tears of the cruciate ligaments) were excluded.

3.1.3.2 Intervention

Arthroscopy with any arthroscopic intervention that includes one or more procedures related to debridement, synovectomy, lavage or any intervention at the synovia, the cartilage of the joint or the menisci.

3.1.3.3 Comparators

Placebo, no treatment, conservative treatment or any other surgical treatment

a. Non-active treatments like:
    - No therapeutic intervention
    - Sham arthroscopy
    - Diagnostic arthroscopy
b. Active treatment like:
    - Lavage (without arthroscopy)
    - Non-surgical treatment (drug treatment, physiotherapy, or acupuncture)
    - Surgical treatment (open surgery or arthroscopic surgery not applied to cartilage, meniscus or synovia, but for example of ligaments or bone)

3.1.3.4 Outcomes

Health outcomes like mortality, morbidity or quality of life and safety outcomes like adverse events and serious adverse events were assessed.

The importance of outcomes was classified according to GRADE, which differentiates critical, important, and less important outcomes. The latter are considered not to be relevant for decision-making, and are therefore not covered in this report. Critical outcomes have a major impact on decision-making and the quality of the evidence available for these outcomes is the basis for judging the overall quality of the evidence for a clinical question.

Critical outcomes:

1. Pain
2. Function
3. Global assessment (e.g. WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index)[6] or KOOS (Knee injury and Osteoarthritis Outcome Score),[7] combined with joint stiffness, physical function)

Important outcomes:

1. Joint stiffness
2. Time to total knee replacement
3. Quality of life (health-related and disease-specific)
4. Adverse events (including subsequent surgeries)
5. Serious adverse events

Subjective outcomes (e.g. health-related quality of life) were only considered if they were assessed with validated measurement instruments.

The relevant time points were defined depending on the available evidence.

3.1.3.5 Study designs

RCTs and quasi-randomised trials; the effectiveness of knee arthroscopy was assessed based on RCTs as studies of this type tend to minimize bias compared to observational studies.[8]

3.1.4 Inclusion and exclusion criteria - PICO 2

3.1.4.1 Population

Patients with symptoms due to degenerative changes of the knee – irrespective of whether they are primarily due to meniscal damage, OA of the knee or a mix of both – were included (see section 3.1.3.1).

3.1.4.2 Intervention

Any therapeutic inpatient arthroscopic intervention was included.

The following arthroscopic interventions at the knee were relevant: debridement, synovectomy, or lavage, interventions at the synovia, the cartilage of the joint or the menisci.

3.1.4.3 Comparator

Any therapeutic outpatient arthroscopic intervention was included (see section 3.1.3.2). The intervention performed in an inpatient setting was compared to the same intervention being performed in an outpatient setting.

3.1.4.4 Outcomes

Health outcomes like mortality, morbidity or quality of life and safety outcomes like adverse events and serious adverse events were the same as defined in section 3.1.3.4 for PICO 1.

The relevant time points were defined depending on the available evidence.

3.1.4.5 Study designs

RCTs and quasi-randomised trials (see section 3.1.3.5).

3.1.5 Data extraction

Data on RCT characteristics and outcomes, by RCT arm where applicable, were extracted into a standardized form by one reviewer and checked by another. Discrepancies were resolved by discussion or third-party arbitration.

Extracted data on RCT characteristics included publication date, country, enrolment period, number of sites, follow-up visits, and descriptions of intervention, comparator, population and key inclusion criteria. Additionally, baseline characteristics of the RCT population were extracted for intervention and comparator groups.

Continuous outcome data were extracted as mean values for each intervention group at follow-up or, if not reported, as mean change from baseline. If more than one measurement instrument was used for the critical outcomes of pain, function and global assessment, a published hierarchy for outcomes was used.[9] In the case of (serious) adverse events, only the number of patients experiencing at least one adverse event were analysed and not the number of events. If only the number of events were reported, this information was extracted, but was not used for the analysis.

Outcome data were extracted for short-term follow-up time (closest to and including six months), an intermediate follow-up time (>6 months and <7 years) and a long-term follow-up time (longest follow-up time including and longer than seven years). Safety outcomes were only reported for the intermediate or long-term follow-up times.

3.1.6 Risk of bias assessment

One reviewer assessed the internal validity (risk of bias) of each RCT, which was checked by a second reviewer. Discrepancies were resolved by discussion or third-party arbitration.

To assess the risk of bias of individual RCTs the following criteria were used:[8 10-25]

- adequate random sequence generation
- adequate concealment of treatment allocation
- adequate blinding of patients, health carers, and outcome assessors
- completeness of outcome data
- reporting bias

Blinding of outcome assessors and completeness of outcome data were judged at the outcome level.

3.1.7 Assessment of the quality of evidence (GRADE)

The quality of the evidence was judged according to GRADE for the critical and important outcomes; i.e. at the outcome level, by considering all available RCTs for the respective outcome. The following criteria were considered to judge the quality of the evidence:[8 11-25]

Criteria for rating down the quality of evidence:

- risk of bias (internal validity), section 3.1.6
- inconsistency
- indirectness
- imprecision (see Judgment of imprecision below)
- publication bias

Criteria for rating up the quality of evidence:

- large magnitude of effect
- dose-response gradient
- all plausible confounders or biases increase the confidence in the estimated effect

Judgment of imprecision

The judgement of imprecision according to GRADE[12] was based on the total sample size for continuous outcomes or on the number of events for binary outcomes. If the total sample size or number of events was judged to be sufficient (optimal information size, OIS), the width of the confidence interval around the point estimate was examined. For continuous outcome measures, the 95% confidence interval (CI) was judged as precise as long as it excluded the line of no effect and a clinically relevant effect at the same time. In order to judge how clinically relevant pooled effect estimates of continuous measures might be for the patient, the concept of minimal clinically important differences (MCID) was used when data were pooled on the original scale. If different instruments were used, data were pooled using a standardized scale.

In case of continuous outcome measures, data were pooled on a standardized scale and an effect was judged to be precise if the 95% CI included 0.5 standard deviations (SDs) either in favour or against combination therapy and excluded the line of no effect.[26 27]

In case of binary data, effects were judged to be precise if the 95% CI included a relative risk increase of greater than 25% either in favour or against arthroscopy and excluded the line of no effect.[10 12]

Using the GRADE software (GRADEpro GDT, https://gdt.guidelinedevelopment.org), results were presented in a summary of findings table. The definition of the four levels of evidence according to GRADE is described in Table 1.[8]

Table 1 Quality levels of evidence

Quality level	Definition
High	We are very confident that the true effect lies close to that of the estimate of the effect
Moderate	We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low	Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
Very low	We have very little confidence in the effect estimate: the true effect is likely to be substantially different form the estimate of effect

3.1.8 Evidence synthesis

RCT characteristics and baseline characteristics were presented in tables and summarised descriptively.

Where possible, outcome data were summarised quantitatively in a meta-analysis by using inverse variance weighting models assuming random effects.[28] Outcome data were pooled for a short-term follow-up time (closest to and including six months), an intermediate follow-up time (longest follow-up time until seven years) and a long-term follow-up time (the longest follow-up time beyond seven years).

The analyses were performed using Review Manager (Version 5.3.5).

In case of multiple-arm RCTs, results of treatment arms were combined for the meta-analysis according to the Cochrane Handbook, where appropriate.[10]

If missing SDs could not be calculated based on other information given in the publication, or were not provided by RCT authors, they were approximated by estimating plausible SDs based on available SDs from the other RCTs reporting on the same or a similar outcome.[10] Estimated SDs were kept sufficiently large to be considered as "conservative assumptions", to avoid inappropriate precision leading to an overestimation of the overall effect estimate or to cause artificial heterogeneity.

Relative risks were calculated for binary outcomes and mean differences for continuous outcomes. Effect estimates (summary estimates and single ones for each RCT) with their corresponding 95% CI were presented in forest plots.

In case of continuous data, mean treatment group differences at follow-up were pooled with differences in mean changes from baseline in the same analysis if necessary, either on the original scale or by using standardised mean differences (SMDs).[10 29] SMDs were pooled if a continuous

outcome was measured on different scales because different instruments were used to assess the outcome. Results were standardised using the following formula:

$SMD = (\text{mean}_{\text{intervention}} - \text{mean}_{\text{comparator}})/SD_{\text{pooled}}$

In other words, the SMD expresses the size of the intervention effect by each trial relative to the variability in that trial.[10] Importantly, a SMD cannot be compared directly with the original scales.

The presence of heterogeneity in the pooled effect estimates was determined by using I^2 and visual examination. The interpretation of I^2 followed the guidance of the Cochrane Handbook for Interventions, where an I^2 of 0% to 40% indicates: there might be no important heterogeneity; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: substantial heterogeneity; 75% to 100%: considerable heterogeneity. The interpretation of the observed I^2 value depended on other measures for heterogeneity, namely the Tau^2 (where a value of Tau^2 of 0.04, 0.09, and 0.16 represented low, moderate and high heterogeneity, respectively), the precision of the individual effect estimates of the included RCTs and visual examination.[10 30]

3.1.9 Subgroup and sensitivity analyses

In case of substantial or considerable heterogeneity in the pooled effect estimates, methodological and clinical factors that might explain heterogeneity were explored in subgroup and sensitivity analyses. Additional sensitivity analyses could be added a posteriori.

The following subgroup analyses were pre-specified:

PICO 1 (therapeutic knee arthroscopy compared to other treatment)

- Patients with symptoms mainly due to meniscal degeneration versus patients with symptoms mainly due to OA of the knee versus mixed population
- Type of comparator (active versus non-active)
- Type of the arthroscopic intervention (e.g. debridement, partial meniscectomy)
- Gender
- Age
- Disease (severity, primary or secondary OA), e.g. severity of osteoarthritis at baseline

Classification of degenerative meniscus of the knee (DMK) versus osteoarthritis (OA)

To assess the effect differences within the population of patients with symptomatic degenerative changes of the knee joint, RCT populations were classified as having symptoms primarily due to degenerative meniscus of the knee (DMK only), osteoarthritis (OA only) or both DMK and OA (mixed). For a clear distinction between OA and DMK, clear information provided in the inclusion/exclusion criteria by the RCT authors was required, stating that DMK was an exclusion criterion for OA or vice versa. A fourth classification of 'mixed unclear' was introduced for RCTs which did not clearly report either or both DMK or OA. Because differentiation between DMK and OA is difficult, classification was based on any information provided by RCT authors (radiologic and clinical evidence, such as MRI; Kellgren-Lawrence and Ahlbäck classifications (see also

Table 2)[31 32]; baseline characteristics; inclusion and exclusion criteria; etc.) and was decided by two reviewers reaching a consensus. The resulting subgroups were compared to those used in existing systematic reviews to ensure RCTs were classified in a similar manner.[33-39]

Table 2 Radiographic grading scales - Kellgren-Lawrence and Ahlbäck classification

Scale[31 32]	**Grade and characteristics**				
Kellgren-Lawrence	0 No joint space narrowing or reactive changes	1 Doubtful joint space narrowing, possible osteophytic lipping	2 Definite osteophytes, possible joint space narrowing	3 Moderate osteophytes, definite joint space narrowing, some sclerosis, possible bone-end deformity	4 Large osteophytes, marked joint space narrowing, severe sclerosis, definite bone ends deformity
Ahlbäck	0 Normal	1 Joint space narrowing is < 3 mm (with or without subchondral sclerosis)	2 Obliteration of joint space	3 Bone defect/loss <5 mm	4 Bone defect and/or loss 5–10 mm

Type of comparator (non-active versus active)

Type of comparator was used to classify RCTs into two subgroups: non-active comparator or active comparator. Non-active comparator included no therapeutic treatment (when compared to intervention), sham or placebo arthroscopy, or diagnostic arthroscopy. Active treatment included lavage without arthroscopy (i.e. closed-needle lavage), non-surgical treatment (drug treatment, physiotherapy, acupuncture, etc.) or surgical treatment (open surgical or arthroscopic).

PICO 2 (inpatient compared to outpatient therapeutic knee arthroscopy)

- DMK versus OA (only if subgroup in PICO 1 showed significant subgroup differences)
- Type of the arthroscopic intervention (debridement, partial meniscectomy)
- Gender
- Age
- Disease (severity, primary or secondary OA), e.g. severity of osteoarthritis at baseline

3.2 Results

3.2.1 Results of the literature search

The electronic literature search yielded 3,657 records (Figure 1). After removing duplicates, 2,850 records were screened at title and abstract level, and 132 potentially relevant studies and 69 relevant reviews, guidelines and/or HTAs were identified. One additional potentially relevant

review was identified from the 2014 IQWiG report. In total 132 potentially relevant full texts were screened for being relevant RCTs; in addition, 70 reviews, guidelines and HTAs were screened to identify relevant RCTs. From the 202 full text articles screened, 21 RCTs were identified with one additional relevant RCT (Biedert 2000) found by screening relevant reviews, guidelines and HTAs. Hence, a total of 22 RCTs (reported in 26 research articles and 18 published abstracts, protocols, comments) were included (Figure 1). There were 21 RCTs (over 2048 patients, in two RCTs the number of patients was not reported) identified for PICO 1 and one RCT (100 patients) identified for PICO 2. Details regarding the search strategy and the number of RCTs and publications included in this report are documented in Appendix 2. The RCT selection process is presented in Figure 1.

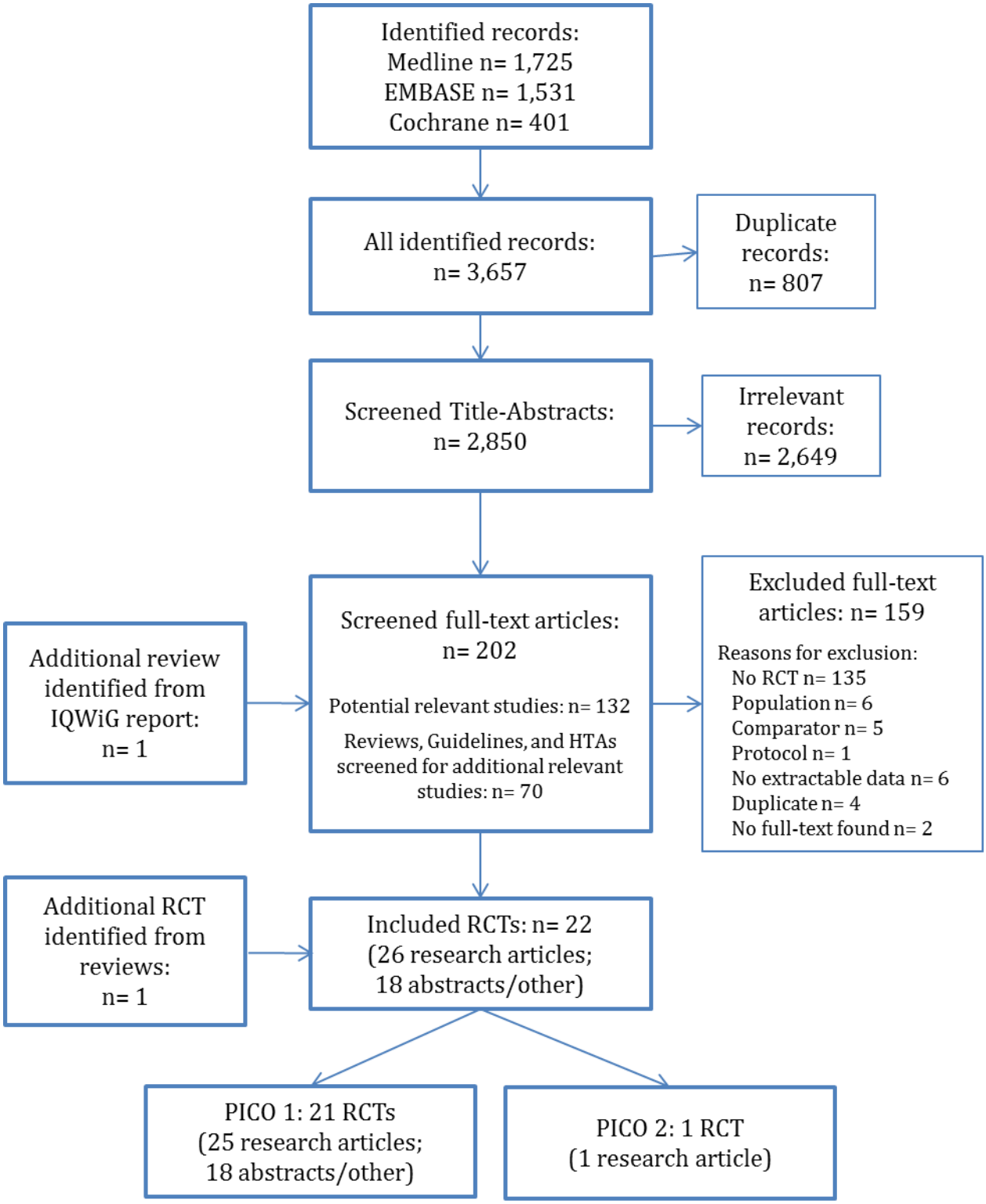


Figure 1 Results of literature search

3.2.2 Overview of included RCTs

Forty-three records relevant for PICO 1 (25 research articles and 18 abstracts, protocols and comments) and one record relevant for PICO 2 (1 research article) were identified, encompassing 21 relevant RCTs for PICO 1 and one RCT for PICO 2, respectively. References for all 26 research articles can be found in Table 3; a complete list of all references is provided in Appendix 3.

An overview of time points of the outcomes analysed from each RCT is given in Table 4. Critical and important outcomes were extracted for short-term follow-up (≤6 months) and for intermediate follow-up (>6 months). The longest follow-up reported was five years; hence, no RCT with long-term follow-up data was identified. RCT characteristics and baseline characteristics are presented for all RCTs jointly, while analyses are presented by PICO.

Table 3 Identification (ID) of included RCTs and corresponding publications

RCT ID	Reference (only relevant publications)
Biedert 2000[40]	Biedert RM. Treatment of intrasubstance meniscal lesions: a randomized prospective study of four different methods. Knee Surg Sports Traumatol Arthrosc 2000;8(2):104-8.
Change 1993[41]	Chang RW, Falconer J, Stulberg SD, et al. A randomized, controlled trial of arthroscopic surgery versus closed-needle joint lavage for patients with osteoarthritis of the knee. Arthritis Rheum 1993;36(3):289-96.
FIDELITY[42]	Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. N Engl J Med 2013;369(26):2515-24.
Forster 2003[43]	Forster MC, Straw R. A prospective randomised trial comparing intra-articular Hyalgan injection and arthroscopic washout for knee osteoarthritis. Knee 2003;10(3):291-3.
Gauffin 2014[44,45]	Gauffin H, Tagesson S, Meunier A, et al. Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: a prospective, randomised, single-blinded study. Osteoarthritis Cartilage 2014;22(11):1808-16. Gauffin H, Sonesson S, Meunier A, et al. Knee Arthroscopic Surgery in Middle-Aged Patients With Meniscal Symptoms: A 3-Year Follow-up of a Prospective, Randomized Study. Am J Sports Med 2017.
Hamberg 1984[46]	Hamberg P, Gillquist J, Lysholm J. A comparison between arthroscopic meniscectomy and modified open meniscectomy. A prospective randomised study with emphasis on postoperative rehabilitation. J Bone Joint Surg Br 1984;66(2):189-92.

RCT ID	Reference (only relevant publications)
Herrlin 2007[47 48]	Herrlin S, Hallander M, Wange P, et al. Arthroscopic or conservative treatment of degenerative medial meniscal tears: a prospective randomised trial. Knee Surg Sports Traumatol Arthrosc 2007;15(4):393-401. Herrlin SV, Wange PO, Lapidus G, et al. Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. Knee Surg Sports Traumatol Arthrosc 2013;21(2):358-64.
Kalunian 2000[49]	Kalunian KC, Moreland LW, Klashman DJ, et al. Visually-guided irrigation in patients with early knee osteoarthritis: a multicenter randomized, controlled trial. Osteoarthritis Cartilage 2000;8(6):412-8.
Kang 2005[50]	Kang JG, Wang ML, Zhang XN. Treatment of knee osteoarthritis with arthroscopic debridement and intra-articular sodium hyaluronate injection. [Chinese]. Journal of Jilin University Medicine Edition 2005;31(5):802-05.
Kirkley 2008[51]	Kirkley A, Birmingham TB, Litchfield RB, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee.[Erratum appears in N Engl J Med. 2009 Nov 12;361(20):2004]. N Engl J Med 2008;359(11):1097-107.
Kise 2016[52 53]	Kise NJ, Risberg MA, Stensrud S, et al. Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. Bmj 2016;354:i3740. Stensrud S, Risberg MA, Roos EM. Effect of exercise therapy compared with arthroscopic surgery on knee muscle strength and functional performance in middle-aged patients with degenerative meniscus tears: a 3-mo follow-up of a randomized controlled trial. Am J Phys Med Rehabil 2015;94(6):460-73.
KIVIS[54]	Arden NK, Reading IC, Jordan KM, et al. A randomised controlled trial of tidal irrigation vs corticosteroid injection in knee osteoarthritis: the KIVIS Study. Osteoarthritis Cartilage 2008;16(6):733-9.
KORAL[55]	Campbell MK, Skea ZC, Sutherland AG, et al. Effectiveness and cost-effectiveness of arthroscopic lavage in the treatment of osteoarthritis of the knee: a mixed methods study of the feasibility of conducting a surgical placebo-controlled trial (the KORAL study). Health Technol Assess 2010;14(5):1-180.
Merchan 1993[56]	Merchan EC, Galindo E. Arthroscope-guided surgery versus nonoperative treatment for limited degenerative osteoarthritis of the femorotibial joint in patients over 50 years of age: a prospective comparative study. Arthroscopy 1993;9(6):663-7.

RCT ID	Reference (only relevant publications)
MeTeOR[57 58]	Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis.[Erratum appears in N Engl J Med. 2013 Aug 15;369(7):683]. N Engl J Med 2013;368(18):1675-84. Katz JN, Wright J, Spindler KP, et al. Predictors and Outcomes of Crossover to Surgery from Physical Therapy for Meniscal Tear and Osteoarthritis: A Randomized Trial Comparing Physical Therapy and Surgery. J Bone Joint Surg Am 2016;98(22):1890-96.
Moseley 1996[59]	Moseley JB, Jr., Wray NP, Kuykendall D, et al. Arthroscopic treatment of osteoarthritis of the knee: a prospective, randomized, placebo-controlled trial. Results of a pilot study. Am J Sports Med 1996;24(1):28-34.
Moseley 2002[60]	Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee.[Summary for patients in J Fam Pract. 2002 Oct;51(10):813; PMID: 12401143]. N Engl J Med 2002;347(2):81-8.
Østerås 2012[61]	Østerås H, Østerås B, Torstensen TA. Medical exercise therapy, and not arthroscopic surgery, resulted in decreased depression and anxiety in patients with degenerative meniscus injury. J Bodywork Mov Ther 2012;16(4):456-63.
Saeed 2015[62]	Saeed K, Khan SA, Ahmed I. Efficacy of intra articular hyaluronic acid versus arthroscopic debridement in terms of improvement in pain score in Kellgran -Lawrence Grading II & III osteoarthritis of knee joint. Pakistan Journal of Medical and Health Sciences 2015;9(3):1011-15.
Vermesan 2013[63]	Vermesan D, Prejbeanu R, Laitin S, et al. Arthroscopic debridement compared to intra-articular steroids in treating degenerative medial meniscal tears. Eur Rev Med Pharmacol Sci 2013;17(23):3192-6.
Weale 1998[64]	Weale AE, Ackroyd CE, Mani GV, et al. Day-case or short-stay admission for arthroscopic knee surgery: a randomised controlled trial. Ann R Coll Surg Engl 1998;80(2):146-9.
Yim 2013[65]	Yim JH, Seon JK, Song EK, et al. A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. Am J Sports Med 2013;41(7):1565-70.

Table 4 Overview of outcomes in identified RCTs and follow-up time points in months

Outcome / RCT ID	Pain	Function	Global assessment	Joint stiffness	Total knee replacement	Quality of life	Adverse events*	Serious adverse events
PICO 1								
Biedert 2000			[>12]				>12	
Chang 1993	3, 12	3, 12	3, 12				*[3-12]*	
FIDELITY	6, 12		6, 12			12		12
Forster 2003	6, 12	6, 12	6, 12		12		*12*	
Gauffin 2014	3, 36	3, 36				3, 36	*3,* 36	36
Hamberg 1984			2					
Herrlin 2007	6, 60	6, 60	6, 60			6, 60	*60*	
Kalunian 2000	3, 12	3, 12	3, 12	3, 12				
Kang 2005			12					
Kirkley 2008	6, 24	6, 24	6, 24	6, 24		6, 24		
Kise 2016	[3, 24]	[3, 24]				[3, 24]	*>6,* 24	24
KIVIS	6	6	6**	6			*0.5*	
KORAL	2							
Merchan 1993			36				*EOS*	36
MeTeOR	6, 12	6, 12			12		*6,* 12, *24*	12
Moseley 1996	[6]							
Moseley 2002	6, 12	6, 12						
Østerås 2012	3		3					
Saeed 2015	[6]						*<5 days*	*EOS*
Vermesan 2013			1,12				*[1]*	
Yim 2013	6, 24		6, 24					
PICO 2								
Weale 1998	[<1]							[EOS]

The numbers in the fields denote the analysed follow-up period in months. Reported, but not suitable for pooling, outcomes are presented in square brackets.
*italics indicate other events (re-operations, cross-overs, surgical procedures, side effects, complications, etc.). **only binary outcome reported
Abbreviations: EOS, End of Study

3.2.3 Characteristics of the included RCTs

PICO 1

Twenty-one RCTs were identified comparing arthroscopic intervention to a comparator treatment in symptomatic patients with degenerative knee disease. A summary of the RCT characteristics and selected baseline characteristics can be found in Table 5 and Table 6. Based on the chosen criteria to define the RCT populations, RCTs for sensitivity analyses were classified as follows: For five RCTs (FIDELITY, Gauffin 2014, Herrlin 2007, Kise 2016, Yim 2013) the RCT populations were classified as having only DMK and no or mild OA (DMK only). In one RCT (Kirkley 2008) the RCT population was classified as having only OA and no DMK (OA only). In one RCT (MeTeOR) the RCT population was classified as 'mixed population' composed of patients having both OA and DMK. For the remaining 14 RCTs (Biedert 2000, Chang 1993, Forster 2003, Hamberg 1984, Kalunian 2000, Kang 2005, KIVIS, KORAL, Merchan 1993, Moseley 1996, Moseley 2002, Østerås 2012, Saeed 2015, Vermesan 2013), RCT reports did not allow for a classification of the RCT populations; therefore, the RCT populations of these RCTs were classified as 'mixed unclear'.

The majority of RCTs were conducted in Europe: three (Gauffin 2014, Hamberg 1984, Herrlin 2007) in Sweden, three (Forster 2003, KIVIS, KORAL) in the UK, two (Kise 2016, Østerås 2012) in Norway, and one each (Biedert 2000, FIDELITY, Merchan 1993) in Switzerland, Finland and Spain. Six RCTs were conducted in North America: five (Chang 1993, Kalunian 2000, MeTeOR, Moseley 1996, Moseley 2002) in the US, and one (Kirkley 2008) in Canada. Three Asian RCTs (Kang 2005, Saeed 2015, Yim 2013) were conducted in China, Pakistan and South Korea, respectively. For the remaining RCT (Vermesan 2013) the country of origin could not be identified. Nine RCTs were conducted at multiple centres, nine were conducted at a single centre and in the remaining three, the setting was not reported. Two RCTs (Kang 2005, Vermesan 2013) randomised knees of patients, while the remaining 19 RCTs randomised patients to up to four treatment arms. Patient enrolment ranged from 10 participants to 351 patients, with length of follow-up ranging from two months to 60 months. Twelve RCTs (Biedert 2000, Chang 1993, Forster 2003, Hamberg 1984, Kang 2005, Kise 2016, KIVIS, Merchan 1993, Østerås 2012, Saeed 2015, Vermesan 2013, Yim 2013) used active comparators consisting of lavage (Chang 1993), exercise therapy (Østerås 2012, Kise 2016, Yim 2013), pharmaceutical therapy (Biedert 2000, Forster 2003, Kang 2005, KIVIS, Merchan 1993, Saeed 2015, Vermesan 2013) or other arthroscopic or surgical procedure (Hamberg 1984). Nine RCTs (FIDELITY, Gauffin 2014, Herrlin 2007, Kalunian 2000, Kirkley 2008, KORAL, MeTeOR, Moseley 1996, Moseley 2002) used non-active comparators, i.e. no therapy (Gauffin 2014, Herrlin 2013, Kirkley 2008 MeTeOR), diagnostic arthroscopy (Kalunian 2000) or placebo/sham surgery (FIDELITY, KORAL, Moseley 1996, Moseley 2002). Three RCTs reported blinding of patients and investigators (FIDELITY, Kalunian 2000, Moseley 1996, Moseley 2002), while in the remaining 18 RCTs either participants or investigators were not blinded or RCTs reports were unclear.

PICO 2

One RCT (Weale 1998) was identified comparing inpatient arthroscopic intervention to outpatient arthroscopic intervention in patients with degenerative disease of the knee. A summary of the RCT characteristics and select baseline characteristics can be found in Table 5 and Table 6. The RCT took place in the UK at one site and had a follow-up period of 3-4 weeks. Of the

100 patients enrolled, 50 were randomised to overnight stay admission (inpatient), while 50 were randomised to day-case admission. Participants and investigators were not blinded.

Table 5 General RCT characteristics

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Biedert 2000 CH (Quasi-RCT)	Apr. 1994 – Aug. 1996 n.r., probably 1 site FU time ranges from 12 – 38 months (mean 26.5 months)	Intrasubstance meniscal lesions OA: Unclear DMK: Yes Mixed: Unclear	Arthroscopic suture repair with access channels (10) Arthroscopic minimal central resection, fibrinclot, suture repair (7) Arthroscopic partial meniscectomy (11)	Anti-inflammatory medication and local physical therapy (ultrasound) (12)	Isolated and painful medial intrasubstance horizontal grade 2 meniscal lesion Clinical symptoms of a meniscal tear MRI linear high grade 2 signal intensity in the medial meniscus
Chang 1993 US	n.r. Multicentre (2 sites) 3 and 12 months	Non-end stage OA OA: Yes, KL I-III DMK: Unclear Mixed: Unclear	Arthroscopic surgery with debridement and lavage (19)	Closed-needle joint lavage (15)	Knee pain for >3 months KL I-III

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
FIDELITY FI	Dec. 2007 – Jan. 2013* Multicentre (5 sites) 2, 6 and 12 months	Degenerative medial meniscus tear with no knee OA OA: No, KL ≤I DMK: Yes, MRI confirm medial meniscus tear Mixed: No	Arthroscopic partial meniscectomy (70)	Sham surgery (76)	Age 35 – 65 years Pain on the medial join line of knee for >3 months MRI and arthroscopically-verified degenerative medial meniscus tear No OA (KL >I)
Forster 2003 UK	n.r. n.r. 6 weeks, 3, 6 and 12 months	Patients with symptomatic knee OA without mechanical symptoms OA: Yes, not further specified DMK: Unclear Mixed: Unclear	Arthroscopy + lavage (19)	Intra articular hyaluronic acid (19)	Symptomatic knee OA with radiographic evidence of some remaining joint space

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Gauffin 2014 SE	Mar. 2010 – Apr. 2012 1 site 3, 12 and 36 months	Middle-aged patients with suspected meniscal symptoms OA: No, >80% of population had KL ≤I DMK: Yes Mixed: No	Arthroscopic partial meniscectomy + exercise programme (75)	Exercise programme (75)	Age 45 – 64 years Symptom duration for >3 months X-ray with Ahlbäck 0 Undergone prior physiotherapy
Hamberg 1984 SE	n.r. 1 site (outpatient) 1, 4 and 8 weeks	Patients with degenerative tears of the medial meniscus OA: Unclear DMK: Yes, not further specified Mixed: Unclear	Arthroscopic partial meniscectomy (10) Arthroscopic total meniscectomy (10)	Open partial meniscectomy (10) Open total meniscectomy (10)	Clinical diagnosis of degenerative meniscal tear Symptoms duration for ≥3 months Weight-bearing radiographs were taken of both knees before operation

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Herrlin 2007 SE	Jun. 2003 – Apr. 2005 1 site 2, 6, 24 and 60 months	Patients suspected and symptomatic non-traumatic, degenerative medial meniscal tears OA: No, author statement: "no or minimal OA", OA ≤I according to the Ahlbäck classification DMK: Yes, MRI confirm medial meniscal tear (41/47 arthroscopically confirmed) Mixed: No	Arthroscopic partial meniscectomy + exercise therapy (47)**	Exercise Therapy (50)**	Age 45 – 64 years Daily medial knee pain for last 2-6 months MRI showing medial meniscus tear Radiographic examination showing OA ≤I according to the Ahlbäck classification
Kalunian 2000 USA	n.r. Multicentre (4 sites) 1, 3 and 12 months	Patients with early knee OA OA: Probably yes, KL ≤II DMK: Unclear, n.r. Mixed: Unclear	Arthroscopy + Lavage (41)***	Diagnostic Arthroscopy (49)***	Age >40 years Knee pain for ≤10 years Normal or minimally abnormal radiographs (KL ≤II)

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Kang 2005 China	Jan. 2001 – Mar. 2003* 2 sites 12 months	Patients with mild and medium knee OA OA: Yes, KL ≤III: Class I: 25.8% (23/89 knees), Class II: 39.3% (35/89), Class III: 34.8% (31/89) DMK: Unclear, n.r. Mixed: Unclear	Arthroscopy + Debridement (32 patients; 41 knees)	Intra-articular sodium hyaluronic injections (2 mL of sodium hyaluronate five times in one week intervals) (37 patients; 48 knees)	Age 42-67 years Diagnosis of knee OA as defined by the ACR Radiographic examination: KL ≤III
Kirkley 2008 CA	Jan. 1999 – Aug. 2007* 1 site 3, 6, 12, 18 and 24 months	Patients with moderate-to-severe OA of the knee OA: Yes, KL ≥II DMK: No, exclusion of patient with large meniscal tears (at clinical detection or MRI) Mixed: No	Arthroscopy + lavage + debridement + optimized physical and medical therapy (94)	Optimized physical and medical therapy (94)	Age ≥18 years Idiopathic or secondary OA of the knee with grade II, III or IV radiographic severity by the modified KL classification No large meniscal tears

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Kise 2016 NO	Oct. 2009 – Sept. 2012 Multicentre (2 sites) 3, 12 and 24 months	Middle-aged patients with degenerative meniscus tears OA: No, KL ≤II, but 95.7% were KL ≤I DMK: Yes, MRI confirmed Mixed: No	Arthroscopic partial meniscectomy (70)	Exercise therapy (70)	Age 35-60 years Unilateral knee pain for >2 months MRI confirmed tear in the medial meniscus KL OA ≤II
KIVIS UK	n.r. Multicentre (2 sites) 2, 4, 12 and 26 weeks	Patients with knee OA OA: Yes, KL 0-I 13%, KL II 68%, KL III-IV 16% and unclear 3% DMK: Unclear, n.r. Mixed: Unclear	Arthroscopy + lavage (tidal irrigation) (71)	Intra-articular corticosteroids injections (40 mg triamcinolone acetonide and 2 ml 1% lignocaine) (79)	Aged 40 – 90 years Clinical diagnosis and radiographic evidence Knee pain for most days of prior months

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
KORAL UK	Jul. 2007 – Nov. 2007 multicentre (2 sites) 2, 6, 12 and 24 months	Patients with knee OA who might be considered for arthroscopic lavage OA: Yes DMK: Unclear Mixed: Unclear	Arthroscopy + Lavage (n.r.)**	Placebo surgery (n.r.)** Non-operative treatment (planned to receive pharmacological, physiotherapy, intra-articular injection, but not received) (3)	Age ≥18 years Radiographic evidence of OA of the knee Fit for general anaesthetic
Merchan 1993 ES	Jan. 1988 – Dec. 1990* 1 site 12, 24 and 36 months	Patients with painful limited degenerative OA of the femorotibial joint OA: Yes, not further specified DMK: Unclear, but in 88.6% of the intervention group meniscal tears were reported Mixed: Unclear	Arthroscopic partial meniscectomy, limited debridement of loose articular cartilage and removal of loose bodies (40)	NSAIDs and decrease in activity (40)	Age >50 years Pain for ≤6 months Radiographic evidence of limited degenerative process (minimal joint space narrowing and formation of small osteophytes) according to Ahlbäck

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
MeTeOR USA	June 2008 – Aug. 2011 Multicentre (7 sites) 3, 6, 12, 18 and 24 months (extended follow-up assessment up to 60 months)	Symptomatic patients with a meniscal tear in the setting of mild to moderate OA OA: Yes, KL ≤III (>46% had KL Grade >I) DMK: Yes, MRI confirm medial meniscal tear Mixed: Yes	Arthroscopic partial meniscectomy (arthroscopy + debridement) + physical therapy (174)	Physical therapy (177)	Age ≥45 years Symptoms for ≥4 weeks managed with medication, activity limitations or physical therapy MRI/radiographic evidence of osteophytes or joint space narrowing No KL IV
Moseley 1996 USA	June 1992 1 site 10 days, 6 weeks, 3 and 6 months	Patients with symptomatic OA of the knee OA: Yes DMK: Unclear Mixed: Unclear	Arthroscopy + Lavage + Debridement (2) Arthroscopy + Lavage (3)	Placebo arthroscopy (5)	Age <70 years Symptomatic OA of the knee in spite of ≥6 months of non-operative treatment Moderate knee pain (≥4 on a 0 to 10 scale over a week)

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Moseley 2002 USA	Oct. 1995 – Sept. 1998 1 site 2 and 6 weeks, 3, 6, 12, 18 and 24 months	Patients with OA of the knee OA: Yes DMK: Unclear Mixed: Unclear	Arthroscopy + lavage + debridement (59) Arthroscopy + lavage (61)	Placebo arthroscopy (60)	Age ≤75 years OA of the knee as defined by the ACR Moderate knee paint (≥4 on VAS from 0-10) despite medical treatment for at least 6 months OA severity grade ≥9 based on radiographic assessment (scale 0-12)
Østerås 2012 NO	n.r. Multicentre (2 sites) 3 months	Middle-aged patients with a non-traumatic meniscus tear OA: Unclear, KL ≤II, not further specified DMK: Yes, MRI showing degenerative meniscus tear Mixed: Unclear	Arthroscopic partial meniscectomy (8)	High-dosage medical exercise therapy (MET) (9)	Age 35-60 years Knee pain >3 months MRI showing degenerative meniscus tear KL ≤II Able to physical activity and exercise

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Saeed 2015 PK	Jan. 2012 – Dec. 2014* 1 site 1, 3 and 6 months	Patients with OA of the knee joint OA: Yes, KL II-III DMK: Unclear Mixed: Unclear	Arthroscopic debridement (60)***	Intra articular hyaluronic acid (weekly, five weeks) (60)***	Age ≥40 years History of pain knee joint KL II-III
Vermesan 2013 n.r.	n.r. n.r. 1 month and 1 year	Symptomatic knees with degenerative lesions of the medial compartment (cartilage and meniscus) OA: Unclear, “early stage medial compartment knee osteoarthritis”, not further specified DMK: Yes, not further specified Mixed: Unclear	Arthroscopy + debridement (n patients: n.r., 60 knees)**	Intra-articular steroid injection (1 ml of betamethasone in 4 ml of lidocaine 1%) (n patients: n.r., 60 knees)**	Non-traumatic symptomatic knees MRI confirmed degenerative lesions of the medial compartment (cartilage and meniscus)

RCT ID **Country**	**Enrolment period** **Setting** **Follow-up visits**	**Population** (Patients with …)	**Intervention (n randomised)**	**Comparator (n randomised)**	**Key Inclusion Criteria**
Weale 1998 UK	n.r. 1 site 3-4 weeks	Patients scheduled for unilateral arthroscopic surgery of the knee	Day-case admission for arthroscopy (50)	Overnight stay admission for arthroscopy (50)	Eligible for unilateral arthroscopic surgery No children and no patients age >65 years
Yim 2013 KR	Jan. 2007 – Jul. 2009 1 site 3, 6, 12 and 24 months	Patients with a degenerative horizontal tear of the posterior horn of the medial meniscus OA: No, KL ≤I DMK: Yes, MRI confirmed horizontal tear Mixed: No	Arthroscopic (partial) meniscectomy (54)	Supervised 3 week rehabilitation program, 8 weeks home exercise, and receiving analgesics, NSAIDs, or muscle relaxants (54)	MRI confirmed degenerative horizontal tear of the posterior horn of the medial meniscus Daily knee pain on medial rise despite management at a primary clinic in the previous month KL <II

*RCT period; **unclear number randomised; ***total number of randomised patients not reported

Abbreviations: ACR, American College of Rheumatology; CH, Switzerland; DMK, degenerative meniscal damage of the knee; ES, Spain; FI, Finland; KL, Kellgren-Lawrence; KR, South Korea; MRI, magnetic resonance imaging; NSAID, nonsteroidal anti-inflammatory drug; NO, Norway; n.r., not reported; OA, osteoarthritis; PK, Pakistan; SE, Sweden; UK, United Kingdom; USA, United States of America; VAS, visual analogue scale.

Table 6 Baseline characteristics of included RCTs

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Biedert 2000	28 randomised to three arthroscopic procedures Female: 19 (47.5%) across all groups Age: 30.4 years (range 16-50) across all groups Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: n.r.	12 randomised Female: 19 (47.5%) across all groups Age: 30.4 years (range 16-50) across all groups Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: n.r.
Chang 1993*	18 analysed (19 randomised) Female: 13 (72% of patients analysed) Age: 61 ± 11 years of patients analysed Knee pain since: 51 ± 51 months Pain: 6.5 ± 2.0 (AIMS) Function: 2.3 ± 1.6 (AIMS) Global Assessment: 4.6 ± 2.6 (VAS)	14 analysed (15 randomised) Female: 10 (71% of patients analysed) Age: 65 ± 13 years of patients analysed Knee pain since: 53 ± 57 months Pain: 6.1 ± 2.1 (AIMS) Function: 1.7 ± 1.0 (AIMS) Global Assessment: 4.6 ± 2.5 (VAS)
FIDELITY	70 randomised Female: 28 (40.0%) Age: 52 ± 7 years Symptom duration: n.r. Pain: 5.8 ± 2.0 (after exercise on NRS) Function: n.r. Global Assessment: 60.2 ± 14.7 (Lysholm)	76 randomised Female: 29 (38.2%) Age: 52 ± 7 years Symptom duration: n.r. Pain: 6.1 ± 2.0 (after exercise on NRS) Function: n.r. Global Assessment: 60.1 ± 14.6 (Lysholm)

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Forster 2003	19 randomised Female: n.r. Age: 63 years Symptom duration: n.r. Pain: 7.5 (VAS) Function: 45 (KF) Global Assessment: 13 (LI)	19 randomised Female: n.r. Age: 60 years Symptom duration: n.r. Pain: 7.6 (VAS) Function: 65 (KF) Global Assessment: 10.5 (LI)
Gauffin 2014	75 randomised Female: 22 (29.3%) Age: 54 ± 5 years Symptom duration: n.r. Pain: 55 (95% CI: 51-59) (KOOS, n=74) Function: 65 (95% CI: 61-69) (KOOS) Global Assessment: n.r.	75 randomised Female: 19 (25.3) Age: 54 ± 6 years Symptom duration: n.r. Pain: 58 (95% CI: 54-62) (KOOS, n=74) Function: 68 (95% CI: 63-73) (KOOS) Global Assessment: n.r.

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Hamberg 1984	10 randomised (only arthroscopic partial meniscectomy) Female: 1 (10.0%) Age: 46.9 years (range 34-60) Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: 64.6 ± 9.1 (Lysholm) And 10 randomised (only arthroscopic total meniscectomy) Female: 1 (10.0%) Age: 46.0 years (range 36-55) Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: 53.9 ± 18.6 (Lysholm)	10 randomised (open partial meniscectomy) Female: 1 (10.0%) Age: 46.3 years (range 34-56) Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: 60.4 ± 9.1 (Lysholm) And 10 randomised (only open total meniscectomy) Female: 1 (10.0%) Age: 52.2 years (range 37-65) Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: 59.8 ± 20.2 (Lysholm)
Herrlin 2007	47 randomised (Initial randomisation n=99, group assignment unclear) Female: 19 (40.4%) Age: 54 ± 5 Symptom duration: n.r. Pain: median 56 (IQR 44-67)(KOOS) Function: median 68 (IQR 54-81) (KOOS) Global Assessment: 61 (IQR 49-70) (Lysholm)	50 randomised (Initial randomisation n=99, assignment unclear), but reported baseline for 49: Female: 19 (38.8%) Age: 56 ± 5.8 Symptom duration: n.r. Pain: median 54 (IQR 50-78) (KOOS) Function: median 76 (IQR 54-87) (KOOS) Global Assessment: 70 (IQR 56-82) (Lysholm)

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Kalunian 2000	41 randomised (total n randomised n.r.) Female: 22 (53.7%) Age: 60.9 years (range 41-88) Symptom duration: 30.0 months (range 2-120) Pain: 9.42 (95% CI 8.79-10.05) Function: 27.5 (95% CI 24.8-31.0) (WOMAC) Global Assessment: 41.09 (range 1-75) (WOMAC)	49 randomised (total n randomised n.r.) Female: 26 (53.1%) Age: 58.3 years (range 40-85) Symptom duration: 34.4 months (range 2-120) Pain: 8.88 (95% CI 8.40-9.36) Function: 28.0 (95% CI 25.5-30.5) (WOMAC) Global Assessment: 40.67 (range 8-86) (WOMAC)
Kang 2005	32 randomised (41 knees) Female: 17 (53.1%) | 24 knees (58.5%) Age: n.r. (range 42-65 years) Duration of condition: 5.5 years (range 0.5-21) Pain: n.r. Function: n.r. Global Assessment: 46.0 (SE 1.9) (Lysholm; n=41 knees)	37 randomised (48 knees) Female: 21 (56.8%) | 29 knees (60.4%) Age: 53 years (range 45-67) Duration of condition: 6.3 years (range 1.5-26) Pain: n.r. Function: n.r. Global Assessment: 46.4 (SE 1.9) (Lysholm; n=48 knees)
Kirkley 2008	92 analysed at baseline (94 randomised) Female: 54 (59%) Age: 58.6 ± 10.2 years Duration of symptoms: 47.1 ± 69.4 months Pain: 239 ± 105 (WOMAC) Function: 830 ± 355 (WOMAC) Global Assessment: 1187 ± 483 (WOMAC)	86 analysed at baseline (94 randomised) Female: 58 (67%) Age: 60.6 ± 9.9 years Duration of symptoms: 40.1 ± 72.6 months Pain: 214 ± 122 (WOMAC) Function: 726 ± 397 (WOMAC) Global Assessment: 1043 ± 542 (WOMAC)

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Kise 2016	70 randomised Female: 27 (39%) Age: 48.9 ± 6.1 years Pain duration: 12.0 ± 15.7 months Pain: 67.6 ± 14.9 (KOOS) Function: 79.6 ± 16.1 (KOOS) Global Assessment: n.r.	70 randomised Female: 27 (39%) Age: 50.2 ± 6.2 years Pain duration: 17.3 ± 21.5 months Pain: 63.4 ± 20.8 (KOOS) Function: 75.0 ± 21.5 (KOOS) Global Assessment: n.r.
KIVIS	71 randomised Female: 52 (73.2%) Age: 64.9 ± 9.7 years Duration of knee OA: median 60 months (IQR 24-120) Pain: 254 ± 88 (WOMAC) Function: 853 ± 312.6 (WOMAC) Global Assessment: n.r.	79 randomised Female: 46 (58.2%) Age: 67.7 ± 9.1 years Duration of knee OA: median 54 months (IQR 29-120) Pain: 247 ± 97 (WOMAC) Function: 831 ± 340.7 (WOMAC) Global Assessment: n.r.
KORAL	1 receiving intervention Female: 4 (44.4%) across all groups Age: 57 (range 43-63) across all groups Symptom duration: n.r. Pain: 3.88 ± 1.36 (VAS) across all groups Function: n.r. Global Assessment: n.r.	4 receiving either placebo or conservative management Female: 4 (44.4%) across all groups Age: 57 (range 43-63) across all groups Symptom duration: n.r. Pain: 3.88 ± 1.36 (VAS) across all groups Function: n.r. Global Assessment: n.r.

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Merchan 1993	35 analysed (40 randomised) Female: 28 (80.0%) Age: 57 years (range 50-63) Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: 26.85 (Knee rating score)	38 analysed (40 randomised) Female: 25 (65.8%) Age: 56 years (50-65) Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: 29.86 (Knee rating score)
MeTeOR	161 analysed (174 randomised) Female: 90 (55.9%) Age: 59.0 ± 7.9 Symptom duration: n.r. Pain: 46.0 ± 15.5 (KOOS) Function: 37.1 ± 17.9 (WOMAC) Global Assessment: n.r.	169 analysed (177 randomised) Female: 97 (57.4%) Age: 57.8 ± 6.8 Symptom duration: n.r. Pain: 47.2 ± 16.4 (KOOS) Function: 37.5 ± 18.3 (WOMAC) Global Assessment: n.r.
Moseley 1996	5 randomised to two arthroscopic procedures (n=3 lavage and n=2 debridement) Female: 0 (00.0%) across all groups Age: 46.4 years (range 30-67) across all groups Symptom duration: n.r. Pain: 5.5 in lavage group and 4.5 in debridement group (average intensity of knee pain, scale 1-10) Function: n.r. Global Assessment: n.r.	5 randomised Female: 0 (00.0%) across all groups Age: 46.4 years (range 30-67) across all groups Symptom duration: n.r. Pain: 5.6 (average intensity of knee pain, scale 1-10) Function: n.r. Global Assessment: n.r.

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Moseley 2002	120 randomised to two arthroscopic procedures Female: 9 (7.5%) Age: 52.4 ± 11.4 years Symptom duration: n.r. Pain: 38.1 ± 17.6 (SF-36) Function: 43.3 ± 22.5 (SF-36) Global Assessment: n.r.	60 randomised Female: 4 (6.7%) Age: 52.0 ± 11.1 years Symptom duration: n.r. Pain: 37.8 ± 17.6 (SF-36) Function: 46.8 ± 22.5 (SF-36) Global Assessment: n.r.
Østerås 2012	8 randomised Female: 3 (37.5%) Age: 52.7 ± 7.2 years Duration of symptoms: 2.1 ± 1.7 years Pain: 3.7 ± 0.9 (VAS) Function: n.r. Global Assessment: 48.4 ± 25.6 (KOOS)	9 randomised Female: 1 (11.1%) Age: 47.0 ± 10.4 years Duration of symptoms: 1.6 ± 1.2 years Pain: 3.5 ± 1.7 (VAS) Function: n.r. Global Assessment: 51.4 ± 24.4 (KOOS)
Saeed 2015	60 randomised Female: 48 (80.0%) Age: n.r. Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: n.r.	60 randomised Female: 50 (83.3%) Age: n.r. Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: n.r.

RCT ID	Intervention (Mean ± SD if not further specified)	Comparator (Mean ± SD if not further specified)
Vermesan 2013	60 randomised knees Female: 49 (n.r.) Age: 59.2 ± 7.5 years Onset of symptoms: 3 ± 1.5 months Pain: n.r. Function: n.r. Global Assessment: 29.1 ± 3.7 (Oxford Knee Score)	60 randomised knees Female: 46 (n.r.) Age: 57.6 ± 7.8 years Onset of symptoms: 3 ± 1.7 months Pain: n.r. Function: n.r. Global Assessment: 30.3 ± 3.5 (Oxford Knee Score)
Weale 1998	50 randomised (inpatient setting) Female: 5 (10.0%) Age: 36.6 years Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: n.r.	50 randomised (outpatient setting) Female: 8 (16.0%) Age: 38.4 years Symptom duration: n.r. Pain: n.r. Function: n.r. Global Assessment: n.r.
Yim 2013	50 analysed (54 randomised) Female: 41 (82.0%) Age: 54.9 ± 10.3 years Onset of symptoms: 8.4 months (range 1.5-123) Pain: 5.2 ± 1.8 (VAS) Function: n.r. Global Assessment: 64.0 ± 11.2 (Lysholm)	52 analysed (54 randomised) Female: 40 (76.9%) Age: 57.6 ± 11.0 years Onset of symptoms: 8.2 months (range 2-81) Pain: 4.9 ± 1.5 (VAS) Function: n.r. Global Assessment: 65.2 ± 10.8 (Lysholm)

*baseline characteristics reported for those receiving treatment

Abbreviations: AIMS, Arthritis Impact Measurement Scales; BMI, body mass index; DMK, degenerative meniscus of the knee; IQR, Inter-quartile range; KF, Knee Society function score; KOOS, Knee injury and Osteoarthritis Outcome Score; LI, Lequesne index; n.r., not reported; NRS,

numerical rating scale; OA, osteoarthritis; SD, standard deviation; SE, standard error; SF-36, Medical Outcomes Study 36-Item Short Form; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

3.2.4 PICO 1

3.2.4.1 Results of the risk of bias assessment

The method for the random sequence generation was adequate in seven of 21 RCTs, unclear in 12 and inadequate in two. Allocation concealment was adequate in eight RCTs, unclear in 11 and the risk of bias was high in two. The risk of performance bias was unclear in one RCT and high in 16. Only four RCTs had a low risk of performance bias. Risk of detection bias was low in four RCTs, unclear in six and high in eleven. For continuous outcome measurements, risk of attrition bias was high in seven RCTs, low in six and unclear in eight. For binary outcome measurements, risk of attrition bias was rated high in four RCTs, unclear in three and low in four. Binary outcome measurements were not reported in ten RCTs; therefore, the risk of attrition bias for binary outcomes was not judged. Finally, the risk of reporting bias was graded as low in two RCTs, high in one, and unclear in the remaining 16. An overview of the risk of bias assessment is shown in Table 7. A detailed description of the risk of bias assessment, including the reasons supporting judgements, is provided in Appendix 6.

Table 7 Results of risk of bias assessment

RCT ID	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete continuous outcome data (attrition bias)	Incomplete binary outcome data (attrition bias)	Selective reporting (reporting bias)
Biedert 2000	High	Unclear	High	Unclear	Unclear	Unclear	Unclear
Chang 1993	Unclear	Unclear	High	High	High	High	Unclear
FIDELITY	Low	Low	Low	Low	Low	Low	Low
Forster 2003	Unclear	Unclear	High	Unclear	High	High	Unclear
Gauffin 2014	Unclear	Low	High	High	High	High	Unclear
Hamberg 1984	Low	High	Unclear	Unclear	Unclear	NA	Unclear
Herrlin 2007	Low	Low	High	High	Unclear	Unclear	Unclear
Kalunian 2000	Low	Unclear	Low	Low	Unclear	NA	Unclear
Kang 2005	Unclear	Unclear	High	Unclear	Low	NA	Unclear
Kirkley 2008	Low	Unclear	High	High	Low	NA	Low
Kise 2016	Low	Low	High	High	Low	Low	Low
KIVIS	Low	Unclear	High	High	High	High	Unclear
KORAL	Unclear	Low	High	High	High	NA	Unclear
Merchan 1993	High	High	High	Unclear	High	Low	Unclear
MeTeOR	Unclear	Low	High	High	Low	Low	Unclear
Moseley 1996	Unclear	Low	Low	Low	High	NA	Unclear
Moseley 2002	Unclear	Low	Low	Low	Unclear	NA	Unclear
Østerås 2012	Unclear	Unclear	High	High	Low	NA	Unclear
Saeed 2015	Unclear	Unclear	High	High	Unclear	Unclear	Unclear
Vermesan 2013	Unclear	Unclear	High	Unclear	Unclear	NA	Unclear
Yim 2013	Unclear	Unclear	High	High	Unclear	NA	High

NA, not applicable

3.2.4.2 Critical Outcomes

3.2.4.2.1 Pain

Pain was assessed in the majority of RCTs and with various instruments (Table 8).

Table 8 Instruments used to assess pain

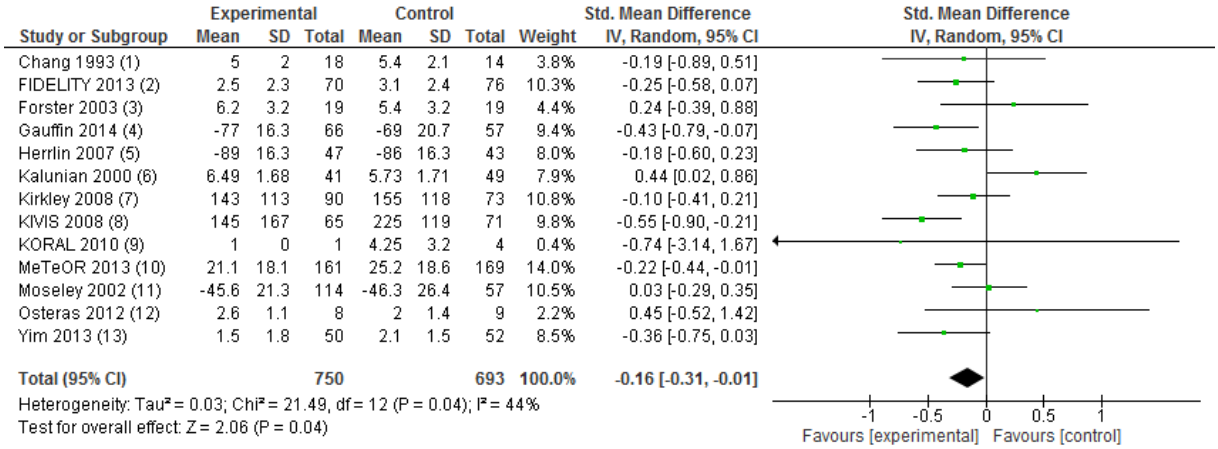
RCT ID	Instrument	Range	Direction
Chang 1993	Arthritis Impact Measurement Scale (AIMS) for pain	0 to 10	Higher score indicates worse pain
FIDELITY	Numerical rating scale for knee pain after exercise	0 to 10	Higher score indicates more extreme pain
Forster 2003	Visual Analogue Scale (VAS) – 10cm	0 to 10	n.r.; assumed as higher score indicates worse pain
KORAL Østerås 2012 Yim 2013	Visual Analogue Scale (VAS) – 10cm	0 to 10	Higher score indicates worse pain
Gauffin 2014 Herrlin 2007 Kise 2016**	Knee Injury and Osteoarthritis Outcome Score (KOOS) pain subscale***	0 to 100	Higher score indicates less pain*
MeTeOR	Knee Injury and Osteoarthritis Outcome Score (KOOS) pain subscale***	0 to 100	Higher score indicates more severe pain
Kalunian 2008 Kirkley 2008 KIVIS	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale****	0 to 20 0 to 500	Higher score indicates worse pain
Moseley 2002	MOS 36-item short-form (SF-36) body pain subscale	0 to 100	Higher score indicates less severe pain*
Saeed 2015**	Knee Society Score System	0 to 50	Higher score indicates more severe pain

*effects multiplied by -1 to invert scale in analyses; **not included in pooled analyses; ***these subscales are usually measured on a Likert scale (0 to 4) and then normalized to a 0 to 100 scale; ****subscale includes pain while walking, stairs, lying, sitting and standing.

Pain - Short-term follow-up

Pain was assessed in 16 RCTs with follow-up times ranging from two to six months. Compared to control, arthroscopy was associated with less knee pain and this effect was statistically significant (SMD -0.16, 95% CI [-0.31, -0.01] and should be interpreted as a small effect, Figure 2; low quality of evidence, Table 13). In sensitivity analyses, when the RCT by Kalunian 2000 was excluded - the only RCT comparing arthroscopic lavage with diagnostic arthroscopy - the heterogeneity decreased to 12% without changing the overall effect estimate substantially (SMD -0.22, 95% CI [-0.34, -0.10], Figure 35 in section 3.2.4.4.1).

Pain results from four RCTs (Moseley 1996, Saeed 2015, Kise 2016, Chang 1993) were not pooled for analysis, because pain was measured as a binary/ordinal variable or due to methodological issues. Results of these RCTs are summarised in section 8.4.1.1 of Appendix 5.

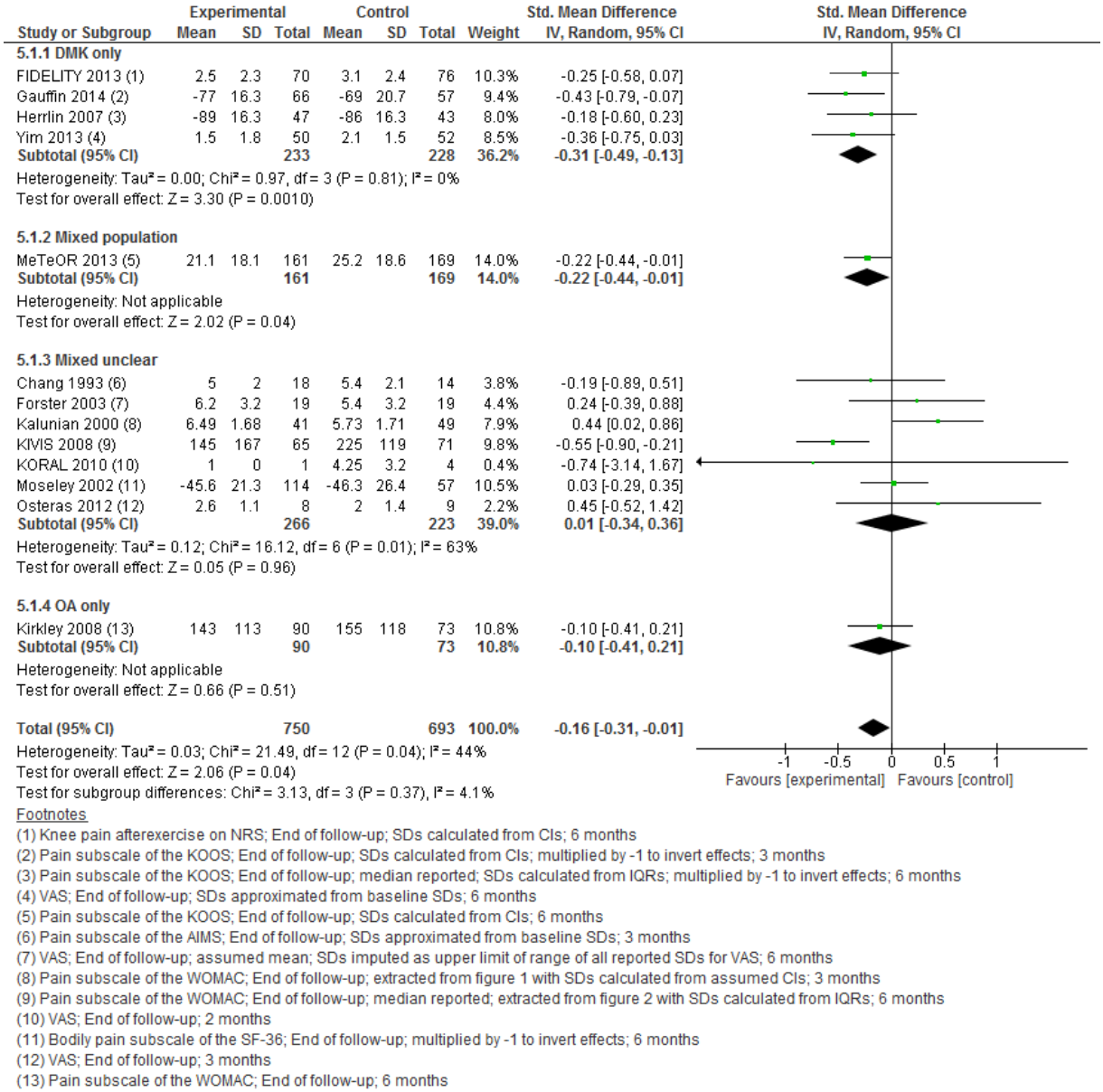


Footnotes
(1) Pain subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 3 months
(2) Knee pain afterexercise on NRS; End of follow-up; SDs calculated from CIs; 6 months
(3) VAS; End of follow-up; assumed mean; SDs imputed as upper limit of range of all reported SDs for VAS; 6 months
(4) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 3 months
(5) Pain subscale of the KOOS; End of follow-up; median reported; SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(6) Pain subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(7) Pain subscale of the WOMAC; End of follow-up; 6 months
(8) Pain subscale of the WOMAC; End of follow-up; median reported; extracted from figure 2 with SDs calculated from IQRs; 6 months
(9) VAS; End of follow-up; 2 months
(10) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; 6 months
(11) Bodily pain subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 6 months
(12) VAS; End of follow-up; 3 months
(13) VAS; End of follow-up; SDs approximated from baseline SDs; 6 months

Figure 2 Pain at short-term follow-up

Subgroups: DMK versus OA

Four RCTs (FIDELITY, Gauffin 2014, Herrlin 2007, Yim 2013) were classified as reporting results for a population with DMK only. One RCT (MeTeOR) reported pain for a population with both DMK and OA (mixed), while in seven RCTs (Chang 1993, Forster 2003, Kalunian 2000, KIVIS, KORAL, Moseley 2002, Østerås 2012), the population was insufficiently described and classified as mixed unclear. Only one RCT (Kirkley 2008) included patients with OA without concurrent DMK (large meniscus tears were excluded). Effect estimates for pain in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. The SMDs were -0.31 (95% CI [-0.49, -0.13]) for DMK only, -0.22 (95% CI [0.44, -0.01]) for patients with mixed OA and DMK, 0.01 (95% CI [-0.34, 0.36]) for patients with mixed OA and DMK of unclear pathology and -0.10 (95% CI [-0.41, 0.21]) for patients with OA only. Figure 3 shows the resulting forest plot with the four subgroup estimates and overall effect estimate.



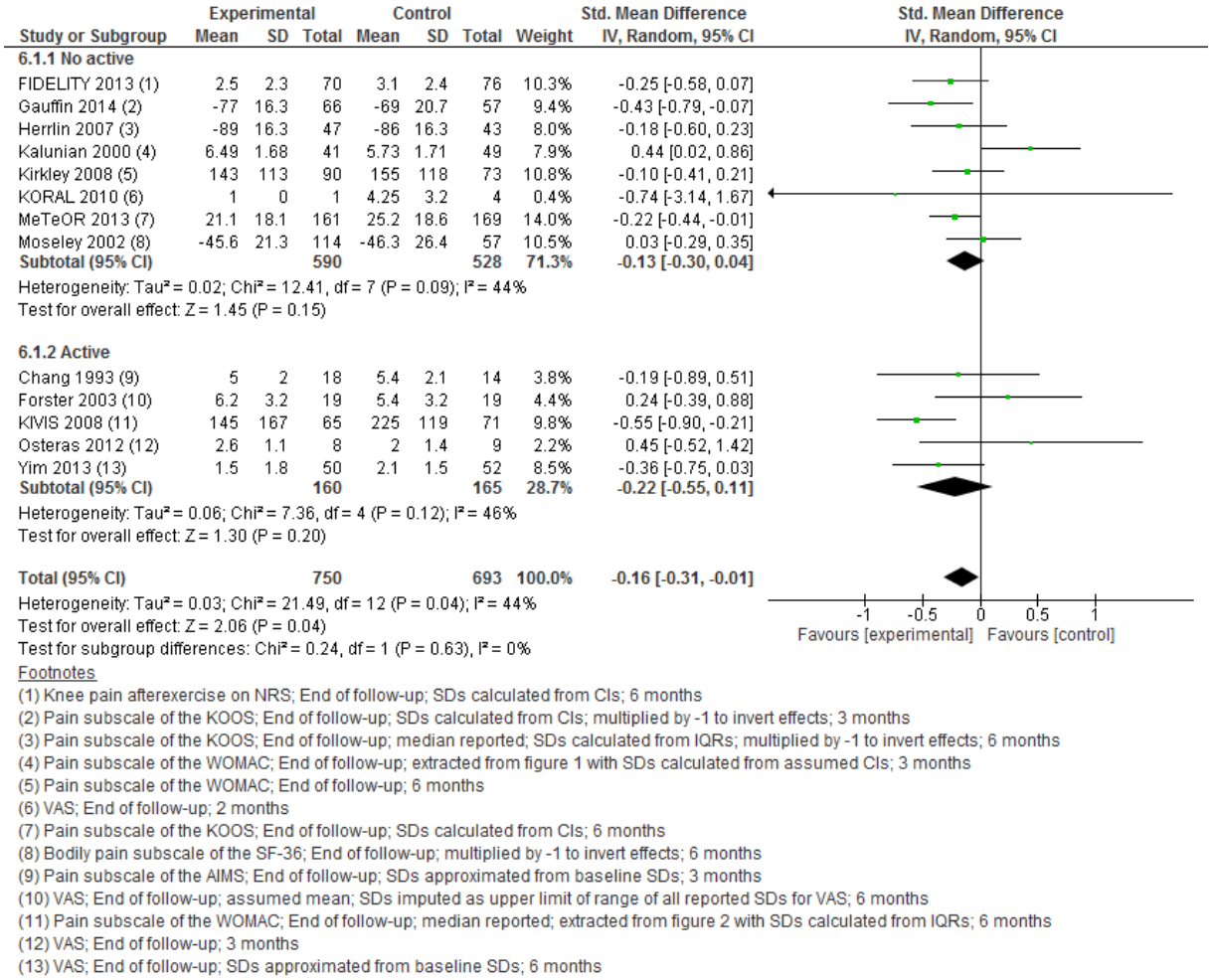
Footnotes

(1) Knee pain afterexercise on NRS; End of follow-up; SDs calculated from CIs; 6 months
(2) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 3 months
(3) Pain subscale of the KOOS; End of follow-up; median reported; SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(4) VAS; End of follow-up; SDs approximated from baseline SDs; 6 months
(5) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; 6 months
(6) Pain subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 3 months
(7) VAS; End of follow-up; assumed mean; SDs imputed as upper limit of range of all reported SDs for VAS; 6 months
(8) Pain subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(9) Pain subscale of the WOMAC; End of follow-up; median reported; extracted from figure 2 with SDs calculated from IQRs; 6 months
(10) VAS; End of follow-up; 2 months
(11) Bodily pain subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 6 months
(12) VAS; End of follow-up; 3 months
(13) Pain subscale of the WOMAC; End of follow-up; 6 months

Figure 3 Pain by DMK/OA classification at short-term follow-up

Subgroups: non-active versus active

In eight RCTs (FIDELITY, Gauffin 2014, Herrlin 2007, Kalunian 2000, Kirkley 2008, KORAL, MeTeOR, Moseley 2002) arthroscopic surgery was compared to non-active comparators, while in five RCTs (Chang 1993, Forster 2003, KIVIS, Østerås 2012, Yim 2013) it was compared to active comparators. There was no statistically significantly different effect in RCTs comparing arthroscopy to non-active treatment (SMD -0.13, 95% CI [-0.30, 0.04]) and in RCTs comparing arthroscopy to active treatment (SMD -0.22, 95% CI [-0.55, 0.11]). Confidence intervals were largely overlapping. Figure 4 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.



Footnotes

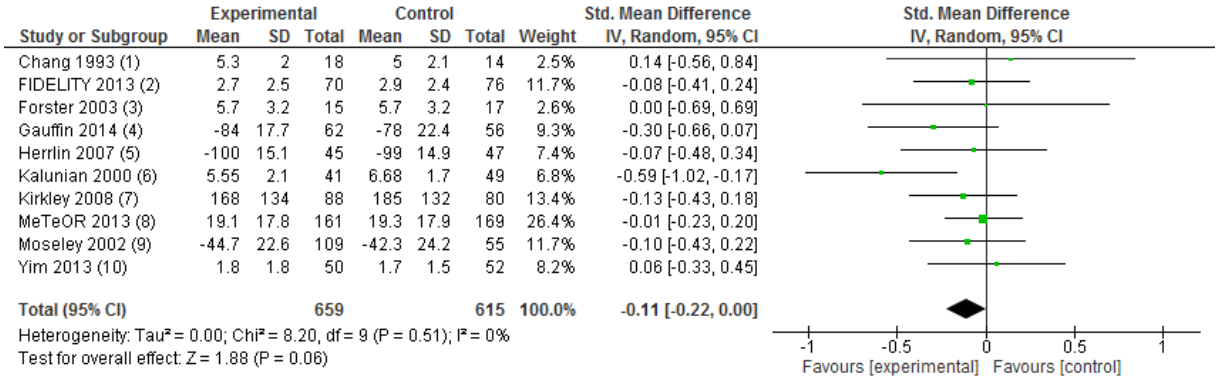
(1) Knee pain afterexercise on NRS; End of follow-up; SDs calculated from CIs; 6 months
(2) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 3 months
(3) Pain subscale of the KOOS; End of follow-up; median reported; SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(4) Pain subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(5) Pain subscale of the WOMAC; End of follow-up; 6 months
(6) VAS; End of follow-up; 2 months
(7) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; 6 months
(8) Bodily pain subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 6 months
(9) Pain subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 3 months
(10) VAS; End of follow-up; assumed mean; SDs imputed as upper limit of range of all reported SDs for VAS; 6 months
(11) Pain subscale of the WOMAC; End of follow-up; median reported; extracted from figure 2 with SDs calculated from IQRs; 6 months
(12) VAS; End of follow-up; 3 months
(13) VAS; End of follow-up; SDs approximated from baseline SDs; 6 months

Figure 4 Pain by comparator type at short-term follow-up

Pain - Intermediate follow-up

Eleven RCTs reported pain at an intermediate follow-up with follow-up times ranging from 12 to 60 months. Compared to control, arthroscopy did not have a statistically significant effect on reducing pain at intermediate follow-up (SMD -0.11, 95% CI [-0.22, 0.00], Figure 5; low quality of evidence, Table 14). Heterogeneity between RCTs was low (I^2=0%).

Additional results (Chang 1993, Kise 2016, Yim 2013) that could not be pooled were descriptively summarised in section 8.4.1.2 of Appendix 5.



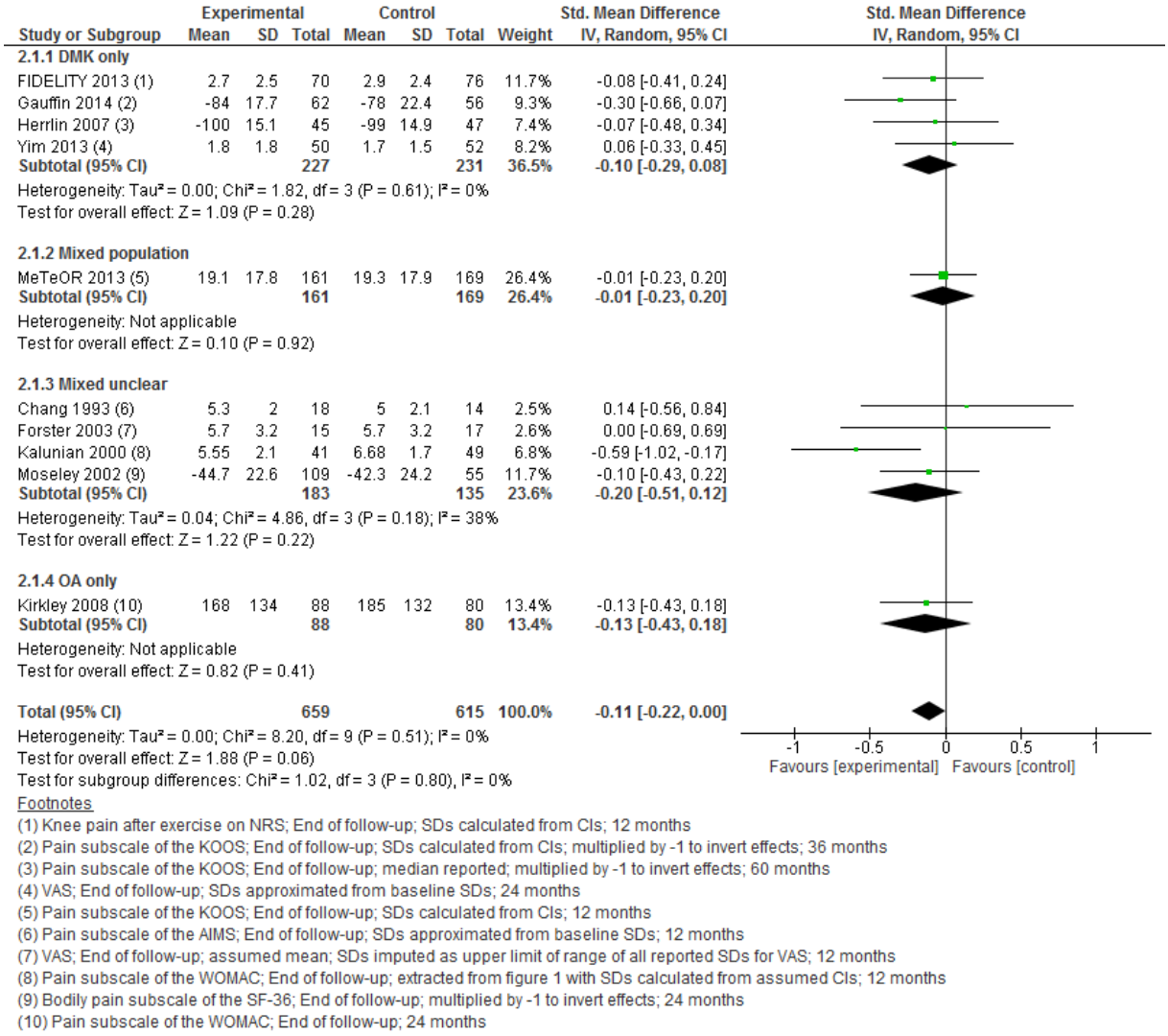
Footnotes

(1) Pain subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 12 months
(2) Knee pain after exercise on NRS; End of follow-up; SDs calculated from CIs; 12 months
(3) VAS; End of follow-up; assumed mean; SDs imputed as upper limit of range of all reported SDs for VAS; 12 months
(4) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 36 months
(5) Pain subscale of the KOOS; End of follow-up; median reported; multiplied by -1 to invert effects; 60 months
(6) Pain subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 12 months
(7) Pain subscale of the WOMAC; End of follow-up; 24 months
(8) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; 12 months
(9) Bodily pain subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 24 months
(10) VAS; End of follow-up; SDs approximated from baseline SDs; 24 months

Figure 5 Pain at intermediate follow-up

Subgroups: DMK versus OA

Four RCTs (FIDELITY, Gauffin 2014, Herrlin 2007, Yim 2013) included patients with DMK only. One RCT (MeTeOR) had patients with both DMK and OA (mixed), and in four RCTs (Chang 1993, Forster 2003, Kalunian 2000, Moseley 2002) the populations were insufficiently described and were classified as mixed unclear. Only one RCT (Kirkley 2008) included patients with OA (OA only) without concurrent DMK (large meniscus tears were excluded). Effect estimates for pain in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. The SMDs were -0.10 (95% CI [-0.29, 0.08]) for DMK only, -0.01 (95% CI [-0.23, 0.20]) for patients with mixed OA and DMK, -0.20 (95% CI [-0.51, 0.12]) for patients with mixed OA and DMK of unclear pathology and -0.13 (95% CI [-0.43, 0.18]) for the OA only subgroup. Figure 6 shows the resulting forest plot with the four subgroup estimates and overall effect estimate.



Footnotes

(1) Knee pain after exercise on NRS; End of follow-up; SDs calculated from CIs; 12 months
(2) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 36 months
(3) Pain subscale of the KOOS; End of follow-up; median reported; multiplied by -1 to invert effects; 60 months
(4) VAS; End of follow-up; SDs approximated from baseline SDs; 24 months
(5) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; 12 months
(6) Pain subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 12 months
(7) VAS; End of follow-up; assumed mean; SDs imputed as upper limit of range of all reported SDs for VAS; 12 months
(8) Pain subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 12 months
(9) Bodily pain subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 24 months
(10) Pain subscale of the WOMAC; End of follow-up; 24 months

Figure 6 Pain by DMK/OA classification at intermediate follow-up

Subgroups: non-active versus active

In seven RCTs (FIDELITY, Gauffin 2014, Herrlin 2007, Kalunian 2000, Kirkley 2008, MeTeOR, Moseley 2002) arthroscopic surgery was compared to non-active comparators and in three (Chang 1993, Forster 2003, Yim 2013) with an active comparator. Compared to non-active treatment, arthroscopy was associated with less knee pain, this effect was statistically significant (SMD -0.14, 95% CI [-0.27, -0.01]), but can only be interpreted as a small effect. There was no statistically significantly different effect in RCTs comparing arthroscopy to active treatment (SMD 0.06, 95% CI [-0.24, 0.37]). Effect estimates for pain in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. Confidence intervals were largely overlapping. Figure 7 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

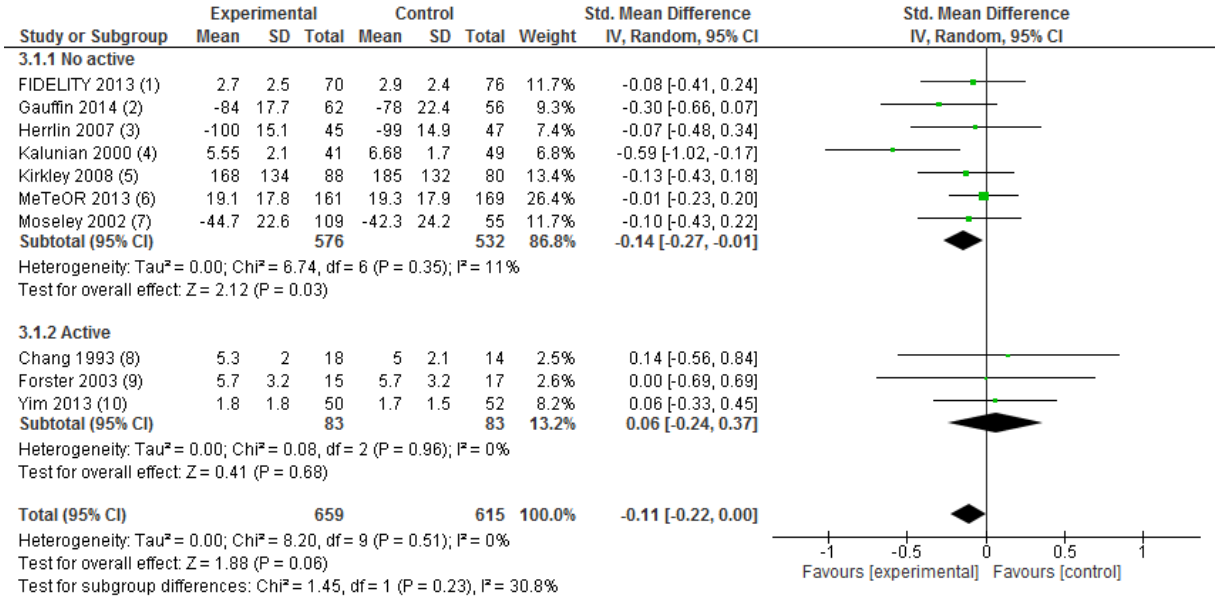


Figure 7 Pain by comparator type at intermediate follow-up

3.2.4.2.2 Function

Function was assessed by various instruments as shown in Table 9. For the purposes of this report, instruments that combined measures of pain and function were defined as global assessments of outcome (section 3.2.4.2.3), while those that only measured function (not pain) were defined as assessing the outcome of function.

Table 9 Instruments used to assess function

RCT ID	Instrument	Range	Direction
Chang 1993	Arthritis Impact Measurement Scale (AIMS) for physical function	0 to 10	Higher score indicates worse physical function
Forster 2003	Knee Society rating system (KSRS) function score	0 to 100	Higher score indicates better knee function*
Gauffin 2014, Herrlin 2007, Kise 2016**	Knee Injury and Osteoarthritis Outcome Score (KOOS) activities of daily living subscale	0 to 100	Higher score indicates better knee function*
Kalunian 2008	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) function subscale	0 to 68	Higher score indicates worse function

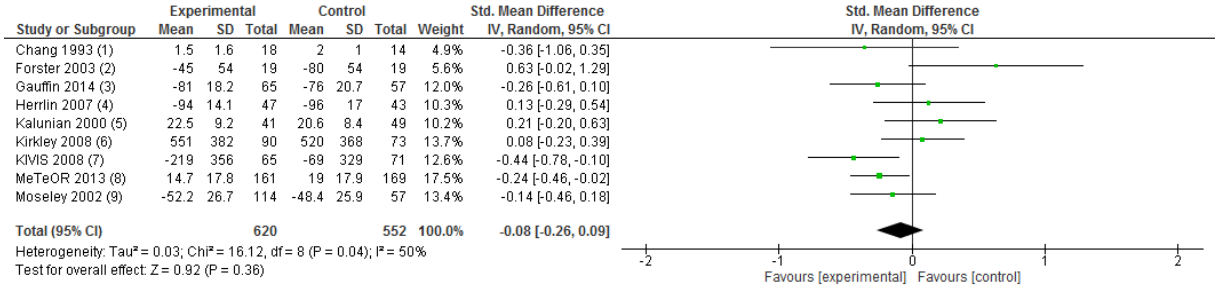
RCT ID	Instrument	Range	Direction
Kirkley 2008, KIVIS	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) function subscale	0 to 1700	Higher score indicates worse function
MeTeOR	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) function subscale	0 to 100	Higher score indicates worse function
Moseley 2002	MOS 36-item short-form (SF-36) physical function subscale	0 to 100	Higher score indicates better function*

*effects multiplied by -1 to invert scale in analyses; **not included in pooled analyses

Function - Short-term follow-up

Ten RCTs reported on the outcome of function at short-term follow-up with follow-up times ranging from three to six months. There was no statistically significant effect on function (SMD -0.08, 95% CI [-0.26, 0.09], Figure 8; very low quality of evidence, Table 13) in favour of arthroscopy when compared to control. Heterogeneity between RCTs was considerable (I^2=50%) and could not be entirely explained in sensitivity analyses. However, the removal of the RCT by Forster 2003 reduced the heterogeneity to I^2=36% (see Figure 36 in section 3.2.4.4.2) without changing the effect estimate.

In addition, MeTeOR reported binary results and Kise 2016 reported mean difference between groups comparison (see section 8.4.2.1 of Appendix 5), which could not be pooled in analyses.



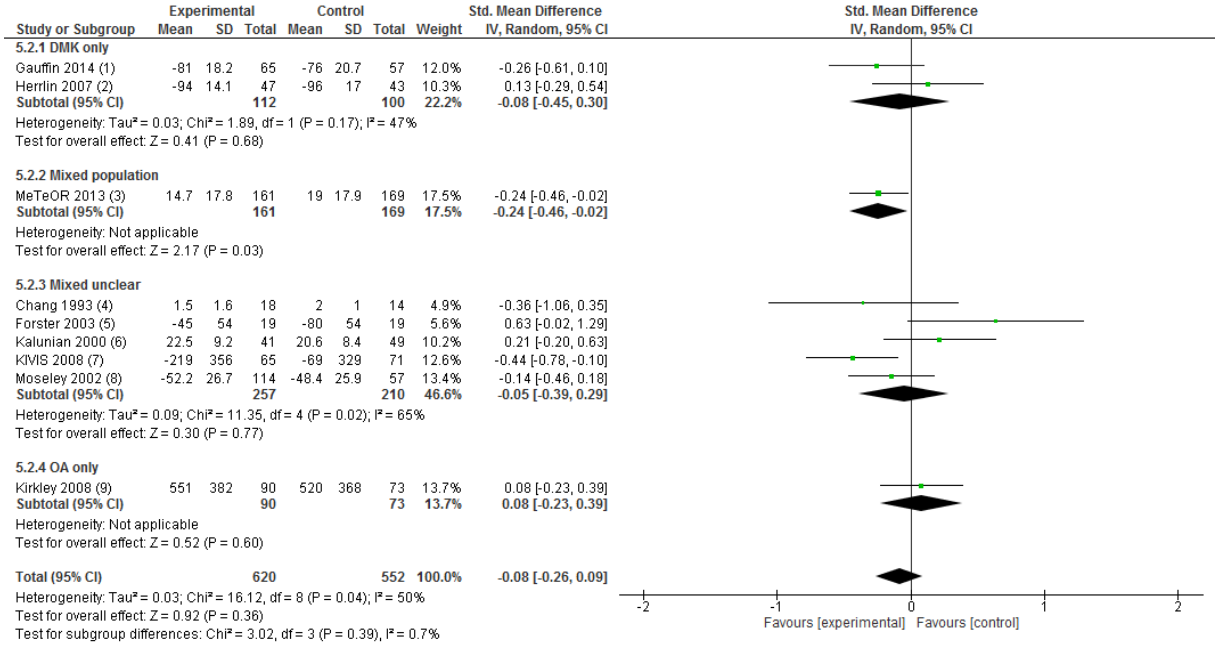
Footnotes
(1) Physical function subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 3 months
(2) Function score of the KSRS; End of follow-up; assumed mean; SDs imputed as most conservative estimate to maintain stastical non-significance; multiplied by -1 to invert effects; 6 months
(3) Activities of daily living (ADL) subscale of KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 3 months
(4) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; median reported; extracted from figure 3 with SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(5) Function subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(6) Physical function subscale of the WOMAC; End of follow-up; 6 months
(7) Physical function subscale of the WOMAC; mean difference from baseline; effect multiplied by -1 to indicate improvement on scale; 6 months
(8) Physical function subscale of the WOMAC; End of follow-up; SDs caluclated from CIs; 6 months
(9) Physical functioning subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 6 months

Figure 8 Function at short-term follow-up

Subgroups: DMK versus OA

Two RCTs (Gauffin 2014, Herrlin 2007) were classified as reporting results for a population with DMK only. One RCT (MeTeOR) reported function for a population with both DMK and OA (mixed), while in five RCTs (Chang 1993, Forster 2003, Kalunian 2000, KIVIS, Moseley 2002), the population was insufficiently described and classified as mixed unclear. Only one RCT (Kirkley 2008) included an OA only population without concurrent DMK (large meniscus tears were

excluded). Although the treatment effect in the mixed population of OA and DMK was statistically significant in favour of arthroscopy, the effect estimates in the subgroups were similar. The SMDs were -0.08 (95% CI [-0.45, 0.30]) for DMK only, -0.24 (95% CI [-0.46, -0.02]) for patients with mixed OA and DMK, -0.05 (95% CI [-0.39, 0.29]) for patients with mixed OA and DMK of unclear pathology and 0.08 (95% CI [-0.23, 0.39]) for OA only subgroup. Figure 9 shows the resulting forest plot with the four subgroup estimates and overall effect estimate.



Footnotes

(1) Activities of daily living (ADL) subscale of KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 3 months
(2) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; median reported; extracted from figure 3 with SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(3) Physical function subscale of the WOMAC; End of follow-up; SDs caluclated from CIs; 6 months
(4) Physical function subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 3 months
(5) Function score of the KSRS; End of follow-up; assumed mean; SDs imputed as most conservative estimate to maintain stastical non-significance; multiplied by -1 to invert effects; 6 months
(6) Function subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(7) Physical function subscale of the WOMAC; mean difference from baseline; effect multiplied by -1 to indicate improvement on scale; 6 months
(8) Physical functioning subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 6 months
(9) Physical function subscale of the WOMAC; End of follow-up; 6 months

Figure 9 Function by DMK/OA classification at short-term follow-up

Subgroups: non-active versus active

In six RCTs (Gauffin 2014, Herrlin 2007, Kalunian 2000, Kirkley 2008, MeTeOR, Moseley 2002), arthroscopic surgery was compared to non-active comparators, while in three RCTs (Chang 1993, Forster 2003, KIVIS) it was compared to active comparators. No statistically significant effect on function was found when comparing arthroscopy to non-active treatment (SMD -0.07, 95% CI [-0.23, 0.08]) or active treatment (SMD -0.08 [95% CI -0.74, 0.57]). Figure 10 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

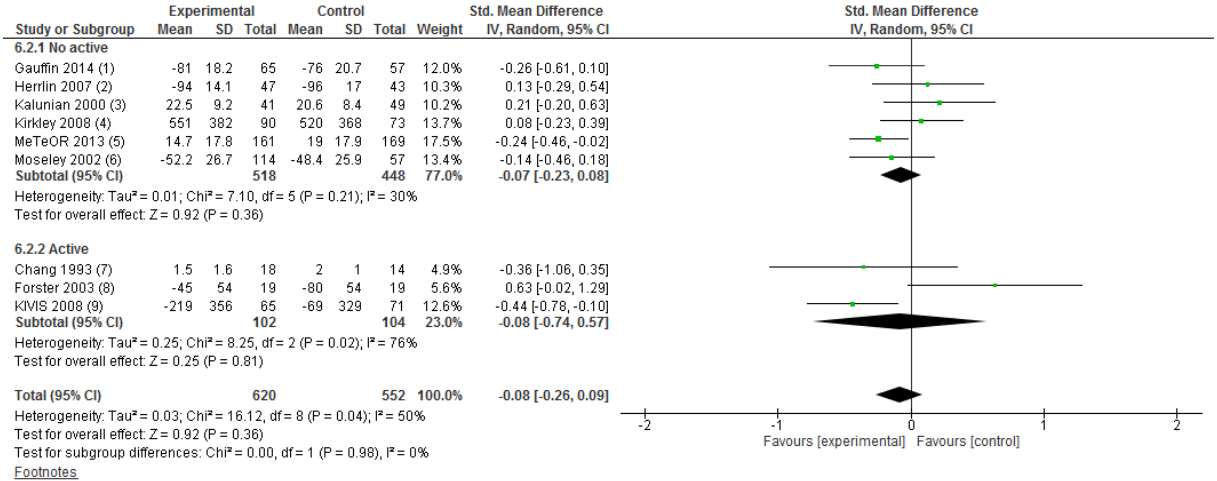
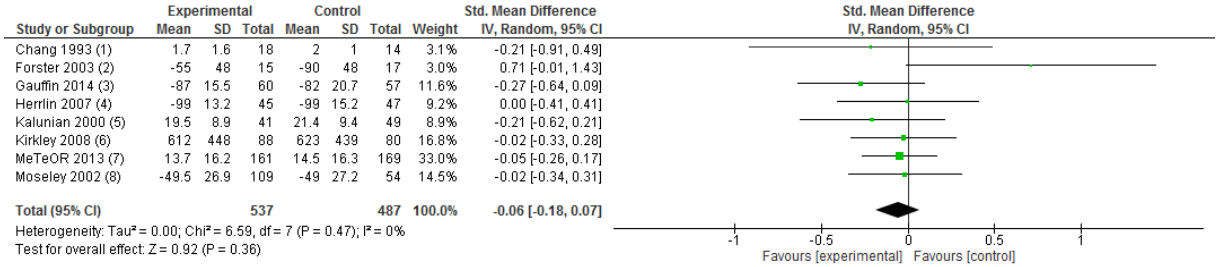


Figure 10 Function by comparator type at short-term follow-up

Function - Intermediate follow-up

Function was assessed in nine RCTs with follow-up times ranging from 12 to 60 months. There was no statistically significant effect on function (SMD -0.06, 95% CI [-0.18, 0.07], Figure 11; low quality of evidence, Table 14) in favour of arthroscopy compared to control. Heterogeneity between RCTs was low (I^2=0%). The observed effect in the RCT by Forster 2003 was in the opposite direction in comparison to the remaining RCTs, probably because function scores at baseline on the Knee Society rating system differed between the arthroscopy and control groups (45 versus 65 [on a zero to 100 scale]); excluding the RCT by Forster 2003 only slightly affected the overall effect estimate (see sensitivity analysis, Figure 37 in section 3.2.4.4.2).

Kise 2016 reported a small mean difference in function of 1.6 (95% CI [-2.9, 6.1]) slightly (but not statistically significantly) favouring the comparator group over the arthroscopy group at 24 months. This result could not be considered in the pooled analysis.



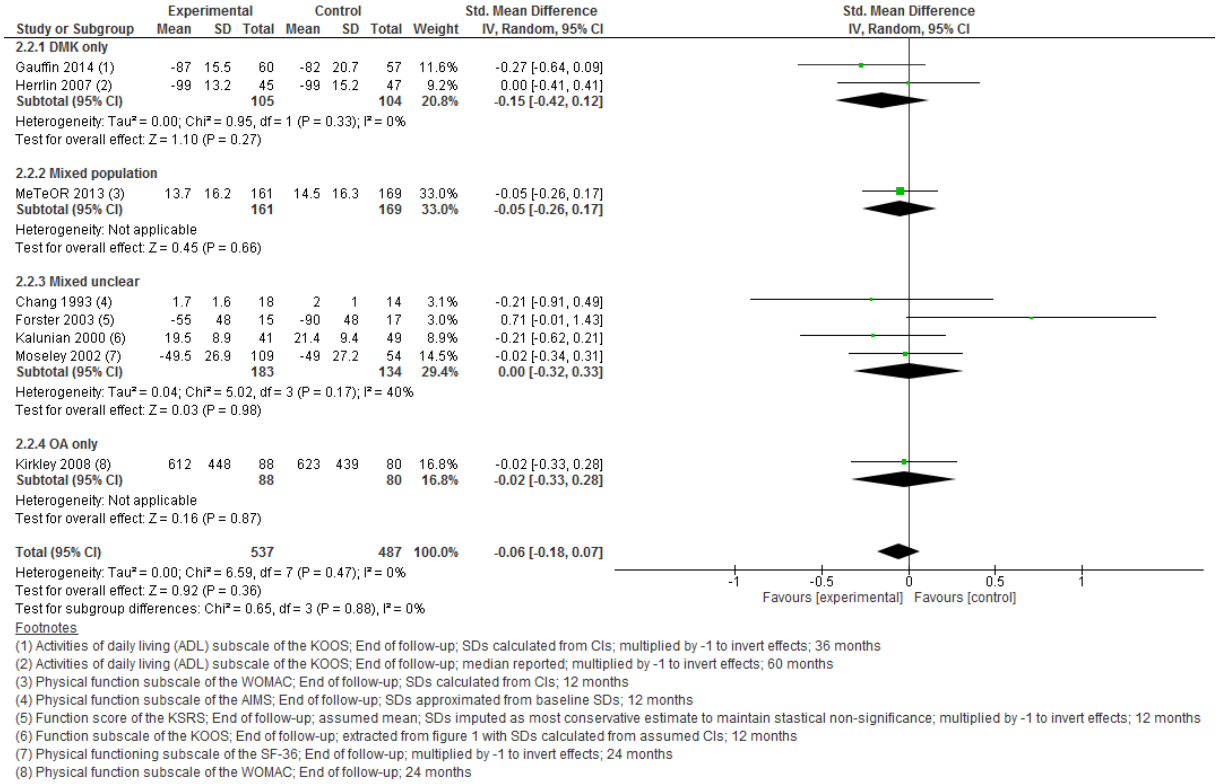
Footnotes

(1) Physical function subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 12 months
(2) Function score of the KSRS; End of follow-up; assumed mean; SDs imputed as most conservative estimate to maintain stastical non-significance; multiplied by -1 to invert effects; 12 months
(3) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 36 months
(4) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; median reported; multiplied by -1 to invert effects; 60 months
(5) Function subscale of the KOOS; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 12 months
(6) Physical function subscale of the WOMAC; End of follow-up; 24 months
(7) Physical function subscale of the WOMAC; End of follow-up; SDs calculated from CIs; 12 months
(8) Physical functioning subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 24 months

Figure 11 Function at intermediate follow-up

Subgroups: DMK versus OA

Two RCTs (Gauffin 2014, Herrlin 2007) included a population with DMK only. One RCT (MeTeOR) included a population with both DMK and OA (mixed), while in four RCTs (Chang 1993, Forster 2003, Kalunian 2000, Moseley 2002), the population was insufficiently described and classified as mixed unclear. Only one RCT (Kirkley 2008) included patients with OA without concurrent DMK (large meniscus tears were excluded). Effect estimates for arthroscopy versus control in the four subgroups were similar and 95% CIs for largely overlapping. The SMDs were -0.15 (95% CI [-0.42, 0.12]) for DMK only, -0.05 (95% CI [-0.26, 0.17]) for patient with mixed OA and DMK, 0.00 (95% CI [-0.32, 0.33]) for patients with mixed OA and DMK of unclear pathology and -0.02 (95% CI [-0.33, 0.28]) for the OA only subgroup. Figure 12 shows the resulting forest plot with the four subgroup estimates and overall effect estimate.



Footnotes

(1) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 36 months
(2) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; median reported; multiplied by -1 to invert effects; 60 months
(3) Physical function subscale of the WOMAC; End of follow-up; SDs calculated from CIs; 12 months
(4) Physical function subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 12 months
(5) Function score of the KSRS; End of follow-up; assumed mean; SDs imputed as most conservative estimate to maintain stastical non-significance; multiplied by -1 to invert effects; 12 months
(6) Function subscale of the KOOS; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 12 months
(7) Physical functioning subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 24 months
(8) Physical function subscale of the WOMAC; End of follow-up; 24 months

Figure 12 Function by DMK/OA classification at intermediate follow-up

Subgroups: non-active versus active

In six RCTs (Gauffin 2014, Herrlin 2007, Kalunian 2000, Kirkley 2008, MeTeOR, Moseley 2002) arthroscopic surgery was compared to non-active comparators, while in two RCTs (Chang 1993, Forster 2003) it was compared to active comparators. There was no statistically significant effect in RCTs comparing arthroscopy to non-active treatment (SMD -0.08, 95% CI [-0.21, 0.05]) and in RCTs comparing arthroscopy to active treatment (SMD 0.25, 95% CI [-0.66, 1.15]). CIs were largely overlapping. Figure 13 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

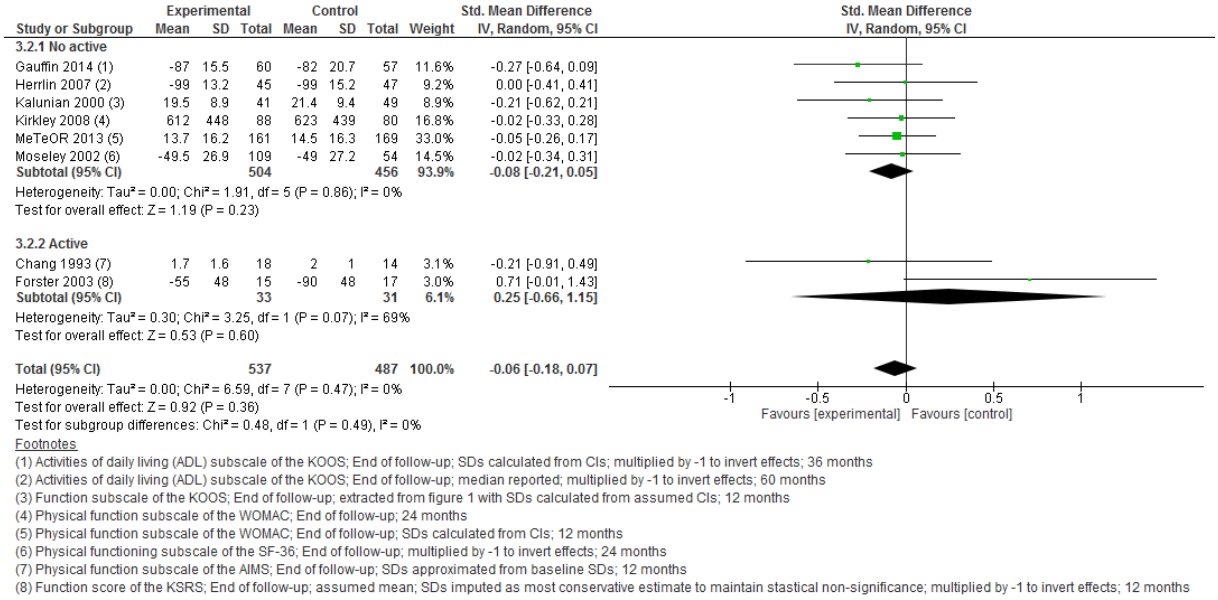


Figure 13 Function by comparator type at intermediate follow-up

3.2.4.2.3 Global assessment

The outcome of global assessment comprises pain, function and other factors, such as quality of life, joint stiffness or disability. Various instruments were used to measure global assessment as seen in Table 10.

Table 10 Instruments used to assess global assessment

RCT ID	Instrument	Range	Direction
Chang 1993	Visual Analogue Scale (VAS) – 10cm	0 to 10	Higher score indicates worse global assessment
FIDELITY, Herrlin 2007, Kang 2005, Yim 2013	Lysholm Knee Scoring Scale	0 to 100	Higher score indicates less severe symptoms*
Hamberg 1984	Lysholm point-scoring scale	0 to 95	
Forster 2003	Lequesne Index (LI)	0 to 24	Higher score indicates more severe symptoms
Kalunian 2008	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)	0 to 96	Higher score indicates worse symptoms
Kirkley 2008	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)	0 to 2400	Higher score indicates worse symptoms

RCT ID	Instrument	Range	Direction
Merchan 1993	Hospital for Special Surgery Knee Rating Score (HSSKRS)	0 to 100	Higher score indicating better condition
Østerås 2012	Knee Injury and Osteoarthritis Outcome Score (KOOS) composite measure**	Assumed 0 to 100	Higher score indicated worse problems
Vermesan 2013	Oxford Knee Score (OKS)	12 to 60	Higher score indicates more difficulties

*effects multiplied by -1 to invert scale in analyses; **not recommended or validates as a composite score according to KOOS User's Guide[66]

Global assessment - Short-term follow-up

Ten RCTs reported on the outcome of global assessment with follow-up times ranging from two to six months with nine RCTs using the number of patients in the denominator and one RCT using the number of knees as the denominator (Vermesan 2013). Therefore, the latter RCT was analysed separately. There was no statistically significant difference in global assessment when comparing arthroscopic intervention to control (SMD 0.03, 95% CI [-0.12, 0.17], Figure 14; low quality of evidence, Table 13). Heterogeneity was low (I^2=0%). The RCT by Vermesan 2013 reported a statistically significant difference in global assessment favouring intra-articular steroid injection (SMD 0.82, 95% CI [0.44, 1.19], Figure 14). Hamberg 1984 was the only RCT where half of the patients (similar distribution in arthroscopy and in control) were randomised to total meniscectomy. If patients with total meniscectomy were excluded in sensitivity analyses the effect estimate was hardly affected (Figure 38 in section 3.2.4.4.3).

Two RCTs (Chang 1993, KIVIS) used binary outcomes for global assessment and the results are reported in section 8.4.3.1 of Appendix 5.

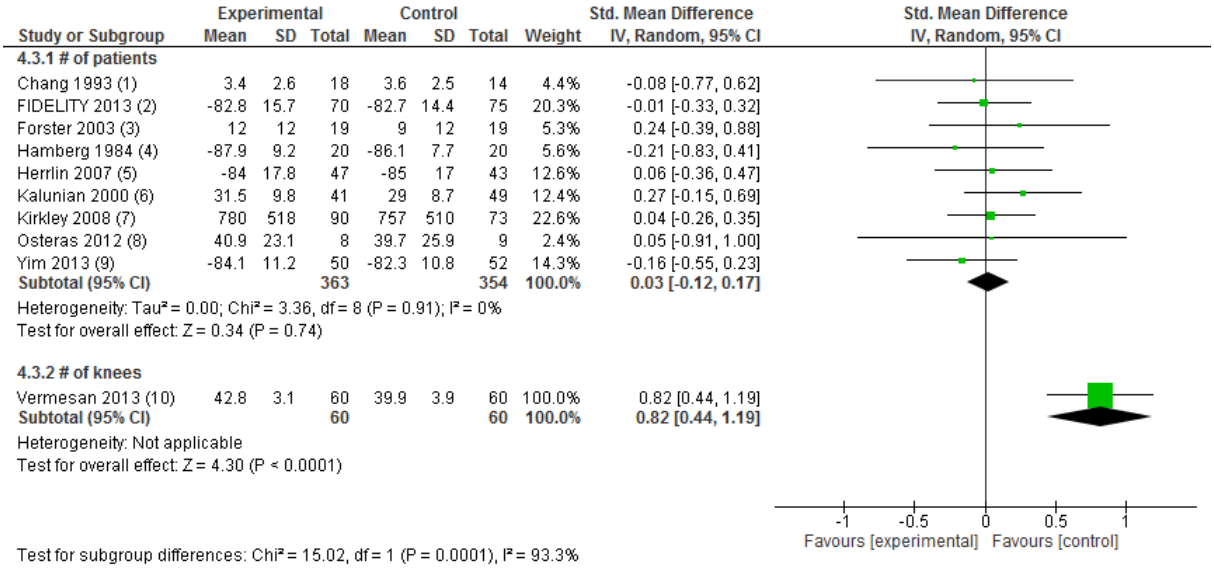


Figure 14 Global assessment at short-term follow-up

Subgroups: DMK versus OA

Three RCTs (FIDELITY, Herrlin 2007, Yim 2013) included patients with DMK only. No RCT included patients with both DMK and OA (mixed), while in five RCTs (Chang 1993, Forster 2003, Hamberg 1984, Kalunian 2000, Østerås 2012) the population was insufficiently described and the populations in these RCTs were classified as mixed unclear. Only one RCT (Kirkley 2008) included patients with OA without concurrent DMK (large meniscus tears were excluded). Effect estimates for global assessment in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. The SMDs were -0.04 (95% CI [-0.25, 0.18]) for patient with DMK only, 0.11 (95% CI [-0.16, 0.38]) for patients with mixed OA and DMK of unclear pathology and 0.04 (95% CI [-0.26, 0.35]) for patients with OA only. Figure 15 shows the resulting forest plot with the four subgroup estimates and overall effect estimate.

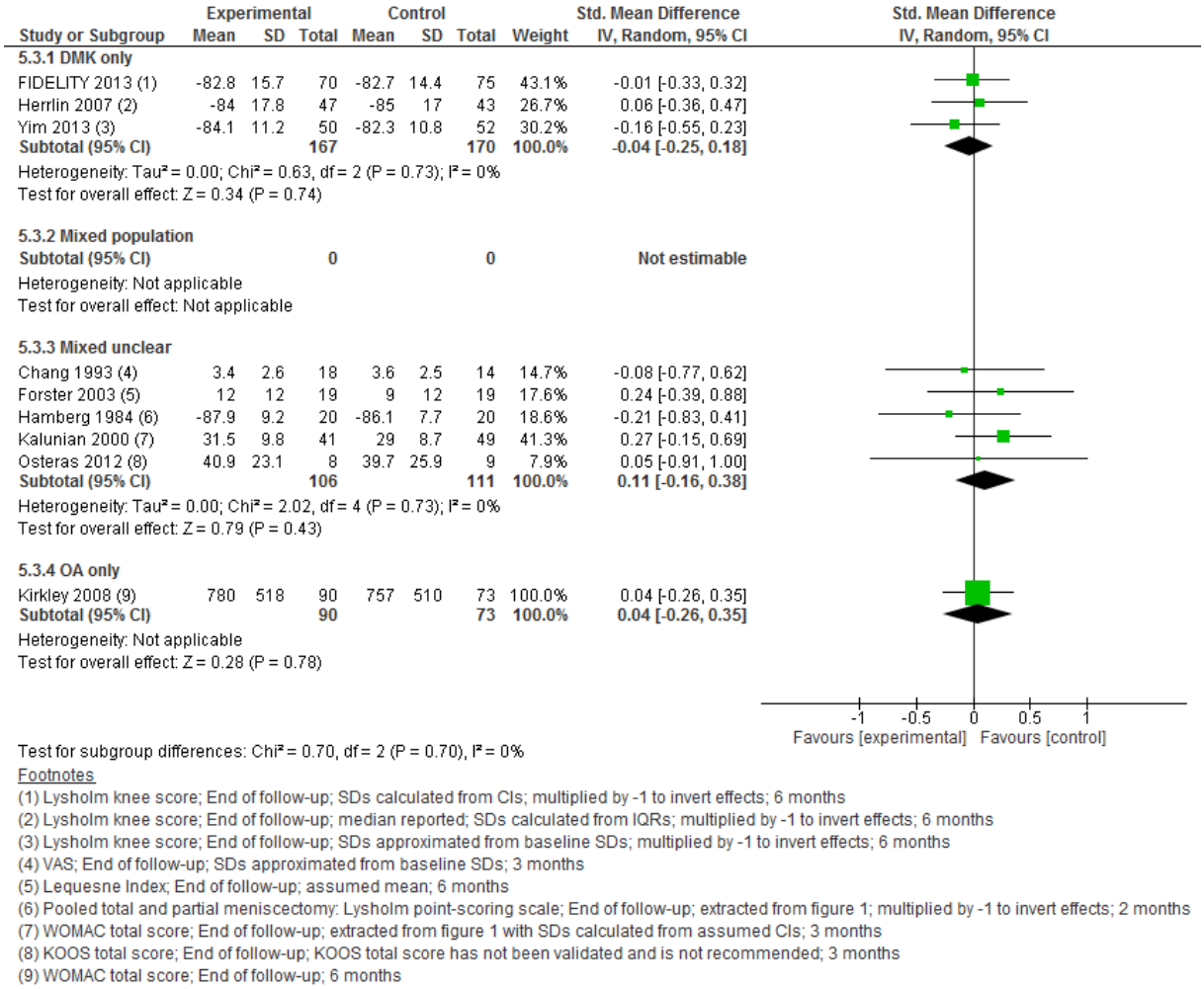


Figure 15 Global assessment by DNK/OA classification at short-term follow-up

Subgroups: non-active versus active

In four RCTs (FIDELITY, Herrlin 2007, Kalunian 2000, Kirkley 2008) arthroscopic surgery was compared to non-active comparators, while in five RCTs (Chang 1993, Forster 2003, Hamberg 1984, Østerås 2012, Yim 2013) it was compared to active comparators. There was no statistically significant difference when comparing arthroscopy to non-active treatment (SMD -0.07, 95% CI [-0.11, 0.25]) and when comparing arthroscopy to active treatment (SMD -0.08 [95% CI -0.34, 0.18]). CIs were largely overlapping. Figure 16 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

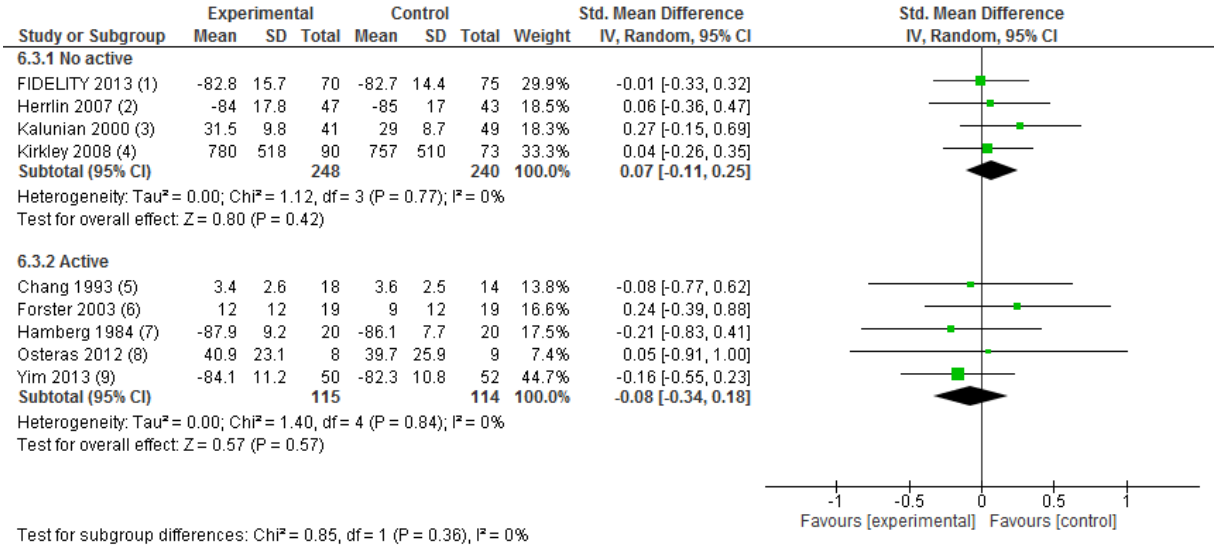


Figure 16 Global assessment by comparator type at short-term follow-up

Global assessment - Intermediate follow-up

Global assessment was assessed in 10 RCTs with follow-up times ranging from 12 to 60 months. In eight of the RCTs the denominator was the number of patients and in two RCTs (Kang 2005, Vermesan 2013) the denominator was the number of knees. There was no statistically significant effect on global assessments for arthroscopy versus control in RCTs using patients as the denominator (SMD 0.04, 95% CI [-0.16, 0.23], Figure 17; low quality of evidence, Table 14) or in RCTs using the number of knees as the denominator (SMD -0.07, 95% CI [-0.95, 0.81], Figure 17). Heterogeneity between RCTs using patients as the denominator was substantial (I^2=39%) and was reduced when removing Herrlin 2007 (I^2=12%) (Figure 39 in section 3.2.4.4.3).

The results of three RCTs (Biedert 2000, Chang 1993, Merchan 1993), where pooling was not possible, are presented in section 8.4.3.2 of Appendix 5.

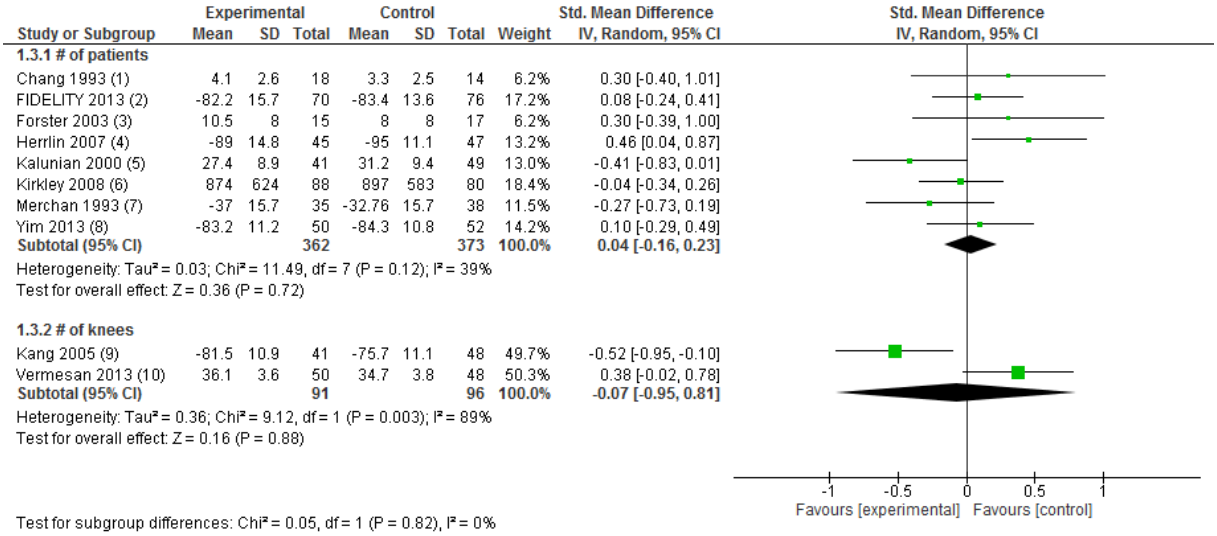


Figure 17 Global assessment at intermediate follow-up

Subgroups: DMK versus OA

Three RCTs (FIDELITY, Herrlin 2007, Yim 2013) included patients with DMK only. No RCT included a population with both DMK and OA (mixed), and in four RCTs (Chang 1993, Forster 2003, Kalunian 2000, Merchan 1993) the populations were insufficiently described and were classified as mixed unclear. Only one RCT (Kirkley 2008) included patients with OA and without concurrent DMK (large meniscus tears were excluded). Effect estimates for global assessment in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. The SMDs were 0.19 (95% CI [-0.04, 0.41]) for DMK only, -0.11 (95% CI [-0.46, 0.24]) for patients with mixed OA and DMK of unclear pathology and -0.04 (95% CI [-0.34, 0.26]) for patients with OA only. Figure 18 shows the resulting forest plot with the four subgroup estimates and overall effect estimate.

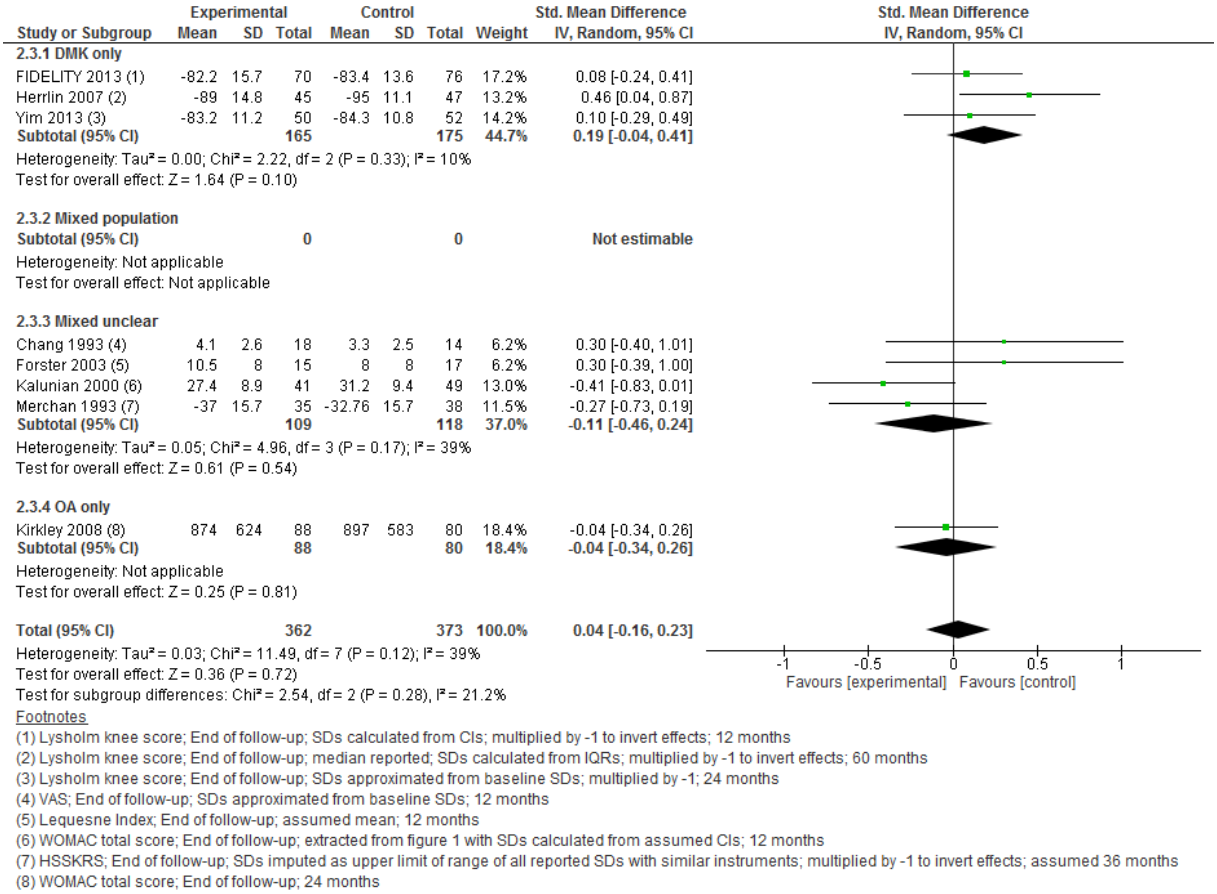


Figure 18 Global assessment by DMK/OA classification at intermediate follow-up

Subgroups: non-active versus active

In four RCTs (FIDELITY, Herrlin 2007, Kalunian 2000, Kirkley 2008) arthroscopic surgery was compared to non-active comparators, while in four RCTs (Chang 1993, Forster 2003, Merchan 1993, Yim 2013) to active comparators. There was no statistically significant effect in RCTs comparing arthroscopy to non-active treatment (SMD 0.02, 95% CI [-0.28, 0.33]) or in RCTs comparing arthroscopy to active treatment (SMD 0.04 [95% CI -0.21, 0.30]). CIs were largely overlapping. Figure 19 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

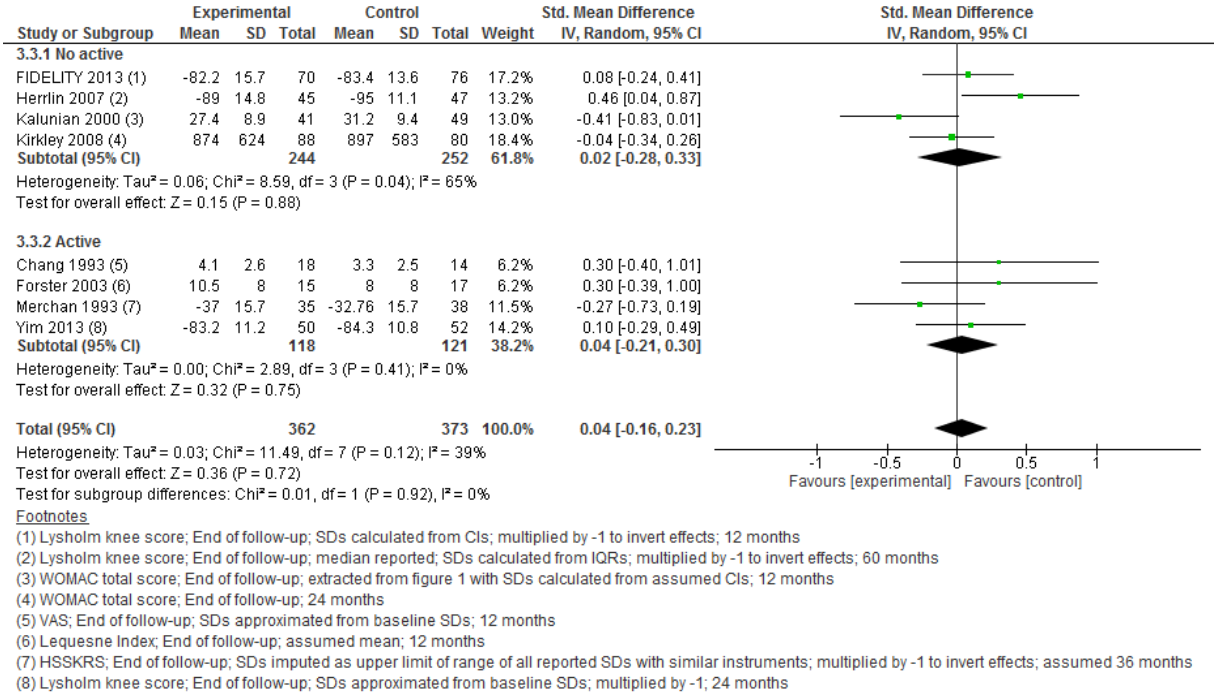


Figure 19 Global assessment by comparator type at intermediate follow-up

3.2.4.3 Important Outcomes

3.2.4.3.1 Joint stiffness

Joint stiffness was measured in three RCTs using the same instrument, but with two different scales. Table 11 gives additional information on the instrument.

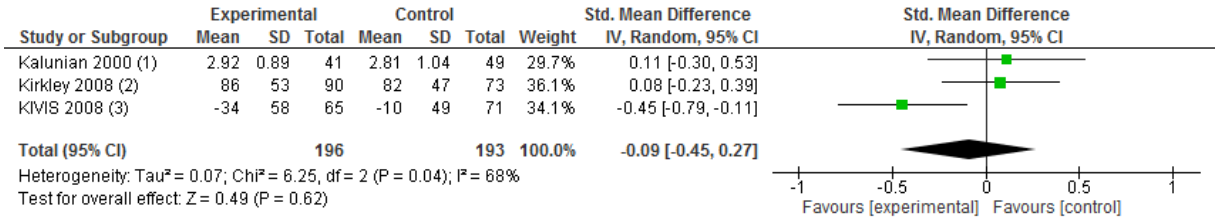
Table 11 Instrument used to assess joint stiffness

RCT ID	Instrument	Range	Direction
Kalunian 2008	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) stiffness subscale	0 to 8	Higher score indicates worse stiffness
Kirkley 2008, KIVIS	Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) stiffness subscale	0 to 200	Higher score indicates worse stiffness

Joint stiffness - Short-term follow-up

Three RCTs reported results for joint stiffness with follow-up times ranging from three to six months. There was no statistically significant effect on joint stiffness found (SMD -0.09, 95% CI [-0.45, 0.27]), Figure 20; very low quality of evidence, Table 13) in favour of arthroscopy when

comparing to control. Heterogeneity for the pooled estimate from three RCTs was substantial (I^2=68%). In the KIVIS RCT, only mean changes from baseline were reported, and baseline values were slightly imbalanced (arthroscopic tidal irrigation: 123±38 versus corticosteroid injection: 112±44), favouring tidal irrigation, which may account for the heterogeneity seen.



Footnotes
(1) Stiffness subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(2) Stiffness subscale of the WOMAC; End of follow-up; 6 months
(3) Stiffness subscale of the WOMAC; mean difference from baseline; effect multiplied by -1 to indicate improvement on scale; 6 months

Figure 20 Joint stiffness at short-term follow-up

Subgroups: DMK versus OA

No RCT included patients with DMK only or patients with both DMK and OA (mixed). In two RCTs (Kalunian 2000, KIVIS 2008) the populations were insufficiently described and the patient populations of these RCTs were classified as mixed unclear. Only one RCT (Kirkley 2008) included patients with OA without concurrent DMK (large meniscus tears were excluded). Effect estimates for joint stiffness in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. The SMDs were -0.18 (95% CI [-0.73, 0.37]) for patients with mixed OA and DMK of unclear pathology and 0.08 (95% CI [-0.23, 0.39]) for patients with OA only. Figure 21 shows the resulting forest plot with the subgroup estimates and overall effect estimate.

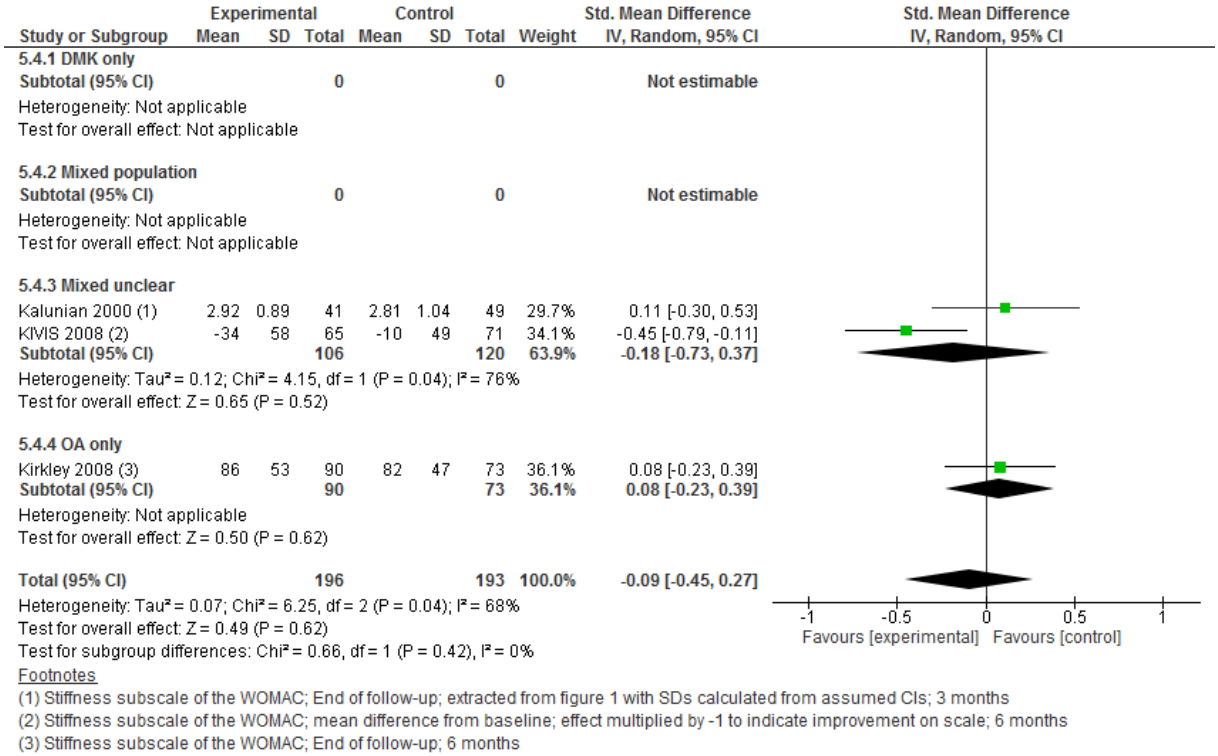


Figure 21 Joint stiffness by DMK/OA classification at short-term follow-up

Subgroups: non-active versus active

In two RCTs (Kalunian 2000, Kirkley 2008) arthroscopic surgery was compared to non-active comparators and in one RCT (KIVIS) with an active comparator. There was no statistically significant effect in the two RCTs comparing arthroscopy to non-active treatment (SMD 0.09, 95% CI [-0.16, 0.34]), but a statistically significant effect estimate was in the one RCT found favouring arthroscopy when compared to active treatment (SMD -0.45, 95% CI [-0.79, -0.11]). In the KIVIS RCT only mean changes from baseline were reported, and the baseline values were slightly imbalanced (arthroscopic tidal irrigation: 123±38 versus corticosteroid injection: 112±44) favouring tidal irrigation. CIs were largely overlapping. Figure 22 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

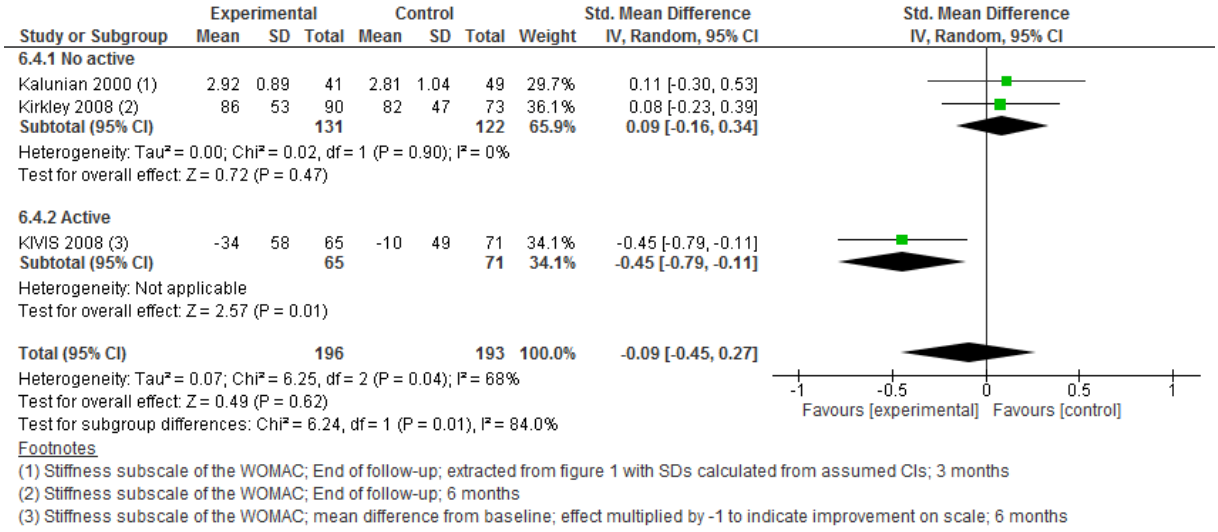


Figure 22 Joint stiffness by comparator type at short-term follow-up

Joint stiffness – Intermediate follow-up

Joint stiffness was assessed by two RCTs, one at 12 months (Kalunian 2000) and the other at 24 months (Kirkley 2008). Compared to control, arthroscopy was found to have no statistically significant effect on joint stiffness (SMD -0.18, 95% CI [-0.75, 0.39], Figure 23; very low quality of evidence, Table 14). Heterogeneity was substantial (I^2=79%). Heterogeneity could not be explained as only two RCTs reported joint stiffness at intermediate follow-up.

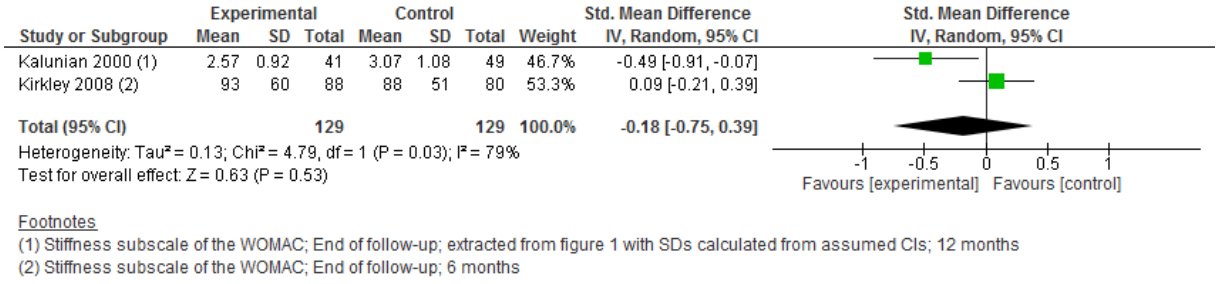


Figure 23 Joint stiffness at intermediate follow-up

Subgroups: DMK versus OA

No subgroup analyses were conducted as only two RCTs reported results for joint stiffness at intermediate follow-up. One RCT (Kalunian 2000) included a population that could not be classified and Kirkley 2008 included patients only with OA without concurrent DMK.

Subgroups: non-active versus active

Both RCTs compared arthroscopy to non-active treatment; hence, no subgroup analyses were conducted.

3.2.4.3.2 Total knee replacement

Total knee replacement (TKR) at 12 months was reported in two RCTs (Forster 2003, MeTeOR). The point estimate indicated a higher relative risk for total knee replacement in the arthroscopic group compared to controls, but this estimate was not statistically significant (RR 1.25, 95% CI [0.38, 4.19], Figure 24; very low quality of evidence, Table 14). Heterogeneity for this outcome was low (I^2=0%).

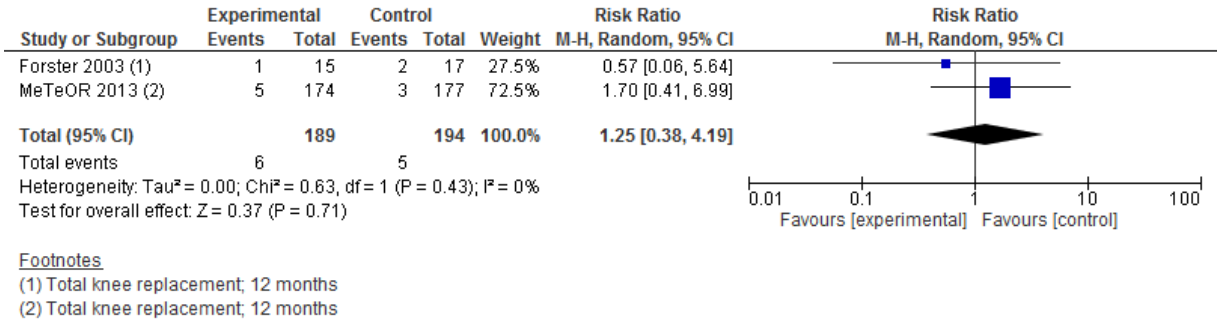


Figure 24 Total knee replacement at 12 months

Subgroups: DMK versus OA

One RCT (MeTeOR) included a mixed population of DMK and OA, and one RCT (Forster 2003) included a population that could not be classified. No subgroup analyses were conducted.

Subgroups: non-active versus active

One RCT (MeTeOR) compared arthroscopy to non-active treatment, while the other RCT (Forster 2003) compared arthroscopy to active treatment. No subgroup analyses were conducted.

3.2.4.3.3 Quality of life

Health-related or disease-specific quality of life was measured with several tools. Three RCTs reported both health-related and disease-specific quality of life. Table 12 shows the various instruments used to assess quality of life.

Table 12 Instruments used to assess quality of life (health-related and disease-specific)

RCT ID	Instrument	Range	Direction
Health-related			
FIDELITY	15D	0 to 1	Higher score indicates better quality of life
Gauffin 2014	EuroQol 5 dimensions Visual Analogue Scale (EQ-5D VAS)	0 to 100	Higher score indicates better imaginable health state
Kirkley 2008	Standard-gamble utility technique	0.0 to 1.0	Higher score indicating better health
Disease-specific			
FIDELITY	Western Ontario Meniscal Evaluation Tool (WOMET)	0 to 100	Higher score indicates absence of symptoms

RCT ID	Instrument	Range	Direction
Herrlin 2007, Gauffin 2014, Kise 2016*	Knee Injury and Osteoarthritis Outcome Score (KOOS) activities of daily living subscale	0 to 100	Higher score indicates better quality of life
Kirkley 2008	MOS 36-item short-form (SF-36) physical component summary	0 to 100	Higher score indicates better quality of life

*not included in pooled analyses

Quality of life - Short-term follow-up

Four RCTs reported quality of life as an outcome with follow-up time ranging from three to six months. Two RCTs (Gauffin 2014, Kirkley 2008) assessed both health-related and disease-specific quality of life, and one RCT (Herrlin 2007) assessed only disease-specific quality of life. Compared to control, there was no statistically significant effect on both health-related and disease-specific quality of life) in favour of arthroscopy compared to control (SMD 0.18, 95% CI [-0.05, 0.42] and SMD 0.18, 95% CI [-0.02, 0.39], respectively, Figure 25; very low qualities of evidence, Table 13). Heterogeneity was low (I^2=0%) for both outcomes.

Kise 2016 reported a between-group mean difference (-4.0, 95% CI [-10.3, 2.2] favouring arthroscopy; therefore, the RCT could not be pooled with results from the other RCTs.

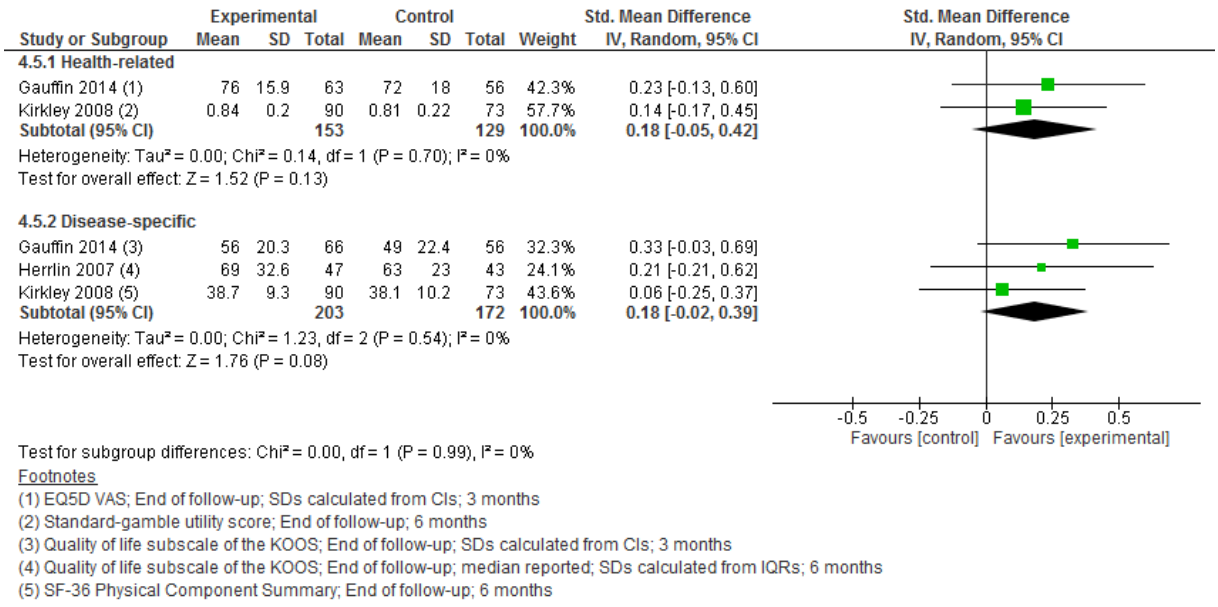


Figure 25 Quality of life at short-term follow-up

Subgroups: DMK versus OA

No subgroup analyses were conducted as too few RCTs reported results for health-related or disease-specific quality of life. Two RCTs (Gauffin 2014, Herrlin 2007) reported a population of DMK only. The other RCT (Kirkley 2008) included patients with OA only without concurrent DMK.

Subgroups: non-active versus active

No subgroup analyses were conducted as too few RCTs reported results for health-related or disease-specific quality of life. All three RCTs (Gauffin 2014, Herrlin 2007, Kirkley 2008) compared arthroscopy to non-active treatment.

Quality of life – Intermediate follow-up

Five RCTs reported on quality of life with follow-up times ranging from 12 to 60 months. Three RCTs (FIDELITY, Gauffin 2014, Kirkley 2008) assessed health-related and disease-specific quality of life, and one RCT (Herrlin 2007) assessed only disease-specific quality of life. Compared to control, arthroscopy did not have a statistically significant effect on health-related and disease-specific quality of life (SMD 0.17, 95% CI [-0.02, 0.36] and SMD 0.06, 95% CI [-0.11, 0.23], respectively, Figure 26; moderate qualities of evidence, Table 14). Heterogeneity was low (I^2=0%) for both outcomes.

The RCT by Kise 2016 reported a between-group mean difference (-1.8, 95% CI [-8.1, 4.5]) favouring arthroscopy and could not be pooled with results from the other RCTs.

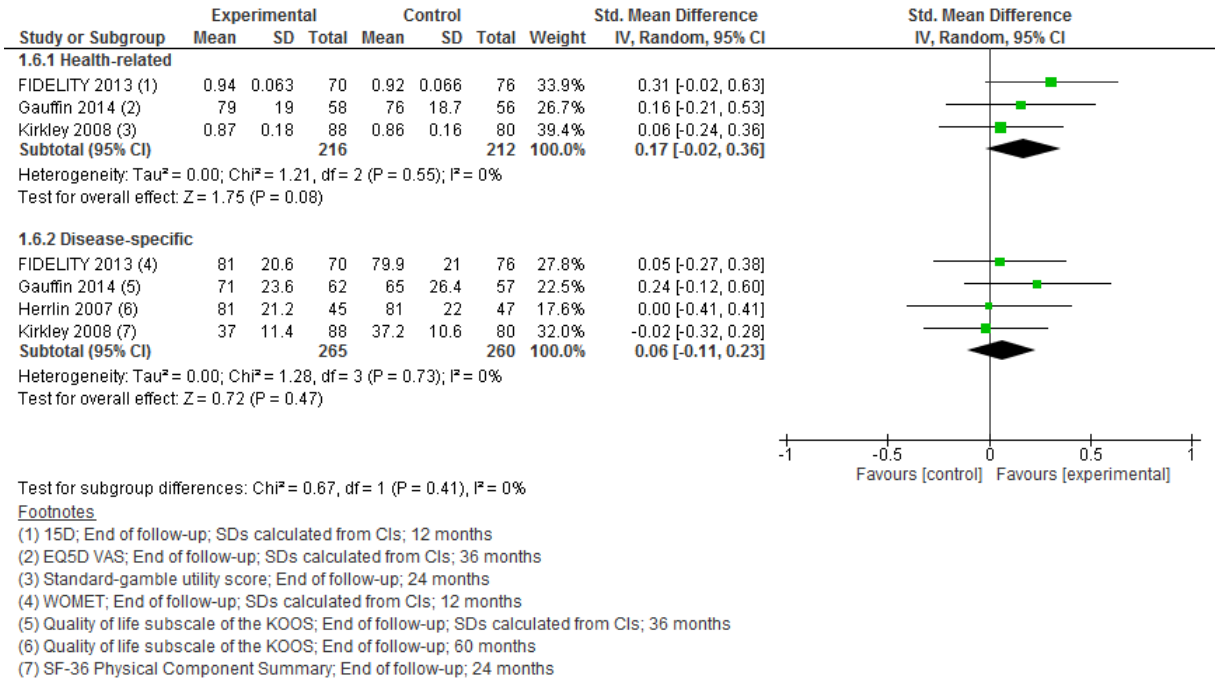


Figure 26 Quality of life at intermediate follow-up

Subgroups: DMK versus OA

Two RCTs (FIDELITY, Gauffin 2014) reported on health-related quality of life in patients with DMK only, and one RCT (Kirkley 2008) in patients with OA only without concurrent DMK (large meniscus tears were excluded). Effect estimates for health-related quality of life in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. The SMDs were 0.24, 95% CI [-0.00, 0.49] for patients with DMK only and 0.06, 95% CI [-0.24, 0.36] for patients with OA only. Figure 27 shows the resulting forest plot with the two subgroup estimates and overall effect estimate. Three RCTs (FIDELITY, Gauffin 2014, Herrlin 2007) reported on disease-specific quality of life included patients with DMK only, and one RCT (Kirkley 2008) included

patients with OA without concurrent DMK (large meniscus tears were excluded). Effect estimates for disease-specific quality of life in these subgroups did not statistically significantly differ and 95% CIs largely overlapped. The SMDs were 0.18, 95% CI [-0.11, 0.47] for DMK only and SMD -0.02, 95% CI [-0.32, 0.28]) for patients with OA only. Figure 28 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

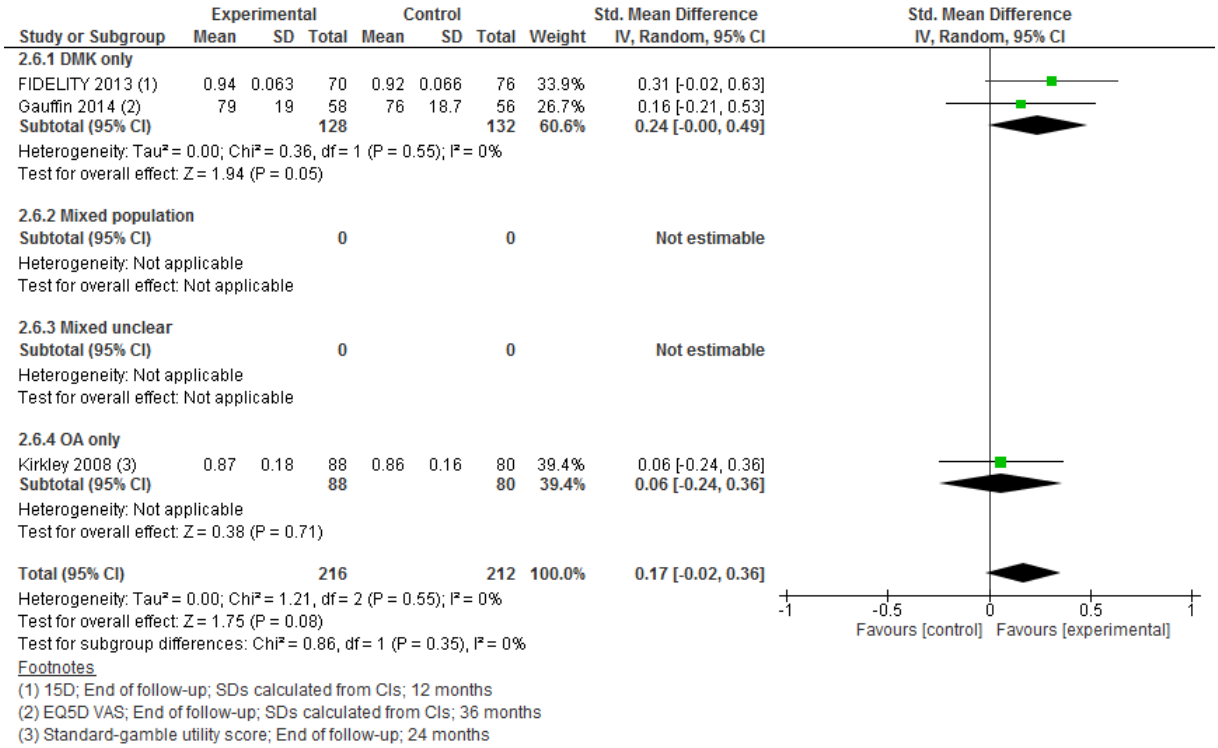


Figure 27 Health-related quality of life by DMK/OA classification at intermediate follow-up

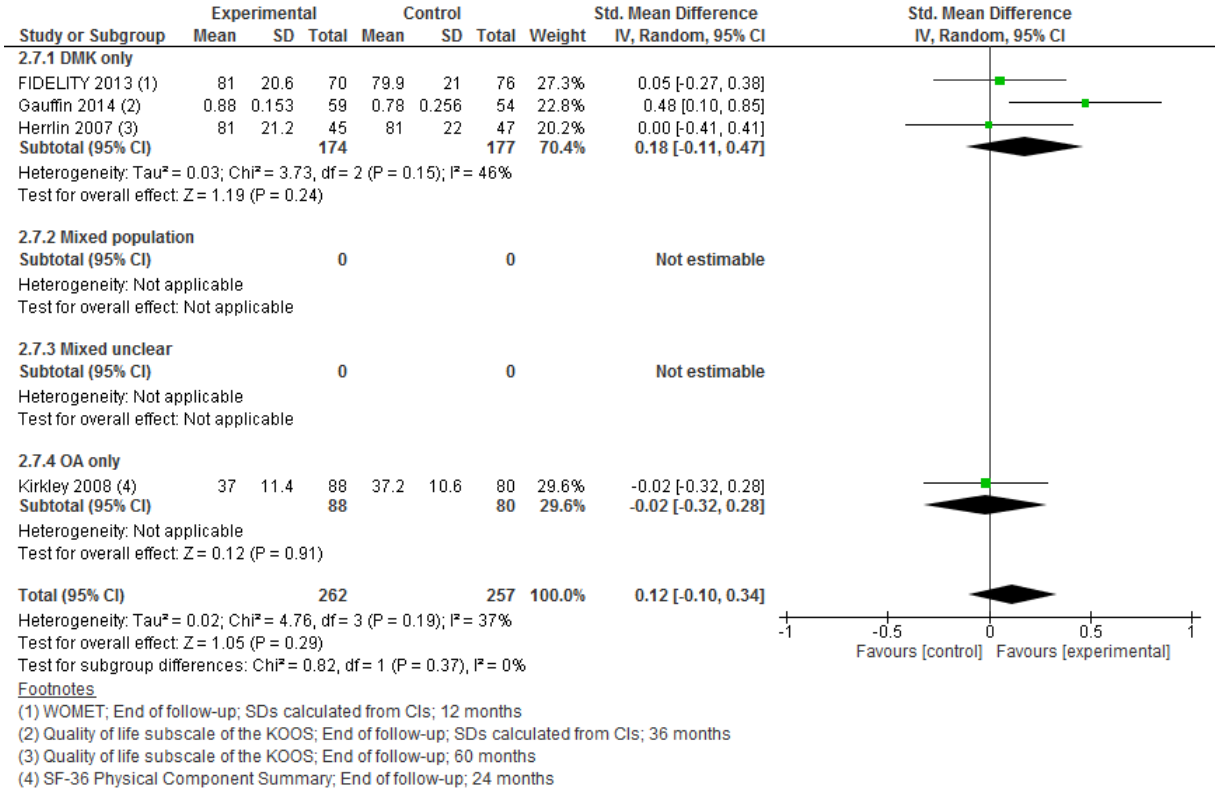


Figure 28 Disease-specific quality of life by DMK/OA classification at intermediate follow-up

Subgroups: non-active versus active

No subgroup analyses were conducted.

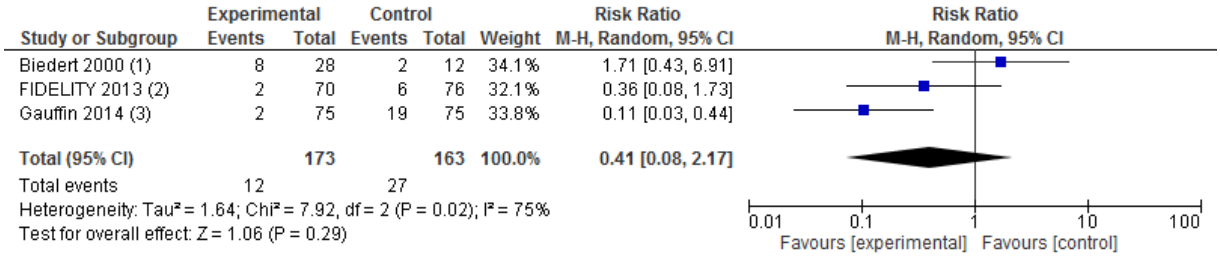
3.2.4.3.4 Adverse events or subsequent surgery

Reporting in RCTs on adverse events and subsequent surgery was inconsistent. RCTs reported number of patients with adverse events without subsequent surgery, or counted subsequent surgery as AEs, or only reported patients with subsequent surgery. Patients with adverse events combined with subsequent surgery were reported in three RCTs (Biedert 2000, FIDELITY, Gauffin 2014); four RCTs (FIDELITY, Gauffin 2014, Kise 2016, MeTeOR) reported patients with adverse events without subsequent surgery. There was no statistical difference between arthroscopy and comparator for adverse events combined with subsequent surgery (RR 0.41, 95% CI [0.08, 2.17], Figure 29, very low quality of evidence, Table 14). Heterogeneity was considerable (I^2=75%) and could not be explained by sensitivity analyses as there were too few RCTs. There was no statistical difference between arthroscopy and comparator for adverse events without subsequent surgery (RR 1.07, 95% CI [0.67, 1.70], Figure 30; low quality of evidence, Table 14). Heterogeneity was low (I^2=0%).

Patients with subsequent surgery were reported in six RCTs (Biedert 2000, FIDELITY, Forster 2003, Gauffin 2014, Herrlin 2007, Kise 2016). The relative risk of subsequent surgery for arthroscopy versus control was 0.24 (95% CI [0.14, 0.44], Figure 31; low quality of evidence, Table 14). (This is explained by a high cross-over rate or patients in control groups who were more

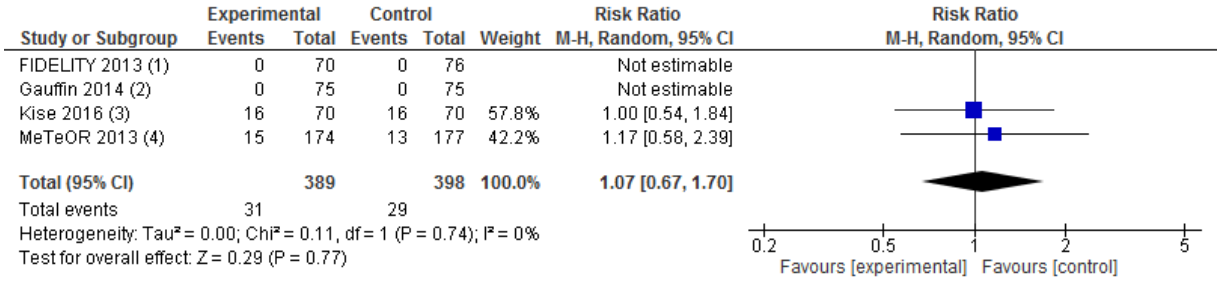
likely to undergo subsequent surgery, RR 4.14 95% CI [2.27, 7.14]). Heterogeneity for this outcome was low (I^2=0%). Eleven RCTs reported on cross-overs; an overview of reported cross-overs is presented in section 8.4.4 of Appendix 5.

Two RCTs reported the number of patients with side effects that occurred within two weeks (KIVIS) and five days (Saeed 2015) after arthroscopy and were not pooled in analyses. The KIVIS RCT reported knee swelling in four patients (5.6%) in the arthroscopy group and in four patients (5.1%) in the comparator group. No other AEs were reported. Saeed 2015 reported pain and mild effusion in 13 patients (21.7%) in the arthroscopy group and pain at injection site in eight patients (13.3%) in the comparator group.



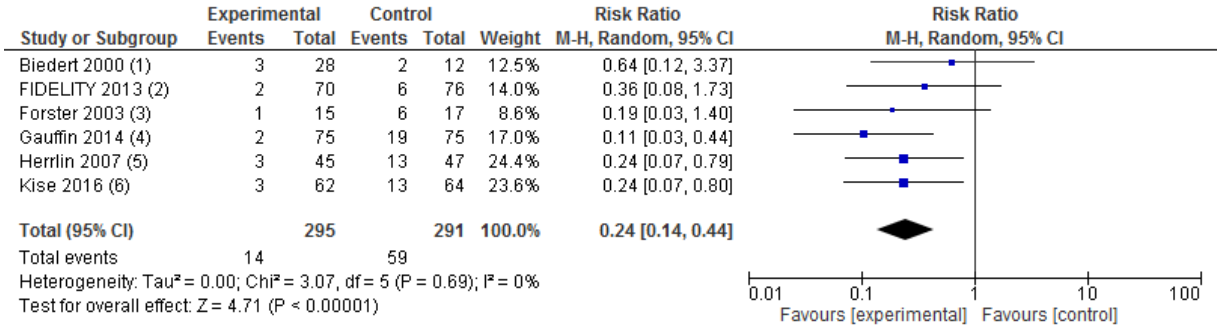
Footnotes
(1) Postoperative problems or complications (including post-intervention surgery); average 26.5 months
(2) Adverse events (including post-intervention surgical interventions); 12 months
(3) Adverse events (including post-intervention arthroscopic procedures); 36 months

Figure 29 Adverse events combined with subsequent surgery at intermediate follow-up



Footnotes
(1) Adverse events (excluding post-intervention surgical interventions); 12 months
(2) Adverse events (excluding post-intervention arthroscopic procedures); 36 months
(3) Adverse events (in index knee and only if treatment sought) (excluding post-intervention surgery); 24 months
(4) Adverse events (excluding post-intervention surgeries); 12 months

Figure 30 Adverse events without subsequent surgery at intermediate follow-up



Footnotes
(1) Post-intervention surgery; average 26.5 months
(2) Post-intervention surgery; 12 months
(3) Post-intervention surgery; 12 months
(4) Post-intervention surgery ; 36 months
(5) Post-intervention surgery; 60 months
(6) Post-intervetion surgery; 24 months

Figure 31 Subsequent surgery at intermediate follow-up

Subgroups: DMK versus OA

Subgroup analyses were not conducted for patients with adverse events combined with subsequent or without surgeries as there were too few RCTs reporting results. For patients with subsequent surgery at intermediate follow-up, four RCTs (FIDELITY, Gauffin 2014, Herrlin 2007, Kise 2016) included patients with DMK only, and in two RCTs (Biedert 2000, Forster 2003) the population was insufficiently described and the populations in these RCTs were classified as mixed unclear. Relative risks for subsequent surgery for arthroscopy versus control were 0.22 (95% CI [0.11, 0.42]) for the DMK only group and 0.39 (95% CI [0.11, 1.40]) for patients with mixed OA and DMK of unclear pathology. Relative risks for subsequent surgery in these subgroups did not statistically significantly differ and 95% CIs were largely overlapping. Figure 32 shows the resulting forest plot with the four subgroup estimates and overall effect estimate.

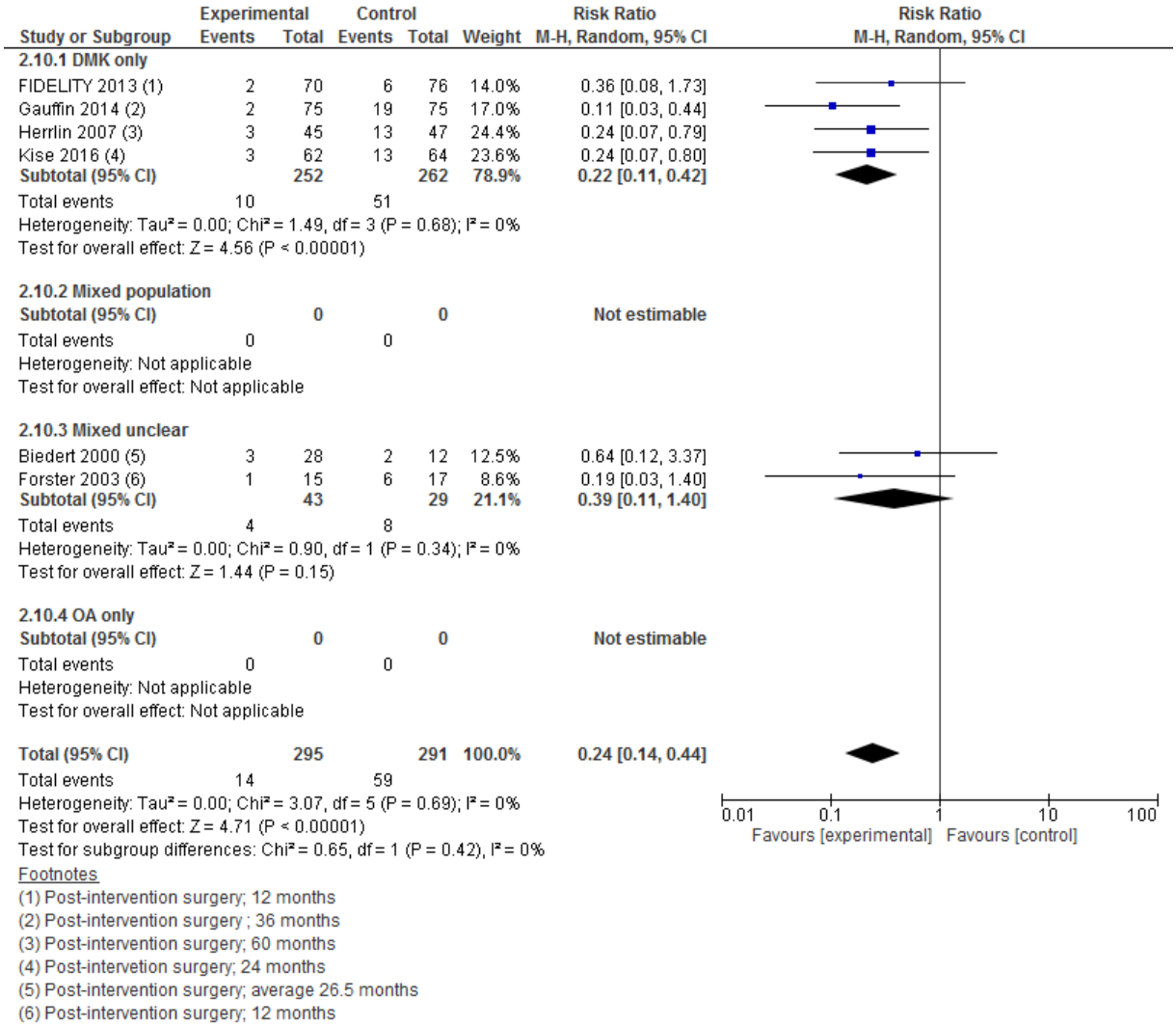


Figure 32 Subsequent surgery by DMK/OA classification at intermediate follow-up

Subgroups: non-active versus active

Subgroup analyses were not conducted for patients with adverse events combined with subsequent or without surgery as there were too few RCTs reporting results. For patients with subsequent surgery at intermediate follow-up, three RCTs (FIDELITY, Gauffin 2014, Herrlin 2007) compared arthroscopic surgery to non-active comparators, and three RCTs (Biedert 2000, Forster 2003, Kise 2016) to active comparators. The relative risk of subsequent surgery of arthroscopy compared to non-active treatment was 0.21 (95% CI [0.09, 0.46]) and the relative risk of subsequent surgery of arthroscopy compared to active treatment was 0.30 (95% CI [0.13, 0.72]). Relative risks for subsequent surgery in these subgroups did not statistically significantly differ and 95% CIs were largely overlapping. Figure 33 shows the resulting forest plot with the two subgroup estimates and overall effect estimate.

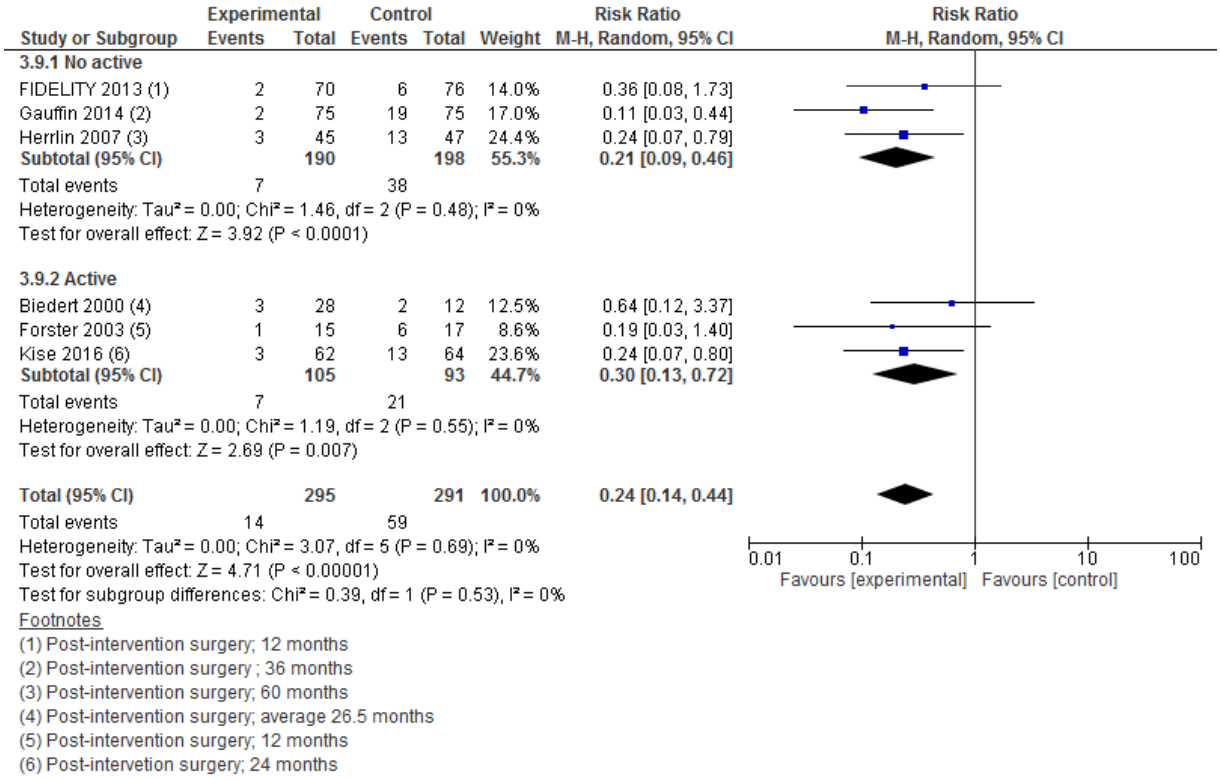


Figure 33 Subsequent surgeries by comparator type at intermediate follow-up

3.2.4.3.5 Serious adverse events

Four RCTs (FIDELITY, Kise 2016, MeTeOR, Saeed 2015) reported SAEs. The relative risk for SAEs of arthroscopy versus control was 1.83 (95% CI [0.39, 8.62], Figure 34; very low quality of evidence, Table 14). Heterogeneity for this outcome was low (I^2=0%).

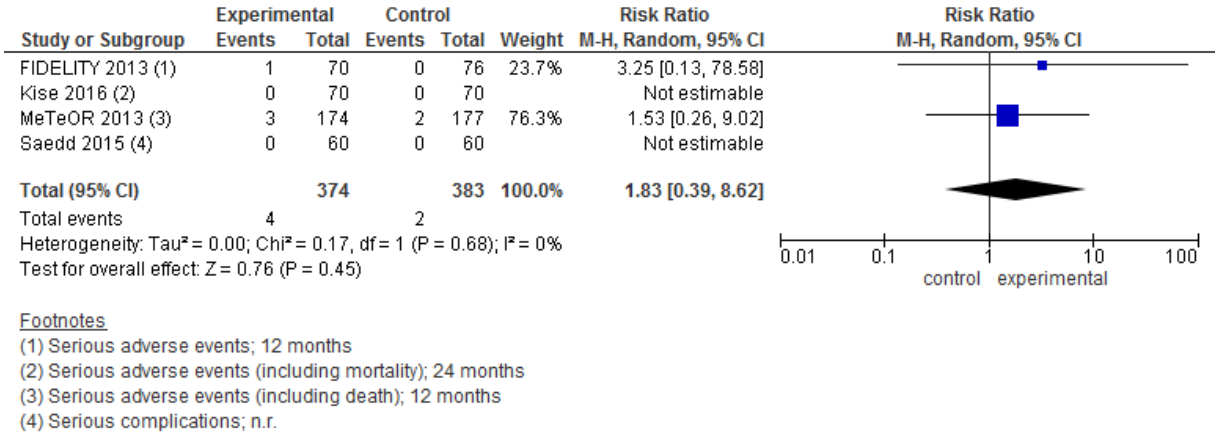


Figure 34 Serious adverse events at intermediate follow-up

Mortality

Five RCTs (Gauffin 2014, Kise 2016, MeTeOR, Merchan 1993, Moseley 2002) reported on mortality. Gauffin 2014 reported one death at 36 months (1.3%) in the arthroscopy group and MeTeOR reported one death at 12 months (0.6%) each from the arthroscopy and comparator groups, Merchan 1993 reported five deaths (12.5%) in the arthroscopy group and two deaths (5%) in the comparator group. Kise 2015 reported no serious adverse events (including deaths) at 24 months and Moseley reported no post-operative deaths. Of the RCTs reporting death, no information was given on if deaths occurred within 30 days of intervention.

Subgroups: DMK versus OA

Too few RCTs reported on serious adverse events to conduct subgroup analyses.

Subgroups: non-active versus active

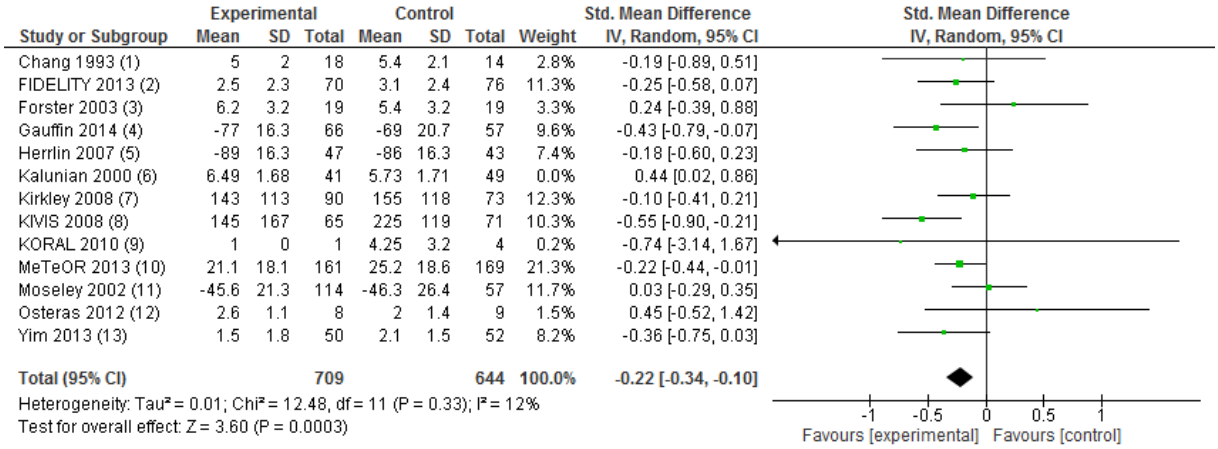
Too few RCTs reported on serious adverse events to conduct subgroup analyses.

3.2.4.4 Sensitivity analyses

Sensitivity analyses were conducted by excluding RCTs with extreme or opposite treatment effects.

3.2.4.4.1 Pain

At short-term follow-up, results for the outcome of pain were pooled across 13 RCTs. Excluding Kalunian 2000 from this analysis reduced the heterogeneity from I^2=44% to I^2=12% with an estimated effect size of SMD -0.22 (95% CI [-0.34, -0.10], Figure 35). Kalunian 2000 was the only RCT comparing arthroscopic lavage (irrigation of knee joint with 3,000 ml of saline) to diagnostic arthroscopy (irrigation of knee joint with 250 ml of saline). At 3 month follow-up, a statistically significant effect (SMD 0.44, 95% CI [0.02, 0.86], Figure 2) was found in favour of diagnostic arthroscopy whereas at was 12 months follow-up a statistically significant effect (SMD -0.59, 95% CI [-1.02, -0.17, Figure 5) in favour of arthroscopic lavage was found.



Footnotes

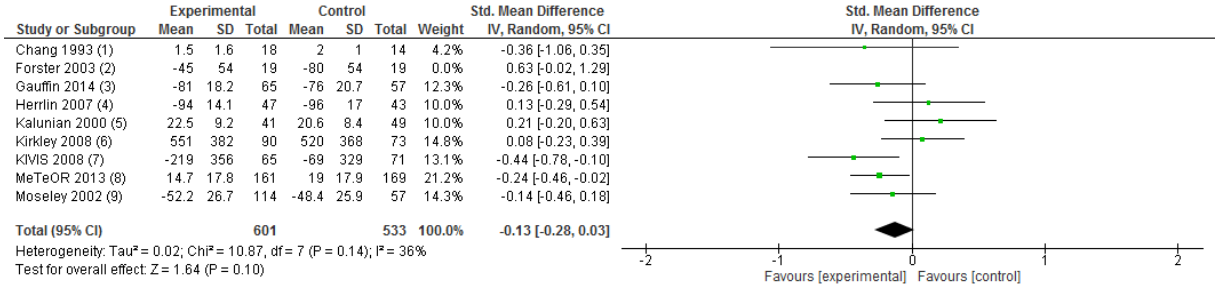
(1) Pain subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 3 months
(2) Knee pain afterexercise on NRS; End of follow-up; SDs calculated from CIs; 6 months
(3) VAS; End of follow-up; assumed mean; SDs imputed as upper limit of range of all reported SDs for VAS; 6 months
(4) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 3 months
(5) Pain subscale of the KOOS; End of follow-up; median reported; SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(6) Pain subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(7) Pain subscale of the WOMAC; End of follow-up; 6 months
(8) Pain subscale of the WOMAC; End of follow-up; median reported; extracted from figure 2 with SDs calculated from IQRs; 6 months
(9) VAS; End of follow-up; 2 months
(10) Pain subscale of the KOOS; End of follow-up; SDs calculated from CIs; 6 months
(11) Bodily pain subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 6 months
(12) VAS; End of follow-up; 3 months
(13) VAS; End of follow-up; SDs approximated from baseline SDs; 6 months

Figure 35 Pain at short-term follow-up, sensitivity analysis

3.2.4.4.2 Function

At short-term follow-up, results for the outcome of function were pooled across nine RCTs. Excluding Forster 2003 from this analysis reduced the heterogeneity from I^2=50% to I^2=36% with an estimated effect size of SMD -0.13, (95% CI [-0.28, 0.03], Figure 36).

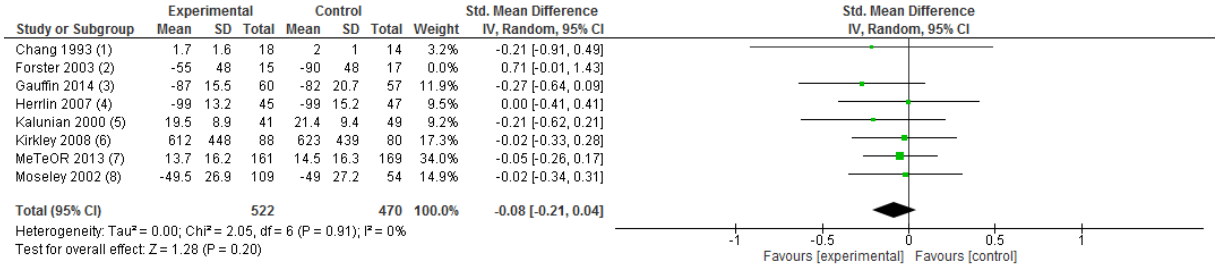
At intermediate follow-up, results from eight RCTs were pooled. Although heterogeneity was low (I^2=0%), to ensure that unequal baseline measures in the RCT by Forster 2003 did not affect the overall effect estimate, a sensitivity analysis was conducted. Excluding Forster 2003 from the analyses did not affect heterogeneity and only resulted in a slight change in the overall effect estimate (SMD -0.08, 95% CI [-0.21, 0.04], Figure 37).



Footnotes

(1) Physical function subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 3 months
(2) Function score of the KSRS; End of follow-up; assumed mean; SDs imputed as most conservative estimate to maintain stastical non-significance; multiplied by -1 to invert effects; 6 months
(3) Activities of daily living (ADL) subscale of KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 3 months
(4) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; median reported; extracted from figure 3 with SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(5) Function subscale of the WOMAC; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(6) Physical function subscale of the WOMAC; End of follow-up; 6 months
(7) Physical function subscale of the WOMAC; mean difference from baseline; effect multiplied by -1 to indicate improvement on scale; 6 months
(8) Physical function subscale of the WOMAC; End of follow-up; SDs caluclated from CIs; 6 months
(9) Physical functioning subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 6 months

Figure 36 Function at short-term follow-up, sensitivity analysis



Footnotes

(1) Physical function subscale of the AIMS; End of follow-up; SDs approximated from baseline SDs; 12 months
(2) Function score of the KSRS; End of follow-up; assumed mean; SDs imputed as most conservative estimate to maintain stastical non-significance; multiplied by -1 to invert effects; 12 months
(3) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 36 months
(4) Activities of daily living (ADL) subscale of the KOOS; End of follow-up; median reported; multiplied by -1 to invert effects; 60 months
(5) Function subscale of the KOOS; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 12 months
(6) Physical function subscale of the WOMAC; End of follow-up; 24 months
(7) Physical function subscale of the WOMAC; End of follow-up; SDs calculated from CIs; 12 months
(8) Physical functioning subscale of the SF-36; End of follow-up; multiplied by -1 to invert effects; 24 months

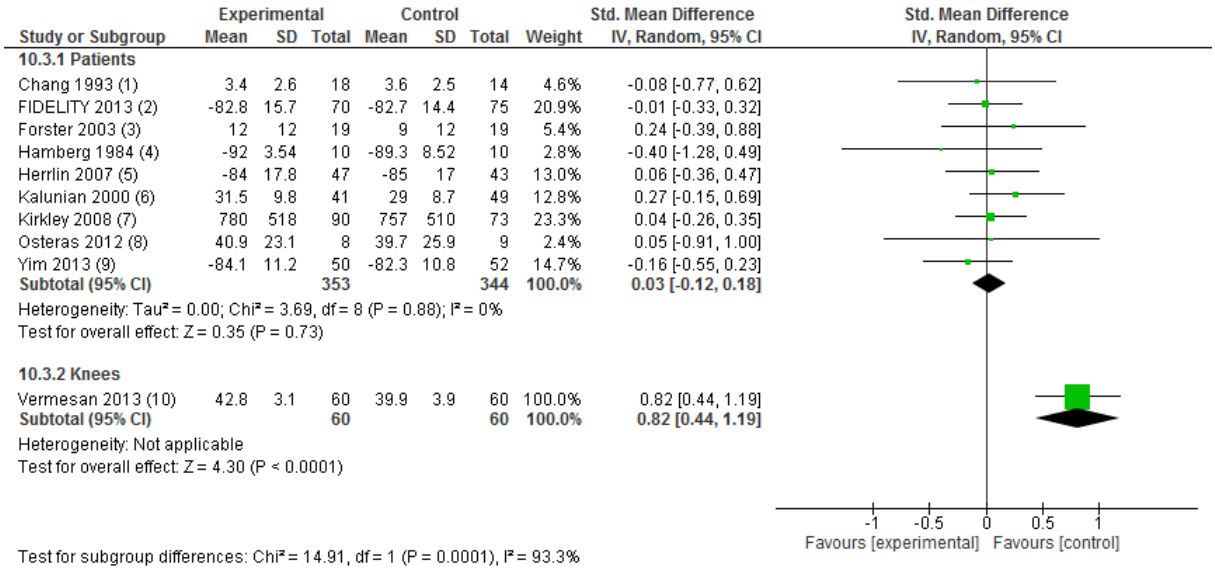
Figure 37 Function at intermediate follow-up, sensitivity analysis

3.2.4.4.3 Global assessment

Results of nine RCTs reporting on global assessment of patients were pooled at short-term follow-up. In the RCT by Hamberg 1984, results were presented for four groups: arthroscopic partial meniscectomy, arthroscopic total meniscectomy, open partial meniscectomy and open total meniscectomy; it was decided to combine results of both arthroscopic groups and both open surgery groups for the purpose of these analyses. To ensure heterogeneity and the overall effect estimate remained unchanged, a sensitivity analysis was performed replacing the combined arthroscopic and open surgery results with results comparing the arthroscopic partial meniscectomy group to the open partial meniscectomy group. When excluding patients with total meniscectomy, I^2 estimates for heterogeneity were unaffected and only slightly changed the overall effect estimate (SMD 0.03 95% CI [-0.12, 0.18], Figure 38).

Eight RCTs reported on global assessment in patients with intermediate follow-up. Excluding the RCT by Herrlin 2007 – the only RCT that had a statistically significant effect on global assessment in favour of comparator (SMD 0.46 [95% CI [0.04, 0.87], Figure 17) - from this analysis reduced

the heterogeneity from I^2=39% to I^2=12%; however, the effect estimate only changed slightly (SMD -0.03, 95% CI [-0.20, 0.14], Figure 39).



Footnotes

(1) VAS; End of follow-up; SDs approximated from baseline SDs; 3 months
(2) Lysholm knee score; End of follow-up; SDs calculated from CIs; multiplied by -1 to invert effects; 6 months
(3) Lequesne Index; End of follow-up; assumed mean; 6 months
(4) Partial meniscectomy: Lysholm point-scoring scale; End of follow-up; extracted from figure 1; multiplied by -1 to invert effects; 2 months
(5) Lysholm knee score; End of follow-up; median reported; SDs calculated from IQRs; multiplied by -1 to invert effects; 6 months
(6) WOMAC total score; End of follow-up; extracted from figure 1 with SDs calculated from assumed CIs; 3 months
(7) WOMAC total score; End of follow-up; 6 months
(8) KOOS total score; End of follow-up; KOOS total score has not been validated and is not recommended; 3 months
(9) Lysholm knee score; End of follow-up; SDs approximated from baseline SDs; multiplied by -1 to invert effects; 6 months
(10) OKS; End of follow-up; 1 month

Figure 38 Global assessment at short-term follow-up, sensitivity analysis

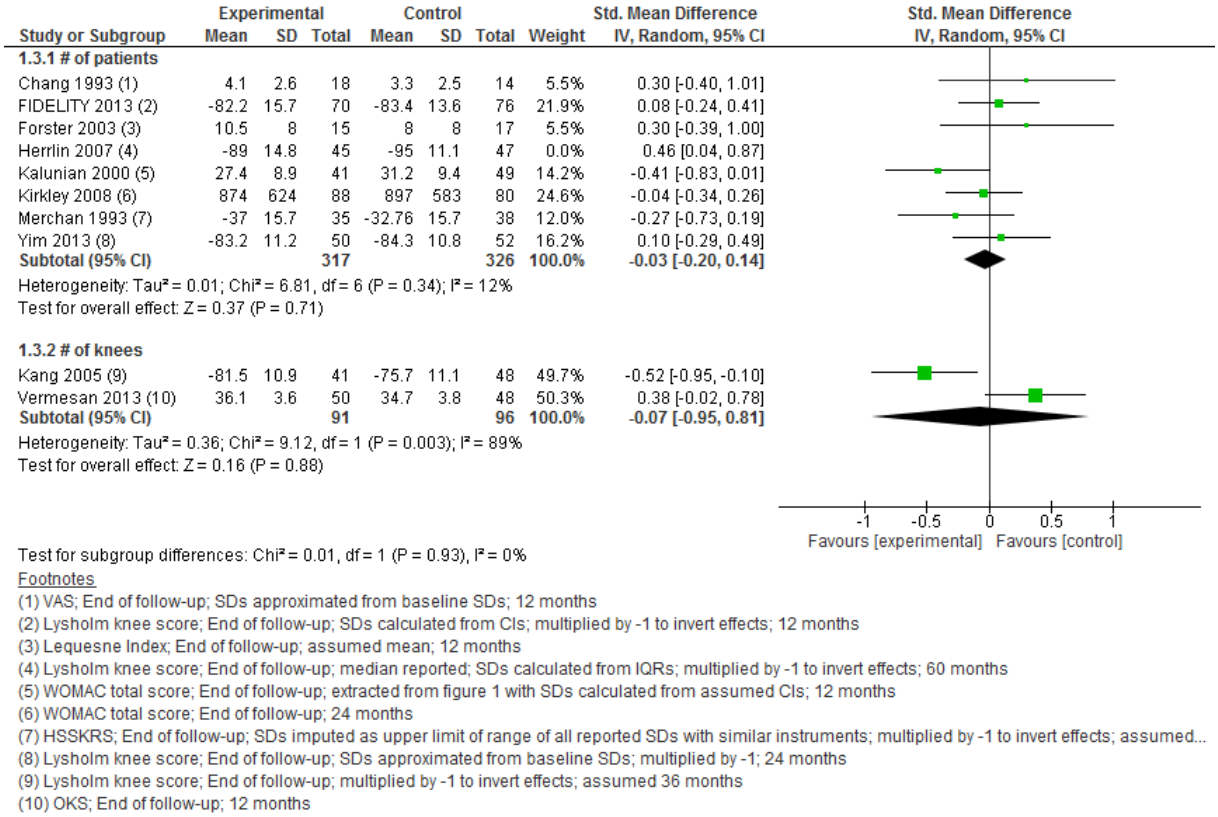


Figure 39 Global assessment at intermediate follow-up, sensitivity analysis

3.2.4.5 Summary of results

Table 13 and Table 14 present the GRADE summary of evidence including the certainty (quality) of evidence along with the effect estimate for each outcome at short-term (≤6 months) and intermediate follow-up (latest time point after six months and <7 years), respectively, for the clinical effectiveness of arthroscopy for treatment of degenerative knee.

3.2.4.5.1 Short-term follow-up (up to and including six months)

Knee pain at short-term follow-up was assessed in a total of 1,443 patients in 13 RCTs. Compared to control, arthroscopy had a statistically significant effect on pain (SMD -0.16, 95% CI [-0.31, -0.01], Figure 2; low quality of evidence, Table 13). Nine RCTs including 1,172 patients reported on the outcome of function and found no statistically significant effect (SMD -0.08, 95% CI [-0.26, 0.09], Figure 8; very low quality of evidence, Table 13). Nine RCTs that included 717 patients found no statistically significant difference in global assessment results, when comparing arthroscopy to control (SMD 0.03, 95% CI [-0.12, 0.17], Figure 14; low quality of evidence, Table 13). A total of 389 patients in three RCTs found no statistically significant effect on joint stiffness (SMD -0.09, 95% CI [-0.45, 0.27]), Figure 20; very low quality of evidence, Table 13) in favour of arthroscopy compared to control. Two RCTs with 282 patients found no statistically significant effect on health-related quality of life in favour of arthroscopy compared to control (SMD 0.18,

95% CI [-0.05, 0.42], Figure 25; very low quality of evidence, Table 13). Additionally, there was no statistically significant effect found in three RCTs with 375 patients on disease-specific quality of life in favour of arthroscopy compared to control (SMD 0.18, 95% CI [-0.02, 0.39], Figure 25; very low quality of evidence, Table 13).

The **overall quality of evidence** was judged to be **very low** because of the very low quality of evidence for the critical outcome of function at short-term follow-up.

3.2.4.5.2 Intermediate follow-up (latest point after six months and up to seven years)

Knee pain at intermediate follow-up was reported for a total of 1,274 patients in 10 RCTs. Compared to control, arthroscopy did not have a statistically significant, beneficial effect on reducing pain (SMD -0.11, 95% CI [-0.22, 0.00], Figure 5; low quality of evidence, Table 14). Eight RCTs including 1,024 patients reported on the outcome of function, and found no statistically significant effect (SMD -0.06, 95% CI [-0.18, 0.07], Figure 11; low quality of evidence, Table 14) in favour of arthroscopy compared to control. Eight RCTs that included 735 patients found no statistically significant effect on global assessment results for arthroscopy versus control (SMD 0.04, 95% CI [-0.16, 0.23], Figure 17; low quality of evidence, Table 14). A total of 258 patients in two RCTs found no statistically significant effect on joint stiffness (SMD -0.18, 95% CI [-0.75, 0.39], Figure 23; very low quality of evidence, Table 14) in favour of arthroscopy compared to control. Two RCTs reporting on 383 patients indicated a higher relative risk for total knee replacement in the arthroscopy group compared to controls, but this estimate was not statistically significant (RR 1.25, 95% CI [0.38, 4.19], Figure 24; very low quality of evidence, Table 14). There was no statistically significant effect in three RCTs with 428 patients on health-related quality of life in favour of arthroscopy compared to control (SMD 0.17, 95% CI [-0.02, 0.36], Figure 26; moderate quality of evidence, Table 14). In addition, there was no statistically significant effect found in four RCTs with 525 patients on disease-specific quality of life in favour of arthroscopy compared to control (SMD 0.17, 95% CI [-0.02, 0.36], Figure 26; moderate quality of evidence, Table 14).

Three RCTs with 336 patients reported on AEs combined with subsequent surgery; there was no statistically significant difference in AE occurrence combined with subsequent surgery in favour of arthroscopy compared to control (RR 0.41, 95% CI [0.08, 2.17], Figure 29, very low quality of evidence, Table 14). There was no statistically significant difference in AE occurrence without subsequent surgery in four RCTs with 787 patients (RR 1.07, 95% CI [0.67, 1.70], Figure 30; low quality of evidence, Table 14). A total of six RCTs with 586 patients found that the relative risk of subsequent surgery for arthroscopy versus control was 0.24 (95% CI [0.14, 0.44], Figure 31; low quality of evidence, Table 14). Four RCTs with 757 patients reported on SAEs; there was no statistically significant difference in SAE occurrence in favour of arthroscopy compared to control (RR 1.83, (95% CI [0.39, 8.62]), Figure 34; very low quality of evidence, Table 14).

The **overall quality of evidence was** judged to be **low** because of the low quality of evidence for the critical outcomes of pain, function and global assessment at intermediate follow-up.

Table 13 Arthroscopy compared to any comparator for the treatment of degenerative changes- short-term follow-up (≤6 months)

Outcomes	№ of participants (RCTs)	Relative effect (95% CI)	Anticipated absolute effects		Certainty of the evidence (GRADE)
			Risk with any comparator	Risk difference with arthroscopy	
Pain	1,443 (13 RCTs)	-	-	SMD **0.16 lower** (0.31 lower to 0.01 lower)	⨁⨁◯◯ LOW [a,b]
Function	1,172 (9 RCTs)	-	-	SMD **0.08 lower** (0.26 lower to 0.09 higher)	⨁◯◯◯ VERY LOW [c,d]
Global assessment	717 (9 RCTs)	-	-	SMD **0.05 higher** (0.1 lower to 0.2 higher)	⨁⨁◯◯ LOW [e,f]
Joint stiffness	389 (3 RCTs)	-	-	SMD **0.09 lower** (0.45 lower to 0.27 higher)	⨁◯◯◯ VERY LOW [g,h,i]
Health-related quality of life	282 (2 RCTs)	-	-	SMD **0.18 higher** (0.05 lower to 0.42 higher)	⨁◯◯◯ VERY LOW [i,j]
Disease-specific quality of life	375 (3 RCTs)	-	-	SMD **0.18 higher** (0.02 lower to 0.39 higher)	⨁◯◯◯ VERY LOW [i,k]

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **SMD:** Standardised mean difference; **RCT:** Randomised controlled trial

Outcomes	№ of participants (RCTs)	Relative effect (95% CI)	Anticipated absolute effects		Certainty of the evidence (GRADE)
			Risk with any comparator	Risk difference with arthroscopy	

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

a. The limitations of the RCTs were very serious because risk of selection bias (random sequence generation) was unclear in 8 RCTs and risk of selection bias (allocation concealment) was unclear in 7 RCTs; risk of performance bias was high in 10 RCTs; risk of detection bias was unclear in 1 RCT and high in 9 RCTs; risk of attrition bias was unclear in 4 RCTs and high in 5 RCTs and risk of selective reporting was unclear in 10 RCTs and high in 1 RCT.
b. Inconsistency was not serious because heterogeneity was explained by Kalunian 2000.
c. The limitations of the RCTs were very serious because risk of selection bias (random sequence generation) was unclear in 5 RCTs and risk of selection bias (allocation concealment) was unclear in 5 RCTs; risk of performance bias was high in 7 RCTs; risk of detection bias was unclear in 1 RCT and high in 6 RCTs; risk of attrition bias was unclear in 3 RCTs and high in 4 RCTs and risk of selective reporting was unclear in 8 RCTs.
d. Inconsistency was serious because heterogeneity could only be partially explained by Forster 2003; however, heterogeneity was still considerable with confidence intervals having little overlap and effects having different directions.
e. The limitations of the RCTs were very serious because risk of selection bias (random sequence generation) was unclear in 4 RCTs and risk of selection bias (allocation concealment) was unclear in 6 RCTs and high in 1 RCT; risk of performance bias was unclear in 1 RCT and high in 6 RCTs; risk of detection bias was unclear in 2 RCTs and high in 5 RCTs; risk of attrition bias was unclear in 4 RCTs and high in 2 RCTs and risk of selective reporting was unclear in 6 RCTs and high in 1 RCT.
f. Inconsistency was not serious because heterogeneity was low and confidence intervals were widely overlapping.
g. The limitations of the RCTs were very serious because risk of selection bias (allocation concealment) was unclear in 3 RCTs; risk of performance bias was high in 2 RCTs; risk of detection bias was high in 2 RCTs; risk of attrition bias was unclear in 1 RCT and high in 1 RCT and risk of selective reporting was unclear in 2 RCTs.
h. Inconsistency was serious because heterogeneity couldn't be explained as too few RCTs were available for sensitivity analyses.
i. Imprecision was serious because the total sample size was below the optimal information size (OIS).

j. The limitations of the RCTs were very serious because risk of selection bias (random sequence generation) was unclear in 1 RCT and risk of selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 2 RCTs; risk of detection bias was high in 2 RCTs; risk of attrition bias was high in 1 RCT and risk of selective reporting was unclear in 1 RCT.
k. The limitations of the RCTs were very serious because risk of selection bias (random sequence generation) was unclear in 1 RCT and risk of selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 3 RCTs; risk of detection bias was high in 3 RCTs; risk of attrition bias was unclear in 1 RCT and high in 1 RCT and risk of selective reporting was unclear in 2 RCTs.

Table 14 Arthroscopy compared to any comparator for the treatment of degenerative changes- intermediate follow-up (≥6 months and ≤7 years)

Outcomes	№ of participants (RCTs)	Relative effect (95% CI)	Anticipated absolute effects		Certainty of the evidence (GRADE)
			Risk with any comparator	Risk difference with arthroscopy	
Pain	1,274 (10 RCTs)	-	-	SMD **0.11 lower** (0.22 lower to 0)	⊕⊕◯◯ LOW [a,b]
Function	1,024 (8 RCTs)	-	-	SMD **0.06 lower** (0.18 lower to 0.07 higher)	⊕⊕◯◯ LOW [b,c]
Global assessment	735 (8 RCTs)	-	-	SMD **0.02 higher** (0.2 lower to 0.23 higher)	⊕⊕◯◯ LOW [d,e]
Joint stiffness	258 (2 RCTs)	-	-	SMD **0.18 lower** (0.75 lower to 0.39 higher)	⊕◯◯◯ VERY LOW [f,g,h]
Total knee replacement	383 (2 RCTs)	**RR 1.25** (0.38 to 4.19)	26 per 1,000	**6 more per 1,000** (16 fewer to 82 more)	⊕◯◯◯ VERY LOW [i,j]
Health-related quality of life	428 (3 RCTs)	-	-	SMD **0.17 higher** (0.02 lower to 0.36 higher)	⊕⊕⊕◯ MODERATE [k]
Disease-specific quality of life	525 (4 RCTs)	-	-	SMD **0.06 higher** (0.11 lower to 0.23 higher)	⊕⊕⊕◯ MODERATE [l]
Adverse events (mixed with/without subsequent surgeries)	336 (3 RCTs)	**RR 0.41** (0.08 to 2.17)	166 per 1,000	**98 fewer per 1,000** (152 fewer to 194 more)	⊕◯◯◯ VERY LOW [g,j,m]

Outcomes	№ of participants (RCTs)	Relative effect (95% CI)	Anticipated absolute effects		Certainty of the evidence (GRADE)
			Risk with any comparator	Risk difference with arthroscopy	
Adverse events (without subsequent surgeries)	787 (4 RCTs)	**RR 1.07** (0.67 to 1.70)	73 per 1,000	**5 more per 1,000** (24 fewer to 51 more)	⨁⨁◯◯ LOW [n,o]
Subsequent surgeries	586 (6 RCTs)	**RR 0.24** (0.14 to 0.44)	203 per 1,000	**154 fewer per 1,000** (174 fewer to 114 fewer)	⨁⨁◯◯ LOW [p,q]
Serious adverse events	757 (4 RCTs)	**RR 1.83** (0.39 to 8.62)	5 per 1,000	**4 more per 1,000** (3 fewer to 40 more)	⨁◯◯◯ VERY LOW [j,r]

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **SMD:** Standardised mean difference; **RR:** Risk ratio; **RCT:** Randomised controlled trial

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

a. The limitations of the RCTs were very serious because the risk of selection bias (random sequence generation) was unclear in 6 RCTs and the risk of selection bias (allocation concealment) was unclear in 5 RCTs; risk of performance bias was high in 7 RCTs; risk of detection bias was unclear in 1 RCT and high in 6 RCTs; risk of attrition bias was unclear in 4 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 7 RCTs and high in 1 RCT.
b. Inconsistency was not serious because heterogeneity was low and confidence intervals were widely overlapping.
c. The limitations of the RCTs were very serious because the risk of selection bias (random sequence generation) was unclear in 5 RCTs and the risk of selection bias (allocation concealment) was unclear in 4 RCTs; risk of performance bias was high in 6 RCTs; risk of detection bias was unclear in 1 RCT and high in 5 RCTs; risk of attrition bias was unclear in 3 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 7 RCTs.

d. The limitations of the RCTs were very serious because the risk of selection bias (random sequence generation) was unclear in 3 RCTs and high in 1 RCT and the risk of selection bias (allocation concealment) was unclear in 5 RCTs and high in 1 RCT; risk of performance bias was high in 6 RCTs; risk of detection bias was unclear in 2 RCTs and high in 4 RCTs; risk of attrition bias was unclear in 3 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 5 RCTs and high in 1 RCT.
e. Inconsistency was not serious because heterogeneity was explained by Herrlin 2007.
f. The limitations of the RCTs were very serious because the risk of selection bias (allocation concealment) was unclear in 2 RCTs; risk of performance bias was high in 1 RCT; risk of detection bias was high in 1 RCT; and risk of selective reporting was unclear in 1 RCT.
g. Inconsistency was serious because heterogeneity couldn't be explained as too few RCTs were available for sensitivity analyses.
h. Imprecision was serious because the total sample size was below the optimal information size (OIS).
i. The limitations of the RCTs were very serious because the risk of selection bias (random sequence generation) was unclear in 2 RCTs and the risk of selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 2 RCTs; risk of detection bias was unclear in 1 RCT and high in 1 RCT; risk of attrition bias was high in 1 RCT and risk of selective reporting was unclear in 2 RCTs.
j. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both appreciable harm or benefit (relative risk increase or decrease greater than 25%) in favour of no arthroscopy and because the total number of events was <300.
k. The limitations of the RCTs were serious because the risk of selection bias (random sequence generation) was unclear in 1 RCT and the risk of selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 2 RCTs; risk of detection bias was high in 2 RCTs; risk of attrition bias was high in 1 RCT and risk of selective reporting was unclear in 1 RCT.
l. The limitations of the RCTs were serious because the risk of selection bias (random sequence generation) was unclear in 1 RCT and the risk of selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 3 RCTs; risk of detection bias was high in 3 RCTs; risk of attrition bias was unclear in 1 RCT and high in 1 RCT and risk of selective reporting was unclear in 2 RCTs.
m. The limitations of the RCTs were serious because the risk of selection bias (random sequence generation) was unclear in 1 RCT and high in 1 RCT, and the risk of selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 2 RCTs; risk of detection bias was unclear in 1 RCT and high in 1 RCT; risk of attrition bias was unclear in 1 RCT and high in 1 RCT, and risk of selective reporting was unclear in 2 RCTs.
n. The limitations of the RCTs were serious because the risk of selection bias (random sequence generation) was unclear in 2 RCTs; risk of performance bias was high in 3 RCTs; risk of detection bias was high in 3 RCTs; risk of attrition bias was high in 1 RCT and risk of selective reporting was unclear in 2 RCTs.
o. Imprecision was very serious because the 95% CI of the effect estimate is sufficiently wide to include both the null and appreciable harm (relative risk increase greater than 50%) in favour of no arthroscopy and because the total number of events was <300.
p. The limitations of the RCTs were serious because the risk of selection bias (random sequence generation) was unclear in 2 RCTs and high in 1 RCT, and the risk of selection bias (allocation concealment) was unclear in 2 RCTs; risk of performance bias was high in 5 RCTs; risk of detection bias was unclear in 2 RCTs and high in 3 RCTs; risk of attrition bias was unclear in 2 RCTs and high in 2 RCTs and risk of selective reporting was unclear in 4 RCTs.
q. Imprecision was serious because the total number of events was <300.
r. The limitations of the RCTs were serious because the risk of selection bias (random sequence generation) was unclear in 2 RCTs and the risk of selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 3 RCTs; risk of detection bias was high in 3 RCTs; risk of attrition bias was unclear in 1 RCT and risk of selective reporting was unclear in 2 RCTs.

3.2.5 PICO 2

3.2.5.1 Results of the risk of bias assessment

The risks of selection bias (allocation concealment) and performance bias (blinding of RCT participants and personnel) were both high in the single available RCT. The risk of detection bias (blinding of outcome assessor) was unclear, while the risk of attrition bias was high. The risk of reporting bias in the RCT was unclear. A summary for the risk of bias assessment is shown in Table 15 and details can be found in Appendix 6.

Table 15 Results of risk of bias assessment, PICO 2

RCT ID	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete continuous outcome data (attrition bias)	Incomplete binary data (attrition bias)	Selective reporting (reporting bias)
Weale 1998	Low	High	High	Unclear	High	High	Unclear

3.2.5.2 Outcomes

Pain was reported after discharge using a 10 cm VAS. The mean pain score after discharge for 45 patients in the short-stay group (inpatient) was 3.4 (SD 2.7), while the mean pain score after discharge for the day-case group (outpatients) was 3.5 (SD 2.6). However, the analgesics use and days off work were longer in the short-stay (3.2 days ± 4.9 and 19.4 days ±10.6, respectively) than in the day-case group (2.6 days ± 3.8 and 11.6 days ± 16.9). The authors found no statistically significant difference between the two groups for all endpoints. Three participants (6%) crossed-over from the short-stay group and discharged themselves on day of surgery, while six participants (12%) crossed-over from the day-case group and were admitted overnight. One patient in the short-stay group required readmission to the hospital due to a wound haematoma.

3.2.5.3 Subgroups

None.

3.2.5.4 Summary of results

Because only one RCT was found, the quality of evidence was not assessed using GRADE. The overall quality of evidence was judged to be very low because RCT data comparing inpatient arthroscopic treatment to outpatient arthroscopic treatment were essentially absent.

4 Cost-effectiveness and budget impact

4.1 Methods

The economic section of this HTA report consists of two main parts.

The first part aims at gaining an understanding of the cost-effectiveness of therapeutic knee arthroscopy compared to any other treatment in patients with degenerative changes of the knee – irrespective of whether they are primarily due to meniscal damage, osteoarthritis of the knee or a mix of both and, to the extent possible, of the cost-effectiveness of inpatient compared to outpatient therapeutic knee arthroscopy in this patient population.

Health economic endpoints considered included costs, QALYs, and incremental cost-effectiveness ratio (i.e. additional cost per life year gained or QALY gained).

The perspectives of interest were the 'KVG perspective' (third party perspective, taking into account the direct medical costs of all healthcare services covered by the Swiss statutory health insurance, irrespective of actual payer) and the societal perspective.

The assessment was based on national and international health economic literature, to the extent available. The analysis included the following steps, which are detailed in subsequent sections.

- Literature search
- Screening of the search results to identify eligible studies and of studies
- Extraction of information and assessment of the quality of reporting in eligible cost-effectiveness studies
- Assessment of the eligible cost-effectiveness studies in terms of transferability to Switzerland
- For the studies found to be qualitatively transferable, adaptation of cost-effectiveness results to Switzerland
- Synopsis of findings

The second part of the economic section aims at estimating the budget impact of meniscus derangement in the Swiss healthcare system from a health insurance system (KVG) perspective.

Following decisions taken during the scoping process, the patient population of primary interest in the clinical and cost-effectiveness parts of this HTA (degenerative changes of the knee primarily due to meniscal damage) differs slightly from that used in the budget impact part (any degenerative changes of the knee).

4.1.1 Cost-effectiveness

4.1.1.1 Literature search

The aim of the literature search was to identify literature on the costs and cost-effectiveness of knee arthroscopy compared to other interventions or conservative treatment, for patients with

degenerative changes of the knee. All types of economic evaluation studies were considered and checked for relevant content: cost-effectiveness-, cost-benefit-, cost-utility- and cost-minimization analyses.

A search strategy was developed to identify all relevant literature in the following electronic databases: Medline and Embase databases including abstracts by using OvidSP (including Ovid MEDLINE(R), Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily Update, Embase), the Cochrane Library and the Centre for Review and Dissemination (CRD) database including the Database of Abstracts of Reviews of Effects (DARE), Cochrane reviews, Health Technology Assessments (HTA) and the Economic Evaluation Database from the UK National Health Service (NHS EED). Search strings for additional databases were not developed because the selection of databases aforementioned has been described as both sufficient and very efficient.[67]

The search strings were obtained by integrating and combining the search strings used in the clinical effectiveness part of this HTA report and published search strings for health economic analyses from the InterTASC Information Specialists' SubGroup (ISSG) website (www.york.ac.uk/inst/crd/intertasc). The following filters, described as highly sensitive in Ovid MEDLINE and EMBASE, were included: The National Health Service Economic Evaluation Database (NHS EED) filter, the NHS Quality Improvement Scotland filter and the Royle filter published in 2003.[67 68] Additional filters such as the Scottish Intercollegiate Guidelines Network (SIGN) filter (http://www.sign.ac.uk/methodology/filters.html), the McKinlay et al.[69] filter, the Wilczynski et al.[70] filter and the Sassi et al.[71] filter were also included. Unspecific abbreviations such as CUA or CBA were not used.

The search strategy and MeSH terms used for each database are described in the search strategy reported in Appendix C, section 9.1 Literature search strategy. The search was performed on 11th July 2017.

4.1.1.2 Screening of the search results

The screening of the literature was divided into three phases. In the first phase, all results of the literature search were screened by title. Titles containing relevant keywords such as knee arthroscopy, knee surgery, costs, value, cost-effectiveness, cost-benefit, cost-utility, quality of life, and burden were considered as potentially relevant.

All papers with potentially relevant titles then proceeded to the second phase, the screening by abstract. In this phase, abstracts were screened for relevant quantitative results (e.g. costs, life year gained, QALYs, or ICERs) or for sentences suggesting potentially relevant content in the full text version.

Potentially relevant abstracts proceeded to the third phase, in which full texts were screened. Articles were then classified as being relevant, partially relevant, or irrelevant.

- Relevant articles needed to meet the following criteria:
    - The article reported a full-scale incremental cost-effectiveness analysis, ideally but not necessarily with an endpoint of cost per QALY gained or cost per life year gained.
    - The 'PIC' of the PICO corresponded to the one defined in the scoping document and used in the systematic review part of this HTA report.

    - The analysis was performed for a jurisdiction with broadly similar socioeconomic characteristics as Switzerland (e.g. North, Central and Western European countries, the USA, Canada, Australia and New Zealand).
- Partially relevant articles were defined as those potentially containing useful additional information concerning effectiveness or costs (but not both). Depending on the quality and quantity of information available from relevant articles, some partially relevant articles were used as an additional source of information and for comparison.
- The remainder of articles were classified as irrelevant.

4.1.1.3 Extraction of information

For the eligible cost-effectiveness studies (i.e. relevant articles as defined above), data extraction was performed, covering the following information:

- Study population (including country, age of patients)
- Intervention
- Comparator(s)
- Setting and perspective of the study
- Cost types included and cost year
- Type of model
- Time horizon
- Discount rate
- Approach to sensitivity analysis
- Effectiveness
- Costs
- Incremental cost-effectiveness ratio (ICER)

A brief, qualitative characterization of each study was prepared in the results section, covering methodological approaches taken, main data sources, methodological issues and potential meaningfulness of the results for Switzerland.

Quality of reporting was assessed against the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 24-item checklist, recommended by the ISPOR Health Economic Evaluations Publication Guidelines Task Force.[72] . The complete CHEERS checklist is reported in the Appendix (section 9.2). Articles were considered of good, average, or poor quality if they addressed respectively more than 80%, between 60% and 80%, or less than 60% of the CHEERS checklist items.

4.1.1.4 Assessment of transferability

International cost-effectiveness studies were assessed for 'qualitative transferability' to Switzerland. A variety of authors have worked on criteria for assessing such transferability between jurisdictions.[73] Methodologic papers published early by O'Brien et al.[74], Welte at al.[75], and Drummond et al.[76] suggested the use of multistep procedures. In the present study, a modified approach as described below and summarised schematically in Figure 40 was adopted. One key reason for using this modified approach is that none of the original models underlying the eligible

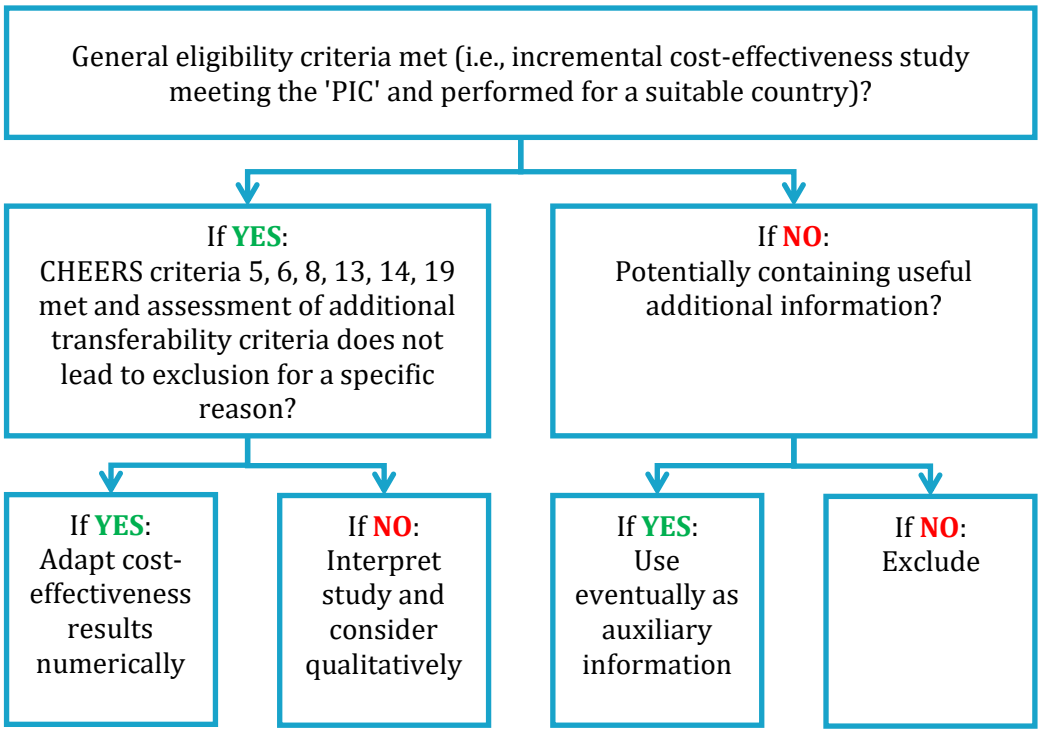
cost-effectiveness studies were available. Therefore, actual model recalculations on the basis of localized input parameters, such as e.g. unit costs, were not possible.

- The most important criteria for qualitative transferability were already covered by the eligibility criteria. Essentially, the assessment of eligibility excluded studies that were not full-scale health economic evaluation studies assessing incremental cost-effectiveness, did not meet the 'PIC', or were performed for countries very different from Switzerland in terms of socioeconomic characteristics. All remaining studies were thus expected to meet CHEERS criteria 4 (population), 7 (intervention / comparator[s]) and 10 (outcome measures).
- Studies not meeting CHEERS items 5, 6, 8, 13, 14 and 19 were regarded as not transferable due to lack of key information. In relation to item 19, the availability of costs and outcomes of interest for both the intervention and the comparator strategies was considered fundamental. Where articles only reported ICERs, the underlying study was considered non-transferable (see Figure 40).
- The remaining studies were considered qualitatively transferable, and underwent numerical adaptation of cost-effectiveness results to Switzerland (see next section), if scrutiny of additional transferability factors taken from O'Brien et al. [74] and Welte et al. [75] did not preclude this for a specific reason. In all other cases, the results of the scrutiny of transferability factors were used qualitatively.

The following additional transferability factors were considered:

- *Methodological characteristics:*
  - Perspective of cost assessment
  - Discount rate
  - Medical cost approach
  - Productivity cost approach
- *Healthcare system characteristics:*
  - Absolute and relative prices in healthcare
  - Clinical practice variation; differences in resource use, incentives and regulations for health-care providers
  - Technology availability
- *Population characteristics:*
  - Demography
  - Disease incidence and prevalence
  - Case-mix
  - Life expectancy
  - Health-status preferences
  - Acceptance, compliance, incentives to the patients
  - Productivity and work-loss time

For most cost-effectiveness studies meeting the general eligibility criteria, severe transferability problems were not expected since methodological and population characteristic were expected to be similar to Switzerland. Regarding healthcare system characteristics, large differences in availability of technology were not expected. Costs of healthcare services were adapted numerically (see section 4.2.1.2). Studies falling in the lower left box of Figure 40 were regarded as being qualitatively transferable.



Criterion 5: State relevant aspects of the system(s) in which the decision(s) need(s) to be made.
Criterion 6: Describe the perspective of the study and relate this to the costs being evaluated.
Criterion 8: State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.
Criterion 13: *Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.
Criterion 14: Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.
Criterion 19: For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.

Figure 40 Steps for study selection and determination of transferability to Switzerland

4.1.1.5 Adaptation of cost-effectiveness results to Switzerland

The adaptation of direct medical cost results to Switzerland was performed in three distinct steps (Figure 41): correction for different levels of resource utilization, correction for different prices of healthcare services, and correction for change in level of resource utilization and prices over time. Subsequently, adapted ICERs were calculated dividing adapted cost differences by originally reported QALY differences. This process was used in a previous HTA report for the SMB[77] and was described in a recent publication.[78] It cannot be interpreted as achieving realistic ICERs for Switzerland, but a certain approximation of cost-effectiveness levels to be expected for

Switzerland. The results of international cost-effectiveness studies, reported for different countries and in different currencies, were made more comparable.

1. *Resource utilization:* The types and quantities of healthcare resources used differ between countries. For the same disease, patients in Switzerland often receive more medical treatments than in other countries (i.e. they are treated more intensively for an equivalent diagnosis). Therefore, a "quantity correction" is necessary. The quantity correction was based on the Organization for Economic Co-operation and Development (OECD) statistics of healthcare expenses per capita, corrected for purchasing power.[79] A correction for differences in resource utilization levels (unaffected by price levels) was thus achieved.
2. *Prices of healthcare services:* The price for the same healthcare service or treatment is often different across countries. A "price correction" was achieved by applying correction factors provided by the OECD. Such purchasing power parities represent the proportional costs for identical products in two countries.[79]
3. *Change in costs over time:* Healthcare costs change over time. For eligible cost-effectiveness studies performed in countries other than Switzerland, the two steps described above achieve an adaptation of reported costs. However, the resulting estimates are valid for the same cost year as in the original study. Additional correction for the development of costs over time was necessary. In the case of a specific disease and set of treatment strategies, costs may change over time due to mere price changes without changes in resource utilization, or resource utilization for the treatment of the disease of interest may also change. In our 'base case' approach, the latter was assumed, and that changes in resource utilization occur with the same cost impact as at the level of total Swiss healthcare expenditures. The resulting correction was based on the yearly growth rates of total Swiss healthcare expenditures, as reported by the Swiss Federal Office of Statistics.[80]

The adaptation of cost data representing indirect costs followed a similar approach. However, the first of the above-described steps is irrelevant in the case of indirect costs. The third step was based on the change in Swiss salaries over time.[81]

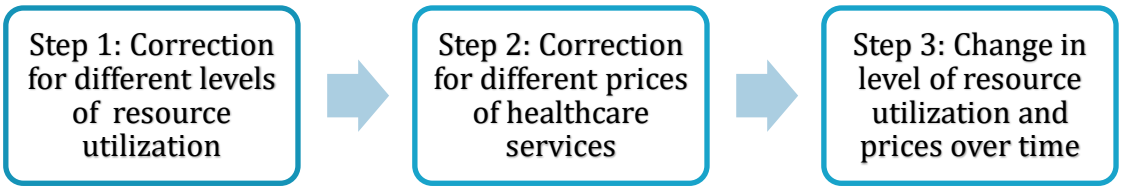


Figure 41 Adaptation of direct medical cost results to Switzerland

4.1.1.6 Synthesis of findings

The resulting different pieces of information were synthesized. This necessarily involved an element of interpretation, but it was an explicit aim to make all related assumptions transparent. The discussion includes a critical review of possible sources of uncertainty. Comparisons of the assumptions and of the data used by the various cost-effectiveness analyses were provided.

4.1.2 Budget impact analyses

The aim of the budget impact analysis was to estimate the budget impact of meniscus derangement in the Swiss healthcare system, from a health insurance system (KVG) perspective. The analysis consisted of two main steps: first, the annual frequency of knee arthroscopy in Switzerland was investigated; second, based on annual frequency, the total annual costs were estimated.

As specified in the scoping, the budget impact analysis was focused on arthroscopic meniscectomy and associated minor arthroscopic interventions for degenerative changes in the knee joint. For this reason, cases that were reimbursed by the accident insurance (i.e. patients with traumatic knee injuries) were excluded from the cost estimations. It should be emphasized that patients with a diagnosis of osteoarthritis of the knee (ICD-10 code M17.xx) were not included in the budget impact assessment because the preliminary analysis showed that this code was mainly used in combination with knee replacement, whereas the combination with arthroscopy or knee derangement occurred rarely. More in detail, in 2014 there were 25,454 hospitalisations receiving an ICD-10 code of M17 as main diagnosis. Of these cases, only 250 were classified with one of the selected SwissDRG codes. The majority was in contrast classified with SwissDRG I43B: Knee replacement implants. This decision not to include ICD-10 code of M17 was done in agreement with the SFOPH.

To assess the frequency arthroscopic knee interventions in the inpatient setting in Switzerland, the Swiss Hospital Statistics of the years 2010 to 2014 provided by the Swiss Federal Statistical Office were analysed. The Swiss Hospital Statistics is a collection of data on all patients that were hospitalized in a Swiss hospital during a specific year. This included a total of 1,345,245 hospitalisations in 2010, 1,363,697 in 2011, 1,353,521 in 2012, 1,374,439 in 2013, and 1,400,830 in 2014. The collected information included patient characteristics, diagnoses, and performed interventions.

As already noticed during the scoping process conducted on behalf of the SFOPH, patients undergoing arthroscopic knee surgery may be identified through three different types of variables reported in the Swiss Hospital Statistics:

- Swiss DRG and All Patients Diagnosis-Related Groups (APDRG) codes indicating the main hospitalisation reason for each patient,
- ICD-10 codes indicating the main diagnosis and secondary diagnoses for meniscus derangement, or
- Swiss classification of the surgical interventions (CHOP) codes indicating which knee arthroscopic treatments were performed.

The Swiss Hospital Statistics only includes patients who were hospitalized (i.e. outpatients were excluded). The frequency of knee arthroscopy in the ambulatory setting, including outpatients who were treated in a hospital, was based on an analysis of medical tariff (TARMED) codes provided by the SFOPH. This analysis provided information about the total number of knee arthroscopies and concomitant arthroscopic knee interventions in the ambulant setting in 2013 and 2014. Other information (e.g. concerning patient characteristics and underlying diagnosis were not available. Nevertheless, the identified TARMED codes allowed a rough estimation of the annual number of ambulant arthroscopic interventions to the meniscus.

To assess costs, data from the diagnosis-related case costs statistics (Statistik diagnosebezogener Fallkosten) provided by the Swiss Federal Office of Statistics, as well as cost estimations of a Swiss insurance company were used. Mean costs per case were applied to the frequency of arthroscopic knee interventions to estimate the annual budget impact.

4.1.2.1 Swiss DRG codes / APDRG codes

The diagnosis-related group (DRG) is a system to classify hospital cases, as a basis for reimbursement. Every inpatient case receives one single DRG classification code. The Swiss DRG system (www.swissdrg.org) was introduced country-wide in January 2012 and is based on the German DRG (www.g-drg.de). The DRG code is assigned according to patient characteristics, ICD-10 diagnoses and the treatments provided to the patients (CHOP codes, www.medcode.ch). In particular, the main diagnosis and the main treatment play a fundamental role, as they are the main drivers of the classification. Thus, depending on how coders in hospitals code the diagnoses and treatments of cases (i.e. depending on their sequence), patients with similar diagnoses and treatments may be assigned to different Swiss DRG codes.

For this analysis, six Swiss DRG codes suggesting potentially relevant knee interventions were identified during the scoping process:

- I04Z: Revision or replacement of the knee with complicating diagnose or arthrodesis
- I23B: Infection/Inflammation of Bone and Joint W Misc Musculoskeletal Procs W Sev or Mod CC
- I18A: Arthroscopy, incl. Biopsy or other interventions on bone or joints, age < 16 years
- I18B: Arthroscopy, incl. Biopsy or other interventions on bone or joints, age > 15 years
- I30Z: Complex knee interventions
- I59Z: Other interventions on Humerus, Tibia, Fibula, ankle joint or relatively complex interventions on knee joint, elbow joint and forearm

The Swiss DRG codes were available in the Swiss Hospital Statistics 2012 to 2014.

For patients hospitalized before 2012, only the APDRG codes were available. Unfortunately, these older codes and the actual Swiss DRG codes were only partially comparable. For the present analyses, following APDRG codes that may be related to knee interventions were identified:

- 221: Synovectomy and ligament reconstruction at the knee, with KK
- 222: Synovectomy and ligament reconstruction at the knee, without KK
- 232: Arthroscopy
- 917: Other knee surgeries, with KK
- 918: Other knee surgeries, without KK
- 1222: Synovectomy and ligament reconstruction at the knee, without KK, with multiple interventions
- 1232: Arthroscopy, with multiple interventions

The APDRG codes were available for the Swiss Hospital Statistics 2010 and 2011.

It should be emphasized that some Swiss DRG codes (I23B, I18A, I18B, and I59Z) and APDRG codes (232, 1232) potentially included interventions at other joints. Moreover, the Swiss DRG code I30Z as well as the APDRG codes 917 and 918 may have included non-arthroscopic

interventions. Therefore, to correctly identify patients with knee problems it was necessary to combine the Swiss DRG codes with the ICD-10 and CHOP codes.

4.1.2.2 ICD-10 codes

The International Classification of Diseases (ICD) is a health care classification system providing a system of diagnostic codes for classifying diseases. The ICD is internationally recognised as a standard diagnostic tool for epidemiology, health management and clinical purposes, is revised periodically, and is currently in its tenth revision (ICD-10, www.who.int/classifications/icd/).

For this analysis, the scoping process identified five ICD-10 codes suggesting a relevant diagnosis for knee/meniscus derangement:

- M23.2: Derangement of meniscus due to old tear or injury
- M23.3: Other meniscus derangements
- M23.8: Other internal derangements of knee
- M23.9: Internal derangement of knee, unspecified
- S83.2: Tear of meniscus, current injury

It is important to remark that a patient may have multiple concurrent diagnoses. For this reason, in the Swiss Hospital Statistics a patient can receive simultaneously one main diagnosis ("Hauptdiagnose") and several secondary diagnoses ("Nebendiagnosen"). Depending on the severity of the diseases, hospital coders will decide which is the principal reason for the hospitalization (the main diagnosis) and which ones are secondary problems (secondary diagnoses). For example, a person involved in a car accident may arrive at the hospital with a cranial fracture and a derangement of the knee. In this case, the cranial fracture would probably be the main diagnosis, whereas the knee derangement would be a secondary diagnosis.

For the present assessment the main diagnosis and up to 10 secondary diagnoses were considered per patient. All diagnoses were considered equally, independent of the sequence.

The total number of relevant diagnoses did not directly reflect the total number of hospitalized patients because many patients received more than one single, relevant diagnosis (e.g. several patients had different M23.2 codes like "M23.21 - Derangement of anterior horn of medial meniscus due to old tear or injury" and "M23.22 - Derangement of posterior horn of medial meniscus due to old tear or injury"). For the cost calculation, all cases with at least one relevant ICD-10 code were considered equally important (i.e. cases with only one relevant diagnosis were considered like those with multiple relevant ICD-10 codes).

4.1.2.3 CHOP codes

CHOP is the Swiss classification of surgical interventions (www.medcode.ch). The following potentially relevant codes were available from the Swiss Hospital Statistics 2011-2014:

- 80.16.10: Arthroskopische Arthrotomie des Kniegelenkes, Gelenkspülung mit Drainage
- 80.16.11: Arthroskopische Arthrotomie des Kniegelenkes, Entfernung freier Gelenkkörper
- 80.16.12: Arthroskopische Arthrotomie des Kniegelenkes, Einlegen oder Entfernen eines Medikamentträgers
- 80.26.00: Arthroskopie des Knies, n.n.bez.

- 80.26.10: Arthroskopie des Knies - diagnostisch
- 80.26.20: Arthroskopisch assistierte Versorgung einer Fraktur am Kniegelenk
- 80.26.99: Arthroskopie des Knies, sonstige
- 80.36.20: Arthroskopische Gelenkbiopsie am Knie
- 80.6X.10: Meniskektomie am Knie, arthroskopisch, partiell -
- 80.6X.11: Arthroskopisch, total - Meniskektomie am Knie
- 80.76.10: Arthroskopische Synovektomie am Kniegelenk
- 80.86.10: Arthroskopische lokale Exzision oder Destruktion am Kniegelenk
- 80.96.10: Arthroskopische Exzision am Kniegelenk
- 80.96.20: Arthroskopisch - Entnahme eine Knorpeltransplantat am Kniegelenk
- 81.47.11: Arthroskopisch - Refixation eines osteochondral Fragment am Kniegelenk
- 81.47.13: Arthroskopisch - Subchondral Spongiosaplastik am Kniegelenk
- 81.47.18: Arthroskopisch - Knorpeltransplantation und Implantation von in-vitro hergestellten Gewebekultur am Kniegelenk
- 81.47.22: Arthroskopisch - Knorpelglättung am Kniegelenk
- 81.47.24: Arthroskopisch - Subchondral Knocheneröffnung am Kniegelenk
- 81.47.25: Arthroskopisch - Subchondral Knocheneröffnung am Kniegelenk mit Einbringen eines azellulär Implantat
- 81.47.51: Arthroskopisch - Knorpeltransplantation mit OATS (osteoarticular transfer system)-Verfahren, Mosaikplastik am Kniegelenk
- 81.47.90: Arthroskopisch - Sonstige Rekonstruktion am Kniegelenk
- 81.99.1A: Arthroskopisch Revision eines Gelenk, Kniegelenk
- 81.99.3A: Arthroskopisch Operation am Gelenkknorpel und an den Meniskus, Kniegelenk
- 81.99.82: Arthroskopisch Operationen an Gelenk und Gelenkstruktur, Kniegelenk - Sonstige

In the year 2010, Swiss Hospital Statistics merged several relevant CHOP codes. The following potentially relevant groups of CHOP codes were identified for this year:

- 80.16: Sonstige Arthrotomie des Knies, merging 80.16.11 and 80.16.12
- 80.26: Arthroskopie des Knies, merging 80.26.00, 80.26.10, 80.26.20, and 80.26.99
- 80.36: Gelenksbiopsie am Knie, including 80.36.20
- 80.6: Knee meniscectomy, merging 80.6X.10 and 80.6X.11
- 80.76: Arthroskopisch Synovektomie am Kniegelenk including 80.76.10
- 80.86: Arthroskopisch lokal Exzision oder Destruktion am Kniegelenk including 80.86.10
- 80.96: Sonstige Exzision am Knie merging 80.96.10 and 80.96.20
- 81.47: Sonstige Rekonstruktion am Kniegelenk, including 81.47.11, 81.47.13, 81.47.18, 81.47.22, 81.47.24, 81.47.25, 81.47.51, and 81.47.90
- 81.99: Sonstige Operationen an Gelenken und Gelenkstrukturen including 81.99.1A. 81.99.3A, and 81.99.82

4.1.2.4 Patient identification in the Swiss Hospital Statistics

Patients with relevant knee problems, diagnoses, or treatments were identified through DRG codes, ICD-10 codes, or CHOP codes. The results of the scoping process have shown that relevant DRG codes, relevant ICD-10 codes, and relevant CHOP-codes only partially overlap. For example, not all patients classified under a relevant DRG code received a relevant diagnosis (ICD-10 code)

or an arthroscopic intervention of the knee (CHOP code). Similarly, not all patients who received some relevant diagnoses (ICD-10 codes) or relevant arthroscopic treatments (CHOP-codes) were classified with one of the DRG codes mentioned above.

For this budget impact analysis, two analytical strategies were used: in the first strategy, only patients who reported at the same time a relevant DRG code, at least one relevant diagnosis (ICD-10 code), and at least one relevant treatment (CHOP code) were included. In the second strategy, all patients who had at the same time at least one relevant diagnosis (ICD-10 code) and one relevant treatment (CHOP code), irrespective of the reported DRG codes, were included.

4.1.2.5 TARMED codes

The TARMED (Tarif Médical) is a standardized tariff system for outpatient medical services launched the first of January 2004. It encompasses about 4,000 tariff positions, which label and assess services provided by doctors (www.fmh.ch/ambulante_tarife/tarmed-tarif.html). The following list provides the codes for knee arthroscopy (24.5610) and all other arthroscopic interventions that may be performed during a surgical procedure to the knee in an ambulatory setting:

- 24.5610: Arthroskopie Kniegelenk
- 24.5615: + Resektion einer Gelenkzyste/Sehnenzyste/tiefen oder oberflächlichen Bursa im Kniegelenkbereich, als Zuschlagsleistung
- 24.5620: + Entfernung freier Gelenkkörper bei Arthroskopie Knie
- 24.5630: + Plicaresektion bei Arthroskopie Knie
- 24.5640: + Synoviektomie subtotal bei Arthrose bei Arthroskopie Knie
- 24.5650: + Synoviektomie subtotal bei {pcP}/postinfektiös bei Arthroskopie Knie
- 24.5660: + Retinakulumspaltung (lateral release) bei Arthroskopie Knie
- 24.5670: + Retinakulumnaht bei Arthroskopie Knie, jede Methode
- 24.5680: + Osteophytenabtragung u/o Notch-Plastik bei Arthroskopie Knie
- 24.5690: + Anlegen einer Spüldrainage bei Arthroskopie Knie
- 24.5700: + Meniskustoilette bei Arthroskopie Knie
- 24.5710: + Resektion Meniscus medialis/Meniscus lateralis, partiell/total, bei Arthroskopie Knie, pro Meniskus
- 24.5720: + Resektion diskoider Meniscus medialis/Meniscus lateralis/ Meniskusganglion, partiell/total bei Arthroskopie Knie
- 24.5730: + Naht Meniscus medialis/Meniscus lateralis bei Arthroskopie Knie
- 24.5735: + Mikrofrakturierung oder Pridie-Bohrung(en)
- 24.5740: + Versorgung bei Osteochondrosis dissecans bei Arthroskopie Knie, Fixation des Dissekates
- 24.5750: + Plastische Versorgung bei Osteochondrosis dissecans bei Arthroskopie Knie
- 24.5760: + Versorgung Ruptur des vorderen Kreuzbandes bei Arthroskopie mittels Naht u/o transossärer Reinsertion
- 24.5770: + Versorgung Ruptur des vorderen Kreuzbandes mittels Naht u/o transossärer Reinsertion sowie Augmentationsplastik bei Arthroskopie Knie, jede Methode
- 24.5780: + Versorgung Ruptur des vorderen Kreuzbandes mittels autoplastischem Ersatz bei Arthroskopie Knie, jede Methode
- 24.5790: + Versorgung Ruptur des vorderen Kreuzbandes mittels alloplastischem Ersatz bei Arthroskopie Knie, jede Methode

- 24.5810: + Versorgung Ruptur hinteres Kreuzband mit. Naht u/o transossäre Reinsertion am femoralen Ursprung sowie Augmentationsplastik bei Arthroskopie Knie, jede Methode
- 24.5820: + Versorgung Ruptur hinteres Kreuzband mittels autoplastischem Ersatz bei Arthroskopie Knie
- 24.5830: + Versorgung Ruptur hinteres Kreuzband mittels alloplastischem Ersatz bei Arthroskopie Knie
- 24.5840: + Zuschlag für tibiale Fixation bei Versorgung Ruptur hinteres Kreuzband bei Arthroskopie Knie, jede Methode
- 24.5850: + Versorgung Tibiakopffraktur bei Arthroskopie Knie mittels Osteosynthese
- 24.5860: + Entfernung des Osteosynthesematerials bei Arthroskopie Knie

The frequencies of use of these codes provided by the Swiss Federal Office of Public Health were based on data from the SASIS Tarifpool (www.sasis.ch) and covered the majority of ambulatory cases in Switzerland (91-93% in 2014 and 79-88% in 2013). The data were projected to the whole Swiss population (see section 4.2.2.7).

Based on TARMED, the differentiation between degenerative knee cases and traumatic knee cases was not possible because no information concerning diagnoses and insurance coverage was available. For the base case cost calculations, the percentage of cases covered by accident insurance calculated for the Swiss Hospital Statistics were applied.

The frequencies provided represented the number of cases (not patients). Thus, it was not possible to assess whether there were patients who had repeated arthroscopic surgery during the same calendar year.

4.1.2.6 Costs

Alongside the two analytical strategies used to estimate the number of inpatients (see section 4.1.2.4), costs of inpatient procedures were estimated using two distinct approaches.

The first approach was based on the diagnosis-related case costs statistics (Statistik diagnosebezogener Fallkosten) provided by the Swiss Federal Office of Statistics, which provides the mean costs per case for each Swiss DRG or APDRG code during a single calendar year.[82] Costs were available for the years 2011 until 2014. For 2010, the costs provided in 2011 were applied.

Table 16 and

Table 17 summarise the mean costs per case for the relevant APDRG codes in 2011 and for the Swiss DRG codes between 2012 and 2014.

These costs per case were combined with the number of cases identified through the presence of at least one DRG code, ICD-10 code, and CHOP code in the Swiss Hospital Statistics 2010 to 2014, according to the first strategy to estimate patient numbers described in section 4.1.2.4. Moreover, as already decided during the scoping, cases that were reimbursed by the accident insurance were excluded from the cost estimations.

The second approach was based on an estimated cost per patient undergoing arthroscopic knee surgery between 2010 and 2014. According to estimations from the insurance company Assura, this cost can range between CHF 3,700 for patients with compulsory insurance coverage (including a cantonal contribution of CHF 1,900) and CHF 13,200 for patients with supplementary private insurance.[83] [84] For the purpose of the main analysis, it was assumed that all hospitalized

patients would be covered by the compulsory health insurance (i.e. that all would cost CHF 3,700), because the aim was to assess budget impact from a KVG perspective. The cost estimate of CHF 3,700 was combined with the number of all patients who had, in combination, at least one relevant diagnosis (ICD-10 code) and one relevant treatment (CHOP code), according to the second strategy to estimate patient numbers described in section 4.1.2.4 Here again, cases with traumatic knee injuries (i.e. cases who were reimbursed be the accident insurance) were excluded.

For the costs of ambulatory procedures, costs per single intervention or sub-intervention were available (e.g. CHF 429 for a knee arthroscopy, CHF 240 for a meniscus resection during a knee arthroscopy, CHF 119 for cleaning the meniscus during a knee arthroscopy). However, these costs did not represent the total costs per intervention, since they did not include, for example, the costs of the consultation and use of the operating theatre, patient documentation, or anaesthesia. Moreover, the combinations of sub-procedures used to treat the patients were unknown (only the total number of ambulatory interventions was available). Therefore, these cost data were not used in estimating the costs of outpatient cases. Instead, an estimate of cost per case by the insurance company Assura was again used. According to this source, the mean cost for an arthroscopic surgery of the meniscus performed in an ambulatory setting would be approximately CHF 2,400.[83] [84] This cost estimate was applied to all ambulatory cases.

It should be emphasised that the cost estimations were based on the number of identified cases (not on the number patients). Although it is relatively rare, during a single calendar year one single same patient may undergo a knee surgery more than once. Patients who were treated repeatedly were considered as separate cases.

Table 16 Mean costs per case according to APDRG codes in 2011

APDRG Code	Description	2011	
		Mean CHF	SD
221	Synovectomy and ligament reconstruction at the knee, with KK	10,901	10,300
222	Synovectomy and ligament reconstruction at the knee, without KK	7,076	2,370
232	Arthroscopy	5,748	4,046
917	Other knee surgeries, with KK	9,498	10,630
918	Other knee surgeries, without KK	4,102	2,492
1222	Synovectomy and ligament reconstruction at the knee, without KK, with multiple interventions	6,251	3,151
1232	Arthroscopy, with multiple interventions	6,764	3,889

Abbreviation: KK, Komorbiditäten und Komplikationen (Comorbidities and Complications)

Table 17 Mean costs per case according to Swiss DRG codes between 2012 and 2014

DRG Code	Description	2012		2013		2014	
		Mean CHF	SD	Mean CHF	SD	Mean CHF	SD
I04Z	Revision or replacement of the knee with complicating diagnose or arthrodesis	31,661	19,731	33,953	17,157	33,092	16,370
I12B	Infection/Inflammation of Bone and Joint W Misc Musculoskeletal Procs W Sev or Mod CC	19,781	11,646	18,964	10,701	18,362	11,224
I18A	Arthroscopy, incl. Biopsy or other interventions on bone or joints, age < 16 years	6,989	4,129	6,674	5,887	6,591	3,205
I18B	Arthroscopy, incl. Biopsy or other interventions on bone or joints, age > 15 years	4,770	2,773	4,988	2,999	5,123	3,811
I30Z	Complex knee interventions	9,083	2,961	9,027	3,018	9,063	2,898
I59Z	Other interventions on Humerus, Tibia, Fibula, ankle joint or relatively complex interventions on knee joint, elbow joint and forearm	5,241	3,099	5,525	3,099	5,640	3,860

4.1.2.7 Sensitivity and scenario analyses

Since the selection of patients for the budget impact analysis required a series of assumptions, a number of sensitivity analyses and scenario analyses were performed.

The first analytical approach to estimate inpatient costs described in section 4.1.2.4 was based on the diagnosis-related case costs statistics. Mean costs per case according to different DRG codes were multiplied with the number of cases that reported at the same time a relevant DRG code, at least one relevant diagnosis, and at least one relevant treatment (excluding patients with traumatic knee injuries). In sensitivity analyses, the mean costs per case were varied by ±20%. Moreover, two additional patient selection scenarios were investigated: first, all cases that were classified with a relevant DRG code were included (regardless of whether they had a relevant ICD-10 code or a relevant CHOP code). Second, all cases with a relevant DRG code except Swiss DRG code I59Z (which potentially includes interventions to other joints or bones) and except Swiss DRG code I18A (which is used for patients younger than 16 years) were considered.

In the second analytical approach to estimate inpatient costs, mean costs per case undergoing an arthroscopic knee surgery estimated by the insurance company Assura were applied to all cases that were reported at the same time with at least one relevant ICD-10 code and one relevant CHOP code. In the base case scenario it was assumed that all eligible cases would have compulsory health insurance (i.e. cost CHF 3,700). In the sensitivity analysis the estimated costs per case varied by 20%. In the scenario analysis all eligible cases were divided according to their insurance status (compulsory, semi-private, or private insurance). For all patients with basic insurance, total costs of CHF 3,700 were applied. The costs for patients with a private insurance were assumed to be CHF 13,200. Finally, for the semi-private cases, the mean between compulsory and private insurance costs were applied (i.e. CHF 8,450). As for the first strategy, cases that were reimbursed by the accident insurance (i.e. with traumatic knee injuries) were excluded from the cost estimations.

In the estimation of the costs of ambulatory interventions, the mean cost for an arthroscopic surgery of the meniscus estimated by the insurance company Assura (CHF 2,400) was combined with the number of knee arthroscopies according to TARMED positions. In the sensitivity analysis the estimated cost per case was varied by ±20%. Moreover, the ambulatory costs were re-estimated assuming that all cases or only half of them were eligible (i.e. that 100% or 50% of the ambulatory cases received the intervention due to a degenerative knee problem relating to the meniscus).

4.2 Results

The first part focuses on the cost-effectiveness of therapeutic knee arthroscopy compared to any other treatment in patients with degenerative changes of the knee according to the published literature. In the second part, the results of the budget impact analysis are reported.

4.2.1 Cost-effectiveness

A total of 512 citations were identified from the electronic databases searches. Following the removal of duplicates (n=79), full citations were reviewed (Figure 42). Based on the title and

abstract, 409 citations were excluded due to inappropriate comparator or non-comparative design; character of a review or commentary piece; inappropriate outcome measure; or no relevant cost information given. A total of 24 citations were included for full text review, of which another 20 were excluded for the same reasons as stated above. The remaining four articles fulfilled the inclusion criteria (Table 18). They were included in the systematic review and assessed using the CHEERS checklist [72] and the algorithm shown in Figure 40. All four studies fulfilled the above-defined criteria for qualitative transferability, which was needed to make a study suitable for numerical adaptation of ICER results to Switzerland.

Table 18 Identification (ID) of included cost-effectiveness analyses and corresponding publications

ID	Reference (only relevant research articles)
Hutt 2015[85]	Hutt JR, Craik J, Phadnis J, Cobb AG. Arthroscopy for mechanical symptoms in osteoarthritis: a cost-effective procedure. *Knee Surg. Sports Traumatol. Arthrosc.*; 2015; 23: 3545–3549.
Losina 2015[86]	Losina E, Dervan EE, Paltiel AD, Dong Y, Wright RJ, Spindler KP, Mandl LA, Jones MH, Marx RG, Safran-Norton CE, Katz JN. Defining the Value of Future Research to Identify the Preferred Treatment of Meniscal Tear in the Presence of Knee Osteoarthritis. *PLoS One* United States; 2015; 10: e0130256.
Lubowitz 2011[87]	Lubowitz JH, Appleby D. Cost-effectiveness analysis of the most common orthopaedic surgery procedures: knee arthroscopy and knee anterior cruciate ligament reconstruction. *Arthroscopy* United States; 2011; 27: 1317–1322.
Marsh 2016[88]	Marsh JD, Birmingham TB, Giffin JR, Isaranuwatchai W, Hoch JS, Feagan BG, Litchfield R, Willits K, Fowler P. Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. *BMJ Open* England; 2016; 6: e009949.

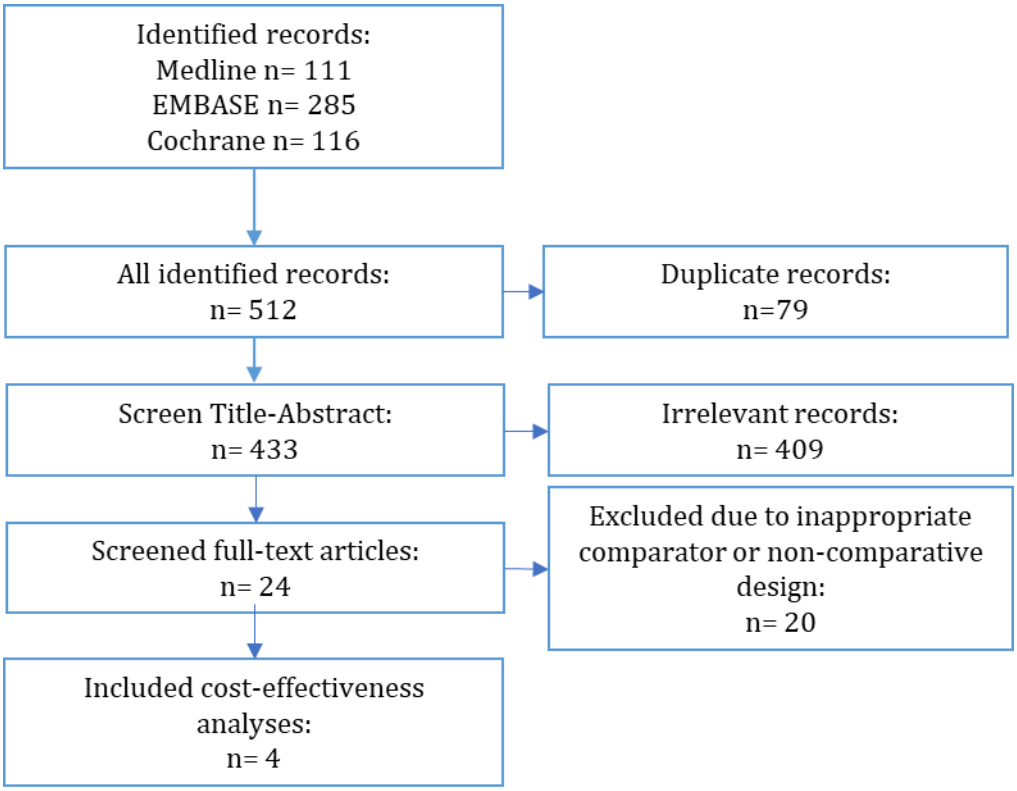


Figure 42 Flow chart describing the systematic process of article selection

4.2.1.1 Review results

4.2.1.1.1 Characteristics and methodology of published cost-effectiveness studies

Of the four included studies, two were from the United States (Losina 2015, Lubowitz 2011), one from Canada (Marsh 2016), and one from the United Kingdom (Hutt 2015). The studies were not funded through private funding sources (i.e. devices or surgical companies).

An overview of the characteristics and demographics of the patient populations modelled or included in the studies is provided in Table 19, together with information on intervention, comparator, setting, time horizons of the analyses, and the type of modelling. Here a summary is given.

Three studies incorporated patients with knee osteoarthritis with a mean age ranging from 58 to 64 years (Hutt 2015, Losina 2015, Marsh 2016). Lubowitz 2011 defined patients undergoing knee arthroscopy as patients having chondroplasty, lateral or medial meniscectomy, lateral or medial meniscus repair, lateral retinacular release, loose body removal, microfracture, or synovectomy. However, the underlying diseases were not specified. In this study, the patient population was definitively younger, with a mean age of 44 years.

Two studies specifically reported the severity of the included/modelled knee osteoarthritis, which was measured with the Kellgren-Lawrence system (KL score) (Marsh 2016, Losina 2015).

In Marsh 2016 all patients had a KL score above 2, whereas in the modelling by Losina 2015 it was assumed that 55.2% of the patients would have a KL score above 2.

The gender distribution ranged from 39% to 56% male patients (in Marsh 2016 and Hutt 2015, respectively). In Losina 2015, a Monte Carlo microsimulation, there was no information concerning gender.

In Hutt 2015, patients were included only if they had failed a trial of non-operative treatment (analgesia, activity modification, and physical therapy). In all other studies patient inclusion was not conditional upon previously failed treatments.

4.2.1.1.2 Intervention and Comparator

Two studies compared arthroscopic surgery to non-operative treatment, which consisted of physical therapy alone (Losina 2015) or physical and medical therapy (Marsh 2016). Losina 2015 modelled two types of arthroscopic interventions: arthroscopic partial meniscectomy (APM) and delayed APM (i.e. PT followed by APM if subjects continued to experience pain). The other two studies assessed the cost-effectiveness of the intervention by comparing the status of the patients before and after knee arthroscopy (Hutt 2015, Lubowitz 2011).

4.2.1.1.3 Main clinical data sources

Marsh 2016 was based on the RCT by Kirkley 2008, evaluating the effectiveness of arthroscopic surgery in addition to optimized physical and medical therapy among patients with symptomatic, radiographic knee OA over a 2-year period.[51] Losina 2015 conducted a decision analysis, which was mainly based on data from the MeTeOR RCT, a multicentre RCT involving symptomatic patients with a meniscal tear and evidence of mild-to-moderate OA on radiographic and MRI imaging. [57] The cost-effectiveness analysis conducted by Hutt 2015 was based on prospectively, self-collected data from 43 patients with radiological OA that were assessed pre- and postoperatively. Similarly, Lubowitz 2011 used data on 93 patients before and after knee arthroscopy.

4.2.1.1.4 Type of economic evaluations

All studies were cost-utility analyses, i.e. cost-effectiveness analyses using QALYs as the measure of clinical benefit.

4.2.1.1.5 Perspective of studies

Two studies reported cost estimations using a societal perspective, meaning that both direct and indirect costs were assessed (Losina 2015, Marsh 2016), whereas the other studies used a healthcare perspective including exclusively direct medical costs (Hutt 2015, Lubowitz 2011).

4.2.1.1.6 Time horizon of studies

Hutt 2015 and Marsh 2016 used a short time horizon (1.5 years and 2 years, respectively). Losina 2015 and Lubowitz 2011 used a 10-year and lifetime horizon, respectively.

Table 19 Population demographics and characteristics of included cost-effectiveness studies

Study	Country Population	Age and gender	Intervention	Comparator	Perspective	Cost types considered Cost year	Approach to analysis Time horizon Discounting
Hutt 2015	UK 43 Patients with radiological OA	64 years (range 38-82) 45% males	[Status after] knee arthroscopy *	Status before knee arthroscopy *	Healthcare perspective Hospital	Direct costs (intervention) n.r.	Prospective cohort study 1.5 years n.r.
Marsh 2016	Canada 168 patients with symptomatic knee OA [KL ≥2] from Kirkley 2008 RCT	59 ± 10 years 61% females	Arthroscopic debridement and partial resection of degenerative knee tissues in addition to non-operative treatments	Optimised non-operative therapy only	Healthcare and societal perspective	Direct costs for arthroscopy included equipment, operating room, and laboratory or other medical tests. Direct costs for non-operative care included physical therapy sessions, medication use. Other healthcare costs included inpatient hospitalisations, physical therapy, medication use, assistive devices, employment time lost, and homemaking or volunteer time lost CAD 2014	Economic evaluation alongside a single-centre RCT 2 years n.r.
Losina 2015	US Adult patients with symptomatic DMK and OA [44.8% KL 0 or 1, 26.4% KL 2, 28.8% KL 3] from MeTeOR RCT	58 ± 7 years No gender distinction	Immediate APM or Delayed APM (i.e. PT followed by APM if subjects continued to experience pain) in addition to PT	PT alone	Societal perspective	Direct medical costs for the treatment of knee pain due to DMK or OA. General medical care. Costs of productivity lost due to treatment and functional disability USD 2013	Monte Carlo micro-simulation 10 years 3%
Lubowitz 2011	US 93 patients undergoing knee arthroscopy	44 years (range 11 to 79) 56% males	[Status after] knee arthroscopy *	Status before knee arthroscopy *	Healthcare perspective Hospital	Direct costs (hospital costs and professional fee) USD 2009	Prospective cohort study Lifetime n.r.

NB: Hutt 2015 and Lubowitz 2011 used a pre-post approach. The status after knee arthroscopy was compared to the status before knee arthroscopy for each patient.
Abbreviations: APM, arthroscopic partial meniscectomy; DMK, degenerative meniscus of the knee; KL, Kellgren-Lawrence; n.r., not reported; OA, osteoarthritis; PT, physical therapy; RCT, randomised controlled trial.

4.2.1.1.7 Discounting

Only one study discounted both costs and QALYs using a 3% discount rate (Losina 2015). For all other studies discounting was not mentioned.

4.2.1.1.8 Measurement of cost and data sources

The types of costs considered varied across studies (Table 20). In two studies, both direct and indirect costs were assessed (Marsh 2016, Losina 2015). In these studies, direct costs included hospitalisation, surgery, medication, and professional's fees. Indirect costs included productivity lost due to treatment and functional disability. In Hutt 2015 and Lubowitz 2011, only direct medical costs, such as hospitalisation and surgery costs, were included.

For the definition of unit costs (i.e. costs for a surgery, an hospital day, medication, etc.), a variety of sources were used in the studies reporting a more comprehensive cost assessment (Marsh 2016, Losina 2015). In the other two studies, cost information came from a single, early cost-effectiveness analysis (in the case of Lubowitz 2011) and from the hospital where the study was performed (in case of Hutt 2015).

Table 20 Types of costs and main sources used in the eligible cost-effectiveness analyses

Article	Type of costs	Sources
Hutt 2015	*Direct costs* Procedure costs	From the NHS trust where the study was performed.
Marsh 2016	*Direct costs* Surgery, hospitalisation, physical therapy, medication, assistive devices *Indirect costs* Employment time lost, homemaking or volunteer time lost	Ontario Case Costing Initiative [89] Ontario Schedule of Benefits [90] Ontario Drug Benefit Formulary [91]
Losina 2015	*Direct costs* Direct medical costs for treatment, physician visits, general medical care including additional management of pain (nonsteroidal anti-inflammatory drug, opioids, injections), travel *Indirect costs* Productivity lost due to treatment and functional disability	MeTeOR RCT [57] Medicare Fee Schedules [92 93] Red Book Online [94] Literature [95-97] National Occupational Employment and Wage Estimates [98]
Lubowitz 2011	*Direct costs* Hospitalisation, surgical professional fee	Cost-effectiveness analysis in knee arthroplasty [99]

4.2.1.1.9 Measurement and sources of clinical effects and health related quality of life

The cost-effectiveness analyses assessed the effects of treatments in different ways and are not easily comparable.

Marsh 2016 assessed the effects of knee arthroscopy using two instruments: the changes in pain, stiffness, and physical function were measured with the Western Ontario McMaster Osteoarthritis Index (WOMAC) total score over 2-year follow-up, whereas utility changes between baseline and 2-year follow-up were measured using a Standard Gambling technique derived from Kirkley 2008 (Table 21). QALYs were calculated as the product of the utility score and the duration of the corresponding health state. The estimated incremental QALYs of arthroscopic surgery versus comparator were -0.02.

Losina 2015 modelled the effects of knee arthroscopy over ten years using quarterly cycles, meaning that transitions between health states were possible every three months. Health states considered were early low pain, early moderate pain, late low pain, and late moderate pain. Distribution of pain relief using a transformed Knee injury and Osteoarthritis Outcome Score (KOOS) Pain scale as well as utilities assessed with an EQ-5D instrument was derived from the MeTeOR RCT.[57] Fluctuation in OA-related pain were derived from the Johnston County Osteoarthritis Project and were stratified according to KL-grade (Table 21, Table 22).[100] The estimated incremental QALYs of arthroscopic surgery versus comparator ranged from 0.086 to 0.095.

Lubowitz 2011 assessed the health-related quality of life before and after knee arthroscopy using the self-administered Quality of Well-Being (QWB) scale.[101] The resulting score was multiplied by the life expectancy of each subject to obtain QALY estimates. The estimated incremental QALYs of arthroscopic surgery versus comparator were 1.091.

Hutt 2015 assessed pre- and postoperative effectiveness outcomes using the Oxford Knee Score (OKS) and a visual analogue score (VAS). The OKS data were converted to EQ-5D scores using a published algorithm.[102] The estimated incremental QALYs of arthroscopic surgery versus comparator were 0.54.

The quality of life results reported in the clinical section suggested that there was a numerical effect but no statistically significant effect on disease-specific quality of life in favour of arthroscopy compared to control, at short-term follow-up (SMD 0.18, 95% CI [-0.02, 0.39], Figure 25; very low quality of evidence, Table 13) and at intermediate follow-up (SMD 0.17, 95% CI [-0.02, 0.36], Figure 26; moderate quality of evidence, Table 14). If compared with these clinical results, it appears that the above mentioned incremental QALY estimates were particularly optimistic in Lubowitz 2011 and Hutt 2015, where they were based on pre-post comparisons. They were less optimistic and potentially compatible in Losina 2015. The slight QALY loss through arthroscopy in Marsh 2016 is also not formally incompatible with the clinical results, given that these were favouring arthroscopy (although estimates were not statistically significant).

Table 21 Effectiveness and utility assumptions for the arthroscopic interventions compared to non-operative treatment in eligible studies

	Arthroscopic intervention	Non-operative treatment
Marsh 2016	WOMAC total score at the 2-year follow-up Baseline: 1222.91 2 years: 1526.45 Utility assessed with a Standard Gambling technique, over 2 years Baseline: 0.79 2 years: 0.84 QALY over 2 years (combining utilities and duration): 1.64	WOMAC total score at the 2-year follow-up Baseline: 1355.26 2 years: 1510.77 Utility assessed with a Standard Gambling technique, over 2 years Baseline: 0.80 2 years: 0.86 QALY over 2 years (combining utilities and duration): 1.66
Losina 2015	*Pain 0-3 months after surgical treatment* Probability of failed pain relief APM: KL grade 0-2: 0.322 / KL grade 3-4: 0.488 APM after PT: KL grade 0-2: 0.400 / KL grade 3-4: 0.667 *3-6 months after surgical treatment* Probability of pain incidence APM: KL grade 0-2: 0.230 /KL grade 3-4: 0.364 APM after PT: KL grade 0-2: 0 /KL grade 3-4: 0 Probability of pain resolution APM: KL grade 0-2: 0.483 /KL grade 3-4: 0.333 APM after PT: KL grade 0-2: 0 /KL grade 3-4: 0 Utility scores by pain state were the same for both study arms Low pain (KOOS <=25): 0.869 High pain (KOOS >25): 0.771	*Pain 0-3 months after treatment* Probability of failed pain relief KL grade 0-2: 0.569 / KL grade 3-4: 0.703 *3-6 months after treatment* Probability of pain incidence KL grade 0-2: 0.227 /KL grade 3-4: 0.182 Probability of pain resolution KL grade 0-2: 0.155 /KL grade 3-4: 0.115 Utility scores by pain state were the same for both study arms Low pain (KOOS <=25): 0.869 High pain (KOOS >25): 0.771
Lubowitz 2011	Postoperative QWB Scale utility score: 0.704	Preoperative QWB Scale utility score: 0.672
Hutt 2015	Postoperative Oxford Knee Score (OKS) converted to EQ-5D score using published algorithm: 0.79	Preoperative Oxford Knee Score (OKS) converted to EQ-5D score using published algorithm: 0.43

Abbreviations: APM, arthroscopic partial meniscectomy; DMK, degenerative meniscus of the knee; KL, Kellgren-Lawrence; n.r., not reported; OA, osteoarthritis; KOOS, Knee injury and Osteoarthritis Outcome Score; OKS, Oxford Knee Score; PT, physical therapy; QALY, quality adjusted life year; QWB: quality of well-being; RCT, randomised controlled trial**;** WOMAC, Western Ontario McMaster Osteoarthritis Index.

Table 22 Effectiveness and utility main sources used in the selected cost-effectiveness studies

Article	Sources of effectiveness	Sources of utility estimates
Marsh 2016	Single centre RCT conducted by Kirkley 2008[51]	Single centre RCT conducted by Kirkley 2008[51]
Losina 2015	Pain relief and incidence derived from the MeTeOR[57] RCT using a transformed Knee injury and Osteoarthritis Outcome Score (KOOS) Changes in pain status attributable to a specific treatment was limited to 6 months following treatment (Expert opinion of few MeTeOR clinical investigator) Annual OA incidence from Losina et al.[103] Total knee arthroplasty uptake rate from the Multicenter Osteoarthritis Study and Osteoarthritis Initiative cohorts[104] Perioperative outcomes and adverse events from several publications.[105-109]	Quality of life utility scores from the MeTeOR RCT and the Johnston County Osteoarthritis Project[57 100]
Lubowitz 2011	Self-administered Quality of Well-Being (QWB) scale multiplied by life expectancy	Self-administered Quality of Well-Being (QWB) scale multiplied by life expectancy
Hutt 2015	Self-administered Oxford Knee Score (OKS) and visual analogue score (VAS)	Preoperative Oxford Knee Score (OKS) converted to EQ-5D score using published algorithm.[102]

4.2.1.2 Adaptation of economic evaluation results to Switzerland and similarities and differences observed between studies

The original results of the four included cost-effectiveness studies are summarised in Table 23. The two studies comparing knee arthroscopy with a non-operative comparator showed discordant results (Losina 2015, Marsh 2016). According to Marsh 2016, knee arthroscopy is dominated by non-operative treatments alone (i.e. arthroscopy was more expensive and less effective), independently from the perspective. In contrast, Losina 2015 reported that immediate and delayed APM may be cost-effective or cost-saving. In particular, delayed APM (i.e. APM after 3 months of physical therapy) seemed to dominate physical therapy alone from a societal perspective (i.e. APM was less expensive and more effective). If compared to physical therapy, immediate APM showed ICERs below USD 25,000 per QALY from a societal perspective and below USD 12,000 per QALY from a healthcare payer perspective. However, when compared to delayed APM, the ICERs for immediate APM increased to USD 103,000 per QALY from a healthcare perspective and USD 72,200 per QALY from a societal perspective.

The two studies comparing patient status before and after knee arthroscopy concluded that arthroscopic knee surgery was cost-effective from a healthcare payer perspective: Lubowitz 2011 reported ICERs of USD 5,783 per QALY, whereas Hutt 2015 reported £ 2,046 per QALY.

According to the CHEERS checklist, the quality of the four assessed studies ranged from good (Marsh 2016, Losina 2015) to poor (Hutt 2015, Lubowitz 2011). Since all studies provided at least sufficient information on costs and effects, and since the information concerning cost-effectiveness of knee arthroscopy was found to be very scarce, all results were adapted for Switzerland (Table 24, Figure 43).

The adapted results for Switzerland were very similar to the original ones. The two studies comparing knee arthroscopy with a non-operative comparator were still going into opposite directions (Marsh 2016, Losina 2015), and the results of the two studies comparing patient status before and after knee arthroscopy were still similar (Hutt 2015, Lubowitz 2011).

Several potential explanatory factors were assessed in order to understand the observed similarities and differences between studies. Due to completely different trial designs, i.e. intervention versus comparator and postoperative versus preoperative, the results of Marsh 2016 were principally compared with those of Losina 2015, whereas the study by Lubowitz 2011 was compared to the study by Hutt 2015.

The studies of Marsh 2016 and Losina 2015 showed opposite results: in Marsh 2016 the comparator (non-operative treatment alone) dominated the intervention. In contrast, Losina 2015 suggested that immediate and delayed APM may be cost-effective (ICER <CHF 30,000) or cost-saving. Except for the age of the included population and the types of costs taken into consideration, the two studies are not easily comparable. Marsh 2016 reported an economic evaluation of a single-centre RCT which included patients with symptomatic, radiographic knee OA,[51] whereas Losina 2015 performed a Monte Carlo microsimulation of patients with both DMK and OA. Moreover, in Marsh 2016 all patients had a KL grade ≥2, whereas almost half of the population in Losina 2015 had KL grade 0-1. The modelling of effectiveness, utility, and quality-adjusted life time were based on different assumptions and sources. The main element leading to discrepant results was the assessment of QALYs. In Marsh 2016, patients receiving non-operative treatments accrued marginally higher QALYs than patients receiving arthroscopy (1.66 versus

1.64). In contrast, the QALY gain reported by Losina 2015, indirectly derived from results of the MeTeOR RCT, [57] showed better results in the intervention group (6.723 versus 6.637). The inclusion of patients with different characteristics may have contributed to this.

Hutt 2015 and Lubowitz 2011 were based on comparisons of patients before and after knee arthroscopy and showed similar results. Considering the differences between the two studies, it was almost surprising to find very similar results (ICERs comparing postoperative status versus preoperative status of CHF 7,200-7,400 per QALY in both studies). Both studies were based on self-collected data and assessed similar types of costs (i.e. exclusively hospitalisation/intervention costs). However, the patients included by Lubowitz 2011 were younger (mean 43.9 years versus 64 years), whereas the time horizon of the analysis was much longer (lifetime versus 1.5 years). It should be emphasized that the quality of both studies, assessed with the CHEERS checklist, was poor. Important information was not or only partially provided.

Table 23 Results of economic evaluations for knee arthroscopy, as originally reported by the authors

Study, perspective (currency)	Costs intervention	Costs comparator	QALY Intervention	QALY comparator	ICER (costs per QALY)
	Incremental cost		Incremental QALYs		
Marsh 2016, Healthcare payer perspective (Can$)	2,633	737	1.64	1.66	Comparator dominates intervention
	1,896		-0.02		
Marsh 2016, Societal perspective (Can$)	3,826	1,614	1.64	1.66	Comparator dominates intervention
	2,211		-0.02		
Losina 2015, Healthcare payer perspective (USD), Immediate APM	12,900	10,800	6.723	6.637	24,419 *
	2,100		0.086		
Losina 2015, Healthcare payer perspective (USD), Delayed APM	11,900	10,800	6.732	6.637	11,579
	1,100		0.095		
Losina 2015, Societal perspective (USD), Immediate APM	38,300	38,200	6.723	6.637	1,163 *
	100		0.086		
Losina 2015, Societal perspective (USD), Delayed APM	37,600	38,200	6.732	6.637	Intervention dominates comparator
	-600		0.095		
Study, perspective (currency)	**Costs post-surgery**	**Costs pre-surgery**	**QALY post-surgery**	**QALY pre-surgery**	**ICER (costs per QALY)**
	Incremental cost		Incremental QALYs		
Lubowitz 2011, Healthcare payer perspective (USD)	6,310	-	24.006§	22.915§	5,783
	6,310		1.0912§		
Hutt 2015, Healthcare payer perspective (£)	1,105	-	0.79	0.43	2,046
	1,105		0.54#		

*The ICERs of immediate APM compared to delayed APM were USD 103,200 per QALY from a healthcare perspective and USD 72,200 per QALY from a societal perspective.

§QWB scores of 0.704 and 0.672 were multiplied by an estimated life expectancy of 34.1 years.

#EQ-5D scores of 0.79 and 0.43 were multiplied by the duration of the follow up (1.5 years)

Table 24 Results of economic evaluations, adapted for Switzerland, cost reported in Swiss Francs

Study, perspective (currency)	Costs intervention	Costs comparator	QALY Intervention	QALY comparator	ICER (costs per QALY)
	Incremental cost		Incremental QALYs		
Marsh 2016, Healthcare payer perspective (CHF)	4,490	1,257	1.64	1.66	Comparator dominates intervention
	3,233		-0.02		
Marsh 2016, Societal perspective (CHF)	6,524	2,753	1.64	1.66	Comparator dominates intervention
	3,771		-0.02		
Losina 2015, Healthcare payer perspective (CHF), Immediate APM	14,400	12,056	6.723	6.637	27,258*
	2,344		0.086		
Losina 2015, Healthcare payer perspective (CHF), Delayed APM	13,284	12,056	6.732	6.637	12,925
	1,228		0.095		
Losina 2015, Societal perspective (CHF), Immediate APM	42,754	42,642	6.723	6.637	1,298*
	112		0.086		
Losina 2015, Societal perspective (CHF), Delayed APM	41,973	42,642	6.732	6.637	Intervention dominates comparator
	-670		0.095		
Study, perspective (currency)	**Costs post-surgery**	**Costs pre-surgery**	**QALY post-surgery**	**QALY pre-surgery**	**ICER (costs per QALY)**
	Incremental cost		Incremental QALYs		
Lubowitz 2011, Healthcare payer perspective (CHF)	7,909.61	-	24.006§	22.915§	7,249
	7,910		1.0912§		
Hutt 2015, Healthcare payer perspective (CHF)	3,977.25	-	0.79	0.43	7,365
	3,977		0.54#		

* The ICERs of immediate APM compared to delayed APM were CHF 115,200 per QALY from a healthcare perspective and CHF 80,600 per QALY from a societal perspective.

§QWB scores of 0.704 and 0.672 were multiplied by an estimated life expectancy of 34.1 years.

#EQ-5D scores of 0.79 and 0.43 were multiplied by the duration of the follow up (1.5 years).

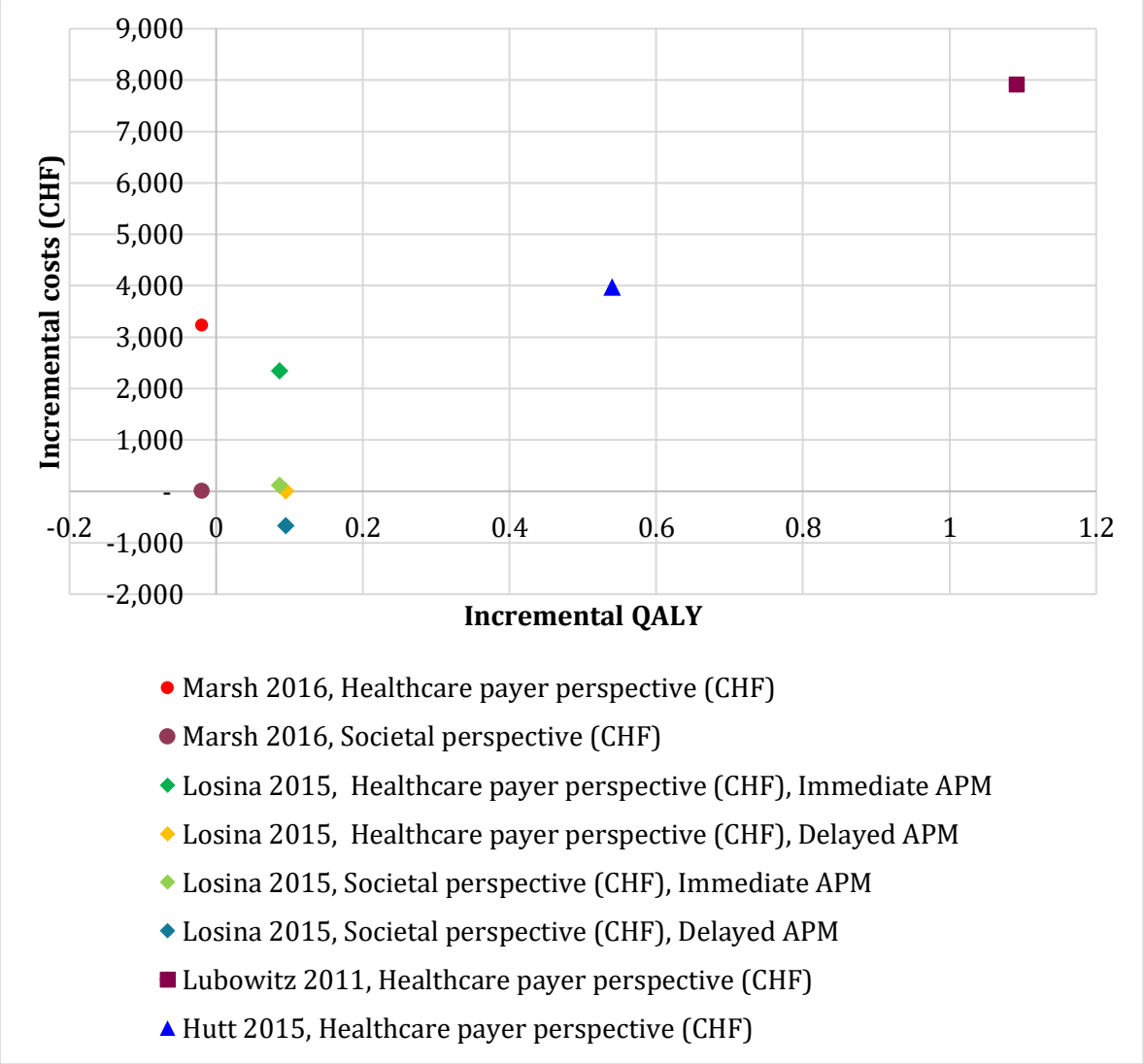


Figure 43 Cost-effectiveness plane, based on costs (in Swiss Francs) adapted to Switzerland and original effect estimates

4.2.1.2.1 Sensitivity analyses within cost-effectiveness studies

Two studies conducted one-way sensitivity analysis (Lubowitz 2011, Marsh 2016), one study performed a probabilistic sensitivity analysis (Losina 2015), and one study reported no sensitivity analysis (Hutt 2015).

Marsh 2016 reported a sensitivity analysis using 95% confidence limits. Even when the largest possible treatment effect was assumed on this basis, the addition of arthroscopic surgery was not economically attractive compared with non-operative treatments only.

Lubowitz 2011 reported that knee arthroscopy would remain highly cost-effective (<CHF 36,700) even after doubling the procedure costs or halving the improvements in QWB scale scores.

Losina 2015 conducted sensitivity analyses that considered alternative scenarios regarding the efficacy of delayed APM and inclusion of indirect costs. Moreover, they also varied simultaneously the potential impact of APM on OA progression and the potential impact of delaying the surgical procedure on its overall efficacy. Finally, they reduced the time horizon of their analysis from 10 to 5 years. The results showed that physical therapy had 3.0% probability of being cost-effective at a willingness-to-pay (WTP) threshold of USD 50,000 per QALY, whereas delayed APM was cost-effective 57.7% of the time at a WTP of USD 50,000 per QALY and 50.2% at a WTP of USD 100,000 per QALY. The probability of immediate APM being cost-effective did not exceed 50% unless WTP exceeded USD 103,000 per QALY (CHF 115,200 per QALY).

4.2.2 Budget impact analyses

The main results, including the estimated number of eligible patients, the estimated annual costs and results of the sensitivity analyses, are presented in the first part of this section. In the second part, details on the frequencies of occurrence of the selected DRG, ICD-10, CHOP, and TARMED codes are provided.

4.2.2.1 Patients eligible for the cost assessment

The total number of eligible cases according to the two selection strategies is shown in Table 25.

In the first strategy, only patients with a relevant DRG code, with at least one relevant diagnosis (ICD-10 code), and at least one relevant treatment (CHOP code) were included. Between 2010 and 2011 the total number of cases did not change (11,922 in 2010 and 11,836 in 2011). Due to the change in the DRG coding system in 2012, it was difficult to estimate a trend concerning the number of cases. In 2012, the number of cases was substantially higher than in the years before (14,816), probably because of misclassification in the new coding system. As illustrated in

Table 26, almost all identified cases were classified under the Swiss DRG code I30Z. The number of cases between 2013 and 2014 were then lower than in 2010 and 2011, and were similar between each other (7,245 and 6,904, respectively).

In the second strategy, all patients with at least one relevant diagnosis (ICD-10 code) and with one relevant treatment (CHOP code) were included, independently from the DRG codes. In contrast to the number of cases identified with the first strategy, the number of cases identified with the second strategy was more consistent over the five years period, most likely because DRGs were not considered. For this reason, the second strategy was considered more realistic than the first one. Additional analyses of the number of eligible cases stratified by relevant ICD-10 diagnoses are provided in the appendix (Additional tables and analyses, Table 45).

Concerning the number of eligible cases in the ambulant setting, available for 2013 and 2014 only, it was assumed that the percentage of cases covered by an accident insurance would be equal to the corresponding percentage in the Swiss Hospital Statistics (37%). Therefore, the number of cases that would meet the PIC criteria would be 6,136 in 2013 and 5,524 in 2014 (see section 4.2.2.7).

Table 25 Estimated number of eligible cases according to two selection strategies between 2010 and 2014

First strategy (for inpatients, based on DRG, ICD-10 and CHOP codes)					
Year	2010	2011	2012	2013	2014
Inpatients (N)	11,922	11,836	14,816	7,245	6,904
Outpatients (N)	-	-	-	6,136	5,524
Total	-*	-*	-*	13,381	12,428
Second strategy (for inpatients, based on ICD-10 and CHOP codes only)					
Year	2010	2011	2012	2013	2014
Inpatients (N)	15,703	15,099	15,794	15,459	14,721
Outpatients (N)	-	-	-	6,136	5,524
Total	-*	-*	-*	21,595	20,245

*Unknown due to lack of information on outpatient procedures

Table 26 Number of eligible cases who reported at the same time a relevant DRG code, at least one relevant ICD-10 code for diagnosis, and at least one relevant CHOP code for treatment between 2010 and 2014 (strategy 1)

Year	2010	2011	Year	2012	2013	2014
APDRG code	**N**	**N**	**Swiss DRG code**	**N**	**N**	**N**
221	65	112	**I04Z**	0	1	0
222	22	37	**I12B**	10	12	15
232	437	218	**I18A**	0	55	61
917	158	145	**I18B**	0	6,586	5,968
918	7,374	6,168	**I30Z**	14,806	584	859
1222	3,825	5,147	**I59Z**	0	7	1
1232	41	9				
Total	11,922	11,836	**Total**	14,816	7,245	6,904

4.2.2.2 Annual costs

The annual inpatient costs according to two selection strategies as well as the annual ambulatory costs are summarised in Table 27. For the first strategy, the resulting estimates of total annual inpatient costs in 2010 and 2011 were CHF 59.31 Mio. and CHF 61.65 Mio., respectively. The costs in 2013 and in 2014 were lower with, CHF 38.79 Mio. and CHF 39.04 Mio., respectively. The results of 2012 were massively higher with CHF 134.68 Mio. The potential reasons for the outlying result for the year 2012 (new DRG coding system) has been described above. The differences in cost estimates between 2010-2011 and 2013-2014 are likely due to the different DRG coding systems (APDRG versus Swiss DRG). This first selection strategy is to be considered not valid for several reasons: first, not all patients who received an ICD-10 diagnosis for meniscus derangement or who received an arthroscopic treatment to the knee (represented by a CHOP code) were classified with one of the investigated DRG codes. Second, not all patients with a relevant DRG code received an arthroscopic treatment to the knee, according to the CHOP codes assigned. For example, in 2014

there were 22,665 cases with a relevant Swiss DRG code. Among them, 13,463 (59%) had a relevant ICD-10 diagnosis and 14,354 (63%) received a relevant treatment (CHOP code). After excluding cases reimbursed by the accident insurance, the total number of cases included in this cost estimation was 6,904, meaning that only 30% of the cases with potentially relevant Swiss DRG codes were included. Moreover, the inconsistency of this first approach is also emphasized in Table 27, where total costs were divided according to the different DRG codes. The inconsistency between 2010 and 2014 can also be seen by calculating the mean costs per inpatient case, which were CHF 4,975 in 2010, CHF 5,209 in 2011, CHF 9,090 in 2012, CHF 5,354 in 2013, and CHF 5,655 in 2014.

The second selection strategy including all cases that reported at least one relevant ICD-10 diagnosis and one relevant treatment (CHOP code), independent of DRG codes, and assuming mean costs per case of CHF 3,700, showed more stable results and was considered as more realistic.

The total annual outpatient costs, assuming costs of CHF 2,400 for each ambulatory case and assuming that the percentage of cases covered by an accident insurance would be the same as in the Swiss Hospital Statistics (37%), were estimated to be around CHF 14.73 Mio. in 2013 and CHF 13.26 Mio. in 2014.

Table 27 Annual inpatient and outpatient costs according to the two selection strategies

First strategy					
Year	2010	2011	2012	2013	2014
Inpatients (Mio. CHF)	59.31	61.65	134.68	38.79	39.04
Outpatients (Mio. CHF)	-	-	-	14.73	13.26
Total (Mio. CHF)	-*	-*	-*	53.52	52.3
Second strategy					
Year	2010	2011	2012	2013	2014
Inpatients (Mio. CHF)	58.1	55.87	58.44	57.2	54.47
Outpatients (Mio. CHF)	-	-	-	14.73	13.26
Total (Mio. CHF)	-*	-*	-*	71.93	67.73

*Unknown due to lack of information on outpatient procedures

Table 28 Total annual inpatients costs for patients hospitalised with relevant Swiss DRG codes between 2010 and 2014 (first inpatient cost estimation strategy)

Year	2010	2011	Year	2012	2013	2014
APDRG code	Mio. CHF	Mio. CHF	Swiss DRG code	Mio. CHF	Mio. CHF	Mio. CHF
221	0.71	1.22	**I04Z**	0.00	0.03	0.00
222	0.16	0.26	**I12B**	0.20	0.23	0.28
232	2.51	1.25	**I18A**	0.00	0.37	0.40
917	1.50	1.38	**I18B**	0.00	32.85	30.57
918	30.25	25.30	**I30Z**	134.48	5.27	7.79
1222	23.91	32.17	**I59Z**	0.00	0.04	0.01
1232	0.28	0.06				
Total	59.31	61.65	**Total**	134.68	38.79	39.04

Figure 44 and Figure 45 illustrate the distribution of total inpatient and outpatient costs for arthroscopic surgeries to the knee in Switzerland in 2013 and 2014 according to the two inpatient cost calculation strategies. According to the first strategy, including all patients who had at the same time at least one relevant DRG code, one relevant ICD-10 code and one relevant CHOP code, the total costs reached CHF 53.52 Mio. in 2013 and CHF 52.30 Mio. in 2014, with outpatient cases accounting for 25-28% of the total costs. The total costs using the eligibility criteria of the second strategy (i.e. considering patients that had at the same time at least one relevant ICD-10 code and one relevant CHOP code) were CHF 71.93 Mio. in 2013 and CHF 67.73 Mio. in 2014 (20% attributable to outpatient cases).

The reported costs were assessed using a KVG perspective. With the available information it was not possible to investigate the total annual costs from a societal perspective. Nevertheless, it is important to emphasize that the total costs from a societal perspective would be much higher: patients undergoing arthroscopic knee interventions are often unable to work for several weeks. Therefore, indirect costs related to the loss of productivity are potentially very high.

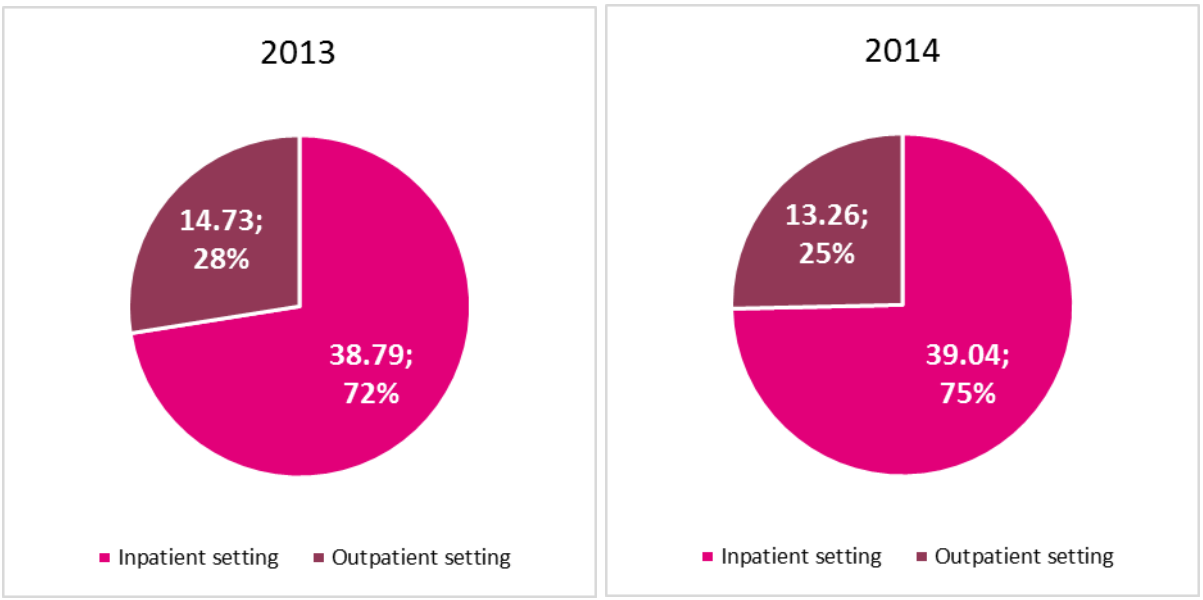


Figure 44 Distribution of total inpatient costs (first strategy) and outpatient costs for arthroscopic surgeries to the knee in Switzerland in 2013 and 2014 (Mio. CHF)

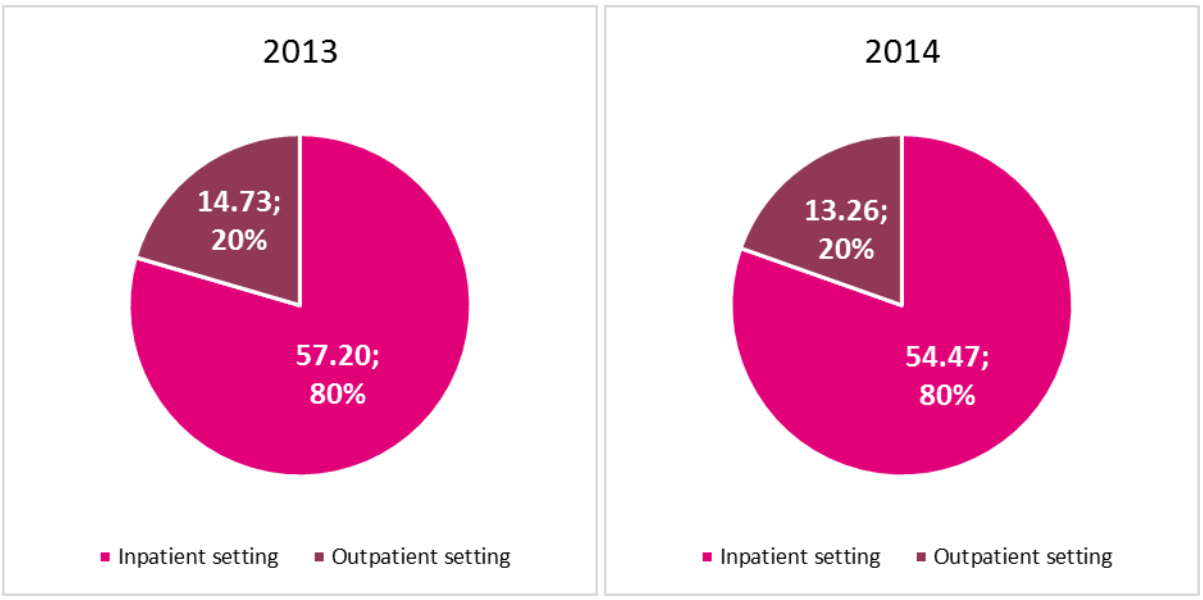


Figure 45 Distribution of total inpatient costs (second strategy) and outpatient costs for arthroscopic surgeries to the knee in Switzerland in 2013 and 2014 (Mio. CHF)

4.2.2.3 Sensitivity analysis and scenario analyses

The first inpatient cost calculation strategy was based on the diagnosis-related case costs statistics. Mean costs per case according to different DRG codes were multiplied with the number of cases that reported at the same time a relevant DRG code, at least one relevant diagnosis, and

at least one relevant treatment. The second inpatient cost calculation strategy was based on mean cost per patients undergoing an arthroscopic knee surgery assuming all patients would have compulsory insurance only. In the sensitivity analyses the mean costs per case was varied by ±20% (for the base case costs, see section 4.1.2.6, Table 16 and

Table 17). The main results were summarised in Table 29.

Table 29 Sensitivity analyses – mean costs per case varied by ±20%

First strategy					
Year	2010	2011	2012	2013	2014
Inpatients (Mio. CHF)	47.45-71.17	49.32-73.98	107.74-161.62	31.03-46.55	31.23-46.85
Outpatients (Mio. CHF)	-	-	-	11.78-17.67	10.61-15.91
Total (Mio. CHF)	47.45-71.17	49.32-73.98	107.74-161.62	42.81-64.22	41.84-62.76
Second strategy					
Year	2010	2011	2012	2013	2014
Inpatients (Mio. CHF)	46.48-69.72	44.69-67.04	46.75-70.13	45.76-68.64	43.57-65.36
Outpatients (Mio. CHF)	-	-	-	11.78-17.67	10.61-15.91
Total (Mio. CHF)	46.48-69.72	44.69-67.04	46.75-70.13	57.54-86.31	54.18-81.27

For the first strategy, two alternative patient selection scenarios were investigated: first, all cases that were classified with a relevant DRG code, irrespective of their ICD-10 and CHOP codes, were included. For this scenario, the total inpatient costs increased dramatically (detailed information are available in the appendix, Table 47). For the years 2010 and 2011, the total inpatient costs increased from approximatively CHF 60 Mio. to more than CHF 90 Mio. (approximately +50%). For the years 2013 and 2014 the increase was even more pronounced, rising from approximatively CHF 40 Mio. to more than CHF 80 Mio. (ca. +100%). These estimations, driven by the limited suitability of the DRG coding system for the given purpose, were obviously not realistic. A second alternative scenario included all cases with a relevant Swiss DRG code but excluded two specific the Swiss DRG codes, namely I59Z and I18A. Code I59Z potentially includes interventions other joints or bones, whereas code I18A relates to patients younger than 16 years. This scenario was calculated for the years 2013 and 2014. The total costs estimations reached then CHF 70.91 Mio. in 2013 and CHF 71.91 Mio. in 2014, and hence, were also dramatically higher.

In the scenario analysis of the second strategy, the inpatient costs were calculated according to the insurance status of the eligible patients, as reported in Table 30.

Table 30 Distribution of cases according to their insurance status (second inpatient cost estimation strategy)

	2010 N (%)	2011 N (%)	2012 N (%)	2013 N (%)	2014 N (%)
Compulsory	8,501 (54%)	8,187 (54%)	8,870 (56%)	9,060 (59%)	8,922 (61%)
Semi-private	4,075 (26%)	4,119 (27%)	4,081 (26%)	3,904 (25%)	3,573 (24%)
Private	3,127 (20%)	2,793 (18%)	2,843 (18%)	2,495 (16%)	2,226 (15%)
Total	15,703	15,099	15,794	15,459	14,721

Table 31 illustrates the total inpatient costs by insurance status and year. The results suggest that between 2010 and 2014 there was a decrease in the total costs (from CHF 107.16 Mio. to CHF 92.59 Mio.). This difference can be explained by two factors: first, the number of cases in 2010 was higher if compared to 2014 (15,703 versus 14,721). Second, the insurance distribution was different: in 2010 there were more patients with a private insurance (20% versus 15% in 2014) and fewer patients with only compulsory insurance (54% versus 61% in 2014). As already mentioned the costs for patients with private insurance were considerably higher than those for patients with compulsory insurance (CHF 13,200 versus CHF 3,700). It should be noted that the results of this scenario analysis cannot be interpreted as costs from a KVG perspective, as they include coverage of services not reimbursement by the compulsory health insurance.

Table 31 Total costs for inpatients with at least one relevant ICD-10 code and one relevant CHOP code between 2010 and 2014, stratified by insurance status (second inpatient cost estimation strategy)

Insurance status	2010 Mio. CHF (%)	2011 Mio. CHF (%)	2012 Mio. CHF (%)	2013 Mio. CHF (%)	2014 Mio. CHF (%)
Compulsory	31.45 (29%)	30.29 (30%)	32.82 (31%)	33.52 (34%)	33.01 (36%)
Semi-private	34.43 (32%)	34.81 (34%)	34.48 (33%)	32.99 (33%)	30.19 (33%)
Private	41.28 (39%)	36.87 (36%)	37.53 (36%)	32.93 (33%)	29.38 (32%)
Total	107.16	101.97	104.83	99.44	92.59

In the outpatient costs estimation, the mean costs for an arthroscopic surgery of the meniscus performed in an ambulatory setting were combined with the number of cases with relevant TARMED positions. In this sensitivity analysis the estimated costs per case was varied by ±20%. The resulting ambulatory costs ranged from CHF 11.78 to CHF 17.67 Mio. in 2013, and from CHF 10.61 to CHF 15.91 Mio. in 2014.

The scenario analysis assuming that all cases with a relevant TARMED code were eligible, i.e. assuming that all ambulatory cases received the intervention due to a degenerative knee problem and not a traumatic problem, resulted in total ambulatory costs of CHF 23.36 Mio. in 2013 and CHF 21.05 Mio. in 2014. In contrast, if only 50% of the patients were to have degenerative knee problems, the total ambulatory costs would have been CHF 11.69 Mio. in 2013, and CHF 10.5 Mio. in 2014.

4.2.2.4 Frequency of Swiss DRG and APDRG codes

The numbers of cases that were hospitalized between 2010 and 2014 with relevant APDRG and Swiss DRG codes are reported in Table 32 and Table 33. The percentage of the cases that were reimbursed by the accident insurance is reported for each year and for each DRG code.

Table 32 Number of cases with relevant APDRG codes hospitalized in 2010 and 2011

APDRG	Description	2010		2011	
		N	% accident	N	% accident
221	Synovectomy and ligament reconstruction at the knee, with KK	227	36%	310	29%
222	Synovectomy and ligament reconstruction at the knee, without KK	1,029	49%	3,527	54%
232	Arthroscopy	2,370	30%	1,068	31%
917	Other knee surgeries, with KK	254	20%	248	18%
918	Other knee surgeries, without KK	12,685	31%	10,175	32%
1222	Synovectomy and ligament reconstruction at the knee, without KK, with multiple interventions	11,256	46%	11,712	42%
1232	Arthroscopy, with multiple interventions	528	29%	236	33%
	Total	28,349	38%	27,276	40%

Abbreviation: KK, Komorbiditäten und Komplikationen (Comorbidities and Complications)

Table 33 Number of cases with relevant Swiss DRG codes hospitalized between 2012 and 2014

DRG Code	Description	2012		2013		2014	
		N	% accident	N	% accident	N	% accident
I04Z	Revision or replacement of the knee with complicating diagnose or arthrodesis	296	10%	359	10%	157	12%
I12B	Infection/Inflammation of Bone and Joint W Misc Musculoskeletal Procs W Sev or Mod CC	358	20%	427	24%	514	23%
I18A	Arthroscopy, incl. Biopsy or other interventions on bone or joints, age <16 years	847	0%	1,183	1%	291	3%
I18B	Arthroscopy, incl. Biopsy or other interventions on bone or joints, age >15 years	69	41%	11,563	33%	11,361	35%
I30Z	Complex knee interventions	33,108	39%	5,446	69%	7,796	69%
I59Z	Other interventions on Humerus, Tibia, Fibula, ankle joint or relatively complex interventions on knee joint, elbow joint and forearm	1,050	49%	1,328	50%	2,546	28%
	Total	35,728	38%	20,306	41%	22,665	45%

A brief review of the two tables suggests that a comparison between the years 2010-2011 and the years 2012-2014 is very difficult since the coding system was completely different. It is also evident that 2012 was a transition year in which the treating physicians and coders still had to fully understand the new Swiss DRG coding system. In fact, in 2012 the great majority of the patients hospitalized for a knee problem received the Swiss DRG code I30Z "Complex knee interventions" (33,108 out of 35,728 cases, 93%), whereas all other Swiss DRG codes were more or less ignored. The coding was then normalized and relatively similar in 2013 and 2014. This suggests that in 2012 many patients were not optimally classified.

Due to this change in the coding system between 2011 and 2012, it is not possible to identify specific trends concerning the increase or decrease of specific DRG codes. In general, from 2010 to 2011, there was a decrease in the overall number of patients hospitalized with relevant APDRG codes (from 28,349 to 27,276, -4%). In contrast, from 2013 to 2014, there was an increase in the overall number of patients hospitalized with relevant Swiss DRG codes (from 20,306 to 22,665, +12%). The increase in the number of complex knee interventions was particularly pronounced (from 5,446 to 7,796, +43%). The difference between the years 2010-2011 and the years 2013-2014 is probably due to the different coding systems and to the fact that some DRG codes potentially included interventions at other joints (e.g. APDRG codes 232, 1232).

4.2.2.5 Frequency of ICD-10 codes

The ICD-10 codes provided in the Swiss Hospital Statistics were particularly detailed. The complete list of relevant ICD-10 codes used between 2010 and 2014 is shown in Table 34. The percentage of cases that were reimbursed by the accident insurance is reported only for 2014, because it was similar in the previous years. For example, the code "M23.2: Derangement of meniscus due to old tear or injury" was divided in "M23.20: Derangement of unspecified meniscus due to old tear or injury", "M23.21: Derangement of anterior horn of medial meniscus due to old tear or injury", "M23.22: Derangement of posterior horn of medial meniscus due to old tear or injury", M23.23: Derangement of other medial meniscus due to old tear or injury", "M23.24: Derangement of anterior horn of lateral meniscus due to old tear or injury", "M23.25: Derangement of posterior horn of lateral meniscus due to old tear or injury", "M23.26: Derangement of other lateral meniscus due to old tear or injury", and "M23.29: Not specified derangement of meniscus".

To allow for easier interpretation, ICD-10 codes were grouped in Table 35 and Figure 46 as described in the methods. In general, a constant increase in the total number of diagnoses can be noticed between 2010 and 2014 (27,300 relevant diagnoses in 2010 to 33,919 diagnoses in 2014). This increase is mainly due to an increase in the diagnoses of derangement of meniscus due to old tear or injury (M23.2), other meniscus derangements (M23.3), and other internal derangements of knee (M23.8). In contrast, unspecified internal derangement of knee (M23.9) and tear of meniscus (S83.2) remained constant through the years. The percentage of cases that were reimbursed by the accident insurance varied from 17% for other meniscus derangements to 70% for tear of meniscus. This suggests that some ICD-10 are mostly related to traumatic problems (e.g. ICD-10 S83.2: tear of meniscus), whereas others are more frequent in patients with degenerative diseases (e.g. ICD-10 code M23.3: other meniscus derangements).

Importantly, the total number of diagnoses (i.e. ICD-10 codes) doesn't reflect the total number of treated patients. In fact, each case potentially received multiple relevant diagnoses (for example

a derangement of anterior horn of medial meniscus due to old tear or injury combined with a derangement of posterior horn of medial meniscus due to old tear or injury). Table 36 and Figure 47 illustrate how many cases received given numbers of relevant ICD-10 codes. In general, most cases received a single diagnosis for knee/meniscus derangement. For example, out of 26,290 cases hospitalized in 2014 with a relevant ICD-10 code, 20,069 (76%) had a single knee-related diagnosis, 5,081 (19%) had two knee-related diagnoses, and 1,140 (4%) had three or more knee-related diagnoses.

Interestingly, the percentage of cases with relevant diagnoses based on ICD-10 codes reimbursed by the accident insurance was higher than in those who had multiple meniscus derangement diagnoses (43% versus 16-27%).

Table 34 Frequency of relevant diagnoses based on ICD-10 codes from 2010 to 2014

Year	2010	2011	2012	2013	2014	
ICD-10 code	N	N	N	N	N	% accident
M23.20	967	1,253	365	168	146	36%
M23.21	241	200	212	324	349	36%
M23.22	2,227	2,583	3,337	4,080	4,647	29%
M23.23	1,230	1,436	1,374	1,331	1,481	36%
M23.24	165	214	360	482	638	27%
M23.25	277	404	613	911	999	34%
M23.26	574	619	803	913	1,129	32%
M23.29	435	281	223	81	50	18%
M23.30	1,871	1,700	824	368	310	15%
M23.31	254	274	428	480	473	18%
M23.32	4,560	5,007	7,024	7,259	6,381	17%
M23.33	1,954	2,023	2,012	1,855	1,868	18%
M23.34	384	508	808	1,005	957	17%
M23.35	515	681	1,080	1,252	1,201	20%
M23.36	789	1,006	1,574	1,699	1,634	16%
M23.39	745	339	154	151	98	16%
M23.80	207	223	213	118	163	33%
M23.81	692	636	710	781	976	54%
M23.82	62	77	53	62	71	46%
M23.83	52	78	135	167	168	49%
M23.84	19	24	32	24	33	52%
M23.85	13	14	18	16	15	40%
M23.86	36	38	52	71	54	44%
M23.87	15	12	25	49	40	35%
M23.89	865	1,060	1,325	1,418	1,660	32%
M23.90	168	92	58	23	25	40%
M23.91	34	64	59	40	34	47%
M23.92	14	5	19	65	24	13%
M23.93	24	17	60	49	32	22%
M23.94	4	1	3	7	8	38%
M23.95	0	8	15	20	11	9%

Year	2010	2011	2012	2013	2014	
ICD-10 code	**N**	**N**	**N**	**N**	**N**	**% accident**
M23.96	13	9	17	25	14	36%
M23.97	3	7	4	6	3	33%
M23.99	242	301	441	298	262	32%
S83.2	7,649	7,591	7,706	7,737	7,965	70%
Total	27,300	28,785	32,136	33,335	33,919	36%

NB: the percentages of cases covered by the accident insurance between 2010 and 2013 were not included in this table since they were very similar those reported in 2014.

Table 35 Frequency of the most relevant group of diagnoses based on ICD-10 codes from 2010 to 2014

Year	2010	2011	2012	2013	2014	
ICD-10 code	N	N	N	N	N	% accident
M23.2	6,116	6,990	7,287	8,290	9,439	31%
M23.3	11,072	11,538	13,904	14,069	12,922	17%
M23.8	1,961	2,162	2,563	2,706	3,180	40%
M23.9	502	504	676	533	413	31%
S83.2	7,649	7,591	7,706	7,737	7,965	70%
Total	27,300	28,785	32,136	33,335	33,919	36%

NB: the percentages of cases covered by the accident insurance between 2010 and 2013 were not included in this table since they were very similar those reported in 2014.

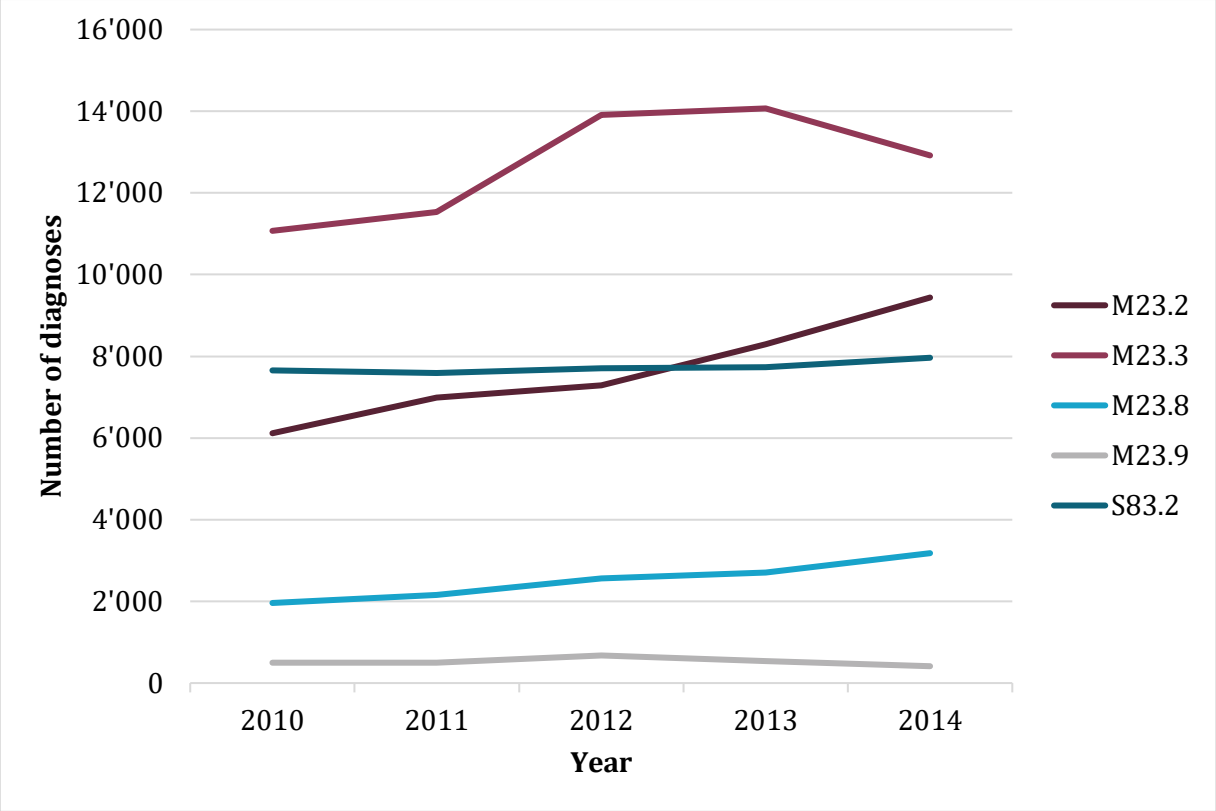


Figure 46 Frequency of the most relevant ICD-10 codes from 2010 to 2014 (grouped)

Table 36 Number of hospitalized cases who received one to 7 relevant diagnoses

Number of diagnoses	2010	2011	2012	2013	2014	
	N	N	N	N	N	% accident
1	22,591	22,499	21,915	20,853	20,069	43%
2	2,039	2,784	4,206	4,908	5,081	27%
3	187	204	511	695	939	20%
4	16	25	55	120	150	20%
5	1	1	10	19	37	16%
6			1	1	12	25%
7					2	0%
Total	24,834	25,513	26,698	26,596	26,290	0%

NB: since the percentages of cases covered by the accident insurance between 2010 and 2013 were not included in this table since they were very similar those reported in 2014.

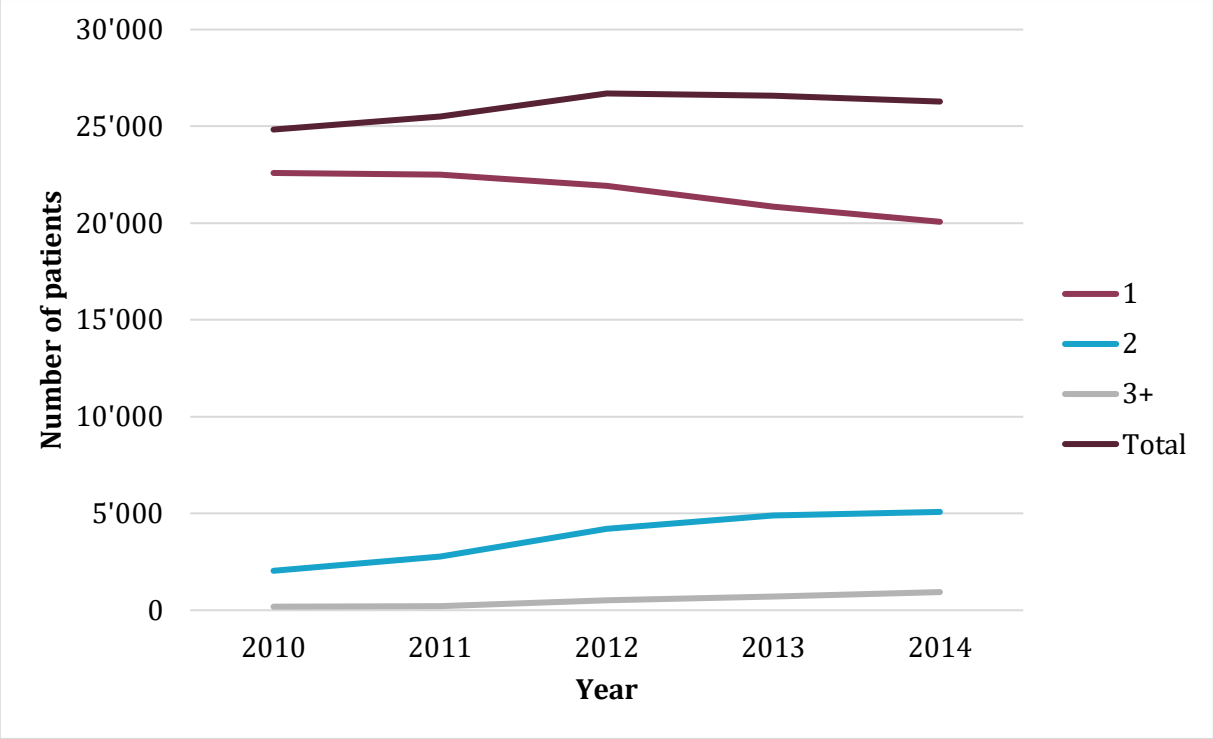


Figure 47 Number of cases receiving one, two, or more meniscus derangement diagnoses

4.2.2.6 Frequency of CHOP codes

The same CHOP codes were available for the years 2011 until 2014, whereas the coding in 2010 was different than in the following years. The codes in 2010 were merged into bigger categories. Table 37 shows the total number of CHOP codes related to knee surgical interventions between 2010 and 2014. The percentage of cases that were reimbursed by the accident insurance is reported only for 2014, because it was similar in the previous years. Importantly, the total number of CHOP codes doesn't reflect the total number of treated patients. A patient could potentially receive multiple knee treatments (CHOP codes) during a single hospitalization (for example a knee arthroscopy followed by a synovectomy and a meniscectomy).

In general, between 2011 and 2013, the total number of CHOP codes increased by over 2,000 per year (from 48,093 in 2011 to 50,131 in 2012 and 52,631 in 2013). In contrast, between 2013 and 2014 the total number of CHOP codes remained stable. The numbers extracted from the Swiss Hospital Statistics for 2010 were only partially interpretable. For the CHOP categories 80.6 and 80.76, the numbers were comparable to those reported in the following years for the CHOP codes 80.6X.10/ 80.6X.11 and 80.76.10, respectively. In contrast, for the other groups of CHOP categories in 2010 the numbers appeared to be higher, suggesting that other CHOP codes non-related to knee arthroscopy were included as well. For example, the CHOP group 80.26 in 2010 included almost 35,000 interventions, whereas the sum of the codes 80.26.00, 80.26.10, 80.26.20, and 80.26.99 in the following years reached 8,590 in 2011, 4,769 in 2012, 3,019 in 2013, and only 1,872 in 2014.

Interestingly, the use of some CHOP codes decreased drastically between 2011 and 2014 (80.26.00, 80.26.10, 80.26.99), whereas others showed a relevant increase (80.16.11, 80.6X.10, 80.76.10, 80.86.10, 81.47.22, and 81.47.24). The trends for the most frequent CHOP codes between 2011 and 2014 are illustrated in Figure 48. Table 38 illustrates how many cases received a given number of relevant CHOP codes. For example, in 2014 there were 15,048 hospitalized cases who received one single relevant CHOP code, 9,150 who received two relevant CHOP codes and 4,322 who received three relevant CHOP codes. Between 2010 and 2011 there were five patients who received during their hospitalisation up to 11 relevant CHOP codes (one main treatment, 10 secondary treatments). For example, one patient in 2011 received following codes describing the surgical intervention: 5x 80.16.10, 3x 80.76.10, 2x 80.36.20, and 1x 80.26.99. From 2011 to 2014 about half of the cases received only one single knee CHOP code, whereas the other half received two or more CHOP codes. Except for 2010, it can be noticed that the total number of hospitalized cases receiving at least one CHOP code remained constant (around 30,000-31,000 per year).

Table 37 Number of CHOP codes between 2010 and 2014

Year	2010		2011	2012	2013	2014	
CHOP group	N	CHOP code	N	N	N	N	% accident
80.16	474	80.16.10	584	852	864	889	26%
		80.16.11	799	1,416	1,562	1,590	36%
		80.16.12	6	8	6	3	0%
80.26	34,918	80.26.00	2,691	501	245	163	28%
		80.26.10	3,290	2,452	1,692	1,111	36%
		80.26.20	131	106	99	110	59%
		80.26.99	2,478	1,710	983	488	34%
80.36	468	80.36.20	375	427	554	536	18%
80.6	23,358	80.6X.10	20,737	22,737	22,788	22,156	36%
		80.6X.11	969	540	273	183	34%
80.76	7,597	80.76.10	7347	8,212	8,953	9,352	33%
80.86	5,028	80.86.10	2,673	3,077	4,518	4,276	43%
80.96	1,867	80.96.10	646	273	331	263	36%
		80.96.20	22	37	30	70	50%
81.47	2,118	81.47.11	45	46	49	51	47%
		81.47.13	48	63	59	29	41%
		81.47.18	7	13	5	11	9%
		81.47.22	4,294	6,576	7,911	8,407	31%
		81.47.24	396	640	1,258	1,957	41%
		81.47.25	27	21	58	79	41%
		81.47.51				22	50%
		81.47.90	114	96	52	58	66%
81.99	711	81.99.1A	77	98	158	56	34%
		81.99.3A	203	149	114	89	39%
		81.99.82	134	81	69	83	34%
Total	76,539	Total	48,093	50,131	52,631	52,032	35%

Figure 48 Frequency of the most common CHOP codes between 2011 and 2014

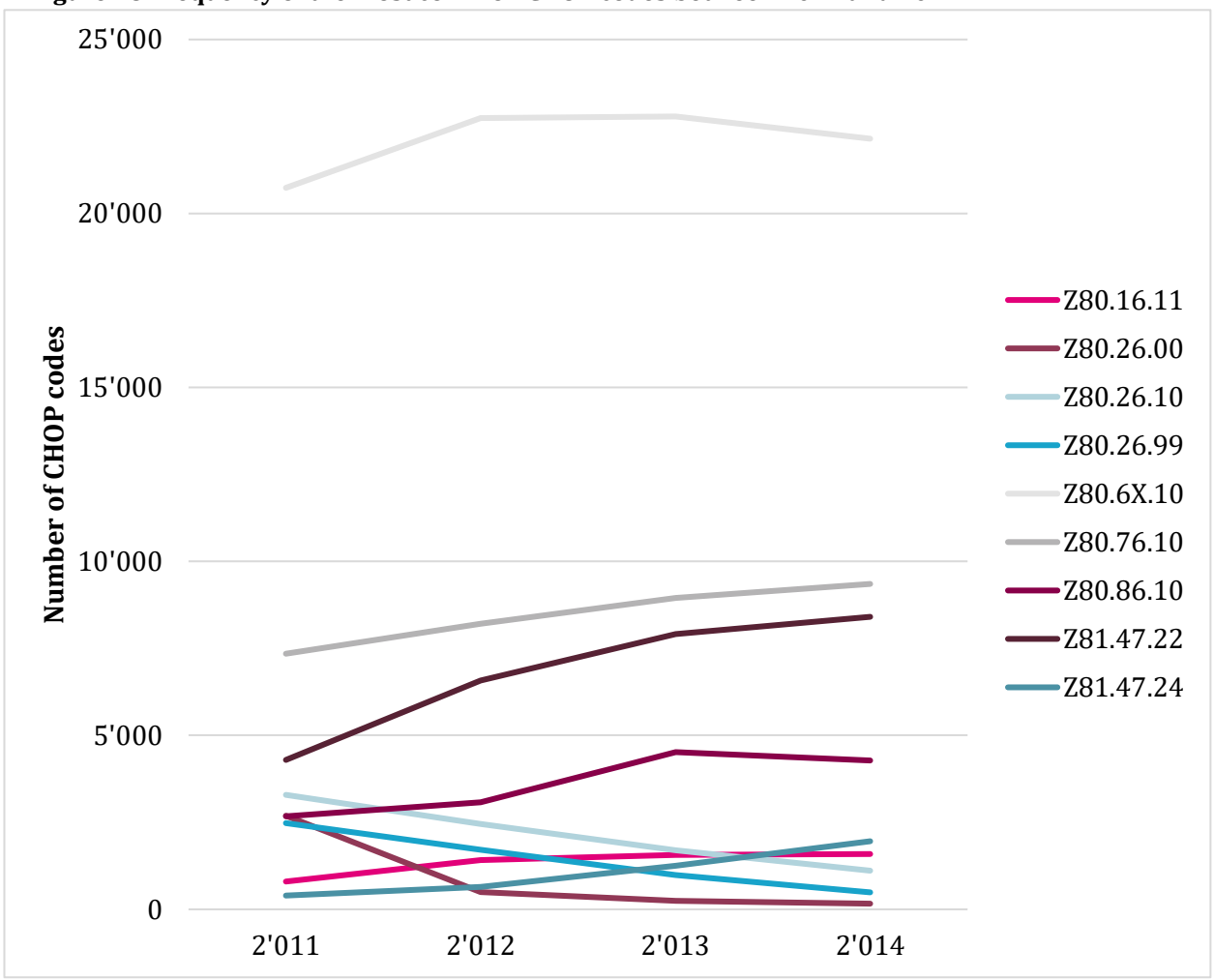


Table 38 Number of hospitalised cases who received one or more relevant CHOP codes

Number of CHOP codes	2010 N	2011 N	2012 N	2013 N	2014 N	2014 % accident
1	10,190	18,489	17,086	15,622	15,048	41%
2	17,656	8,983	9,183	9,583	9,150	35%
3	8,194	2,906	3,596	4,122	4,322	31%
4	1,192	529	745	909	1,066	30%
5	156	101	114	234	218	36%
6	101	30	30	73	36	42%
7	12	8	9	13	7	57%
8	12	4	4	12	9	33%
9	3	0	4	4	3	0%
10	5	2	3	1	0	0%
11	4	1	0	0	0	0%
Total	37,525	31,053	30,774	30,573	29,859	37%

4.2.2.7 Frequency of TARMED codes

A total of 9,740 and 8,769 cases received an ambulatory arthroscopic knee intervention in 2013 and 2014, respectively. As illustrated in Table 39, meniscus resections (TARMED code 24.5710), resection of the plica (TARMED code 24.5630) and removal of joint bodies (TARMED code 24.5620) were the most frequent interventions performed during an ambulatory knee arthroscopy. Assuming that the percentage of cases covered by an accident insurance would be similar to the percentage seen in the Swiss Hospital Statistics (37%), the number of cases meeting the ambulatory PIC criteria was estimated to be 6,136 in 2013 and 5,524 in 2014.

Table 39 Number of ambulant arthroscopic interventions to the knee/meniscus (TARMED codes)

TARMED code, description	2013	2014
24.5610, Arthroskopie Kniegelenk	**9,740**	**8,769**
24.5615, + Resektion einer Gelenkzyste/Sehnenzyste/tiefen oder oberflächlichen Bursa im Kniegelenkbereich, als Zuschlagsleistung	142	176
24.5620, + Entfernung freier Gelenkkörper bei Arthroskopie Knie	1,280	1,259
24.5630, + Plicaresektion bei Arthroskopie Knie	2,505	2,336
24.5640, + Synoviektomie subtotal bei Arthrose bei Arthroskopie Knie	896	808
24.5650, + Synoviektomie subtotal bei {pcP}/postinfektiös bei Arthroskopie Knie	46	40
24.5660, + Retinakulumspaltung (lateral release) bei Arthroskopie Knie	232	205
24.5670, + Retinakulumnaht bei Arthroskopie Knie, jede Methode	11	8
24.5680, + Osteophytenabtragung u/o Notch-Plastik bei Arthroskopie Knie	576	524
24.5690, + Anlegen einer Spüldrainage bei Arthroskopie Knie	263	276
24.5700, + Meniskustoilette bei Arthroskopie Knie	812	623

TARMED code, description	2013	2014
24.5710, + Resektion Meniscus medialis/Meniscus lateralis, partiell/total, bei Arthroskopie Knie, pro Meniskus	8,570	8,001
24.5720, + Resektion diskoider Meniscus medialis/Meniscus lateralis/Meniskusganglion, partiell/total bei Arthroskopie Knie	217	185
24.5730, + Naht Meniscus medialis/Meniscus lateralis bei Arthroskopie Knie	543	222
24.5735, + Mikrofrakturierung oder Pridie-Bohrung(en)	389	439
24.5740, + Versorgung bei Osteochondrosis dissecans bei Arthroskopie Knie, Fixation des Dissekates	26	32
24.5750, + Plastische Versorgung bei Osteochondrosis dissecans bei Arthroskopie Knie	34	19
24.5760, + Versorgung Ruptur des vorderen Kreuzbandes bei Arthroskopie mittels Naht u/o transossärer Reinsertion	4	3
24.5770, + Versorgung Ruptur des vorderen Kreuzbandes mittels Naht u/o transossärer Reinsertion sowie Augmentationsplastik bei Arthroskopie Knie, jede Methode	0	4
24.5780, + Versorgung Ruptur des vorderen Kreuzbandes mittels autoplastischem Ersatz bei Arthroskopie Knie, jede Methode	27	31
24.5790, + Versorgung Ruptur des vorderen Kreuzbandes mittels alloplastischem Ersatz bei Arthroskopie Knie, jede Methode	0	3
24.5810, + Versorgung Ruptur hinteres Kreuzband mit. Naht u/o transossäre Reinsertion am femoralen Ursprung sowie Augmentationsplastik bei Arthroskopie Knie, jede Methode	0	1
24.5820, + Versorgung Ruptur hinteres Kreuzband mittels autoplastischem Ersatz bei Arthroskopie Knie	0	4
24.5830, + Versorgung Ruptur hinteres Kreuzband mittels alloplastischem Ersatz bei Arthroskopie Knie	0	1
24.5840, + Zuschlag für tibiale Fixation bei Versorgung Ruptur hinteres Kreuzband bei Arthroskopie Knie, jede Methode	1	0
24.5850, + Versorgung Tibiakopffraktur bei Arthroskopie Knie mittels Osteosynthese	0	0
24.5860, + Entfernung des Osteosynthesematerials bei Arthroskopie Knie	52	19

NB: the TARMED code 24.5610 represents the total number of knee arthroscopies. All other TARMED codes illustrate the interventions that may be performed during a knee arthroscopy. As a combination of several interventions is possible, the total number of knee arthroscopies is lower than the sum of all interventions.

5 Discussion

This report assessed the clinical effectiveness and economic consequences in patients with knee pain due to osteoarthritis and meniscal degeneration undergoing knee arthroscopy compared to control. The results of the clinical effectiveness and safety assessment are discussed first, followed by the discussion of the results of the cost-effectiveness and budget impact analyses. In the final section, results from both parts of the HTA are discussed jointly.

5.1 Clinical effectiveness and safety

5.1.1 PICO 1

Twenty-one RCTs were included in the report on the effectiveness and safety of arthroscopy in patients with degenerative changes of the knee. There is no evidence that arthroscopic interventions have any benefit on outcomes measured at short- or intermediate follow-up with the exception of a small effect on the reduction of pain at short follow-up. Therefore, it remains unclear whether knee arthroscopy has an effect on the assessed outcomes. Long-term follow-up data were not available. Reporting on harm was scarce and no conclusion can be drawn regarding the benefit-harm balance. Cross-over rates to surgery in control patients were considerable. This may have affected the results and under the intention-to-treat (ITT) approach this might have reduced treatment effects for some outcomes in particular for intermediate or long-term follow-up. This bias could only be compensated for in an individual patient data meta-analysis, with censoring of patients at the time of switching. However, this would also imply that patient relevant outcomes would have been measured at the time of switching, allowing for an adequate time-updated analysis. The overall quality of evidence at short- and intermediate follow-up was judged very low and low, respectively. In addition, there is no evidence that the subgroup of patients with solely DMK has a benefit from arthroscopic treatment.

The quality of evidence was judged very low and low for most of the outcomes, mainly because of limitations of the included RCTs. Effect estimates of subjective outcomes (most outcomes of the present report are patient reported) are likely to be overestimated if blinding is missing[110-112], which causes substantial uncertainty of the empirical evidence. Blinding of patients and outcome assessors may be challenging in this context. Sham interventions may put patients at risk and therefore have serious ethical implications, which nowadays might be considered as unacceptable.[113] Of the 21 RCTs, only four (FIDELITY 2013, Kalunian 2000, Moseley 1996, Moseley 2002) reported blinding (sham/placebo surgery) of patients and personnel. Interestingly, cross-over rate in one of the blinded RCTs (FIDELITY 2013) was less than 5% (the other three did not report cross-over), and hence was lower than in several other non-blinded RCTs where cross-over rates varied between 0% to 35% (see section 8.4.4 of Appendix 5). Overall, eleven RCTs reported cross-over (defined as patients in the comparator arm who received subsequent arthroscopic surgery) in 118 out of 637 (18.5%) patients in the control group. It appears that cross-over rates are lower in adequately blinded studies, but this is not the only reason for the variation of cross-over rates between trials. For instance, patients with preference for arthroscopy might only accept trial participation if there is the opportunity to switch interventions. From the trial protocol of MeTeOR (highest cross-over rate), it is stated that control patients who received physical therapy

were at “any time” allowed to switch to arthroscopy; hence, patients with preference for arthroscopy were probably more likely to accept the trial conditions including the risk to be randomized into the physical therapy group. This is in contrast of the blinded FIDELITY study where cross-over was also allowed, but no earlier than 6 months after sham surgery. This might have prevented patients with preference for arthroscopy from participating. In general, the reporting of whether, or when, cross-over was allowed was very inconsistent and could not be assessed further. To conclude this paragraph and as mentioned above, high cross-over rates are problematic in the ITT analysis, and may lead to underestimation of treatment effects of the intervention. However, in daily routine, it is frequent that patients receiving conservative treatment will switch to arthroscopic intervention, and hence the RCT findings may well reflect a realistic scenario.

A further limitation of the RCTs was the high risk of attrition bias. Attrition was high or unclear in most of the included RCTs. Missing outcome data impact effect estimates; therefore, a high proportion of missing data can lead to biased estimates. This is important for rare binary outcomes, especially if the ratio of participants with missing data to participants with events is high.[10] Attrition in combination with the scarce reporting of harm (adverse events and serious adverse events) make it impossible to draw any conclusions about potential harm caused by the intervention. Of note, in two RCTs (Kang 2005, Vermesan 2013) the unit of randomisation was unclear, or knees were randomised; thus, in both RCTs it was unclear whether global assessment was reported per knee or per patient. If both knees were operated in some patients, these observations cannot be analysed as independent observations. Ignoring this fact in the analysis will lead to biased estimates because the data are clustered.[114] Therefore, the findings of these RCTs were reported separately.

The patient populations with pain due to degenerative changes of the knee were very heterogeneous and did not allow for a clear distinction between patients with OA and DMK. First, patients presenting with pain from knee OA often have OA in combination with DMK or vice versa. From the existing literature it was expected that the prevalence of DMK in patients with symptomatic OA might range between 39% and 76%.[115-118] Second, clinical diagnosis does not allow for a clear distinction of pain due to OA or DMK. The populations of the included RCTs were typically recruited because of persistent knee pain. Whereas radiographic evidence was used in most RCTs to assess OA, only few used MRI or arthroscopic evidence to assess potential meniscus degenerations. Degenerative meniscal lesions are increasing with age and many individuals with such lesions are asymptomatic. Diagnosis of both OA and DMK is thus not entirely dependent on MRI or x-ray, which only reveal structural changes of the knee joint, but must be made by integrating physical examination, patient history, duration and severity of symptoms and findings from imaging techniques.[116 119-121] The radiographic evidence in combination with physical examination often provides an incomplete picture. For example, degenerative meniscus is often revealed only post-hoc during arthroscopy. It should be underlined that structural damage of the knee is not always associated with dysfunctional and/or painful symptoms in the general population.[116 122 123]

In the present report, the populations of the included RCTs were classified into four groups. This classification system may have some limitations, but was considered the best available compromise. Our approach is in line with existing literature.[33-39] Fourteen of the 21 included RCTs were classified as having included patients without allowing for a clear classification and were rated as ‘mixed unclear’. Moreover, in the included RCTs, arthroscopic interventions often revealed additional pathologies in addition to previously known radiographic signs of OA. For

instance, in the only RCT recruiting patients with OA only and which excluded patients with large meniscal tears (Kirkley 2008), around 80% of patients received meniscal debridement during arthroscopic intervention due to smaller lesions that had not been identified by MRI. These observations underline that a clear distinction between DMK and OA may be difficult to achieve in daily routine. Based on the observed effects within sub-groups of RCTs representing DMK only and OA only patients, no conclusions can be drawn on whether an OA or a DMK population would benefit from therapeutic arthroscopy.

Several RCTs reported that a considerable number of eligible patients refused to participate. In the seven RCTs with more than 150 patients (FIDELITY, Gauffin 2014, Kirkley 2008, Kise 2016, KIVIS, MeTeOR, Moseley 2002), between 3.2% and 59% of eligible patients declined to participate. Where reported, the main reason was patient preference for one of the randomised treatments. Patients' preference for one treatment (arthroscopy or comparator) does affect overall generalizability of findings.[124 125]

Some instruments for assessing OA are not suitable for patients with any kind of disorder of the knee.[126] For instance, the frequently reported KOOS was initially developed for a population at risk of developing OA; currently, only the Swedish and Dutch versions of the instrument are validated for a mixed patient population with osteoarthritis and/or meniscal tears. The English version is not validated for assessing a mixed population.[127-129] Both RCTs (Gauffin 2014, Herrlin 2007) using the KOOS instrument in patients with only DMK were conducted in Sweden; however, a recent systematic review reported flaws in the Swedish KOOS version, questioning the instrument's validation for meniscal tears.[127] This systematic review critically appraised the instruments used to measure patient-reported outcomes in patients with meniscal tears and the authors concluded that the evidence for the validity of the instruments available for this population was of poor quality and incomplete.[127] This is a matter of concern mainly for the critical outcomes of pain, function and global assessment (measured by KOOS, Lysholm, WOMAC, and International Knee Documentation Committee), and WOMET, a disease specific quality of life measure, but less so for the EQ-5D, which is a generic instrument that has been widely validate in many different language and patient populations. Measuring an outcome with a wrong or inappropriate instrument can affect the judgment of the quality of evidence according to GRADE. Though instruments used were not always validated to assess a mixed population, the quality of evidence was not downgraded for serious indirectness because of the uncertainty of the distinction between imminent causes of knee degeneration and pain (see above OA and DMK) in most RCTs; hence, the magnitude of indirectness was not assessable.

Adverse events were inconsistently reported in most RCTs. Three RCTs (Biedert 2000, FIDELITY, Gauffin 2014) counted and reported patients undergoing subsequent surgery as patients having an adverse event. In these RCTs it was not possible to distinguish between patients crossing over and undergoing subsequent surgery, and those with an adverse event. This approach may have led to an increased numbers of patients with adverse event in the comparator group (arthroscopy 12/173 [6.9%] versus control 29/163 [17.8%]). However, in four RCTs, patients with adverse events were reported separately from those with subsequent surgery. In these trials, the number of patients with adverse events was balanced between groups (arthroscopy 31/389 [8.0%] versus control 29/398 [7.3%]). In general, safety outcomes were rarely reported and might be biased due to high attrition in several RCTs (see discussion above). Hence the quality of evidence was low to very low for adverse events and very low for serious adverse events. Consequently, too few data were available to weigh potential benefits against potential harms. A recent systematic review compared arthroscopy with conservative treatment in patients with degenerative knee

disease while assessing complications reported in retrospective cohort studies and RCTs. The authors concluded that within a three-month time frame there is probably a small risk of mortality, venous thromboembolism, infection or nerve damage due to knee arthroscopy.[130] In addition, cohort studies reported that adverse events or complications due to arthroscopy occurred in 0.02% to 1% of patients within a three-month time frame.[131-133]

An abundance of literature has been published on therapeutic knee arthroscopy; therefore, for the purposes of this discussion, it was decided to focus on reviews published after 2012 since more than a third of the RCTs included in this report were published after that time. Sixteen reviews (both narrative and systematic reviews including meta-analyses) were found from 2012 onwards[33-39 130 134-141], along with two HTA reports.[4 142] The most relevant systematic reviews with meta-analyses are briefly addressed. Thereof, the IQWiG assessed the effect of arthroscopy in an OA population in 2014, Khan et al. and van de Graaf et al. in patients with mainly degenerative meniscus in 2014 and 2016, respectively, and Brignardello-Petersen et al. in patients with both OA and DMK in 2017.[4 35 39 130] Compared to IQWiG, two RCTs conducted in OA patients were added in the present report (Østerås 2012, Saeed 2015); Østerås 2012 was not found by the IQWiG search and Saeed 2015 was published thereafter. No additional RCTs with degenerative meniscus patients were identified compared to the reviews by Khan et al. or van de Graaf et al., but the present report added a substantial number of additional RCTs that were conducted in patients with degenerative osteoarthritic knee disease compared to the review by Brignardello-Petersen et al. In the present report, nine more RCTs were included. In general, the findings of the present report are in line with those from these recent evidence syntheses.

For endpoints at short-term follow-up, these systematic reviews reported either no or only small statistically significant differences between arthroscopy and comparators for pain and function outcomes.[4 35 39] The IQWiG HTA found no indication of benefit for the treatment of OA when comparing arthroscopy to non-active or active comparators for the outcomes pain, function, global assessment and quality of life.[4] Khan et al. evaluated the effectiveness of arthroscopy in populations with degenerative tears of the meniscus with mild or no OA and reported no statistically significant difference in pain scores at short-term follow-up when comparing arthroscopy to non-operative treatment (MD 0.20, 95% CI [-0.67, 0.26]).[35] A small, statistically significant difference in favour of arthroscopy was found for function at short-term follow-up (SMD 0.25, 95% CI [0.02, 0.48]). However, this difference was judged by the authors as not clinically relevant (0.5 SD).[35] The more recent review by van de Graaf et al. compared arthroscopic partial meniscectomy to conservative treatment for patients with non-obstructive meniscal tears.[39] Statistically significant differences favouring arthroscopic surgery were found at six months for pain (KOOS MD 3.56, 95% CI [0.18, 6.95] or VAS/numerical rating scale MD 0.56, 95%CI [0.28, 0.83]) and physical function (SMD 0.17, 95% CI [0.01, 0.32]). These differences were considered clinically irrelevant.[39]

Overall, short-term results differed little from the results of the present report and the reported differences were too small to be potentially clinically relevant. The small difference in short-term function outcomes between this report and the reports by Khan et al. and van de Graaf et al. is probably attributable to methodological differences, such as different inclusion and exclusion criteria.

At intermediate follow-up, findings on pain and function outcomes presented in this report were consistent with those of recently published reviews, where no statistically significant effects were found between arthroscopy and control.[4 35 39]

A systematic review including RCTs and non-randomised studies published by Brignardello-Petersen et al. in 2017 evaluated the benefits and harms of knee arthroscopy versus conservative management in patients with degenerative knee disease.[130] Though the present report found nine additional RCTs compared to that review, the reported results are very similar. Small statistically significant differences were found at short-term follow-up of three months for pain (MD 5.38, 95% CI [1.95, 8.81]) and function (MD 4.94, 95% CI [1.50, 8.38]), while no statistically significant differences between arthroscopy and conservative treatment were found at long-term follow-up of up to two years for pain and function.[130] In addition, no statistically significant beneficial effects were found for quality of life and knee replacement, which is consistent with the findings of this report.[130] The Brignardello-Petersen et al. review did not look at OA only and DMK only populations, but assessed subgroups based on OA status (based on radiographic evidence) and reported that OA status did not affect the interpretation of the effect estimates for pain and function.[130]

All systematic reviews, including the present HTA report, found that safety outcomes (adverse events and serious adverse events) were poorly reported. The 2014 IQWiG report did not draw any conclusions regarding harm, because of insufficient data, while the review by van de Graaf et al. was only able to summarise the reported adverse events of two RCTs.[4 39] Brignardello-Petersen et al. included non-randomised studies in addition to RCTs to assess harms, though heterogeneity was high and the assessed time frame was, despite non-randomised studies, limited to three months.[130]

In conclusion, there is no evidence that arthroscopic interventions have any benefit on outcomes measured at short- or intermediate follow-up, with the exception of a small effect on pain at short follow-up. Therefore, it remains unclear whether knee arthroscopy has an effect on the assessed outcomes. Long-term follow-up data were not available. Reporting on harm was scarce and no conclusions can be drawn regarding the benefit-harm balance. The overall quality of evidence at short- and intermediate follow-up was judged to be very low and low, respectively. In addition, there is no evidence that the subgroup of patients with solely DMK has a benefit from arthroscopic treatment. Findings of this assessment may be generalizable to a broader population experiencing knee pain due to a degenerative knee disorder. The findings are consistent with recently published reviews.

5.1.2 PICO 2

Arthroscopic procedures on the knee can be performed in the inpatient or outpatient setting. The inpatient setting in theory facilitates immediate post intervention care or pain management, while patient preference and reduced costs are potential advantages of the outpatient setting. Only one RCT (Weale 1998) identified compared the inpatient setting with the outpatient setting.

This single RCT had several limitations, such as high risk of bias, reporting of pain only in the week after discharge and low sample size. About 30% of the participants in both arms received only a diagnostic arthroscopy and no intervention. Only 10% and 16% of the patients in the outpatient or short-stay groups, respectively, were diagnosed with OA. Half of the patients in either group with meniscal tears received meniscectomy. Moreover, three participants (6%) crossed-over from the outpatient to the inpatient group, while six participants (12%) crossed-over from the inpatient to the outpatient group. Because over a third of the RCT population received only

diagnostic arthroscopy and no follow-up outcomes were reported, no generalizable conclusions can be made.

5.2 Cost-effectiveness and budget impact

The systematic search of economic studies identified four eligible cost-effectiveness studies. Two studies compared knee arthroscopy to non-operative treatment (Losina 2015, Marsh 2016). Two other studies were based on an assessment of patient status before and after arthroscopy (Hutt 2015, Lubowitz 2011). All studies were cost-utility analyses, i.e. cost-effectiveness analyses using QALYs as the measure of benefit. Three studies based their calculations on prospectively collected information (Hutt 2015, Lubowitz 2011, Marsh 2016). The study by Marsh 2016 was based on the RCT by Kirkley 2008, evaluating the effectiveness of arthroscopic surgery in addition to optimized physical and medical therapy in patients with symptomatic, radiographic knee OA over a 2-year period.[51] The decision analytic modelling conducted by Losina 2015 was mainly informed by data from the MeTeOR RCT, a multicentre study involving symptomatic patients with a meniscal tear and evidence of mild-to-moderate OA.[57] The cost-effectiveness analysis conducted by Hutt 2015 was based on data from a cohort of 43 patients with radiological OA that were assessed pre- and postoperatively. Similarly, Lubowitz 2011 analysed data on a cohort of 93 patients before and after knee arthroscopy.

Marsh 2016 reported that knee arthroscopy is more expensive and less effective than non-operative interventions alone. The three other studies were in favour of knee arthroscopy. The results of Losina 2015 suggested that arthroscopic surgery may be cost-effective (ICER <CHF 30,000 per QALY gained) or even cost-saving if compared to physical therapy alone. The studies conducted by Lubowitz 2011 and Hutt 2015 showed very similar results, suggesting that knee arthroscopic intervention ameliorated the condition of patients, with an ICER of CHF 7,200-7,400 per QALY gained.

The quality of reporting of Lubowitz 2011 and Hutt 2015 assessed with the CHEERS checklist was poor, showing significant reporting issues that may hint at underlying methodological issues (affecting the design, methods and results). Comparing preoperative status with postoperative status does not represent a reliable basis for deciding whether an intervention is cost-effective or not. In contrast to RCTs, pre-post clinical studies provide, in general, very low-quality evidence of clinical effectiveness (see section 3.1.7). Losina 2015 and Marsh 2016 were of higher reporting quality. They were based on different designs, assumptions, and sources and led to discordant results. Several methodological issues were identified.

In particular, Losina 2015 appears to be affected by several methodological flaws which will be detailed. Losina 2015 was the only study partially based on a RCT with contemporary controls (i.e. comparing arthroscopic surgery with a control group) and showing results in favour of knee arthroscopy. These results were inconsistent with those of the present assessment of the clinical effectiveness and safety of arthroscopy (see section 3.2); hence, a more detailed discussion is needed. The effectiveness assumptions used to populate the model were scrutinized. The authors reported that distributions of pain relief three months after initial treatment (arthroscopy or physical therapy) were derived from the MeTeOR RCT using a transformed Knee injury and Osteoarthritis Outcome Score (KOOS) Pain scale.[57] Moreover, based on the expert opinion of a panel of MeTeOR clinical investigators, the authors assumed that changes in pain status attributable to a specific treatment would be limited to the first six months following that

treatment (i.e. six months after arthroscopy or after initiation of physical treatment). For use in the health economic model, pain results from MeTeOR were dichotomized at a KOOS score of 25, with scores above 25 characterized as 'Moderate Pain' and scores at or below 25 as 'Low Pain'. Transition probabilities for pain and pain resolution were stratified by OA severity.[57 100] Utilities for low and high pain were also derived from the MeTeOR RCT and were assumed to be equal across study groups. The change in overall QALYs across groups was driven by different probabilities of failed pain relief (patients undergoing physical therapy only had a higher probability of failed pain relief if compared to patients undergoing surgery), and different probabilities of pain incidence and pain resolution (higher for patients undergoing surgery).

A closer look at the MeTeOR RCT published in 2013 revealed several potential issues. First, contrasting Losina 2015's favourable conclusions regarding the cost-effectiveness of arthroscopy, the overall results of the RCT indicated only a small positive effect of arthroscopy compared to no surgical treatment in WOMAC physical function and KOOS pain scores, both in favour of knee arthroscopy (see Figure 2A and 2B in Katz 2013) that was only visible after the first three months of follow-up. At six months after randomisation differences were very small and non-significant (probably below the minimal important difference for the measurement instruments used). After twelve months of follow-up, both study groups showed almost identical reductions in WOMAC physical function and KOOS pain scores. Unfortunately, the derivation of the approaches for after three months (i.e. when the difference in favour of knee arthroscopy were less pronounced or even inexistent) was insufficiently described. This could have led to biased results in favour of arthroscopy. Second, despite reporting similar WOMAC physical function scores and KOOS pain scores, the two patient groups in the MeTeOR RCT were slightly different concerning the severity of OA. In particular, the arthroscopy group included patients with more severe OA (28.0% KL grade 3 versus 23.1% in the physical therapy group).[57] The inclusion of more severe patients may have had an impact on overall score changes after arthroscopy. In fact, in their supplementary table 3, Katz 2013 reported a higher mean change in WOMAC score in the arthroscopy group only for KL grades 0-2 (21.9 versus 17.2 in the physical therapy group). In contrast, patients with KL grade 3 in the arthroscopy group reported lower improvements in WOMAC scores if compared to the physical therapy group (19.0 versus 21.9). Including less severe patients in the physical therapy group may have favoured non-operative treatment.

Additional issues in the paper by Losina 2015 include lack of clarity in the description of how the probabilities of pain incidence and pain resolution were derived. For example, for patients undergoing delayed arthroscopy after physical therapy, the probabilities of pain incidence or pain resolution three months after treatment were set to zero (compared to a range of 0.115-0.227 for physical therapy and 0.230-0.483 for immediate arthroscopy). It remains unclear if for this patient group, the authors started the modelling only at the time point of the delayed arthroscopy, which would not be state-of-the-art. Moreover, the assumption concerning the effectiveness of delayed arthroscopy was based on only 24 patients that crossed-over between three and six months. Given the limited number of patients, the risk of inaccurate results is particularly high. A lack of clarity was also found regarding the link between utilities and pain status: the authors reported that RCT-based utilities were calculated for two pain status groups, based on EQ-5D results (of note, EQ-5D results were not reported in the publications of the MeTeOR RCT, as the EQ-5D results were specific to economic analyses as outlined in the trial protocol). However, the approach to calculation was not described. Also, the authors mentioned that the utilities were derived from the MeTeOR study, where participants were asked to fill out the EQ-5D instrument. However, in the mentioned publication,[57] the collection of these data was not reported (and is therefore absent

from the effectiveness section in this report). A lack of explanation was also noticed regarding the derivation of several other inputs parameters. For example, the authors did not explain how the clinical end economic parameters for the periods following the first year of follow-up were derived from OA cohort studies. It is therefore impossible to judge whether their calculations were methodologically correct and whether the assumptions were realistic.

A final major issue in the underlying MeTeOR RCT was the high cross-over rate from conservative treatment (physical therapy) to arthroscopy. Out of 167 patients in the physical therapy group, 62 subjects (37%) finally underwent arthroscopy (26 subjects in the first three months after randomisation, 24 subjects between three and six months follow-up, and 12 patients after six months follow-up). Exploratory results (Figure 2C and supplement page 2 in Katz 2013) suggest more substantial improvements in the cross-over patients compared to those who remained on conservative treatment.[57] How the effects of cross-over were considered in the derivation of the input parameters in the model of Losina 2015 remains unclear.

The above-mentioned aspects suggest that the results reported by Losina 2015 are not sufficiently valid. Given how the methodological approaches are presented, the mechanism behind the reported favourable cost-effectiveness results (which contrast the much less clear-cut overall results of MeTeOR) remains unclear.

The study by Marsh 2016 concluded that arthroscopic surgery in addition to non-operative treatment (i.e. physical therapy) for knee OA is not an economically attractive treatment option if compared to non-operative treatment alone. The results were based on a self-conducted, single centre RCT including 168 patients (represented in the assessment of the clinical effectiveness and safety as Kirkley 2008).[51] As in the case of Losina 2015, several issues have to be considered. First, despite randomisation, the patient characteristics in the intervention and non-operative groups differed with regard to gender distribution (39% versus 28% in the non-operative group), Kellgren-Lawrence grade (48% KL grade 2 and 47% KL grade 3 versus 42% KL grade 2 and 53% KL grade 3 in the non-operative group), and WOMAC total scores (1222.9 versus 1354.1 points in the non-operative group; the WOMAC score was rescaled here so that a higher number indicated a better outcome). After 24 months follow-up, the WOMAC scores for the two groups were 1526.4 and 1510.8, respectively. This implied that patients undergoing arthroscopy had a higher improvement than patients in the non-operative group (+303.5 versus +155.5 points). Interestingly, the differences in WOMAC scores between the patient groups were not reflected in terms of utilities (utilities were measured with a standard gamble technique). In fact, the baseline utilities as well as the utilities reported after 24 months follow-up were almost identical: 0.79±0.22 in the arthroscopy group versus 0.80±0.21 in the non-operative group at baseline and 0.84±0.23 versus 0.86±0.16 at 24 months. Second, in the cost-effectiveness analysis the authors assumed a QALY increase of 1.64±0.40 in the arthroscopy group and an increase of 1.66±0.30 in the non-operative group. The very small QALY difference in favour of the non-operative treatment, combined with its lower costs if compared to arthroscopy, led to a situation in which physical therapy dominated arthroscopy. Still, considering the differences in the group characteristics and the limited differences in utilities and QALYs (alongside considerably larger standard deviations), a clear conclusion on a final judgement on the cost-effectiveness of the alternative treatments is not possible.

Given the very limited health economic evidence, it is difficult to draw firm conclusions on the cost-effectiveness of arthroscopic surgery in patients with degenerative changes of the knee. Although three out of four eligible studies reported arthroscopy to be cost-effective, this cannot

be regarded as convincing, given the methodological issues described above and also in light of the results of the clinical part of this HTA (see section 5.3). On the other hand, it cannot be ruled out that arthroscopic surgery may be an economically sensible option for some groups of patients, e.g. in light of cross-over observed in the available RCTs.

The objective of the budget impact analysis was to investigate the total in- and outpatient costs of arthroscopic surgery in patients with DMK in Switzerland. According to the scope, the population included in this analysis was meant to be slightly different from the population addressed in the clinical review and in the assessment of cost-effectiveness, which considered patients with any degenerative knee problems. The results of the budget impact analysis suggested that the total expenditure for knee/meniscus derangement in Switzerland, aimed at approximating a KVG perspective, ranged from CHF 53.52 Mio. to CHF 71.93 Mio. in 2013 and from CHF 52.30 Mio. to CHF 67.73 Mio. in 2014. Outpatient costs accounted for 20-28% of the total costs. The results of the budget impact analysis according to our second patient selection strategy, i.e. a strategy based on ICD-10 codes and CHOP codes which were considered as more realistic than the first strategy also considering DRG codes, suggested total inpatient costs of CHF 58.10 Mio. in 2010, CHF 55.87 Mio. in 2011, CHF 58.44 Mio. in 2012, CHF 57.20 Mio. in 2013 and CHF 54.47 Mio. in 2014. Total outpatient costs were estimated to be CHF 14.73 Mio. in 2013 and CHF 13.26 Mio. in 2014. The total expenditure for meniscus derangement in Switzerland was thus estimated to be CHF 71.93 Mio. in 2013 and CHF 67.73 Mio. in 2014.

For the sake of completeness, the results of the budget impact analysis according to the first patient selection strategy suggested that the total inpatient costs were CHF 59.31 Mio. in 2010, CHF 61.65 Mio. in 2011, CHF 134.68 Mio. in 2012, CHF 53.52 Mio. in 2013 and CHF 52.30 Mio. in 2014. Within this set of results, the cost estimations for the years 2010, 2011 and 2012 seemed to be particularly uncertain, for two reasons: first, a different DRG coding system (APDRG, not yet SwissDRG) was used until 2010 and 2011. Second, the distribution of patients in 2012, the first year after the introduction of the new DRG coding system (SwissDRG was unrealistically high, probably due to misclassifications.

The results of the budget impact analysis are in line with the results reported in a very recent publication of the Swiss Health Observatory (Obsan).[143] Assuming approximately 14,000 meniscectomies per year and inpatient costs of CHF 4,889 per case, the authors of the Obsan estimated total inpatient costs of CHF 55.6 Mio. in 2016. In the present study, the assumed unit costs were lower (CHF 3,700), whereas the estimated number of arthroscopic surgeries to the knee were higher (ranging between 15,000 and 16,000). Several reasons may explain these discrepancies. First, the unit costs used in the Obsan report consisted in the mean costs registered by the CSS insurance in 2016. Although it was not clearly stated, these costs may have included cases with private insurance (and consequently higher costs). Second, the costs estimations also included examinations before and after surgery. The magnitude of these costs was not described, but it is reasonable to assume that this may have also led to higher unit costs per patient. Third, the list of CHOP codes used to identify patients undergoing knee arthroscopy in the Obsan report was slightly different if compared to the present work, which included a larger variety of arthroscopic knee interventions. This could explain the higher estimated number of cases.

In a recently published observational study investigating the use of arthroscopic meniscal surgery in degenerative knee disease in Switzerland and using administrative claims data of a major Swiss health insurance company, the authors reported that the incidence of arthroscopic partial meniscectomy, debridement, and lavage in patients over the age of 40 was 388 per 100,000

person-years in 2012 and 352 per 100,000 person-years in 2015.[2] To identify inpatient surgeries, the CHOP codes 80.6X.10 and 80.86.11 were used, whereas outpatient procedures were identified through TARMED codes (24.5710 and 24.5700). On average, 68% of all identified patients were inpatients.

These numbers are comparable with those found in the present analysis. The Swiss Hospital Statistics 2014 showed that there were 22,156 CHOP codes 80.6X.10 (36% of them reimbursed by the accident insurance) and 4,276 CHOP codes 80.86.11 (43% reimbursed by the accident insurance). Considering only the cases that were not reimbursed by the accident insurance (n=16,617) and referring them to the Swiss population over 40 years old in 2014 (n= 4,374,989), a rate of 380 per 100,000 person-years could be calculated. The rates for the previous years would be 407 per 100,000 person-years for 2013, 398 per 100,000 person-years for 2012, and 368 per 100,000 person-years for 2011. Incidence rates based on the number of cases identified through the second patient selection strategy (considering all patients with at least one relevant ICD-10 code and one relevant CHOP code; see above) would be 336 per 100,000 person-years in 2014, 358 per 100,000 person-years in 2013, 371 per 100,000 person-years in 2012, 360 per 100,000 person-years in 2011, and 380 per 100,000 person-years in 2010. Thus, the estimated incidence rates as well as the estimated percentage of inpatient cases (72-80%) were very similar to those reported by Muheim et al.

Recently published analyses of the Obsan investigating the frequency of arthroscopic meniscectomy in inpatients (CHOP codes 80.6X.10 and 80.6X.11) showed that the standardized intervention rate in Switzerland was 332 per 100,000 persons in 2013, 311 per 100,000 persons in 2014, and 306 per 100,000 persons in 2015 (http://versorgungsatlas.ch/index.php/de/MENK/). Again, these numbers are comparable with those reported in this assessment. The numbers are a little bit lower, at least partially due to the fact that the incidence in the Obsan analyses was calculated using the Swiss population above 17 years (versus the Swiss population over 40 years in our analyses and in the paper by Muheim et al.).

The present budget impact analysis has several strengths: firstly, to identify patients with degenerative knee problems two different approaches were used: i) combining relevant DRG codes with ICD-10 codes and CHOP codes, which led to a very strict selection of the patients and resulted in a too conservative cost estimation. It turned out that due to limited suitability of DRG codes for the purpose of patient identification, given real-life coding practices, and due to specific issues resulting from the introduction of the SwissDRG coding system in 2012, this approach yielded results of questionable validity. ii) focussing on a combination of ICD-10 codes and CHOP codes permitted a broader and more accurate selection of patients, resulting in a more realistic cost estimation. Since the selection in the second approach was independent from DRG codes, it was possible to observe that the number of treated patients seemed to remain stable between 2010 and 2014. Another strength was the calculation of total costs of inpatient arthroscopies using two different sources of unit costs: the mean costs per case reported in the diagnosis-related case costs (DRG) statistics and mean costs per case according to different insurance coverage estimated by the insurance company Assura.

One limitation of the present assessment is the identification of eligible inpatients, as already addressed above. Three different coding systems (DRG, ICD-10, and CHOP) were used. The analyses suggested that many patients classified with a relevant DRG code or receiving a relevant ICD-10 diagnosis did not receive any arthroscopic intervention (CHOP code). Inversely, some

patients undergoing an arthroscopic intervention were not classified or diagnosed accordingly. It could be interesting to investigate whether additional selection strategies might be possible, beyond the two strategies discussed above. A further limitation concerns the calculation of ambulatory arthroscopy costs, where the total number of arthroscopic interventions was available, but the effective costs per patient were not known (e.g., costs of the consultation, costs of the attending physicians, costs of the patient record, costs for medication or costs for anaesthesia were not available). For the cost calculation, an estimate from an insurance company was used. [83] [84] However, it is not clear which costs exactly were included in this estimate. Moreover, the proportion of arthroscopic interventions performed due to trauma was unknown. An additional limitation concerns the assumption of the percentage of cases covered by an accident insurance in the ambulant setting. The available data did not provide a differentiation between accident and non-accident cases. For this reason, we used the same percentage as for the Swiss Hospital Statistics. A further limitation concerns the representability of the Assura estimation for the entire country. Assura has around one million clients in Switzerland, and according to their reports the Swiss-German portfolio is now as big as the French portfolio. Whether the overrepresentation of French speaking clients may have led to biased cost estimates is not clear since the methods used to calculate the average costs per patient were not reported in detail. An additional limitation affecting the cost estimates for both inpatients and ambulatory patients concerns the lack of information about costs occurring before and after surgery. For example, before opting for a surgical intervention, patients may receive physical therapies or pain medication. Similarly, after surgery patients usually require specific rehabilitation treatments and pain medications. Depending on the severity of the disease and on patient condition (including health status and motivation), the number and costs of additional therapies could be substantial and quite variable between patients. Finally, it should be remembered that with the available information it was not possible to investigate into total annual costs from a societal perspective. Nevertheless, it is important to emphasize that the total costs from a societal perspective would be much higher than direct medical costs alone: patients undergoing arthroscopic knee interventions are often unable to work for several weeks. Therefore, indirect costs related to loss of productivity are potentially very high.

A final remark concerns the rationality of arthroscopic interventions to the knee. According to published analyses of Obsan, there is a very large variation in the frequency of arthroscopies between different Swiss cantons (Figure 49, http://versorgungsatlas.ch/index.php/de/MENK/). For example, in 2015, the incidence rate of arthroscopic meniscectomy in the Cantons of Geneva and Vaud was below 184 per 100,000 persons, whereas the incidence rate in the Cantons of St. Gallen, Schwyz, Appenzell Innerrhoden was above 452 per 100,000 persons.

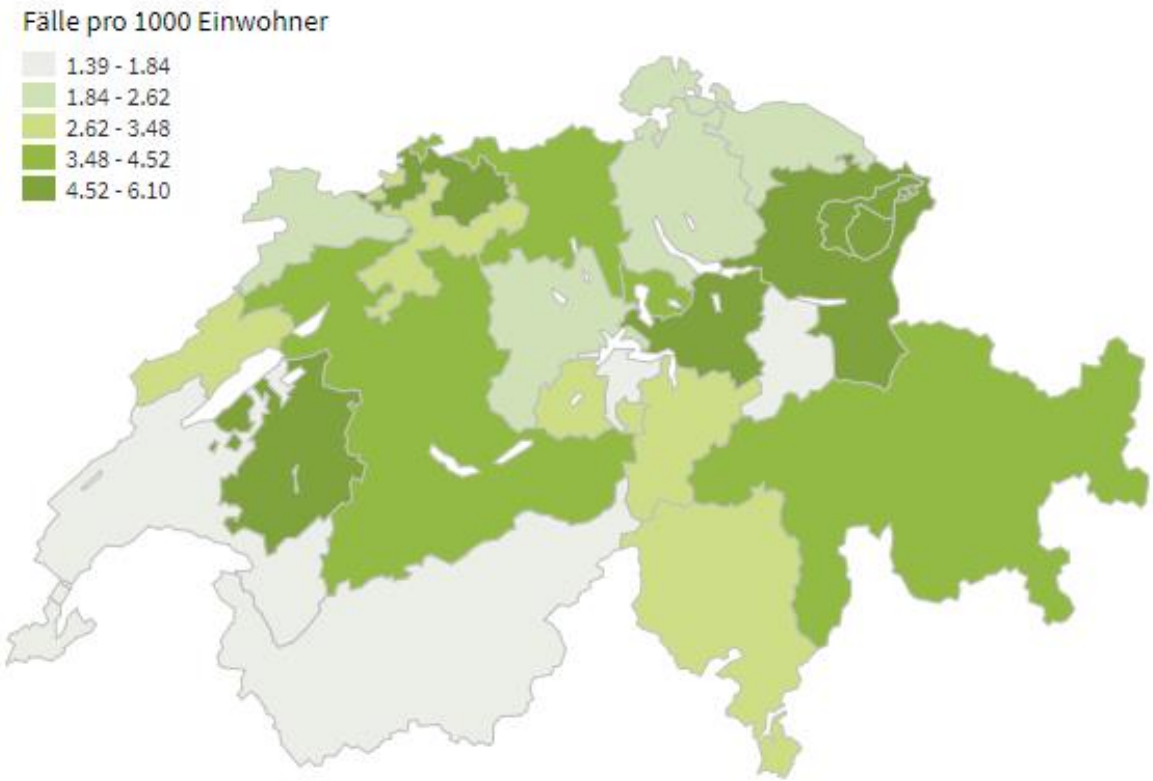


Figure 49 Incidence of arthroscopic meniscectomy by Swiss Canton in 2015 (Source: OBSAN/ISPM 2017)

5.3 Joint discussion

The assessment of clinical effectiveness and the health economic analysis were based on separate systematic reviews, i.e. the health economic analysis was not a de novo modelling based on data from the review of clinical effectiveness. The joint discussion focuses mainly on the similarities and discrepancies between clinical effectiveness and health economic analysis. PICO 2 (inpatient versus outpatient setting) will not be addressed in the joint discussion because limited evidence was found in both the clinical effectiveness and the health economic analyses.

For the assessment of the clinical effectiveness of arthroscopy in patients with degenerative changes of the knee, 21 RCTs (>2000 patients) were identified. There is no evidence that patients with degenerative knee benefit from therapeutic arthroscopy and it remains unclear whether knee arthroscopy has an effect on the assessed outcomes. Only two of the 21 RCTs, which were identified for the clinical effectiveness assessment, were used in two of the four identified cost-effectiveness studies. Of the other 19 RCTs, four more (FIDELITY, Herrlin 2007, Gauffin 2014, Kise 2008) reported on quality of life but cost-effectiveness studies did not make use of these results.

Marsh 2016 used the results by Kirkley 2008 and concluded that physical therapy was dominant (cost-saving) in comparison with arthroscopy. This is in line with the results of the assessment of clinical effectiveness. Losina 2015 reported knee arthroscopy to be dominant (cost-saving), and

thus came to an opposite conclusion. Losina 2015 used the MeTeOR RCT as a key clinical data source, but the methodological approaches used and that assumption made by Losina 2015 were insufficiently described and appeared partially problematic. Therefore, the validity of this economic analysis should be considered as uncertain.

As the cost-effectiveness analyses by Marsh 2016 and Losina 2015 made use of the RCTs Kirkley 2008 and MeTeOR, they were central for the discussion of the clinical and health economic assessment. However, these RCTs differed in several aspects:

First, Kirkley 2008 recruited patients with radiographic confirmed OA and excluded those with large meniscal tears detected with MRI. Kirkley 2008 was also the only RCT classified into the OA only group in clinical assessment. Still, around 80% of patients received meniscal debridement during arthroscopic intervention. The MeTeOR RCT included only patients with MRI-confirmed meniscal tears, but OA was not required as a selection criterion, hence only around 60% of the MeTeOR RCT population had radiographic confirmed OA. Consequently, the interventions were different, i.e. arthroscopic lavage and debridement in Kirkley 2008 and arthroscopic partial meniscectomy in MeTeOR. Irrespective of intentions, both RCTs represent mixed populations of OA and DMK, which underlines everyday clinical practice, where a clear distinction between OA and DMK is usually not made.

Second, Kirkley 2008 reported that none of the patients crossed-over, whereas MeTeOR reported that over 35% of the population crossed-over from physical therapy to arthroscopy. In Losina 2015, the authors call this population "delayed arthroscopic partial meniscectomy" and the analytical approach in this part of the study may be particularly problematic. The analytical approach in this part of the study may be particularly problematic. Most of the RCTs did not report on cross-over or post-randomisation arthroscopy. However, some RCTs reported cross-over to be substantial. For instance, Forster 2003 and Herrlin 2007 reported cross-over rates in the comparator group of almost 30%. High cross-over rates with ITT analysis increase the risk of type 2 errors; consequently, a potential benefit of arthroscopy might be underestimated. To what extent cross-overs have affected the effect estimates of the present assessment is unclear. There were two RCTs that reported small cross-over rates, one of them was blinded (sham-surgery) (FIDELITY); both RCTs (FIDELITY, Kirkley 2008) reported no greater benefit by arthroscopic treatment than the other RCTs. Still, given a substantial amount of cross-over in several of the identified RCTs, it cannot be ruled out that arthroscopic surgery may be a clinically and economically sensible option for some groups of patients.

Despite the different interventions and different cross-over rates in Kirkley 2008 and MeTeOR, the observed effects in the clinical assessment were very much alike and did not sufficiently explain the discrepant cost-effectiveness results of Losina 2015 (favouring arthroscopy) and Marsh 2016 (disfavouring arthroscopy) in the health economic analysis. Methodological differences in the approaches to cost-effectiveness analyses were most likely stronger drivers of the observed discrepancy. In Losina 2015, several methodological aspects are insufficiently reported and some approaches taken may be questionable. The findings of this study are considered very uncertain.

Assessment of clinical effectiveness suggested a numerical but no statistically significant effect on disease-specific quality of life in favour of arthroscopy compared to control. This contradicts the very optimistic estimates of QALY gains through arthroscopy by Lubowitz 2011 and Hutt 2015. Both the modest QALY gain estimated by Losina 2015 and the small QALY loss estimated by Marsh 2016 may be regarded as consistent with the clinical quality of life findings.

Given very limited health economic evidence, it is difficult to draw firm conclusions on the cost-effectiveness of arthroscopic surgery in patients with degenerative changes of the knee. Although three of four available studies reported arthroscopy to be cost-effective, this cannot be regarded as convincing, given the methodological issues described and also in light of the results of the assessment of effectiveness.

6 References

1. Pellegrini S, Kohler DO, S. Variations géographiques dans les soins de santé. La situation en Suisse (Obsan Dossier 42). Secondary Variations géographiques dans les soins de santé. La situation en Suisse (Obsan Dossier 42). http://www.obsan.admin.ch/sites/default/files/publications/2015/obsan_dossier_42.pdf.
2. Muheim LLS, Senn O, Fruh M, et al. Inappropriate use of arthroscopic meniscal surgery in degenerative knee disease. Acta Orthop 2017;**88**(5):550-55.
3. Swiss Federal Office of Statistics. *Krankenhausstatistik*, 2014.
4. IQWiG. *[Arthroscopy of the knee joint for gonarthrosis]*. Cologne: Institut fuer Qualitaet und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2014.
5. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;**29**(8):1039-49.
6. Wolfe F, Kong SX. Rasch analysis of the Western Ontario MacMaster questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. Ann Rheum Dis 1999;**58**(9):563-8.
7. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;**28**(2):88-96.
8. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011.
9. Juhl C, Lund H, Roos EM, et al. A hierarchy of patient-reported outcomes for meta-analysis of knee osteoarthritis trials: empirical evidence from a survey of high impact journals. Arthritis 2012;**2012**:136245.
10. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. www.cochrane.org/training/cochrane-handbook The Cochrane Collaboration last edited 20 March 2011
11. Guyatt G, Oxman AD, Akl E, et al. GRADE guidelines 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2010.
12. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 2011;**64**(12):1283-93.
13. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol 2012.
14. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines 12. Preparing Summary of Findings tables-binary outcomes. J Clin Epidemiol 2012.
15. Guyatt G, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2010.
16. Guyatt G, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence-study limitations (risk of bias) and publication bias. J Clin Epidemiol 2011.
17. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol 2011;**64**(12):1277-82.
18. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol 2011;**64**(12):1294-302.
19. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol 2011;**64**(12):1303-10.

20. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. JClinEpidemiol 2011;**64**(12):1311-16.
21. Brunetti M, Shemilt I, Pregno S, et al. Grade guidelines: 10. Considering resource use and rating the quality of economic evidence. J Clin Epidemiol 2012.
22. Guyatt GH, Thorlund K, Oxman AD, et al. GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles - Continuous outcomes. Journal of Clinical Epidemiology 2013;**66**(2):173-83.
23. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol 2013;**66**(7):719-25.
24. Andrews JC, Schunemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;**66**(7):726-35.
25. Guyatt GH, Oxman AD, Schünemann HJ, et al. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. Journal of Clinical Epidemiology 2011;**64**(4):380-82.
26. Cohen J. *Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, NJ: Lawrence Earlbaum Associates*, 1988.
27. Jüni P, Reichenbach S, Dieppe P. Osteoarthritis: rational approach to treating the individual. Best Pract Res Clin Rheumatol 2006;**20**(4):721-40.
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;**7**(3):177-88.
29. da Costa BR, Nuesch E, Rutjes AW, et al. Combining follow-up and change data is valid in meta-analyses of continuous outcomes: a meta-epidemiological study. J Clin Epidemiol 2013;**66**(8):847-55.
30. Rucker G, Schwarzer G, Carpenter JR, et al. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol 2008;**8**:79.
31. Kohn MD, Sassoon AA, Fernando ND. Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. Clinical Orthopaedics and Related Research® 2016;**474**(8):1886-93.
32. Wright RW. Osteoarthritis Classification Scales: Interobserver Reliability and Arthroscopic Correlation. J Bone Joint Surg Am 2014;**96**(14):1145-51.
33. Azam M, Shenoy R. The Role of Arthroscopic Partial Meniscectomy in the Management of Degenerative Meniscus Tears: A Review of the Recent Literature. The open orthopaedics journal 2016;**10**:797-804.
34. Ha AY, Shalvoy RM, Voisinet A, et al. Controversial role of arthroscopic meniscectomy of the knee: A review. World Journal of Orthopedics 2016;**7**(5):287-92.
35. Khan M, Evaniew N, Bedi A, et al. Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. CMAJ : Canadian Medical Association Journal 2014;**186**(14):1057-64.
36. Lamplot JD, Brophy RH. The role for arthroscopic partial meniscectomy in knees with degenerative changes: a systematic review. Bone & Joint Journal 2016;**98-B**(7):934-8.
37. Monk P, Garfjeld Roberts P, Palmer AJ, et al. The Urgent Need for Evidence in Arthroscopic Meniscal Surgery. American Journal of Sports Medicine 2017;**45**(4):965-73.
38. Thorlund JB, Juhl CB, Roos EM, et al. Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms.[Reprint of BMJ. 2015;350:h2747; PMID: 26080045]. British Journal of Sports Medicine 2015;**49**(19):1229-35.
39. van de Graaf VA, Wolterbeek N, Mutsaerts EL, et al. Arthroscopic Partial Meniscectomy or Conservative Treatment for Nonobstructive Meniscal Tears: A Systematic Review and Meta-analysis of Randomized Controlled Trials. Arthroscopy 2016;**32**(9):1855-65.e4.
40. Biedert RM. Treatment of intrasubstance meniscal lesions: a randomized prospective study of four different methods. Knee Surg Sports Traumatol Arthrosc 2000;**8**(2):104-8.

41. Chang RW, Falconer J, Stulberg SD, et al. A randomized, controlled trial of arthroscopic surgery versus closed-needle joint lavage for patients with osteoarthritis of the knee. Arthritis & Rheumatism 1993;**36**(3):289-96.
42. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. New England Journal of Medicine 2013;**369**(26):2515-24.
43. Forster MC, Straw R. A prospective randomised trial comparing intra-articular Hyalgan injection and arthroscopic washout for knee osteoarthritis. Knee 2003;**10**(3):291-3.
44. Gauffin H, Sonesson S, Meunier A, et al. Knee Arthroscopic Surgery in Middle-Aged Patients With Meniscal Symptoms: A 3-Year Follow-up of a Prospective, Randomized Study. American Journal of Sports Medicine 2017.
45. Gauffin H, Tagesson S, Meunier A, et al. Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: a prospective, randomised, single-blinded study. Osteoarthritis & Cartilage 2014;**22**(11):1808-16.
46. Hamberg P, Gillquist J, Lysholm J. A comparison between arthroscopic meniscectomy and modified open meniscectomy. A prospective randomised study with emphasis on postoperative rehabilitation. Journal of Bone & Joint Surgery - British Volume 1984;**66**(2):189-92.
47. Herrlin S, Hallander M, Wange P, et al. Arthroscopic or conservative treatment of degenerative medial meniscal tears: a prospective randomised trial. Knee Surgery, Sports Traumatology, Arthroscopy 2007;**15**(4):393-401.
48. Herrlin SV, Wange PO, Lapidus G, et al. Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. Knee Surgery, Sports Traumatology, Arthroscopy 2013;**21**(2):358-64.
49. Kalunian KC, Moreland LW, Klashman DJ, et al. Visually-guided irrigation in patients with early knee osteoarthritis: a multicenter randomized, controlled trial. Osteoarthritis & Cartilage 2000;**8**(6):412-8.
50. Kang JG, Wang ML, Zhang XN. Treatment of knee osteoarthritis with arthroscopic debridement and intra-articular sodium hyaluronate injection. [Chinese]. Journal of Jilin University Medicine Edition 2005;**31**(5):802-05.
51. Kirkley A, Birmingham TB, Litchfield RB, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee.[Erratum appears in N Engl J Med. 2009 Nov 12;361(20):2004]. New England Journal of Medicine 2008;**359**(11):1097-107.
52. Stensrud S, Risberg MA, Roos EM. Effect of exercise therapy compared with arthroscopic surgery on knee muscle strength and functional performance in middle-aged patients with degenerative meniscus tears: a 3-mo follow-up of a randomized controlled trial. American Journal of Physical Medicine & Rehabilitation 2015;**94**(6):460-73.
53. Kise NJ, Risberg MA, Stensrud S, et al. Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. BMJ 2016;**354**:i3740.
54. Arden NK, Reading IC, Jordan KM, et al. A randomised controlled trial of tidal irrigation vs corticosteroid injection in knee osteoarthritis: the KIVIS Study. Osteoarthritis & Cartilage 2008;**16**(6):733-9.
55. Campbell MK, Skea ZC, Sutherland AG, et al. Effectiveness and cost-effectiveness of arthroscopic lavage in the treatment of osteoarthritis of the knee: a mixed methods study of the feasibility of conducting a surgical placebo-controlled trial (the KORAL study). Health Technology Assessment (Winchester, England) 2010;**14**(5):1-180.
56. Merchan EC, Galindo E. Arthroscope-guided surgery versus nonoperative treatment for limited degenerative osteoarthritis of the femorotibial joint in patients over 50 years of age: a prospective comparative study. Arthroscopy 1993;**9**(6):663-7.

57. Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. New England Journal of Medicine 2013;**368**(18):1675-84.
58. Katz JN, Wright J, Spindler KP, et al. Predictors and Outcomes of Crossover to Surgery from Physical Therapy for Meniscal Tear and Osteoarthritis: A Randomized Trial Comparing Physical Therapy and Surgery. Journal of Bone & Joint Surgery - American Volume 2016;**98**(22):1890-96.
59. Moseley JB, Jr., Wray NP, Kuykendall D, et al. Arthroscopic treatment of osteoarthritis of the knee: a prospective, randomized, placebo-controlled trial. Results of a pilot study. American Journal of Sports Medicine 1996;**24**(1):28-34.
60. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee.[Summary for patients in J Fam Pract. 2002 Oct;51(10):813; PMID: 12401143]. New England Journal of Medicine 2002;**347**(2):81-8.
61. Osteras H, Osteras B, Torstensen TA. Medical exercise therapy, and not arthroscopic surgery, resulted in decreased depression and anxiety in patients with degenerative meniscus injury. Journal of Bodywork & Movement Therapies 2012;**16**(4):456-63.
62. Saeed K, Khan SA, Ahmed I. Efficacy of intra articular hyaluronic acid versus arthroscopic debridement in terms of improvement in pain score in Kellgran -Lawrence Grading II & III osteoarthritis of knee joint. Pakistan Journal of Medical and Health Sciences 2015;**9**(3):1011-15.
63. Vermesan D, Prejbeanu R, Laitin S, et al. Arthroscopic debridement compared to intra-articular steroids in treating degenerative medial meniscal tears. European Review for Medical & Pharmacological Sciences 2013;**17**(23):3192-6.
64. Weale AE, Ackroyd CE, Mani GV, et al. Day-case or short-stay admission for arthroscopic knee surgery: a randomised controlled trial. Annals of the Royal College of Surgeons of England 1998;**80**(2):146-9.
65. Yim JH, Seon JK, Song EK, et al. A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. American Journal of Sports Medicine 2013;**41**(7):1565-70.
66. *The 2012 User's Guide to: Knee injury and Osteoarthritis Outcome Score KOOS* 2012.
67. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. Health Technol Assess 2003;**7**(34):iii, ix-x, 1-51.
68. Royle PaNW. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. Health Technol Assess 2003; **7**(34):p. iii, ix-x, 1-51.
69. McKinlay RJ, Wilczynski NL, Haynes RB, et al. Optimal search strategies for detecting cost and economic studies in EMBASE. BMC Health Serv Res 2006;**6**:67.
70. Wilczynski NL, Haynes RB, Lavis JN, et al. Optimal search strategies for detecting health services research studies in MEDLINE. CMAJ 2004;**171**(10):1179-85.
71. Sassi F, Archard L, McDaid D. Searching literature databases for health care economic evaluations: how systematic can we afford to be? Med Care 2002;**40**(5):387-94.
72. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Int J Technol Assess Health Care 2013;**29**(2):117-22.
73. Goeree R, He J, O'Reilly D, et al. Transferability of health technology assessments and economic evaluations: a systematic review of approaches for assessment and application. Clinicoecon Outcomes Res 2011;**3**:89-104.
74. O'Brien BJ. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. Am J Manag Care 1997;**3 Suppl**:S33-9.

75. Welte R, Feenstra T, Jager H, et al. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics 2004;**22**(13):857-76.
76. Drummond M, Barbieri M, Cook J, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. Value Health 2009;**12**(4):409-18.
77. Swiss Medical Board. Bariatric surgery vs. conservative treatment for obesity and overweight. 2016, January 29.
78. Ademi Z, Tomonaga Y, van Stiphout J, et al. Adaptation of cost-effectiveness analyses to a single country: the case of bariatric surgery for obesity and overweight. Swiss Med Wkly 2018;**148**:w14626.
79. OECD. Health Statistics 2015 - Purchasing Power Parities for GDP and related indicators. Secondary Health Statistics 2015 - Purchasing Power Parities for GDP and related indicators. http://stats.oecd.org/.
80. Swiss Federal Office of Statistics. Kosten der Gesundheitswesen seit 1960. Secondary Kosten der Gesundheitswesen seit 1960 2015 http://www.bfs.admin.ch/bfs/portal/de/index/themen/14/01/new/nip_detail.html?gnpID=2014-095.
81. Swiss Federal Office of Statistics. Schweizerischer Lohnindex aufgrund der Daten der Sammelstelle für die Statistik der Unfallversicherung (SSUV) 2014. Secondary Schweizerischer Lohnindex aufgrund der Daten der Sammelstelle für die Statistik der Unfallversicherung (SSUV) 2015. http://www.bfs.admin.ch/bfs/portal/de/index/themen/03/04.html.
82. Swiss Federal Office of Statistics. Statistik diagnosebezogener Fallkosten. https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/erhebungen/fks.html. Accessed: February 2018.
83. Balmer H. Fehlanreize, arge Wettbewerbsverzerrungen und steigende Kosten, die aus dem Ruder laufen. clinicum 2017;**1/17**:57-61.
84. Woodtli N. Stationär statt ambulant: So kassieren Spitäler. 2017.
85. Hutt JR, Craik J, Phadnis J, et al. Arthroscopy for mechanical symptoms in osteoarthritis: a cost-effective procedure. Knee Surg Sports Traumatol Arthrosc 2015;**23**(12):3545-9.
86. Losina E, Dervan EE, Paltiel AD, et al. Defining the Value of Future Research to Identify the Preferred Treatment of Meniscal Tear in the Presence of Knee Osteoarthritis. PloS one 2015;**10**(6):e0130256-e56.
87. Lubowitz JH, Appleby D. Cost-effectiveness analysis of the most common orthopaedic surgery procedures: knee arthroscopy and knee anterior cruciate ligament reconstruction. Arthroscopy 2011;**27**(10):1317-22.
88. Marsh JD, Birmingham TB, Giffin JR, et al. Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. BMJ Open 2016;**6**(1):e009949.
89. Ontario Case Costing Initiative (OCCI). Version 7.1 Acute Inpatient Databases for fiscal year 2010/11. Toronto: Ministry of Health and Long-Term Care, 2013.
90. Ontario Schedule of Benefits. Physician services under the Health Insurance Act. 2013. Toronto: Ministry of Health and Long-Term Care, 2013.
91. Ontario Drug Benefit Formulary 2013. Ministry of Health and Long-Term Care. Toronto: Government of Ontario, Canada, 2013. https://www.healthinfo.moh.gov.on.ca/formulary/ (accessed Jan 2013).
92. Medicare Fee Schedules. Center for Medicare and Medicaid Services (2012).
93. Medicare Hospital Outpatient Prospective Payment System, Addendum B. Centers for Medicare & Medicaid Services (2012).
94. Red Book Online. Truven Health Analytics Inc (2014).

95. Healthcare Cost and Utilization Project. Nationwide Inpatient Sample. Rockville, MD: Agency for Healthcare Research and Quality (2014).
96. Buntin MB, Deb P, Escarce JJ, et al. Comparison of Medicare Spending and Outcomes for Beneficiaries with Lower Extremity Joint Replacements. Santa Monica, CA: RAND Corporation, 2005. https://www.rand.org/pubs/working_papers/WR271.html. 2005.
97. Nahin RL, Barnes PM, Stussman BJ, et al. Costs of complementary and alternative medicine (CAM) and frequency of visits to CAM practitioners: United States, 2007. Natl Health Stat Report 2009(18):1-14.
98. National Occupational Employment and Wage Estimates, United States. Washington, D.C.: US Bureau of Labor Statistics, Division of Occupational Employment Statistics (May 2013). .
99. Lavernia CJ, Guzman JF, Gachupin-Garcia A. Cost effectiveness and quality of life in knee arthroplasty. Clin Orthop Relat Res 1997(345):134-9.
100. Jordan JM, Helmick CG, Renner JB, et al. Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in African Americans and Caucasians: the Johnston County Osteoarthritis Project. J Rheumatol 2007;**34**(1):172-80.
101. Siebert WJ GE, David KM, Ganiats TG, Kaplan RM. Quality of Well Being Self-Administered (QWB-SA) scale. San Diego: Health Services Research Center, University of California, 2008. Available from: http://hoap.ucsd.edu/qwbinfo/QWB-Manual.pdf. Accessed Febraury 2018. Secondary Quality of Well Being Self-Administered (QWB-SA) scale. San Diego: Health Services Research Center, University of California, 2008. Available from: http://hoap.ucsd.edu/qwbinfo/QWB-Manual.pdf. Accessed Febraury 2018.
102. Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. Qual Life Res 2013;**22**(3):683-94.
103. Losina E, Weinstein AM, Reichmann WM, et al. Lifetime risk and age at diagnosis of symptomatic knee osteoarthritis in the US. Arthritis Care Res (Hoboken) 2013;**65**(5):703-11.
104. Weinstein AM, Rome BN, Reichmann WM, et al. Estimating the burden of total knee replacement in the United States. J Bone Joint Surg Am 2013;**95**(5):385-92.
105. Goldstein JL, Silverstein FE, Agrawal NM, et al. Reduced risk of upper gastrointestinal ulcer complications with celecoxib, a novel COX-2 inhibitor. Am J Gastroenterol 2000;**95**(7):1681-90.
106. Hame SL, Nguyen V, Ellerman J, et al. Complications of arthroscopic meniscectomy in the older population. Am J Sports Med 2012;**40**(6):1402-5.
107. Katz JN, Barrett J, Mahomed NN, et al. Association between hospital and surgeon procedure volume and the outcomes of total knee replacement. J Bone Joint Surg Am 2004;**86-A**(9):1909-16.
108. Silverstein FE, Graham DY, Senior JR, et al. Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs. A randomized, double-blind, placebo-controlled trial. Ann Intern Med 1995;**123**(4):241-9.
109. Solomon SD, McMurray JJ, Pfeffer MA, et al. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. N Engl J Med 2005;**352**(11):1071-80.
110. Hrobjartsson A, Emanuelsson F, Skou Thomsen AS, et al. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. Int J Epidemiol 2014;**43**(4):1272-83.
111. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;**273**(5):408-12.
112. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;**336**(7644):601-5.

113. Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 2008;**148**(4):295-309.
114. Dell-Kuster S, Droeser RA, Schäfer J, et al. Systematic review and simulation study of ignoring clustered data in surgical trials. British Journal of Surgery 2018;**105**(3):182-91.
115. Bhattacharyya T, Gale D, Dewire P, et al. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. J Bone Joint Surg Am 2003;**85-A**(1):4-9.
116. Englund M, Guermazi A, Gale D, et al. Incidental meniscal findings on knee MRI in middle-aged and elderly persons. N Engl J Med 2008;**359**(11):1108-15.
117. Englund M, Niu J, Guermazi A, et al. Effect of meniscal damage on the development of frequent knee pain, aching, or stiffness. Arthritis Rheum 2007;**56**(12):4048-54.
118. Link TM, Steinbach LS, Ghosh S, et al. Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. Radiology 2003;**226**(2):373-81.
119. Beaufils P, Becker R, Kopf S, et al. Surgical management of degenerative meniscus lesions: the 2016 ESSKA meniscus consensus. Knee Surg Sports Traumatol Arthrosc 2017;**25**(2):335-46.
120. Solomon DH, Simel DL, Bates DW, et al. Does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. JAMA 2001;**286**(13):1610-20.
121. Zhang W, Moskowitz RW, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis & Cartilage 2008;**16**(2):137-62.
122. Kornaat PR, Bloem JL, Ceulemans RY, et al. Osteoarthritis of the knee: association between clinical features and MR imaging findings. Radiology 2006;**239**(3):811-7.
123. Tornbjerg SM, Nissen N, Englund M, et al. Structural pathology is not related to patient-reported pain and function in patients undergoing meniscal surgery. Br J Sports Med 2017;**51**(6):525-30.
124. Halpern SD. Evaluating preference effects in partially unblinded, randomized clinical trials. J Clin Epidemiol 2003;**56**(2):109-15.
125. Katz JN, Losina E, Lohmander LS. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of surgical interventions for osteoarthritis. Osteoarthritis Cartilage 2015;**23**(5):798-802.
126. Wang D, Jones MH, Khair MM, et al. Patient-reported outcome measures for the knee. J Knee Surg 2010;**23**(3):137-51.
127. Abram SG, Middleton R, Beard DJ, et al. Patient-reported outcome measures for patients with meniscal tears: a systematic review of measurement properties and evaluation with the COSMIN checklist. BMJ Open 2017;**7**(10):e017247.
128. Roos EM, Roos HP, Ekdahl C, et al. Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports 1998;**8**(6):439-48.
129. van de Graaf VA, Wolterbeek N, Scholtes VA, et al. Reliability and Validity of the IKDC, KOOS, and WOMAC for Patients With Meniscal Injuries. Am J Sports Med 2014;**42**(6):1408-16.
130. Brignardello-Petersen R, Guyatt GH, Buchbinder R, et al. Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. BMJ Open 2017;**7**(5).
131. Yacub JN, Rice JB, Dillingham TR. Nerve injury in patients after hip and knee arthroplasties and knee arthroscopy. Am J Phys Med Rehabil 2009;**88**(8):635-41; quiz 42-4, 91.
132. Bohensky MA, deSteiger R, Kondogiannis C, et al. Adverse outcomes associated with elective knee arthroscopy: a population-based cohort study. Arthroscopy 2013;**29**(4):716-25.
133. Maletis GB, Inacio MC, Reynolds S, et al. Incidence of symptomatic venous thromboembolism after elective knee arthroscopy. J Bone Joint Surg Am 2012;**94**(8):714-20.

134. Barlow T, Downham C, Griffin D. Arthroscopy in knee osteoarthritis: a systematic review of the literature. Acta Orthopaedica Belgica 2015;**81**(1):1-8.
135. Bollen SR. Is arthroscopy of the knee completely useless?: Meta-analysis - A reviewer's nightmare. Bone and Joint Journal 2015;**97B**(12):1591-92.
136. Buchbinder R, Richards B, Harris I. Knee osteoarthritis and role for surgical intervention: lessons learned from randomized clinical trials and population-based cohorts. Current Opinion in Rheumatology 2014;**26**(2):138-44.
137. Carr AJ, Price AJ, Glyn-Jones S, et al. Advances in arthroscopy-indications and therapeutic applications. Nature Reviews Rheumatology 2015;**11**(2):77-85.
138. Cote MP, Apostolakos JM, Voss A, et al. A Systematic Review of Meta-analyses Published in Arthroscopy: The Journal of Arthroscopic and Related Surgery. Arthroscopy - Journal of Arthroscopic and Related Surgery 2016;**32**(3):528-37.
139. Lee DJ, Elfar JC. Utility of Arthroscopic Surgery for Osteoarthritis of the Knee. Geriatric Orthopaedic Surgery & Rehabilitation 2015;**6**(1):47-9.
140. Mezhov V, Teichtahl AJ, Strasser R, et al. Meniscal pathology - the evidence for treatment. Arthritis Research & Therapy 2014;**16**(2):206.
141. Spahn G, Hofmann GO, Klinger HM. The effects of arthroscopic joint debridement in the knee osteoarthritis: results of a meta-analysis. Knee Surgery, Sports Traumatology, Arthroscopy 2013;**21**(7):1553-61.
142. Evidence D, Standards Branch HQO. Arthroscopic Debridement of the Knee: An Evidence Update. Ontario Health Technology Assessment Series 2014;**14**(13):1-43.
143. Roth S, Pellegrini S. Le potentiel de transfert du stationnairevers l'ambulatoire. Analyse pour une sélection d'interventions chirurgicales. Étude réalisée sur mandat de l'Office fédéral de la santé publique (OFSP) (Obsan Dossier 63). Neuchâtel: Observatoire suisse de la santé. 2018.
144. Sihvonen R, Paavola M, Malmivaara A, et al. Finnish Degenerative Meniscal Lesion Study (FIDELITY): a protocol for a randomised, placebo surgery controlled trial on the efficacy of arthroscopic partial meniscectomy for patients with degenerative meniscus injury with a novel 'RCT within-a-cohort' study design. BMJ Open 2013;**3**(3):09.
145. Sihvonen R, Englund M, Turkiewicz A, et al. Mechanical Symptoms and Arthroscopic Partial Meniscectomy in Patients With Degenerative Meniscus Tear: A Secondary Analysis of a Randomized Trial.[Summary for patients in Ann Intern Med. 2016 Apr 5;164(7). doi: 10.7326/P16-9008 Note: ; PMID: 26856887]. Annals of Internal Medicine 2016;**164**(7):449-55.
146. Jarvinen T, Sihvonen R, Paavola M, et al. Arthroscopic partial meniscectomy vs sham surgery for degenerative meniscus tear. Arthroscopy - Journal of Arthroscopic and Related Surgery 2014; 30(6 suppl. 1). http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/679/CN-01063679/frame.html.
147. Jarvinen T, Sihvonen R, Englund M, et al. Mechanical symptoms and arthroscopic partial meniscectomy in patients with degenerative meniscus tear: a secondary analysis of a randomized, placebocontrolled trial. Arthroscopy - journal of arthroscopic and related surgery Conference: 35th annual meeting of the arthroscopy association of north america Boston, MA united states Conference start: 20160414 Conference end: 20160416 Conference publication: (varpagings) 2017; 32(6 suppl. 1). http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/485/CN-01267485/frame.html.
148. Brophy R. Arthroscopic partial meniscectomy was not better than sham surgery for medial meniscal tear. Journal of Bone & Joint Surgery - American Volume 2014;**96**(16):1396.
149. Campbell MK, Skea ZC, Sutherland AG, et al. Effectiveness and cost-effectiveness of arthroscopic lavage in the treatment of osteoarthritis of the knee: a mixed methods study

of the feasibility of conducting a surgical placebo-controlled trial (the KORAL study) (Structured abstract). Health Technology Assessment Database 2010; (4). http://onlinelibrary.wiley.com/o/cochrane/clhta/articles/HTA-32010000106/frame.html.
150. Tuakli-Wosornu YA, Selzer F, Losina E, et al. Predictors of Exercise Adherence in Patients With Meniscal Tear and Osteoarthritis. Archives of Physical Medicine and Rehabilitation 2016;**97**(11):1945-52.
151. Skoniecki DJ, Palmisano J, Losina E, et al. Factors associated with refusal to participate in a randomized controlled trial of surgery vs non-operative therapy for meniscal tear in knee osteoarthritis. Arthritis and Rheumatism 2009;**60**:1936.
152. MacFarlane L, Yang HY, Collins JE, et al. Influence of baseline magnetic resonance imaging features on outcomes of operative and non-operative treatment of meniscal tear in patients > 45. Arthritis and rheumatology Conference: american college of rheumatology/association of rheumatology health professionals annual scientific meeting, ACR/ARHP 2016 United states Conference start: 20161111 Conference end: 20161116 2017; 68.

http://onlinelibrary.wiley.com/doi/10.1002/art.39977/abstract.
153. Katz JN, Wright J, Mandl LA, et al. Influence of mechanical symptoms on treatment outcomes for meniscal tear in the setting of osteoarthritis. Arthritis and Rheumatism 2013;**65**:S1224.
154. Katz JN, Spindler K, Safran-Norton C, et al. Predictors and outcomes of cross-over to surgery in a randomized trial of surgery vs. physical therapy for meniscal tear and osteoarthritis. Osteoarthritis and cartilage 2015; 23. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/898/CN-01084898/frame.html.
155. Katz JN, Chaisson CE, Cole B, et al. The meteor trial: Preliminary results of an RCT of arthroscopic partial meniscectomy vs physical therapy in patients greater than 45. Arthritis and rheumatism 2012; 64. http://onlinelibrary.wiley.com/doi/10.1002/art.37735/abstract.
156. Katz JN, Chaisson CE, Cole B, et al. The MeTeOR trial (Meniscal Tear in Osteoarthritis Research): rationale and design features. Contemporary Clinical Trials 2012;**33**(6):1189-96.
157. Englund M, Zhang F, Guermazi A, et al. The effect of arthroscopic partial meniscectomy in patients with osteoarthritis on meniscal body extrusion. Arthritis and Rheumatology Conference: American College of Rheumatology/Association of Rheumatology Health Professionals Annual Scientific Meeting, ACR/ARHP 2015;**67**(no pagination).
158. Blacher RS. Arthroscopic surgery for osteoarthritis of the knee. The New England journal of medicine 2002;**347**(21):1717-19; author reply 17-19.
159. Moseley JB, O'Malley K, Petersen NJ, et al. Arthroscopic surgery was not effective for relieving pain or improving function in osteoarthritis of the knee: Commentary. Evidence-Based Medicine 2003;**8**(2):56.
160. Wray NR, Moseley JB, O'Malley K. Arthroscopic treatment of osteoarthritis of the knee [1]. Journal of Bone and Joint Surgery - Series A 2003;**85**(2):381.

7 Appendix A – Introduction and Methods

7.1 Introduction

Appendix 1: Introductory tables

Table 40 Frequencies of main ICD-10 diagnoses for knee/meniscus derangement in 2014 in Switzerland.[3]

ICD-10 Code	Description	Main Diagnosis	% Accident	1. Secondary diagnosis	2. Secondary diagnosis	3. Secondary diagnosis	Total
M23.2	Derangement of meniscus due to old tear or injury	5'683	28.8	2'221	897	343	9'173
M23.3	Other meniscus derangements	7'745	16.4	3'202	1'148	459	12'570
M23.8	Other internal derangements of knee	822	50.9	1'093	719	315	3'000
M23.9	Internal derangement of knee, unspecified	85	35.3	163	99	36	418
S83.2	Tear of meniscus, current injury	4'755	69.9	2'703	365	88	7'911
Total	Derangement or tear of meniscus	19'090	34.0	9382	3228	1241	33'072

Since many patients received more than one single, relevant diagnosis per admission (e.g. 2 different M23.2 codes like "M23.21 - Derangement of anterior horn of medial meniscus due to old tear or injury" and "M23.22 - Derangement of posterior horn of medial meniscus due to old tear or injury"), the total number of ICD-10 codes (N=33'072) does not directly reflect the total number of hospitalized patients. To reduce multiple counting additional analyses were performed.

8 Appendix B – Clinical effectiveness and safety

8.1 Search strategy

Appendix 2: Search strategies

Medline Search

17/05/2017 09:41AM

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)

Search Strategy:

---

1 Arthroscopy/ (20289)
2 arthroscop*.ti,ab. (25321)
3 meniscectom*.ti,ab. (2688)
4 or/1-3 (31017)
5 Osteoarthritis, Knee/ (15055)
6 Osteoarthritis/ (33322)
7 (osteoarthrit* or osteoarthro* or oa or degenerative joint disease* or menisc*).ti,ab. (79777)
8 exp Meniscus/ (6418)
9 exp Knee Joint/ (52957)
10 knee*.ti,ab. (122669)
11 (6 or 7 or 8) and (9 or 10) (33872)
12 4 and (5 or 11) (6611)
13 randomized controlled trial.pt. (462669)
14 controlled clinical trial.pt. (94066)
15 randomized.ab. (404458)
16 placebo.ab. (189133)
17 drug therapy.fs. (1993753)
18 randomly.ab. (280868)
19 trial.ab. (423549)
20 groups.ab. (1730053)
21 or/13-20 (4106680)
22 (animals not (humans and animals)).sh. (4366946)
23 21 not 22 (3551682)
24 search*.tw. (348457)
25 meta analysis.mp,pt. (130739)
26 review.pt. (2296927)
27 or/24-26 (2558900)
28 or/23,27 (5611357)
29 12 and 28 (1725)

EMBASE Search

17.05.2017 09:58

Database: Embase <1974 to 2017 May 16>

Search Strategy:

---

1 exp arthroscopic surgery/ (9364)
2 arthroscopic*.ti,ab. (20965)
3 meniscectom*.ti,ab. (3079)
4 or/1-3 (25586)
5 Knee Osteoarthritis/ (23408)
6 (osteoarthrit* or osteoarthro* or oa or degenerative joint disease* or menisc*).ti,ab. (103206)
7 Joint Degeneration/ (2171)
8 knee*.ti,ab. (150287)
9 Knee Arthroscopy/ (5481)
10 6 and (7 or 8) (39453)
11 6 and 9 (1853)
12 4 and (5 or 10) (4857)
13 11 or 12 (5699)
14 (meta analysis or systematic review or MEDLINE).tw. (240562)
15 random*.tw. (1185417)
16 clinical trial*.mp. (1437192)
17 exp health care quality/ (2467849)
18 or/15-17 (4171600)
19 or/14,17 (2638111)
20 13 and 19 (1531)

Cochrane Search

Date Run: 06.06.2017 08:44:23

ID	Search	Hits
#1	MeSH descriptor Arthroscopy explode all trees in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	47
#2	arthroscop*:ti,ab in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	2589
#3	meniscecto*:ti,ab in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	268
#4	(#1 or #2 or #3) in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	2729
#5	MeSH descriptor Osteoarthritis, Knee explode all trees in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	116
#6	MeSH descriptor Osteoarthritis explode all trees in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	187

#7	(osteoarthrit* or osteoarthro* or oa or degenerative joint disease* or menisc*):ti,ab in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	8413
#8	MeSH descriptor meniscus explode all trees in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	16
#9	Mesh descriptor Knee Joint explode all trees in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	210
#10	knee*:ti,ab in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	15042
#11	((#6 or #7 or #8) and (#9 or #10)) in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	5145
#12	(#4 and (#5 or #11)) in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments	401

ClinicalTrials.gov Search

Date: 16/05/2017

91 studies found for: ("knee arthroscopy" OR "arthroscopy" OR "arthroscopic" OR "meniscectomy" OR "arthroscopic surgery") AND ("random" OR "randomised" OR "randomized" OR "randomly") AND ("knee" OR "knee joint" OR "joint") AND ("osteoarthritis" OR "meniscus" OR "meniscal tear" OR "degenerative" OR "degenerative disease" OR "menisci" OR "arthritis")

International Clinical Trials Registry Platform Search

Date: 17/05/2017

87 studies found for: menisc* OR osteoarthr* OR oa* OR degenerat* in Condition field; arthrosc* OR menisc* OR knee arthroscop* in Intervention field

8.2 List of references of included RCTs

Appendix 3: List of all references of included RCTs

RCT ID	Reference
Biedert 2000[40]	Biedert RM. Treatment of intrasubstance meniscal lesions: a randomized prospective study of four different methods. Knee Surg Sports Traumatol Arthrosc 2000;8(2):104-8.
Chang 1993[41]	Chang RW, Falconer J, Stulberg SD, et al. A randomized, controlled trial of arthroscopic surgery versus closed-needle joint lavage for patients with osteoarthritis of the knee. Arthritis Rheum 1993;36(3):289-96.

RCT ID	Reference
FIDELITY[42 144-148]	Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. N Engl J Med 2013;369(26):2515-24. Sihvonen R, Paavola M, Malmivaara A, et al. Finnish Degenerative Meniscal Lesion Study (FIDELITY): a protocol for a randomised, placebo surgery controlled trial on the efficacy of arthroscopic partial meniscectomy for patients with degenerative meniscus injury with a novel 'RCT within-a-cohort' study design. BMJ Open 2013;3(3):09. Sihvonen R, Englund M, Turkiewicz A, et al. Mechanical Symptoms and Arthroscopic Partial Meniscectomy in Patients With Degenerative Meniscus Tear: A Secondary Analysis of a Randomized Trial.[Summary for patients in Ann Intern Med. 2016 Apr 5;164(7). doi: 10.7326/P16-9008 Note: ; PMID: 26856887]. Ann Intern Med 2016;164(7):449-55. Jarvinen T, Sihvonen R, Paavola M, et al. Arthroscopic partial meniscectomy vs sham surgery for degenerative meniscus tear. Arthroscopy - Journal of Arthroscopic and Related Surgery 2014; 30(6 suppl. 1). Jarvinen T, Sihvonen R, Englund M, et al. Mechanical symptoms and arthroscopic partial meniscectomy in patients with degenerative meniscus tear: a secondary analysis of a randomized, placebocontrolled trial. Arthroscopy - journal of arthroscopic and related surgery Conference: 35th annual meeting of the arthroscopy association of north america Boston, MA united states Conference start: 20160414 Conference end: 20160416 Conference publication: (varpagings) 2017; 32(6 suppl. 1). Brophy R. Arthroscopic partial meniscectomy was not better than sham surgery for medial meniscal tear. J Bone Joint Surg Am 2014;96(16):1396.
Forster 2003[43]	Forster MC, Straw R. A prospective randomised trial comparing intra-articular Hyalgan injection and arthroscopic washout for knee osteoarthritis. Knee 2003;10(3):291-3.
Gauffin 2014[44 45]	Gauffin H, Tagesson S, Meunier A, et al. Knee arthroscopic surgery is beneficial to middle-aged patients with meniscal symptoms: a prospective, randomised, single-blinded study. Osteoarthritis Cartilage 2014;22(11):1808-16. Gauffin H, Sonesson S, Meunier A, et al. Knee Arthroscopic Surgery in Middle-Aged Patients With Meniscal Symptoms: A 3-Year Follow-up of a Prospective, Randomized Study. Am J Sports Med 2017.
Hamberg 1984[46]	Hamberg P, Gillquist J, Lysholm J. A comparison between arthroscopic meniscectomy and modified open meniscectomy. A prospective randomised study with emphasis on postoperative rehabilitation. J Bone Joint Surg Br 1984;66(2):189-92.

RCT ID	Reference
Herrlin 2007[47 48]	Herrlin S, Hallander M, Wange P, et al. Arthroscopic or conservative treatment of degenerative medial meniscal tears: a prospective randomised trial. Knee Surg Sports Traumatol Arthrosc 2007;15(4):393-401. Herrlin SV, Wange PO, Lapidus G, et al. Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. Knee Surg Sports Traumatol Arthrosc 2013;21(2):358-64.
Kalunian 2000[49]	Kalunian KC, Moreland LW, Klashman DJ, et al. Visually-guided irrigation in patients with early knee osteoarthritis: a multicenter randomized, controlled trial. Osteoarthritis Cartilage 2000;8(6):412-8.
Kang 2005[50]	Kang JG, Wang ML, Zhang XN. Treatment of knee osteoarthritis with arthroscopic debridement and intra-articular sodium hyaluronate injection. [Chinese]. Journal of Jilin University Medicine Edition 2005;31(5):802-05.
Kirkley 2008[51 88]	Kirkley A, Birmingham TB, Litchfield RB, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee.[Erratum appears in N Engl J Med. 2009 Nov 12;361(20):2004]. N Engl J Med 2008;359(11):1097-107. Marsh JD, Birmingham TB, Giffin JR, et al. Cost-effectiveness analysis of arthroscopic surgery compared with non-operative management for osteoarthritis of the knee. BMJ Open 2016;6(1):e009949.
Kise 2016[52 53]	Kise NJ, Risberg MA, Stensrud S, et al. Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. Bmj 2016;354:i3740. Stensrud S, Risberg MA, Roos EM. Effect of exercise therapy compared with arthroscopic surgery on knee muscle strength and functional performance in middle-aged patients with degenerative meniscus tears: a 3-mo follow-up of a randomized controlled trial. Am J Phys Med Rehabil 2015;94(6):460-73.
KIVIS[54]	Arden NK, Reading IC, Jordan KM, et al. A randomised controlled trial of tidal irrigation vs corticosteroid injection in knee osteoarthritis: the KIVIS Study. Osteoarthritis Cartilage 2008;16(6):733-9.

RCT ID	Reference
KORAL[55 149]	Campbell MK, Skea ZC, Sutherland AG, et al. Effectiveness and cost-effectiveness of arthroscopic lavage in the treatment of osteoarthritis of the knee: a mixed methods study of the feasibility of conducting a surgical placebo-controlled trial (the KORAL study). Health Technol Assess 2010;14(5):1-180. Campbell MK, Skea ZC, Sutherland AG, et al. Effectiveness and cost-effectiveness of arthroscopic lavage in the treatment of osteoarthritis of the knee: a mixed methods study of the feasibility of conducting a surgical placebo-controlled trial (the KORAL study) (Structured abstract). Health Technology Assessment Database 2010; (4).
Merchan 1993[56]	Merchan EC, Galindo E. Arthroscope-guided surgery versus nonoperative treatment for limited degenerative osteoarthritis of the femorotibial joint in patients over 50 years of age: a prospective comparative study. Arthroscopy 1993;9(6):663-7.

RCT ID	Reference
MeTeOR[57 58 150-157]	Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis.[Erratum appears in N Engl J Med. 2013 Aug 15;369(7):683]. N Engl J Med 2013;368(18):1675-84. Katz JN, Wright J, Spindler KP, et al. Predictors and Outcomes of Crossover to Surgery from Physical Therapy for Meniscal Tear and Osteoarthritis: A Randomized Trial Comparing Physical Therapy and Surgery. J Bone Joint Surg Am 2016;98(22):1890-96. Tuakli-Wosornu YA, Selzer F, Losina E, et al. Predictors of Exercise Adherence in Patients With Meniscal Tear and Osteoarthritis. Archives of Physical Medicine and Rehabilitation 2016;97(11):1945-52. Katz JN, Chaisson CE, Cole B, et al. The MeTeOR trial (Meniscal Tear in Osteoarthritis Research): rationale and design features. Contemp Clin Trials 2012;33(6):1189-96. Skoniecki DJ, Palmisano J, Losina E, et al. Factors associated with refusal to participate in a randomized controlled trial of surgery vs non-operative therapy for meniscal tear in knee osteoarthritis. Arthritis and Rheumatism 2009;60:1936. Katz JN, Chaisson CE, Cole B, et al. The meteor trial: Preliminary results of an RCT of arthroscopic partial meniscectomy vs physical therapy in patients greater than 45. Arthritis and rheumatism 2012; 64. Katz JN, Wright J, Mandl LA, et al. Influence of mechanical symptoms on treatment outcomes for meniscal tear in the setting of osteoarthritis. Arthritis and Rheumatism 2013;65:S1224. Englund M, Zhang F, Guermazi A, et al. The effect of arthroscopic partial meniscectomy in patients with osteoarthritis on meniscal body extrusion. Arthritis and Rheumatology Conference: American College of Rheumatology/Association of Rheumatology Health Professionals Annual Scientific Meeting, ACR/ARHP 2015;67(no pagination). Katz JN, Spindler K, Safran-Norton C, et al. Predictors and outcomes of cross-over to surgery in a randomized trial of surgery vs. physical therapy for meniscal tear and osteoarthritis. Osteoarthritis and cartilage 2015; 23. MacFarlane L, Yang HY, Collins JE, et al. Influence of baseline magnetic resonance imaging features on outcomes of operative and non-operative treatment of meniscal tear in patients > 45. Arthritis and rheumatology Conference: american college of rheumatology/association of rheumatology health professionals annual scientific meeting, ACR/ARHP 2016 United states Conference start: 20161111 Conference end: 20161116 2017; 68.

RCT ID	Reference
Moseley 1996[59]	Moseley JB, Jr., Wray NP, Kuykendall D, et al. Arthroscopic treatment of osteoarthritis of the knee: a prospective, randomized, placebo-controlled trial. Results of a pilot study. Am J Sports Med 1996;24(1):28-34.
Moseley 2002[60 158-160]	Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee.[Summary for patients in J Fam Pract. 2002 Oct;51(10):813; PMID: 12401143]. N Engl J Med 2002;347(2):81-8. Blacher RS. Arthroscopic surgery for osteoarthritis of the knee. The New England journal of medicine 2002;347(21):1717-19; author reply 17-19. Moseley JB, O'Malley K, Petersen NJ, et al. Arthroscopic surgery was not effective for relieving pain or improving function in osteoarthritis of the knee: Commentary. Evidence-Based Medicine 2003;8(2):56. Wray NR, Moseley JB, O'Malley K. Arthroscopic treatment of osteoarthritis of the knee [1]. Journal of Bone and Joint Surgery - Series A 2003;85(2):381.
Østerås 2012[61]	Osteras H, Osteras B, Torstensen TA. Medical exercise therapy, and not arthroscopic surgery, resulted in decreased depression and anxiety in patients with degenerative meniscus injury. J Bodywork Mov Ther 2012;16(4):456-63.
Saeed 2015[62]	Saeed K, Khan SA, Ahmed I. Efficacy of intra articular hyaluronic acid versus arthroscopic debridement in terms of improvement in pain score in Kellgran -Lawrence Grading II & III osteoarthritis of knee joint. Pakistan Journal of Medical and Health Sciences 2015;9(3):1011-15.
Vermesan 2013[63]	Vermesan D, Prejbeanu R, Laitin S, et al. Arthroscopic debridement compared to intra-articular steroids in treating degenerative medial meniscal tears. Eur Rev Med Pharmacol Sci 2013;17(23):3192-6.
Weale 1998[64]	Weale AE, Ackroyd CE, Mani GV, et al. Day-case or short-stay admission for arthroscopic knee surgery: a randomised controlled trial. Ann R Coll Surg Engl 1998;80(2):146-9.
Yim 2013[65]	Yim JH, Seon JK, Song EK, et al. A comparative study of meniscectomy and nonoperative treatment for degenerative horizontal tears of the medial meniscus. Am J Sports Med 2013;41(7):1565-70.

8.3 Eligibility criteria in the included RCTs

Appendix 4 Eligibility criteria of the included RCTs

RCT ID	Inclusion criteria	Exclusion criteria
Biedert 2000	"…patients…with an isolated and painful medial intrasubstance meniscal lesion were included in this prospective study. All patients had clinical symptoms of a meniscal tear and a MRI linear high grade 2 signal intensity in the medial meniscus."	n.r.
Chang 2003	"All patients with the following characteristics were eligible for the study: 1) persistent knee pain for longer than 3 months, despite conservative medical and rehabilitation management, which restricted work, athletic, or self-care activities to an extent unacceptable to the patient, 2) weight bearing knee radiographs showing grade 1,2, or 3 changes as described by Kellgren and Lawrence (8), 3) age >20 years, 4) willingness to attend followup visits at 3 and 12 months, and 5) willingness to give written informed consent. In patients with bilateral disease, the more symptomatic knee was designated the study knee."	"Exclusion criteria were: 1) knee surgery within 6 months of study entry, 2) total knee replacement, 3) any concurrent illness which would influence functional assessment of the knee or preclude arthroscopic surgery, e.g., severe intermittent claudication or cardiac disease, and 4) Kellgren class 4 changes or radiographs, as determined by[authors]."

RCT ID	Inclusion criteria	Exclusion criteria
FIDELITY	"Inclusion criteria: 1. Age: 35 to 65 years 2. Persistent (> 3 months) pain on the medial joint line of the knee 3. Pain provoked by palpation or compression (forced flexion) of the medial tibiofemoral joint line or a positive McMurray sign 4. MRI showing signals characteristic of medial meniscus injury 5. Arthroscopically-verified degenerative medial meniscus tear"	"Exclusion criteria: 1. Obvious trauma-induced onset of symptoms 2. Locked knee (that cannot be straightened normally) 3. Previous surgical procedure on the affected knee 4. Clinical knee OA (ACR Criteria) 5. Radiographic knee OA (Kellgren-Lawrence grade > 1)* 6. Acute (within the previous year) fracture of the affected extremity 7. Decreased range of motion of the knee 8. Instability of the knee 9. MRI assessment shows pathology other than degenerative knee disease requiring treatment other than APM 10. Arthroscopic examination reveals pathology other than a degenerative injury to the medial meniscus requiring intervention other than APM" * "The Kellgren–Lawrence scale evaluates the radiographic severity of osteoarthritis of the knee. Grade 0 denotes normal; grade 1 doubtful narrowing of joint space and possible osteophytic lipping. Patients were excluded if they had definite narrowing of the joint line or an osteophyte in weight-bearing posteroanterior knee radiography with the use of a fixed-flexion protocol (knees in 20° of flexion and the beam oriented 10° above the horizontal axis)."
Forster 2003	"To be included, the patient had to have symptomatic knee osteoarthritis with radiographic evidence of some remaining joint space on weight bearing films and be fit for regional or general anaesthesia."	"Those patients who had mechanical symptoms, intra-articular injection within the last 6 months, previous arthroscopic surgery or hypersensitivity to avian proteins were excluded."
Gauffin 2014	"Inclusion criteria were: age 45-64, symptom duration more than 3 months, standing X-ray with Ahlbäck 0 (less than 50% reduction of the joint space, without consideration of possible osteophytes), had undergone prior physiotherapy, and could understand the Swedish language."	"Patients were excluded when they had a locked knee or joint lockings for more than 2 s more often than once a week, rheumatic or neurological disease, fibromyalgia, replacement of hip- or knee joints, or a contraindication for day-surgery at the current unit (BMI > 35 or a serious medical illness)."
Hamberg 1984	"Patients with degenerative tears of the medial meniscus but with no history of previous injury or operation on the affected knee were selected."	n.r.

RCT ID	Inclusion criteria	Exclusion criteria
Herrlin 2007	"Inclusion criteria at the start of the study were: (a) age 45-64; (b) daily medial knee pain during the last 2-6 months and clinical signs giving suspicion of medial meniscal tear without any history of trauma; (c) MRI showing medial meniscal tear; (d) understanding of the Swedish language."	"Exclusion criteria were: (a) traumatic meniscal injury; (b) radiographic examination showing osteoarthritis > 1 according to the Ahlbäck classification; (c) neurological or rheumatic diseases; (d) loose bodies, ligaments injuries, osteochondral defects and tumors (MRI); (e) knee surgery during the last year; (f) prosthetic replacements of the hip or knee joint; (g) fractures of the lower extremities less than 1 year earlier; (h) contraindications to physical training."
Kalunian 2000	"Inclusion criteria were age greater than 40 years, knee pain for 10 years or less, unsatisfactory pain relief as assessed by both the patient and their primary physician despite at least 6 weeks of supervised physical therapy (isometric exercises and joint protection techniques) and two or more different non-steroidal antiinflammatory drugs (NSAIDs) and/or analgesics given for 3 or more weeks each. If the patient was unable to tolerate NSAIDs and/or analgesics, then the criterion for failure to respond to these agents was waived. If the patient was unable to undergo supervised physical therapy because of third-party payor limitations, then the criterion for failure to respond to these modalities was waived. Patients had to demonstrate a willingness to attend follow-up visits and were required to give written informed consent…" "All patients were required to have normal or minimally abnormal radiographs (Kellgren/ Lawrence grades 0–2). All patients were required to fulfill American College of Rheumatology (ACR) criteria for the classification of knee OA using either clinical and radiographic, traditional clinical or clinical and laboratory methods or classification tree clinical or clinical and laboratory methods."	"Exclusion criteria included: back/hip or ankle/foot disease of significant severity to confuse the clinical assessment of the patient's knee pain; intraarticular corticosteroid injection into the affected knee within 1 month prior to enrollment; significantly abnormal radiographs (Kellgren/ Lawrence grades 3–4); body mass index greater than 35 kg/m2; sensitivity to amide anesthetic agents; any serious medical illness that would, in the opinion of the investigators, place the patient at increased risk should the patient participate in the study; and a recent history of substance abuse."
Kang 2005	"All knee osteoarthritis (KOA) patients were diagnosed prior to treatment by a specialist in this field in accordance with the 1995 KOA diagnostic criteria of the American College of Rheumatology…"	"During these observations no grade IV were included among the study subjects."

RCT ID	Inclusion criteria	Exclusion criteria
Kirkley 2008	"Eligible patients were 18 years of age or older with idiopathic or secondary osteoarthritis of the knee with grade 2, 3, or 4 radiographic severity, as defined by the modified Kellgren–Lawrence classification."	"Patients were excluded if they had large meniscal tears ("bucket handle" tears), as detected by clinical examination or, in a minority of cases, by magnetic resonance imaging. Other exclusion criteria were inflammatory or postinfectious arthritis, previous arthroscopic treatment for knee osteoarthritis, more than 5 degrees of varus or valgus deformity, previous major knee trauma, Kellgren–Lawrence grade 4 osteoarthritis in two compartments (the medial or lateral compartments of the tibiofemoral joint or the patellofemoral compartment) in persons over 60 years of age, intraarticular corticosteroid injection within the previous 3 months, a major neurologic deficit, serious medical illness (life expectancy of less than 2 years or high intraoperative risk), and pregnancy. Patients who were unable to provide informed consent or who were deemed unlikely to comply with follow-up were also excluded."
Kise 2016	"The inclusion criteria were (1) unilateral knee pain for more than 2 mos without a history of significant trauma, where "significant trauma" was defined as a single event of sufficient impact provoking the initial knee pain and problems; (2) a tear in the medial meniscus confirmed by magnetic resonance imaging (MRI); (3) a Kellgren-Lawrence OA grade 2 or less, graded with a standing anterior-posterior radiograph of the injured knee held in a fixed flexed position, using a Plexiglas frame (SynaFlexer); (4) between 35 and 60 yrs of age; (5) eligible for arthroscopic surgery; and (6) able to perform physical activities and exercise. Eligibility for surgery was defined as a clinical diagnosis of a symptomatic meniscus tear, which consisted of the treating orthopedic surgeon's clinical opinion based upon physical examination, history, and MRI."	"Exclusion criteria were acute locked knee, ligament injury, or knee surgery within the previous 2 yrs. Patients had to meet all six inclusion criteria and none of the exclusion criteria to be eligible."
KIVIS	"Patients were eligible if they: had a clinical diagnosis of knee OA, had knee pain for most days of the prior month, had radiographic evidence consistent with knee OA and were between 40 and 90 years of age."	"Exclusion criteria included: symptomatic hip OA, co-existent inflammatory or crystal arthritis, prior knee surgery, injury to the knee in the preceding 6 months or any intra-articular injection in the preceding 3 months or inability to provide informed consent."

RCT ID	Inclusion criteria	Exclusion criteria
KORAL	"Patients were eligible for inclusion if they were: (1) adults (18 years or older) with radiological evidence of osteoarthritis of the knee who might be considered for arthroscopic lavage; (2) fit for general anaesthetic – defined by the ASA grade 1 and 2; and (3) able to give informed consent."	"Excluded patients were those for whom the orthopaedic surgeon judged that arthroscopic lavage was clearly indicated; for whom arthroplasty was clearly indicated; who had clear contraindication to general anaesthesia; who were unable to speak English; and who had an inability to complete follow-up questionnaires."
Merchan 1993	"Only those patients with a limited degenerative process (minimal joint space narrowing and formation of small osteophytes) were accepted for treatment."	"Contraindications for inclusion in this study included duration of pain >6 months, patient body weight >85 kg in men and >70 kg in women, and history of previous surgery." "Patients with an appreciable instability or an angular deformity of more than 15 ° were excluded from the study, as were patients with any previous surgery of the affected knee. Those patients with femoropatellar joint involvement were also excluded from the study."
MeTeOR	"Age 45 years or greater; Symptoms for at least 4 weeks, managed with one or more of: medications, activity limitations or PT; Symptoms consistent with torn meniscus (at least one of: clicking, catching, popping, giving way, pain with pivot or torque, pain that is episodic, pain that is acute and localized to one joint line); Availability of knee radiograph and MRI; Evidence on knee MRI of osteophytes or full-thickness cartilage defect; or plain radiographic evidence of osteophytes or joint space narrowing; Evidence on knee MRI of a meniscal tear that extends to the surface of the meniscus; Willingness to undergo randomization and ability to understand and sign an informed consent document."	"A chronically locked knee (e.g. patient cannot reduce locking; a clear-cut indication for (APM); Kellgren–Lawrence grade 4 (far advanced OA); Inflammatory arthritis or clinically symptomatic chondrocalcinosis; Injection with viscosupplementation in past 4 weeks in index knee; Contraindication to surgery or physical therapy; Bilateral symptomatic meniscal tears; Prior surgery on same knee."
Moseley 1996	"The entry criteria for the pilot study were 1) symptomatic osteoarthritis of the knee in spite of a minimum of 6 months of nonoperative treatment including nonsteroidal antiinflammatory medication; 2) at least moderate knee pain (2'4 on a 0 to 10 scale) on average over a week's time; 3) age under 70; and 4) no medical problems that placed the patient at significant risk for complications from a general anesthetic."	n.r.

RCT ID	Inclusion criteria	Exclusion criteria
Moseley 2002	"Patients were eligible if they were 75 years old or younger, had osteoarthritis of the knee as defined by the American College of Rheumatology, reported at least moderate knee pain on average (≥4 on a visual-analogue scale ranging from 0 to 10) despite maximal medical treatment for at least six months, and had not undergone arthroscopy of the knee during the previous two years."	"The severity of osteoarthritis in the study knee (that with the greatest pain-induced limitation of function) was assessed radiographically and graded on a scale of zero to four. The scores for the three compartments were added together to generate a severity grade of 0 to 12. Criteria for exclusion were a severity grade of 9 or higher, severe deformity, and serious medical problems."
Østerås 2012	"The inclusion criteria were subjects with knee pain for more than 3 months, who were 35-60 years old and eligible for an arthroscopic partial meniscectomy and MRI showing a degenerative meniscus tear. The magnetic resonance imaging (MRI) included coronal T1-weighted turbo spin echo (TSE) and T2-weighted fat saturated TSE, transversal T2-weighted gradient echo and oblique sagittal T2-weighted fat saturated imaging sequences performed on a Siemens 1.5 Tesla Magnet (Symphony) before starting treatment."	"The exclusion criteria were ACL rupture for individuals requiring acute trauma surgeries, including high-energy traumas with ligament injuries, osteoarthritis grade 3-4 (Kellgren-Lawrence classification), haemarthroses and acute cases of locking knee and symptomatic pain in contrary extremities, as well as other musculoskeletal comorbidities severely affecting lower extremity muscle function that override the symptoms from the knee, and comorbidities excluding physical activities and exercise that are not able to speak or read the language of interest."
Saeed 2015	"One hundred and twenty patients of either sex above the age of 40 with history of pain knee joint were selected from the outpatient department. " "Only K-L grade II & III patients were included in the study."	"Patients below 40 years and with history of injury or accident, prior intervention like intraarticular steroid injections within three months were excluded."
Vermesan 2013	"For this purpose we took…patients with non traumatic symptomatic knees which had degenerative lesions of the medial compartment (cartilage and meniscus) on MRI's."	n.r.
Weale 1998	"Consecutive patients scheduled for unilateral arthroscopic surgery of the knee were considered eligible for inclusion in the study. Patient selection was based on The Royal College of Surgeons of England guidelines."	"Children and patients aged over 65 years were excluded from the study. Also excluded were the unfit, those living alone, those with inadequate social support and those without transport."

RCT ID	Inclusion criteria	Exclusion criteria
Yim 2013	"Patients with a degenerative horizontal tear of the posterior horn of the medial meniscus on magnetic resonance imaging (MRI), who were referred to the Center for Joint Disease at our institution between January 2007 and July 2009 for the treatment of nontraumatic knee pain, were asked to participate in this study. The inclusion criteria included daily knee pain on the medial side with mechanical symptoms affecting daily living activities despite management at a primary clinic during the previous month."	"The exclusion criteria included a history of definite trauma, previous knee surgery, ligament deficiency, systeinic arthritis, and osteonecrosis. In addition, patients showing a marked degenerative change with grade ≥2, according to the Kellgren-Lawrence classification, were excluded."

8.4 Additional results for outcomes of PICO 1

Appendix 5 Additional results for outcomes of PICO 1

8.4.1 Pain

8.4.1.1 Short-term

Moseley 1996 was a small (N=10) pilot RCT which reported 3 pain outcomes at six months: intensity of worst knee pain, average intensity of knee pain and intensity of today's pain. As eight of the 10 patients in the pilot RCT had posttraumatic osteoarthritis, the outcomes were not pooled with other RCTs. Table 41 shows the reported results of the RCT. Saeed 2015 reported a pain score using the Knee Society Score System before and six months after procedure. Scores can range from zero (severe) to 50 (none) in 10 point increments.

Table 42 shows the pain score before and six months after procedure. Kise 2016 reported pain at 3 months using the pain subscale of the Knee Injury and Osteoarthritis Outcome Score (KOOS) ranging from zero to 100 with higher scores indicating less pain. A mean difference in pain of -1.8 (95% CI [-7.1, 3.5]) was reported favouring the arthroscopy group compared to the comparator group.

One RCT (Chang 1993) reported pain improvement defined as ≥1 cm decrease on 10 cm pain subscale of AIMS at 3. In the arthroscopy groups, 56% of patients improved at 3 and in the comparator group, 43% of patients improved at 3 months.

Table 41 Pain outcomes (means) at baseline and six months, Moseley 1996

	Placebo		**Lavage**		**Debridement**	
	Baseline N=5	**6 months N=5**	**Baseline N=3**	**6 months N=2**	**Baseline N=2**	**6 months N=2**
Intensity worst knee pain	6.8	8.4	8	8.5	7	9
Average intensity of knee pain	5.6	6.8	5.5	7.5	4.5	7
Intensity of today's pain	5.2	6.8	5	5.5	3	6

*Range 1 (no pain) to 10 (severe pain)

Table 42 Number and percent of patients by pain score on the Knee Society Score System, Saeed 2015

Knee Society Score*	Arthroscopy (n=60)		Intra-articular injection of hyaluronic acid (n=60)	
	Pre-procedure	6 months	Pre-procedure	6 months
10	4 (6.7%)	-	8 (13.4%)	-
20	42 (70.0%)	16 (26.6%)	26 (43.3%)	-
30	14 (23.3%)	22 (36.7%)	26 (43.3%)	24 (40.0%)
40	-	22 (36.7%)	-	14 (23.3%)
45	-	-	-	22 (36.7%)

*Range 0 (severe) to 50 (none)

8.4.1.2 Intermediate

Kise 2016 reported pain at 24 months using the pain subscale of the Knee Injury and Osteoarthritis Outcome Score (KOOS) ranging from zero to 100 with higher scores indicating less pain. A mean difference in pain of 1.4 (95% CI [-3.9, 6.8]) was reported favouring the comparator group compared to the arthroscopy group.

Two RCTs (Chang 1993, Yim 2013) reported pain improvement or relief as binary outcomes at an intermediate follow-up time. Chang 1993 reported pain improvement defined as ≥1 cm decrease on 10 cm pain subscale of AIMS at 12 months. In the arthroscopy groups, 56% of patients improved, while in the comparator group, 43% of patients improved at 12 months. In Yim 2013, complete relief of pain was defined as a value of zero or one point on the VAS, while improved pain was >2 point decrease on the VAS. In the arthroscopy group, 68% and 26% of patients reported complete and improved pain relief, respectively, at 24 months follow-up, while 67% and 23% of patients in the comparator group reported complete and improved pain, respectively.

8.4.2 Function

8.4.2.1 Short-term

Kise 2016 reported function at 3 months using the activities of daily living subscale of the Knee Injury and Osteoarthritis Outcome Score (KOOS) ranging from zero to 100 with higher scores indicating better physical function. A mean difference in function of 1.4 (95% CI [-3.0, 5.9]) was reported favouring the comparator group compared to the arthroscopy group.

The MeTeOR RCT defined treatment success as ≥8 point decrease on function subscale of WOMAC with no cross-over. At six months, 108 of 161 patients in the arthroscopy group had treatment success compared to the 74 of 169 patients in the comparator group.

8.4.3 Global assessment

8.4.3.1 Short-term

Two RCTs (Chang 1993, KIVIS) reported global assessment as a binary outcome at short-term follow-up. Chang 1993 reported both physician global assessment using a 4-point ordinal scale and a patient global assessment using a VAS. Physician-assessed global improvement was defined as ≥1 point decrease on ordinal scale ranging from one (no disease) to four (very severe disease). At 3 months, 47% of patients in the arthroscopy group and 48% of patients in the comparator group were improved. Patient-assessed global improvement was defined as ≥1 cm decrease on VAS from zero (best) to 10 (worst). At 3 months, 50% or patients in the arthroscopy group and 42% of patients in the comparator group were improved. In KIVIS, 64% of patients in the arthroscopy group and 29% of patients in the comparator group reported improvement in symptoms at six months.

8.4.3.2 Intermediate

One RCT (Biedert 2000) reported global assessment based on the International Knee Documentation Committee form, full weight-bearing radiography in extension and control MRI. Based on the International Knee Documentation Committee evaluation, patients were classified as normal, nearly normal, abnormal or severely abnormal. Table 43 shows the adapted results from *Table 1* in the publication.[40]

Table 43 Number and percent of patients by International Knee Documentation Committee evaluation, Biedert 2000

	Intervention		Comparator	
Normal	17	60.7%	3	25.0%
Nearly normal	6	21.4%	6	50.0%
Abnormal	4	14.3%	3	25.0%
Severely abnormal	1	3.6%	0	0.0%

Two RCTs (Chang 1993, Merchan 1993) reported global assessment as a binary outcome at short-term follow-up. Chang 1993 reported both physician global assessment using a 4-point ordinal scale and a patient global assessment using a VAS. Physician-assessed global improvement was defined as ≥1 point decrease on ordinal scale ranging from one (no disease) to four (very severe disease). At 12 months, 47% of patients in the arthroscopy group and 48% of patients in the comparator group were improved. Patient-assessed global improvement was defined as ≥1 cm decrease on VAS from zero (best) to 10 (worst). At 3 months, 44% or patients in the arthroscopy group and 58% of patients in the comparator group were improved. In Merchan 1993, 75% of patients in the arthroscopy group and 16% of patients in the comparator group reported improvement in symptoms at the time of their last evaluation (range 12-36 months).

8.4.4 Cross-overs

Cross-overs

Eleven RCTs (Biedert 2000, Chang 1993, FIDELITY, Forster 2003, Gauffin 2014, Herrlin 2007, Kirkley 2008, Kise 2016, MeTeOR, Østerås 2012, Yim 2013) reported on cross-overs from the comparator group to the arthroscopy group or post-treatment arthroscopy in the comparator group. Table 44 shows the number and percent cross-overs along with the follow-up time.

Table 44 Cross-overs to arthroscopy

RCT ID	Follow-up time	Cross-over (n)	Comparator (n)	Cross-over (%)	Definition
Biedert 2000	mean 26.5 months	2	12	17	Post-treatment arthroscopic procedure
Chang 1993	12 months	2	14	14	Post-treatment arthroscopic procedure
FIDELITY	12 months	4	76	5	Additional arthroscopic procedure*
Forster 2003	12 months	5	17	29	Post-treatment arthroscopic procedure
Gauffin 2014	3 months	2	75	3	Cross-over**
Gauffin 2014	12 months	16	75	21	Cross-over**
Gauffin 2014	36 months	19	75	25	Cross-over**
Herrlin 2007	60 months	13	47	28	Post-treatment arthroscopic procedure
Kirkley 2008	24 months	0	94	0	Cross-over
Kise 2016	24 months	13	70	19	Cross-over***
MeTeOR	6 months	51	169	30	Cross-over**
MeTeOR	12 months	59	169	35	Cross-over**
Østerås 2012	3 months	0	9	0	Cross-over
Yim 2013	24 months	1	54	2	Cross-over

*Before randomization, blinded patients were informed at entry into the study of the opportunity to cross-over if relief of symptoms was not achieved, but not before 6 months after randomization. **Unblinded patients were informed of the opportunity to cross-over at any time (MeTeOR) after randomization. ***Cross-over based on clinical evaluation by orthopaedic surgeon and initiated by participant or physiotherapist, but the authors stated that cross-over criteria was not strict enough and sample size calculation accounted for 20% cross-over.

8.5 Results of risk of bias and support of judgment

Appendix 6: Results of risk of bias assessments and support of judgment

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Biedert 2000	High "The patients were randomly assigned by birthdate to one of the four treatment groups."	Unclear n.r.	High n.r.; patients in control group of anti-inflammatory medication would know they did not receive arthroscopic interventions; additionally post-intervention rehabilitation is different amongst groups	Unclear n.r.	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	Unclear protocol not found

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Chang 1993	Unclear "Subjects who answered “yes” were randomly assigned to arthroscopy or lavage and then asked to accept the assigned therapy."	Unclear "Subjects who answered “yes” were randomly assigned to arthroscopy or lavage and then asked to accept the assigned therapy."	High "...patients were asked not to reveal their treatment"; impossible to blind procedures to patients	High "At each study site, a single assessor not associated with the procedures and blinded to the patient’s treatment regimen (patients were asked not to reveal their treatment and to place a bandaid over actual or potential arthroscopy scars) examined the patients." However, patients were unblinded and most outcomes were self-reported; therefore, the risk of detection bias is high.	High Missing data 10-20% and not comparable among RCT arms (i.e. number of missing and reasons for missing data)	High Missing data 10-20% and not comparable among RCT arms (i.e. number of missing and reasons for missing data)	Unclear protocol not found; all outcomes in methods reported in results (table and text)

FIDELITY	Low The sequentially numbered, opaque, sealed envelopes were prepared by a statistician with no clinical involvement in the execution of the trial using a computer-generated schedule and the envelopes were kept in a secure, agreed location at each centre. To minimise the risk of predicting the treatment assignment of the next eligible patient (to ensure concealment), randomisation was performed in unfixed blocks (block size known only to the statistician)."	Low "To enter a patient into the study, a research/staff nurse opened an envelope containing the treatment assignment and revealed it to the surgeon by showing the paper, but the allocation was not expressed verbally. The sequentially numbered, opaque, sealed envelopes were prepared by a statistician with no clinical involvement in the execution of the trial using a computer-generated schedule and the envelopes were kept in a secure, agreed location at each centre. To minimise the risk of predicting the treatment assignment of the next eligible patient (to ensure concealment), randomisation was performed in unfixed blocks (block size known only to the statistician)."	Low "During the diagnostic arthroscopic procedure, if a patient was confirmed to be eligible for the trial, the surgeon asked a research nurse to open an envelope containing the study-group assignment (arthroscopic partial meniscectomy or sham surgery) and reveal it to the surgeon; the assignment was not revealed to the patient." "To ensure the blinding of the patients at the four study sites using spinal anaesthesia, the blinding of the patient was further ensured by shielding the patients' view with a vertical drape and aiming the arthroscopy monitors away from the patient's line of vision."	Low "Only the orthopaedic surgeon and other staff in the operating room were made aware of the group assignment, and they did not participate in further treatment or follow-up of the patient." "The writing committee developed and recorded two interpretations of the results on the basis of a blinded review of the primary outcome data (treatment A compared with treatment B), one assuming that treatment A was arthroscopic partial meniscectomy, and the other assuming that treatment A was sham surgery. Only after the committee members had agreed that there would be no further changes in the interpretation was the randomization code broken, the correct interpretation chosen, and the manuscript finalized (see the Supplementary Appendix). "The staff delivering the care was blind to the treatment allocation."	Low Missing data ≤5%	Low Missing data ≤5%	Low Protocol published in BMJ and as part of supplementary materials. Clinical Trials: NCT00549172 found. Protocol amendments published within supplementary materials protocol and on Clinical Trials.gov. All outcomes at all time-points specified were reported.
Forster 2003	Unclear	Unclear	High	Unclear n.r.	High	High	Unclear

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
	"...each patient was randomised by sealed envelope to receive either a course of Hyalgan injections or an arthroscopic washout."	"...each patient was randomised by sealed envelope to receive either a course of Hyalgan injections or an arthroscopic washout."	Probably not blinded RCT		Missing data 10-20% and not comparable among RCT arms (i.e. number of missing and reasons for missing data)	Missing data 10-20% and not comparable among RCT arms (i.e. number of missing and reasons for missing data)	protocol not found; all outcomes in methods reported in results (table and text)
Gauffin 2014	Unclear Random sequence generation unknown - "The allocations were placed in sequentially numbered, opaque, sealed envelopes in 15 blocks, block size 10. Envelopes were opened after the enrolment by the patient and a nurse."	Low "The allocation sequence was concealed from the orthopaedic surgeon that enrolled and assessed participants. The allocations were placed in sequentially numbered, opaque, sealed envelopes in 15 blocks, block size 10. Envelopes were opened after the enrolment by the patient and a nurse."	High "Envelopes were opened after the enrolment by the patient and a nurse." "Immediately after randomisation, when the participants were aware of the treatment they would receive, patients were asked to report their expectation of the treatment."	High No clear statement, but patients are aware of their group assignment	High missing data >20% in either RCT arm	High Missing data 10-20% and no adequate method used to deal with missing data in the analysis (ex. Multiple imputation, but not last observation carried forward)	Unclear Clinical Trials: NCT01288768 found; Original and Final protocols found in Supplementary Materials; Amendments published with original article, but unclear risk of reporting bias.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Hamberg 1984	Low "The patient was allocated by a table of random numbers to one of four different groups."	High The patient was allocated by a table of random numbers to one of four different groups.	Unclear n.r.	Unclear n.r.; unclear whether Lysholm was self-administered or clinician-administered	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	NA	Unclear protocol not found; all outcomes in methods reported in results (table and text)
Herrlin 2007	Low The patients were randomised by drawing a sealed opaque envelope numbered and prepared according to a computer-generated randomisation schedule.	Low "The patients were randomized by drawing a sealed opaque envelope numbered and prepared according to a computer-generated randomization schedule."	High "No blinding was possible."	High "No blinding was possible."	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	Unclear Protocol not found. Outcome measured in methods, reported in results.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Kalunian 2000	Low "The simple randomization program resulted in 41 patients randomized to full volume irrigation and 49 patients to minimal irrigation." "Patients were assigned to treatment groups by simple randomization using a random number generator."	Unclear n.r.	Low "Patients were blinded to their treatment group..."	Low Patients...were evaluated by blinded assessors before arthroscopy and at follow-up visits. The blinded assessors were rheumatologists who did not participate in the arthroscopic irrigation procedures.	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	NA	Unclear protocol not found; all outcomes in methods reported in results (table and text)
Kang 2005	Unclear n.r.	Unclear n.r.	High patients received either surgery or injections-blinding of patient was not possible	Unclear n.r.	Low Missing data ≤5%	NA	Unclear protocol not found; all outcomes in methods reported in results (table and text)

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Kirkley 2008	Low "The patients were randomly assigned, with the use of a computer-generated schedule, to receive optimized physical and medical therapy alone (control group) or to receive both optimized physical and medical therapy and arthroscopic treatment."	Unclear "To minimize the risk of predicting the treatment assignment of the next eligible patient, randomization was performed in permuted blocks of two or four with random variation of the blocking number."	High patients received either surgery or physical therapy - blinding of patient was not possible	High "The investigators who assessed outcomes were unaware of treatment assignments." "To preserve blinding, each patient wore a neoprene sleeve over the knee so that the study nurse could not identify a surgical scar." Though investigators were blinded, since patients were not blinded, PRO's are at high risk of detection bias.	Low Missing data 10-20% , comparable among RCT arms (i.e. number of missing and reasons for missing data) and adequate method used to deal with missing data in the analysis (ex. Multiple Imputation, but not “last observation carried forward”)	NA	Low Clinical Trials: NCT00158431; all outcome measures on registry site and methods section reported in published paper

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
KIVIS	Low "patients were randomised, using sealed envelopes, by random number generation, stratified by centre to either TI or intra-articular steroid injection"	Unclear envelopes were sealed but unknown if they were opaque - "patients were randomised, using sealed envelopes, by random number generation, stratified by centre to either TI or intra-articular steroid injection"	High "It was only single blind, which may have led to a bias in favour of TI if patients felt that it was a superior treatment."	High "Single blind, blind observer", but patient was not blinded, but outcome were self-administered, therefore high. "All patients were given standard dressings to apply to the wound before each follow-up visit in order to fully cover the procedure site and maintain the assessor's (study nurse) blinding. In addition they were advised not to inform the blinded assessor of their group allocation in order to maintain blinding." "...research nurse, who was blind to the treatment received by the patient..."	High Missing data 10-20% and no adequate method used to deal with missing data in the analysis (ex. Multiple imputation, but not last observation carried forward)	High Missing data 10-20% and no adequate method used to deal with missing data in the analysis (ex. Multiple imputation, but not last observation carried forward)	Unclear protocol not found; all outcomes in methods reported in results (table and text)

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
KORAL	Unclear "Patients were randomised to one of the three trial groups using a fully automated computerised telephone randomisation... Allocation incorporated minimisation on centre and key prognostic factors…"	Low "Patients were randomised to one of the three trial groups using a fully automated computerised telephone randomisation... Allocation incorporated minimisation on centre and key prognostic factors…"	High comparator of conservative management cannot be blinded	High Not all RCT staff was blinded; unclear is statistician was blinded."	High missing data >20% in either RCT arm	NA	Unclear Registered with ISRCTN, retrospectiv ely (ISRCTN023 28576).

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Merchan 1993	High "Randomization was performed as patients with limited degenerative osteoarthritis of the FT joint presented to the outpatient clinic by pulling consecutively numbered envelopes that had previously been randomly placed on a bulletin board."	High Envelopes were numbered but unknown if they were opaque and envelopes not kept in a secure location - "Randomization was performed as patients with limited degenerative osteoarthritis of the FT joint presented to the outpatient clinic by pulling consecutively numbered envelopes that had previously been randomly placed on a bulletin board."	High patients received either surgery or physical therapy - blinding of patient was not possible	Unclear n.r.	High Missing data 10-20% and no adequate method used to deal with missing data in the analysis (ex. Multiple imputation, but not last observation carried forward)	Low Missing data ≤5%	Unclear Protocol not found. Outcome measured in methods, reported in results.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
MeTeOR	Unclear "Patients were then randomly assigned in a 1:1 ratio to a treatment group with the use of a secure program on the trial website. Randomization was conducted in blocks of varying size within each site."	Low "Randomization was performed in real time on MeTeOR's secure website. Subjects were randomized in blocks of varying size within each site."	High "After randomization, the patient was informed about the treatment assignment; the surgeon was informed as part of the surgical booking process."	High "...our study was not blinded."	Low Missing data 10-20% , comparable among RCT arms (i.e. number of missing and reasons for missing data) and adequate method used to deal with missing data in the analysis (ex. Multiple Imputation, but not “last observation carried forward”)	Low Missing data 10-20% , comparable among RCT arms (i.e. number of missing and reasons for missing data) and adequate method used to deal with missing data in the analysis (ex. Multiple Imputation, but not “last observation carried forward”)	Unclear Protocol and registry (https://clinicaltrials.gov/ct2/show/results/NCT00597012) found. All health-related measures reported; But only 12 months results reported, protocol pre-specified also follow-up time points at 24 until 60 months.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Moseley 1996	Unclear Random sequence generation unknown - "Each patient was taken to the operating room and the randomization envelope was opened to reveal which procedure the patient was to receive."	Low "Each patient was taken to the operating room and the randomization envelope was opened to reveal which procedure the patient was to receive."	Low "The physicians performing the postoperative assessment and the patients remained blinded as to treatment." "All postoperative hospital care was performed by orthopaedic residents, nurses, and other personnel who were blinded to the type of treatment that the patient received."	Low "Follow-up examinations were performed by an orthopaedic surgeon (not the surgeon who performed the procedure) who was blinded to the treatment the patient received."	High missing data >20% in either RCT arm	NA	Unclear protocol not found; some outcomes (general well-being) was not reported in results

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Moseley 2002	Unclear "Sealed, sequentially numbered, stratum-specific envelopes containing treatment assignments were prepared and given to the research assistant. After the patient was in the operating suite, the surgeon was handed the envelope."	Low Sealed, sequentially numbered, stratum-specific envelopes containing treatment assignments were prepared and given to the research assistant. After the patient was in the operating suite, the surgeon was handed the envelope.	Low "The treatment assignment was not revealed to the patient."	Low "Both patients and assessors of outcome were blinded to the treatment assignments."	Unclear Missing data 10-20% and unclear whether comparable among RCT arms (i.e. number of missing in each RCT group and reasons for missing data were not reported)	NA	Unclear Protocol not found. Outcome measured in methods, reported in results.
Østerås 2012	Unclear n.r.	Unclear "The tester was not blinded to which intervention the patient received." "…so there would not be a blinding to group allocation." No additional reference to allocation concealment in publication; therefore, unclear risk of selection bias in regards to allocation concealment.	High "The tester was not blinded to which intervention the patient received." "…so there would not be a blinding to group allocation."	High "The outcome measurements were also not obtained by a blinded assessor, which is a major limitation, as a blinded assessment is considered essential to help prevent bias and assure internal validity in a clinical trial."	Low Missing data ≤5%	NA	Unclear Protocol not found. Outcome measured in methods, reported in results.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Saeed 2015	Unclear n.r.	Unclear n.r.	High "They were divided in two equal groups, each group comprised of 60 patients and the respective procedure was explained to each group.	High n.r.	Unclear Missing data ≤10% and unclear if comparable between RCT arms (i.e. number of missing in each RCT groups and reasons for missing data were not reported)	Unclear Missing data ≤10% and unclear if comparable between RCT arms (i.e. number of missing in each RCT groups and reasons for missing data were not reported)	Unclear Protocol not found. Outcome measured in methods, reported in results.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Stenstrud 2015	Low "Computer generated randomisation sequence, stratified by sex in blocks of eight, and these were concealed from the surgeons who enrolled and assessed the participants."	Low "Randomization was carried out immediately after baseline testing by drawing a sealed opaque envelope, numbered and prepared by an investigator not otherwise involved in the randomization procedure, according to a computer-generated randomization schedule." "computer generated randomisation sequence, stratified by sex in blocks of eight, and these were concealed from the surgeons who enrolled and assessed the participants."	High Patients received either surgery or exercise therapy - blinding of patient was not possible; "Following the informed consent and completion of the baseline measures, the envelopes were opened by the patients and the allocation was revealed."	High "The post-intervention test was performed by another physical therapist blinded to group allocation." "To preserve blinding, each patient wore long pants or neoprene sleeves over both knees so that any surgical scars were not identified."	Low Missing data 10-20% , comparable among RCT arms (i.e. number of missing and reasons for missing data) and adequate method used to deal with missing data in the analysis (ex. Multiple Imputation, but not “last observation carried forward”)	Low Missing data 10-20% , comparable among RCT arms (i.e. number of missing and reasons for missing data) and adequate method used to deal with missing data in the analysis (ex. Multiple Imputation, but not “last observation carried forward”)	Low Clinical Trials: NCT01002794; all outcome measures (in allotted time frame) on registry site and methods section reported in published paper. Though the secondary outcome of hop test/knee bend was supposed to be measured at two years, instead of 12 reporting bias is considered low.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Vermesan 2013	Unclear n.r.	Unclear n.r.	High patients received either surgery or steroid injections - blinding of patient was not possible	Unclear n.r.	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	NA	Unclear Protocol not found. Outcome measured in methods, reported in results.
Weale 1998	Low "Patients were allocated randomly to day-case or „overnight stay admission at the time of attendance in the pre-assessment clinic. A random number table was used for this purpose."	High "A random number table was used for this purpose."	High patients knew if they were a day case or an overnight case	Unclear n.r.	High Missing data 10-20% and not comparable among RCT arms (i.e. number of missing and reasons for missing data)	High Missing data 10-20% and not comparable among RCT arms (i.e. number of missing and reasons for missing data)	Unclear Protocol not found. Outcome measured in methods, reported in results.

RCT ID	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Incomplete outcome data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Yim 2013	Unclear n.r.	Unclear Envelopes were sealed but unknown if they were opaque - "Subsequent treatment was decided by randomization using a closed-envelope technique and dividing the participants into 2 different groups."	High patients received either surgery or exercise program - blinding of patient was not possible	High "Clinical outcome measures and physical examinations were conducted by independent authors (J.-I.C. and M.-C.K) not involved in the treatment at 3 months, 1 year, and 2 years in the outpatient consulting room." Assessed at high risk for detection bias, because most outcomes are patient-reported and patients are not blinded."	Unclear Missing data ≤10% and unclear if comparable between RCT arms (i.e. number of missing in each RCT groups and reasons for missing data were not reported)	NA	High Protocol not found. Some time-points of outcome measured in methods not reported.

9 Appendix C – Health economic analysis

9.1 Literature search strategy

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R). Search performed on: 11 July 2017

	Terms	Results
1	Arthroscopy/	20548
2	Arthroscop*.ti,ab.	25774
3	Meniscectom*.ti,ab.	2727
4	or/1-3	31528
5	Osteoarthritis, Knee/	15374
6	Osteoarthritis/	33778
7	(osteoarthrit* or osteoarthro* or oa or degenerative joint disease* or menisc*).ti,ab.	81309
8	exp Meniscus/	6540
9	exp Knee Joint/	5735
10	(6 or 7 or 8) and (9 or 10)	26870
11	4 and (5 or 11)	3189
12	(animals not (humans and animals)).sh.	4405530
13	11 not 12	3029
14	Afford$	89201
15	Budget$	31271
16	Capital expenditures	2076
17	Cost$	584029
18	Cost-benefit analyses	695
19	Cost-benefit analysis	73696
20	Cost-consequences analyses	4
21	Cost-consequences analysis	54
22	Cost-effectiveness analyses	2044
23	Cost-effectiveness analysis	8231
24	Cost-minimization analyses	36
25	Cost-minimization analysis	480
26	Cost-utility	3822
27	Cost-utility analyses	549
28	Cost-utility analysis	1984
29	Economic$	278037
30	Economic-evaluation	7828
31	Expenditure$	61836
32	Fee$	618039
33	Finance$	9837
34	Financial	95583
35	Financing	45156
36	Health expenditures	17634

	Terms	Results
37	Health resource allocation	122
38	Health resource utilization	345
39	Health-economic$	6727
40	Medical savings accounts	615
41	Monetary	6816
42	Pharmaco-economic analyses	12
43	Pharmaco-economic analysis	22
44	Pharmacoeconomic$	3588
45	Pharmacoeconomic-analyses	283
46	Pharmacoeconomic-analysis	426
47	Price$	30487
48	Socioeconomic$	188333
49	or/14-48	1709786
50	13 AND 49	111

Database: Embase 1974 to 2017 July 10. Search performed on: 11 July 2017

	Terms	Results
1	exp arthroscopic surgery/	9686
2	Arthroscopic*.ti,ab.	21549
3	Meniscectom*.ti,ab.	3189
4	or/1-3	26362
5	Knee Osteoarthritis/	24156
6	(osteoarthrit* or osteoarthro* or oa or degenerative joint disease* or menisc*).ti,ab.	106722
7	Joint Degeneration/	2239
8	Knee*.ti,ab.	154994
9	Knee Arthroscopy/	5602
10	6 and (7 or 8)	41148
11	6 and 9	1899
12	4 and (5 or 10)	5051
13	11 or 12	5912
14	(animals not (humans and animals)).sh.	385
15	13 not 14	5912
16	Afford$	109265
17	Budget$	4637
18	Capital expenditures	198
19	Cost$	864308
20	Cost-benefit analyses	858
21	Cost-benefit analysis	76910
22	Cost-consequences analyses	4
23	Cost-consequences analysis	76

	Terms	Results
24	Cost-effectiveness analyses	2618
25	Cost-effectiveness analysis	128302
26	Cost-minimization analyses	50
27	Cost-minimization analysis	3173
28	Cost-utility	9402
29	Cost-utility analyses	747
30	Cost-utility analysis	8462
31	Economic$	581343
32	Economic-evaluation	18800
33	Expenditure$	70749
34	Fee$	749118
35	Finance$	23272
36	Financial	184517
37	Financing	25161
38	Health expenditures	1888
39	Health resource allocation	158
40	Health resource utilization	648
41	Health-economic$	41866
42	Medical savings accounts	144
43	Monetary	8408
44	Pharmaco-economic analyses	12
45	Pharmaco-economic analysis	37
46	Pharmacoeconomic$	79627
47	Pharmacoeconomic-analyses	365
48	Pharmacoeconomic-analysis	786
49	Price$	41843
50	Socioeconomic$	184429
51	or/16-50	2327635
52	15 AND 51	285

Database: The Cochrane Library . Search performed on: 11 July 2017

	Terms	Results
1	MeSH descriptor Arthroscopy explode all trees	1437
2	Arthroscop*:ti,ab	2635
3	Meniscecto*:ti,ab	265
4	(#1 or #2 or #3)	2953
5	MeSH descriptor Osteoarthritis, Knee explode all trees	2332
6	MeSH descriptor Osteoarthritis explode all trees	4499
7	(osteoarthrit* or osteoarthro* or oa or degenerative joint disease* or menisc*):ti,ab	8530

	Terms	Results
8	MeSH descriptor meniscus explode all trees	174
9	Mesh descriptor Knee Joint explode all trees	2984
10	Knee*:ti,ab	15242
11	((#6 or #7 or #8) and (#9 or #10))	5722
12	(#4 and (#5 or #11))	486
13	Cost and Cost Analysis	46977
14	Economics	26268
15	Quality of Life	68382
16	Quality-adjusted life years	7141
17	Health	206886
18	Healthcare	19492
19	Financing, Health	514
20	Value of Life	12668
21	Health Resources	15304
22	Budgets	316
23	Health Status	66728
24	#13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23	257813
25	#12 and #24	116

9.2 CHEERS checklist

Section/item	Item No	Recommendation	Reported on page No/line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	
Methods			

Section/item	Item No	Recommendation	Reported on page No/line No
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	
Measurement of effectiveness	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	
Estimating costs and resources	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	
Currency, price date and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	

Section/item	Item No	Recommendation	Reported on page No/line No
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	
Characterising uncertainty	20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	
Discussion			

Section/item	Item No	Recommendation	Reported on page No/line No
Study findings, limitations, generalizability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalizability of the findings and how the findings fit with current knowledge.	
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	

9.3 Additional tables and analyses

9.3.1 Number of eligible cases stratified by ICD-10 and CHOP codes

Table 45 Number of eligible cases stratified by relevant ICD-10 diagnoses (one, two, or more) and relevant treatments (one, two, or more CHOP codes) between 2010 and 2014

2010		Number of cases with one, two, or more ICD-10 diagnoses			
		1	2	3+	Total
Number of cases with one, two, or more treatments (CHOP code)	1	1,289	47	4	1,340
	2	8,008	526	58	8,592
	3+	4,879	808	84	5,771
	Total	14,176	1,381	146	15,703
2011		**Number of cases with one, two, or more ICD-10 diagnoses**			
		1	**2**	**3+**	**Total**
Number of cases with one, two, or more treatments (CHOP code)	1	7,224	607	37	7,868
	2	4,185	802	64	5,051
	3+	1,606	507	67	2,180
	Total	13,015	1,916	168	15,099
2012		**Number of cases with one, two, or more ICD-10 diagnoses**			
		1	**2**	**3+**	**Total**
Number of cases with one, two, or more treatments	1	6,559	1,054	122	7,735
	2	3,902	1,160	167	5,229
	3+	1,746	920	164	2,830

(CHOP code)	**Total**	12,207	3,134	453	15,794
2013		**Number of cases with one, two, or more ICD-10 diagnoses**			
		1	**2**	**3+**	**Total**
Number of cases with one, two, or more treatments (CHOP code)	**1**	5,528	1,130	127	6,785
	2	3,768	1,343	241	5,352
	3+	1,897	1,147	278	3,322
	Total	11,193	3,620	646	15,459
2014		**Number of cases with one, two, or more ICD-10 diagnoses**			
		1	**2**	**3+**	**Total**
Number of cases with one, two, or more treatments (CHOP code)	**1**	4,951	1,025	206	6,182
	2	3,389	1,339	309	5,037
	3+	1,830	1,284	388	3,502
	Total	10,170	3,648	903	14,721

9.3.2 Sensitivity and scenario analyses – additional tables

Table 46 Sensitivity analysis of the first inpatient strategy - mean costs per case varied by ±20%

Year APDRG code	2010 Mio. CHF	2011 Mio. CHF	Swiss DRG code	2012 Mio. CHF	2013 Mio. CHF	2014 Mio. CHF
221	0.57-0.85	0.98-1.47	**I04Z**	0.00	0.03-0.04	0.00
222	0.12-0.19	0.21-0.31	**I12B**	0.16-0.27	0.18-0.27	0.22-0.33
232	2.01-3.01	1.00-1.50	**I18A**	0.00	0.29-0.44	0.32-0.48
917	1.20-1.80	1.10-1.65	**I18B**	0.00	26.28-39.42	24.46-36.69
918	24.20-36.30	20.24-30.36	**I30Z**	107.59-161.38	4.22-6.33	6.23-9.34
1222	19.13-28.69	25.74-38.61	**I59Z**	0.00	0.03-0.05	0.00-0.01
1232	0.22-0.33	0.05-0.07				
Total	47.45-71.17	49.32-73.98	**Total**	107.74-161.62	31.03-46.55	31.23-46.85

Table 47 First alternative scenario analysis including all relevant DRG codes (strategy 1)

Year	2010	2011		2012	2013	2014
APDRG code	Mio. CHF	Mio. CHF	Swiss DRG code	Mio. CHF	Mio. CHF	Mio. CHF
221	1.59	2.39	**I04Z**	8.39	10.97	4.57
222	3.68	11.46	**I12B**	5.70	6.14	7.29
232	9.54	4.21	**I18A**	5.91	7.79	1.85
917	1.94	1.94	**I18B**	0.20	38.71	37.87
918	35.75	28.39	**I30Z**	181.97	15.08	22.19
1222	38.19	42.19	**I59Z**	2.81	3.67	10.37
1232	2.54	1.06				
Total	93.23	91.63	**Total**	204.98	82.36	84.13

9.3.3 Frequency of Swiss DRG, ICD-10 and CHOP codes combined

This section focuses on the relationships between the Swiss DRG, the ICD-10, and the CHOP codes for the year 2014. A hospitalized patient receives one Swiss DRG (main reason for hospitalization), but he can receive multiple ICD-10 codes (primary diagnosis and secondary diagnoses) and multiple CHOP codes (main treatment and secondary treatments). Consequently, a patient classified with a Swiss DRG code that has nothing to do with meniscus problems may still receive a knee-related diagnosis or treatment.

Table 48 shows if and how many knee related diagnoses were assigned to cases with the relevant Swiss DRG codes in 2014. Up to 41% (9,202 out of 22,665) of the hospitalized cases classified with one relevant Swiss DRG code didn't receive an ICD-10 code indicating a diagnosis of meniscus derangement. Only for the Swiss DRG code I18B "Arthroscopy, incl. biopsy or other interventions on bone or joints, age > 15 years" there was a clear majority of persons who received a diagnosis for knee/meniscus derangement (9,247 out of 11,361, 81%). This suggests that many patients hospitalized with a Swiss DRG code for knee problems did not necessarily receive a meniscus related diagnosis based on ICD-10.

Inversely, only 13,463 of the 26,290 (51%) cases that received one or more ICD-10 diagnoses for meniscus derangement were automatically identifiable through relevant Swiss DRG codes.

Table 48 Number of knee/meniscus related diagnoses (ICD-10 codes) by relevant Swiss DRG codes in the year 2014

	Number of ICD-10 codes indicating meniscus diagnoses				
Swiss DRG code	None	1	2	3+	Total
I04Z	156	1	0	0	157
I12B	495	15	3	1	514
I18A	222	61	7	1	291
I18B	2,114	7,509	1,478	260	11,361
I30Z	3,671	3,639	413	73	7,796
I59Z	2,544	0	2	0	2,546
Sub-total	**9,202**	**11,225**	**1,903**	**335**	**22,665**
Other Swiss DRG codes		8,844	3,178	805	

Swiss DRG codes and CHOP codes are also only partially overlapping (Table 49). Around 37% (8,311 of 22,665) of the cases that were registered with a relevant Swiss DRG code did not receive any knee arthroscopic surgery based on CHOP codes. Only for the Swiss DRG code I18B "Arthroscopy, incl. biopsy or other interventions on bone or joints, age > 15 years", there was a clear majority of persons who underwent a knee arthroscopic surgery (10,349 out of 11,361, 91%). Consequently, many patients hospitalized with knee problems didn't necessarily receive an arthroscopic knee surgery according to the CHOP codes.

Out of 29,859 cases that received at least one arthroscopic knee surgery according to CHOP codes in 2014, only 14,354 (48%) were classified with a relevant Swiss DRG code. Inversely, 4,499 cases that, according to one or several CHOP codes, received an arthroscopic knee treatment were classified with other Swiss DRG codes.

Table 49 Number of arthroscopic knee surgeries (CHOP codes) by relevant DRG codes in 2014

	Number of relevant treatments based on CHOP codes				
DRG code	**none**	**1**	**2**	**3+**	**Total**
I04Z	142	7	6	2	157
I12B	355	39	57	63	514
I18A	117	145	26	3	291
I18B	1,012	7,700	2,456	193	11,361
I30Z	4,141	2,656	756	243	7,796
I59Z	2,544	2	0	0	2,546
Subtotal	**8,311**	**10,549**	**3,301**	**504**	**22,665**
Other DRG codes		4,499	5,849	5,157	

The frequency of relevant ICD-10 codes compared to the frequency of relevant CHOP codes is shown in Table 50. Only 8% of the cases (2,766 out of 32,625) that underwent an arthroscopic knee surgery according to CHOP codes did not receive an ICD-10 diagnosis for meniscus derangement. In contrast, 19% (6,335 out of 32,625) of the cases that received an ICD-10 diagnosis for meniscus derangement did not receive any arthroscopic knee surgery according to CHOP codes.

Table 50 Number of knee related diagnoses (ICD-10 codes) versus number of arthroscopic knee surgeries (CHOP codes) in 2014

		Number of arthroscopic knee surgeries				
		none	**1**	**2**	**3+**	**Total**
Number of diagnoses for knee/meniscus derangement	**none**		4,299	1,420	616	6,335
	1	2,611	9,078	5,497	2,883	20,069
	2	137	1,414	1,840	1,690	5,081
	3+	18	257	393	472	1,140
	Total	2,766	15,048	9,150	5,661	32,625