PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Arthroscopic meniscectomy versus non-surgical or sham treatment in patients with MRI confirmed degenerative meniscus lesions: a protocol for an individual participant data meta-analysis
AUTHORS	Wijn, Stan; Rovers, Maroeska; Rongen, Jan; Østerås, Håvard; Risberg, May Arna; Roos, Ewa; Hare, Kristoffer; van de Graaf, VA; Poolman, Rudolf; Englund, Martin; Hannink, Gerjon

VERSION 1 – REVIEW

REVIEWER	Antti Malmivaara National Institute for Health and Welfare (THL)
REVIEW RETURNED	20-Jun-2019

GENERAL COMMENTS	Review on paper “Arthroscopic meniscectomy versus non-surgical or sham treatment in patients with degenerative meniscus lesions: a protocol for an individual participant data meta-analysis Thank you for having the opportunity to review this paper, which describes a protocol for an IPD meta-analysis, that has not been performed earlier on this topic. An IPD meta-analysis is in general superior to a traditional meta-analysis, which uses aggregated data reported in the original studies. In the introduction the authors state that there are currently no trials which would have shown effectiveness of arthroscopic meniscectomy in patients with degenerative meniscus lesions. However, a RCT by Gauffin et al. have indeed indicated that arthroscopic meniscectomy is an effective treatment for carefully selected patients. The current paper does not refer to the RCT by Gauffin et al., of which there are two publications: one year follow-up, and a three year follow-up. I have myself been an author on one of the RCTs on effectiveness of arthroscopic partial meniscectomy (APM) for a degenerative knee (Sihvonen et al.), and at the time we published the results, I thought myself, that this sham controlled trial was a proof of concept study, and as there was no effectiveness, it gave a strong indication, that APM really does not work. However, I have later on been in a position, where I have had to assess carefully the whole evidence on this matter. The work has led to two scientific publications (references below). These two papers show that the description of patient selection, patient characteristics, pretreatments before randomization,

	intervention comparisons, and how fully interventions were carried out in the actual experiments, were poorly reported in all the seven RCTs. Additionally, the RCTs were clinically heterogeneous: - The way patients were selected to the RCT is known only in the paper by Gauffin et al. In this setting around 95% of patients of the catchment area ended up with having been considered for eligibility, and the refusal rate was very low. In all other RCTs only a small minority of catchment area patients were included, and there is no information on how the selection process did happen. As there is no documentation of the clinical path prior to randomization, it is not possible to infer which patients actually entered the trial. - The concomitant osteoarthrosis vs no osteoarthrosis differ between the RCTs. - Conservative treatment before randomization varied: in three RCTs there was no treatment, in Gauffin's study 3 months conservative treatment was a prerequisite for randomization, in the other RCTs some conservative treatment was tried before patients being eligible. - There was considerable crossover from the conservative treatment arm to the APM arm in most of the RCTs (c. 20 to 35%). - The intervention contrast differs in the previous RCTs on APM. E.g. in some studies APM is compared to exercises, in some APM + exercise is compare to mere exercise. The two sham trials form a group of their own. - The outcome measures differ between the studies. The heterogeneity (patient selection, patient indication, presence of concomitant osteoarthrosis, treatment before assessing eligibility, intervention contrasts, outcome measures) is of such magnitude that all the RCTs published on effectiveness of arthroscopic meniscectomy have indeed answered to different study questions (see the reference below). Additionally, as the cross-over from conservative to operative treatment in RCTs showing no effectiveness was substantial, the clinical conclusions should be related to the intervention actualized in the experiment, i.e. whether to operate all patients or 20-35 % of the patients (those who do not recover). I think that the premises of the research group do not coincide with the current evidence. I think that the proof of concept study, that I was involved with, is very important, but because of high selection it cannot verify that there could not be effectiveness in other patient groups. The empirical evidence from RCTs currently available, indicate in that one should use strict inclusion criteria for those being offered APM (Gauffin et al.) or offering surgery to those whose symptoms do not alleviate (trials like Katz et al with 35% cross-over from conservative to surgical group). In my opinion, the challenges even for an IPD meta-analysis on these RCTs assessing effectiveness of APM are considerable. The present intended IPD meta-analysis is not able to reach those patients that have not been involved in current RCTs, as the only RCT with representative patient population is that by Gauffin et al. Moreover, because of the poor reporting of PICO characteristics in the original studies, the statistical adjustments in the meta-analyses

	remain uncertain. In addition, the clinical heterogeneity of PICOs in the different RCTs, questions whether IPD is at all feasible. I suggest that the researchers would take these points into consideration and write a revised protocol. However, the main question may still remain open: whether it is possible to create a design for an IPD meta-analysis, which could add on the current evidence. If it could add, no definite conclusions from the IPD meta-analysis will be warranted, because of the poor reporting and clinical heterogeneity of the RCTs published on this topic. There will remain a clear need for further sham controlled and open RCTs reporting comprehensively patient characteristics, preferably in a representative sample of patients living in the recruitment area, and taking adherence to interventions into account when making conclusions on effectiveness. Antti Malmivaara, MD, PhD, Professor References: Malmivaara A. Pure Intervention Effect or Effect in Routine Health Care - Blinded or Non-blinded Randomized Controlled Trial. BMC Medical Research Methodology, 2018;18:91 Malmivaara A. Validity and Generalizability of Findings of Randomized Controlled Trials on Arthroscopic Partial Meniscectomy of the Knee.Scand J Med Sci Sports. 2018 May 16.

REVIEWER	Xiang-Dong Wu Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, China.
REVIEW RETURNED	24-Sep-2019

GENERAL COMMENTS	This study is a protocol for an individual participant data meta-analysis to compare the efficacy and safety of arthroscopic meniscectomy versus non-surgical or sham treatment in patients with degenerative meniscus, the authors will try to tease out whether there are subgroups of patients who suffer from degenerative meniscus lesions will benefit from arthroscopic partial meniscectomy. The design of this protocol was guided by the PRISMA-P. This is an interesting topic of great clinical significance, which would help to minimize the research to practice gap. However, some issues in this protocol require attention. 1. It is a little contradictory in the Introduction section. As arthroscopic partial meniscectomy (APM) has been well-established, it should not be whether or not, but should be what kind of subgroup of patients. Maybe it would be better to rephrase.

	“Arthroscopic partial meniscectomy (APM) is a well-established surgical procedure intended to treat symptoms believed to be caused by degenerative meniscus lesions.” “Therefore, the objective of this IPDMA is to assess whether there are subgroups of patients with degenerative meniscus lesions who benefit from APM in comparison with non-surgical or sham treatment.” 2. The individual participant data include baseline characteristics, radiographic information, clinical symptoms, surgical or non-surgical managements, if important original data is not available, will these trials be excluded? 3. The “Data collection and transfer” and “Data check” sections will be time-intensive tasks. The collected data (may even include radiographic image) will be rather complicated, how will you manage such diverse data? Is there any predefined data format? What software will be used? Please described the Methods section sufficiently to allow the study to be repeated. 4. My biggest concern is the “Outcomes variables”. The primary outcomes include KOOS, EQ5D, SF-36. I am deeply concerned whether the other factors that might affect the surgical effects would be take into consideration, such as surgeon experience, type of hospital, or whether the surgery was performed as part of a training or outreach programme. In addition, will patient related factors such as mental health or psychological status be included in analysis if available? The patient primary symptoms and diagnosis, surgical treatment, as well as postoperative functional evaluations are all susceptible contextual factors that might influence the analysis and subgroup analyses, how will the authors cope with this in the individual participant data meta-analysis? 5. The Additional files are duplicated.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

In the introduction the authors state that there are currently no trials which would have shown effectiveness of arthroscopic meniscectomy in patients with degenerative meniscus lesions. However, a RCT by Gauffin et al. have indeed indicated that arthroscopic meniscectomy is an effective treatment for carefully selected patients. The current paper does not refer to the RCT by Gauffin et al., of which there are two publications: one year follow-up, and a three year follow-up.

Author response:
Dear reviewer, thank you for your comments. We agree that the paper of Gauffin et al. shows that arthroscopic surgery reduces pain in patients with meniscal symptoms and that this might be the

population that is most prevalent in the day-to-day clinic. However, meniscal symptoms are known to have a weak association with meniscal tears (Englund, 2007). Therefore, we decided to focus on the effectiveness of arthroscopic meniscectomy in patients with MRI confirmed degenerative meniscal tears, and had to exclude the study by Gauffin et al. as their patient population had no MRI verified degenerative meniscal tears (also listed as one of our primary inclusion criteria).
We have, however, added this issue to the discussion of our paper to make clear that there is a paper that shows effectiveness of meniscectomy in middle-age patients with degenerative meniscal symptoms.

Englund M, Niu J, Guermazi A, Roemer FW, Hunter DJ, Lynch JA, et al. Effect of meniscal damage on the development of frequent knee pain, aching, or stiffness. Arthritis Rheum. 2007 Dec 1;56(12):4048–54.

Author action:
We have added the following text to the manuscript:
Page 11, lines 273 – 275 (underlined was added)

In the last decade, several RCTs found no or very little benefit of APM compared to non-operative or sham treatment in patients with MRI confirmed degenerative meniscus tears,[43–52] although there is circumstantial evidence that it is effective in middle-aged patients with degenerative meniscal symptoms.[53]

I have myself been an author on one of the RCTs on effectiveness of arthroscopic partial meniscectomy (APM) for a degenerative knee (Sihvonen et al.), and at the time we published the results, I thought myself, that this sham controlled trial was a proof of concept study, and as there was no effectiveness, it gave a strong indication, that APM really does not work. However, I have later on been in a position, where I have had to assess carefully the whole evidence on this matter. The work has led to two scientific publications (references below).

These two papers show that the description of patient selection, patient characteristics, pretreatments before randomization, intervention comparisons, and how fully interventions were carried out in the actual experiments, were poorly reported in all the seven RCTs. Additionally, the RCTs were clinically heterogeneous:

The way patients were selected to the RCT is known only in the paper by Gauffin et al. In this setting around 95% of patients of the catchment area ended up with having been considered for eligibility, and the refusal rate was very low. In all other RCTs only a small minority of catchment area patients were included, and there is no information on how the selection process did happen. As there is no documentation of the clinical path prior to randomization, it is not possible to infer which patients actually entered the trial. The concomitant osteoarthrosis vs no osteoarthrosis differ between the RCTs.

Author response:
We agree that selection of patients may be important regarding the generalizability and applicability of a study. However, studying applicability is exactly the reason to perform an IPDMA as is also stated in several tutorials by the IPDMA methods group of the Cochrane collaboration (Stewart LA, Tierney JF, 2002 / Higgins et al. 2019). That is, we will study whether differences in patient characteristics modify the treatment effect.

Furthermore, Gauffin et al. reported a catchment area of 172,316 inhabitants and ultimately included 179 patients. They noted that “In Sweden, more than 95% of the population is directed to the public medical service.” This, however, does not imply that 95% of the catchment area was considered for eligibility. Moreover, other trials included in our meta-analysis have also included patients that were directed through to public medical service. Nevertheless, as stated above we had to exclude the study by Gauffin et al. as their patient population had no MRI verified degenerative meniscal tears.

Stewart LA, Tierney JF. To IPD or not to IPD? Eval. Health Prof. 2002;25(1):76–97.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

Conservative treatment before randomization varied: in three RCTs there was no treatment, in Gauffin’s study 3 months conservative treatment was a prerequisite for randomization, in the other RCTs some conservative treatment was tried before patients being eligible.

Author response:
We thank the reviewer for this comment and we do agree that the treatment before randomization varied. We will study whether this variation influences the results in a sensitivity analysis.

There was considerable crossover from the conservative treatment arm to the APM arm in most of the RCTs (c. 20 to 35%).

Author response:
Crossover is inevitable in surgical trials like these, and similar rates have been reported in other surgical trials. Generally, only patients in the control/sham group with persistent complaints change treatment group (one-way crossover). Per protocol analyses that exclude patients who change groups will therefore underestimate the effect of treatment. Conversely, analysing patients on the basis of time spent in a treatment arm might overestimate or underestimate this effect. For these reasons it is generally recommended to perform an intention to treat analysis. This was not directly stated in our manuscript and has been added.

Author action:
We have added the following text to the manuscript:
Page 10, lines 262

All analyses will be performed according to the intention-to-treat principle.

The intervention contrast differs in the previous RCTs on APM. E.g. in some studies APM is compared to exercises, in some APM + exercise is compare to mere exercise. The two sham trials form a group of their own.

Author response:
Our goal is to pragmatically evaluate the effectiveness of arthroscopic meniscectomy compared to no meniscectomy. In the included studies, some indeed compare arthroscopic meniscectomy to sham

surgery and others to physical therapy, we, unfortunately, cannot change that. However, we can perform sensitivity analyses on the different comparators to detect if there are any differences when using different comparators, as described in the Cochrane Handbook (2019). We will analyse if there is a difference and if so report the results separately for the different comparators.

Tierney JF et al. Chapter 26.4.3: Individual participant data. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019

The outcome measures differ between the studies.

Author response:
We agree that the outcomes differ between studies, and this is also one of the well-known pitfalls of an IPDMA. The main advantage of an IPDMA, next to studying relevant subgrouping effects, is however, that due to the raw data we can re-analyse and pool data in a more sophisticated way than in a conventional meta-analysis. In this IPDMA we defined new outcomes based on 4 categories: pain, function, quality of life and mental effects (as also described in the protocol). By standardizing the measures, we are able to pool them and are able to make inferences.

The heterogeneity (patient selection, patient indication, presence of concomitant osteoarthrosis, treatment before assessing eligibility, intervention contrasts, outcome measures) is of such magnitude that all the RCTs published on effectiveness of arthroscopic meniscectomy have indeed answered to different study questions (see the reference below). Additionally, as the cross-over from conservative to operative treatment in RCTs showing no effectiveness was substantial, the clinical conclusions should be related to the intervention actualized in the experiment, i.e. whether to operate all patients or 20-35 % of the patients (those who do not recover).

Author response:
We agree that there is some heterogeneity between the included RCTs as they had slightly different inclusion criteria, outcomes and control arm. However, this heterogeneity is exactly what we need to study potential relevant subgrouping effects, i.e. by means of detecting a treatment-covariate interaction. By combining the trials in a one- and two-staged IPD meta-analysis, subgroups that were too small to be distinguished in the RCTs can potentially be detected. As suggested by the reviewer, we will also study whether the persistence of complaints is a relevant subgrouping variable.

Author action:
We have added the following text to the manuscript:
Page 10, lines 260 - 261

Third, we will study whether the persistence of complains is a relevant subgrouping variable.

I think that the premises of the research group do not coincide with the current evidence. I think that the proof of concept study, that I was involved with, is very important, but because of high selection it cannot verify that there could not be effectiveness in other patient groups. The empirical evidence from RCTs currently available, indicate in that one should use strict inclusion criteria for those being offered APM

(Gauffin et al.) or offering surgery to those whose symptoms do not alleviate (trials like Katz et al with 35% cross-over from conservative to surgical group).

In my opinion, the challenges even for an IPD meta-analysis on these RCTs assessing effectiveness of APM are considerable. The present intended IPD meta-analysis is not able to reach those patients that have not been involved in current RCTs, as the only RCT with representative patient population is that by Gauffin et al. Moreover, because of the poor reporting of PICO characteristics in the original studies, the statistical adjustments in the meta-analyses remain uncertain. In addition, the clinical heterogeneity of PICOs in the different RCTs, questions whether IPD is at all feasible.

Author response:
As mentioned above, the heterogeneity of patient characteristics is actually an advantage in IPDMA as it offers the opportunity to study whether these factors do modify the treatment effect. We agree that PICO characteristics are often poorly reported, but the advantage of doing an IPDMA is that we will have access to the raw data, which will enable us to study these characteristics in more detail, and to perform adjusted analyses that are not possible in a conventional meta-analysis. That is, we will analyse the raw individual patient data and therefore, poor reporting of the results and statistical adjustments of the published RCTs do not affect our analyses.

I suggest that the researchers would take these points into consideration and write a revised protocol. However, the main question may still remain open: whether it is possible to create a design for an IPD meta-analysis, which could add on the current evidence. If it could add, no definite conclusions from the IPD meta-analysis will be warranted, because of the poor reporting and clinical heterogeneity of the RCTs published on this topic.

Author response:
Thank you again for your comments. We hope that our answers above, or otherwise the relevant literature regarding the merits of IPDMA will convince the reviewer that is it at least worthwhile to give our approach a try. Currently, an IPDMA is considered the gold standard to detect potential subgroups of patients in existing RCT data. If we ignore the potential presence of subgroups there is a change that meniscectomy will be deemed ineffective, while there might still be a subgroup that benefits from the surgery (as described by Rongen et al. 2017; BMJ-Open). Together with the original trialists, we believe that we can combine the existing RCT data to evaluate if (with the current knowledge) we can find a subgroup that does benefit from meniscectomy in this population. This can help future research as this can narrow the inclusion of patients, decreasing heterogeneity and distortion of an existing treatment effect in a subgroup.

Rongen JJ, Hannink G, Rovers MM. Response: Don’t throw the baby out with the bath water. 2017 Jun 7. BMJ Open. Available at: https://bmjopen.bmj.com/content/7/5/e016114.responses

There will remain a clear need for further sham controlled and open RCTs reporting comprehensively patient characteristics, preferably in a representative sample of patients living in the recruitment area, and taking adherence to interventions into account when making conclusions on effectiveness.

Author response:
We hope that our study will provide some new evidence on relevant subgroups and can be used to inform future sham controlled RCTs.

Antti Malmivaara, MD, PhD, Professor

References:
Malmivaara A. Pure Intervention Effect or Effect in Routine Health Care - Blinded or Non-blinded Randomized Controlled Trial. BMC Medical Research Methodology, 2018;18:91

Malmivaara A. Validity and Generalizability of Findings of Randomized Controlled Trials on Arthroscopic Partial Meniscectomy of the Knee.Scand J Med Sci Sports. 2018 May 16.

Reviewer: 2

1. It is a little contradictory in the Introduction section. As arthroscopic partial meniscectomy (APM) has been well-established, it should not be whether or not, but should be what kind of subgroup of patients. Maybe it would be better to rephrase.
“Arthroscopic partial meniscectomy (APM) is a well-established surgical procedure intended to treat symptoms believed to be caused by degenerative meniscus lesions.”
“Therefore, the objective of this IPDMA is to assess whether there are subgroups of patients with degenerative meniscus lesions who benefit from APM in comparison with non-surgical or sham treatment.”

Author response:
Dear reviewer, thank you for your comments. As suggested by the reviewer we have rephrased the sentence.

Author action:
We have added the following text to the manuscript:
Page 4, lines 78 - 79 (simple markup)

“Arthroscopic partial meniscectomy (APM) is a regularly performed surgical procedure intended to treat symptoms believed to be caused by degenerative meniscus lesions.”

2. The individual participant data include baseline characteristics, radiographic information, clinical symptoms, surgical or non-surgical managements, if important original data is not available, will these trials be excluded?

Author response:
In the initial systematic review we select trials based on our inclusion criteria, otherwise trials are excluded. If some key variables are not available in the full datasets these trials will NOT be excluded. We will both perform a one- and two-stage analysis, which will allow us to handle systematic missing values with a two-level structure using hierarchical multiple imputations by chained equations and still use the valuable data that is present. (as is described in the missing data section of our manuscript)

3. The “Data collection and transfer” and “Data check” sections will be time-intensive tasks. The collected data (may even include radiographic image) will be rather complicated, how will you manage such diverse

data? Is there any predefined data format? What software will be used? Please described the Methods section sufficiently to allow the study to be repeated.

Author response:
We agree with the reviewer that data collection and data checking is the most time-intensive task of our project. We will collect all trial information in a database format, and there is no pre-defined format because every researcher is using a different format. As it is common practice in IPDMA we will collect all data in its original format and then transform that data using R-software. For the data check, we will also use R-software to redo all the analyses published in the original trials to detect any discrepancies. Radiographic images are not included, but the outcomes are. After checking the separate trail sets, all datasets will be merged in one large dataset. We have added this explanation to our method section as suggested.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

4. My biggest concern is the “Outcomes variables”. The primary outcomes include KOOS, EQ5D, SF-36. I am deeply concerned whether the other factors that might affect the surgical effects would be take into consideration, such as surgeon experience, type of hospital, or whether the surgery was performed as part of a training or outreach programme. In addition, will patient related factors such as mental health or psychological status be included in analysis if available? The patient primary symptoms and diagnosis, surgical treatment, as well as postoperative functional evaluations are all susceptible contextual factors that might influence the analysis and subgroup analyses, how will the authors cope with this in the individual participant data meta-analysis?

Author response:
We agree with the reviewer that other aspects like type of hospital might affect the (surgical) effects. Therefore, we can include type of hospital in our sensitivity analyses as random-effects of the mixed effect model to determine if it influences the treatment effect. We will also look into mental health and psychological status as a trial have shown that surgery can have an effect on mental health. (Østerås et al, 2012) For the contextual factors, we rely on the factors that have been collected by the included trials and we aim to take these factors into account.

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 Bodyw Mov Ther. 2012;16(4):456–63.

5. The Additional files are duplicated.

Author response:
Thank you for noticing, this has been corrected.

VERSION 2 – REVIEW

REVIEWER	Antti Malmivaara National Institute for Health and Welfare (THL)

REVIEW RETURNED	22-Nov-2019

GENERAL COMMENTS	- The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.

REVIEWER	Xiang-Dong Wu Peking Union Medical College Hospital
REVIEW RETURNED	22-Nov-2019

GENERAL COMMENTS	Accept with Minor Revision The “Ethics and dissemination ” in the “ABSTRACT” section need to be addressed more appropriately since it is an individual participant data meta-analysis.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Authors:
Reviewer: 1

Reviewer 1:
Thank you for this clarification. However, this exclusion is so important that it should be mentioned in the title of the paper: “.. with MRI confirmed degenerative meniscal lesions…”. There are only seven RCTs published altogether and exclusion of any of these will lead to misinterpretations unless made clear from the start.
You have now added to the discussion that “in patients with MRI confirmed degenerative meniscus tears,[43–52] although there is circumstantial evidence that it is effective in middleaged patients with degenerative meniscal symptoms.” However, the word ‘circumstancial’ refers to indirect evidence, but this was not the case in Gauffin et al’s trial, as the evidence was direct. You have to use a word that maintains the nature of a direct evidence, e.g. some evidence.

Authors’ response:
Thank you for your response and suggestions. We agree that the addition to the title can increase interpretation. We have included “..MRI confirmed degenerative meniscal lesion..” in the title of the paper and changed the word circumstantial in the discussion to better reflect the results of Gauffin et al.

Authors action:
We have added the following text to the manuscript:
Page 1, lines 1-2 (underlined was added)

“Arthroscopic meniscectomy versus non-surgical or sham treatment in patients with <u>MRI confirmed</u> degenerative meniscus lesions: a protocol for an individual participant data meta-analysis”

Page 11, line 283-285 (underlined was changed)

“In the last decade, several RCTs found no or very little benefit of APM compared to non-operative or sham treatment in patients with MRI confirmed degenerative meniscus tears,[43–52] although <u>there is some evidence</u> that it is effective in middle-aged patients with degenerative meniscal symptoms.[53]

Reviewer 1:
Thank you for this reply. The problem with all the other RCTs except Gauffin et al’s is that they

do not report issues related to selection at all or the eligibility to the trial has been selective: (1) what was patients' clinical path prior to assessment of eligibility, (2) share of patients declining participation was up to 55%, (3) conservative therapy was not at all given in three out of six trials and for three months only in the Gauffin study, (4) the number of patients recruited per year per hospital varied from 6 patients to 82 patients. Please see Table 2 in the paper "Malmivaara A. Pure Intervention Effect or Effect in Routine Health Care - Blinded or Nonblinded Randomized Controlled Trial. BMC Medical Research Methodology, 2018;18:91". It is obvious that all the patients except those in the Gauffin's trial have been highly selected, and moreover in a way which is impossible to track afterwards. It means that any analysis however sophisticated of the empirical data available is not able to make any definite conclusions about effectiveness.

This varying and major selection of patients to the RCTs has to be addressed as bringing considerable uncertainty in assessing the treatment modifying effects. This uncertainty is amplified because of poor reporting of patient characteristics in all seven RCTs.
As the study objects of a systematic review are the RCTs, their characteristics must be described appropriately, see final conclusion in the end of this appraisal.

Authors' response:
Thank you for your response. We completely agree that the selection of patients together with the main characteristic of the RCTs (as you have described in table 2 of your paper) has to be addressed to reflect the uncertainty of our final results when we pool these studies. However, we would like to include this information (or lack thereof) in our final manuscript instead of the protocol as we are currently still collecting data from the different principal investigators. Therefore, it is not certain of which trials we can include the individual participant data yet.

We have built a table that includes the published information of four studies of which we have received the individual participant data, but we have not analyzed the data yet. It is possible that the actual trial data will provide additional information on the characteristics of the RCTs that are not included in the publication itself.

The table is located at the bottom of this letter and a brief description was added to the **"**Data check and risk of bias" paragraph. The table will include the following topics (or lack thereof):
1) selection of participants, 2) inclusion- and exclusion criteria, 3) description of clinical path prior to inclusion, 4) number of participants that declined participation, 5) diagnostic characteristics, 6) work characteristic, 7) socioeconomic characteristics, 7) intervention and control treatment, 8) cross-over, 9) other health care services during follow-up and 10) outcome measures.

Malmivaara A. Pure Intervention Effect or Effect in Routine Health Care - Blinded or Nonblinded Randomized Controlled Trial. BMC Medical Research Methodology, 2018;18:91

Reviewer 1:
Thank you for this reply. However, the problem remains, that we do not have data on which patients were left in the trial and which patients were left out. E.g. in three trials the patients were from the beginning randomized to surgical vs conservative treatment. Consequently we do not know e.g. if patients who were considered to fare well regardless of whether they get surgery or exercise were recruited. If so, it may be that no subgroup analysis or modifying factor analysis would bring forward any group that could favor of the arthroscopy. Again, reporting of these characteristics in the original RCTs have to be described in your systematic review.

Authors' response:
Thank you for expressing your valid concerns. We will report these characteristics of the original RCTs (including presence or absence of information on the selection of patients) in our final IPDMA manuscript. Please see our example table located at the bottom of this letter.

Reviewer 1:
Thank you for this reply. I agree that intention to treat analysis is the method to base inferences on. But the interpretation from the intention to treat analysis must be according to what happened in the experiment, e.g. in the trial by Katz et al. 35% of the patients crossed over from the conservative treatment arm to the surgical treatment arm. The obvious conclusion is that instead of operating all patients, it is enough to operate 35% (those who did not recover during the follow-up). A conclusion that there would not be effectiveness is not based on the empirical findings. You have to describe adherence to interventions, cross-overs and use of health care services in each of the eligible RCTs, and record also instances where data was not presented in the RCTs.

Authors' response:
We will describe the characteristics of the RCTs in a table, including the adherence to the intervention, cross-overs, use of health services and describe when this data has not been collected or described in detail in our IPDMA as in our example table located at the bottom of this letter.

Reviewer 1:
Thank you for your response. I agree with you that we cannot change what has been the study question in previous RCTs. But we have to take it as it is: in my interpretation based on careful description of all the RCTs, none of the seven previous trials have studied the same thing. Please see Figure 3 in the paper "Malmivaara A. Pure Intervention Effect or Effect in Routine Health Care - Blinded or Non-blinded Randomized Controlled Trial. BMC Medical Research Methodology, 2018;18:91".

It is not just differences in the intervention contrasts e.g. what is the effectiveness between surgery vs exercise vs additional benefit of surgery beyond exercise, which are definitely different study questions. Other differences relate to effectiveness of surgery per se, without placebo effect (Sihvonen et al.) vs effectiveness of surgery with placebo effect related to exercise with its placebo effect. Furthermore the patients differ in terms of presence/degree of concomitant osteoarthrosis, in terms of traumatic onset and in terms of prior conservative treatment. In conclusion, all the seven RCTs have studied different PICOs.
The study object of a systematic review on randomized trials is the RCT itself. Therefore it is mandatory that the PICO of each RCT will be reported as stated in the aims of the researchers. However, this will not suffice: the researcher (and the reader) of a systematic review have to know how the RCTs were actually about: how were the patients selected to each RCT; how comprehensively were the patient characteristics described in each RCT and what the characteristics were; what was the adherence to the interventions, what were the shares of patients crossing over, and what was the use of other health care services during the follow-up; and what were all the outcomes that were documented in the articles. Furthermore, one must ascertain that an appropriate statistical analysis has been undertaken. You may want to look at this paper: "Malmivaara A. Generalizability of findings from Randomized Controlled Trials is limited in the Leading General Medical Journals. J Clin Epidemiol. 2019;107:36-41".

Authors' response:
Thank you for suggesting your relevant and interesting paper regarding the generalizability of findings from RCTs. It was a pleasure to read your paper that stresses the importance to include an extensive set of RCT characteristics including PICOs in the publication of RCTs, but also in any kind of review as our current study.

We will describe the characteristics of the RCTs in our IPDMA in a table including the (or lack of) adherence to the intervention, cross-overs and use of other health services. Please see our example table located at the bottom of this letter.

Malmivaara A. Generalizability of findings from Randomized Controlled Trials is limited in the Leading General Medical Journals. J Clin Epidemiol. 2019;107:36-41

Reviewer 1:
Thank you for this reply. I agree that the differences in outcomes do not pose such a hindrance to entering a modifier analysis than other PICO characteristics mentioned above.

Authors' response:
Thank you. We will include the PICO characteristics of the trials under comparison in our IPDMA. Please see our example table located at the bottom of this letter.

Reviewer 1:
Thank you for your response. I must respectfully disagree with your view that "there is some heterogeneity between the included RCTs …" In my opinion, based on careful description of the RCTs, all the seven RCTs have studied different study questions. Please see Figure 3 in the paper "Malmivaara A. Pure Intervention Effect or Effect in Routine Health Care - Blinded or Non-blinded Randomized Controlled Trial. BMC Medical Research Methodology, 2018;18:91". To provide empirical evidence for the heterogeneity between the RCTs, a proper description of the characteristics of the included RCTs must be included in the protocol.

Authors' response:
Thank you for your comment. The heterogeneity between RCTs is certainly an issue that we have to describe in our final IPDMA study. Therefore, we will include an adequate table describing the characteristics of the RCTs that are included in our study (e.g. the PICO, (missing information about) patients' trajectory prior to inclusion) so the reader is able to interpret the uncertainty / generalizability of our results. Please see our example table located at the bottom of this letter.

Reviewer 1:
Thank you for your reply. I agree with you that this is the point theoretically if the study questions based on PICO would not be grossly different. Again, in order to show the degree of heterogeneity of the RCTs, you must include a description of their characteristics in your protocol.

Authors' response:
Thank you for your comment. We will include the characteristics of the included studies in our final paper displaying the degree of heterogeneity of the RCTs.

Reviewer 1:
Thank you for your reply. I am in favor of using all the methods available to advance our understanding of effectiveness of APM. However, we have to critically assess when our methods are suitable for assessing our study question and when they are not. The major challenge in my opinion is that all the RCTs published hitherto on APM have described poorly the PICO as it has been realized in the experiment. Please see also the two papers, which I referred to already in my first reviewer round.
The poor reporting of RCTs is not limited to APM research, please see a paper on reporting of RCTs in medicine at large: "Malmivaara A. Generalizability of findings from Randomized Controlled Trials is limited in the Leading General Medical Journals. J Clin Epidemiol.

2019;107:36-41"

Authors' response:
Thank you for your comment. We will certainly describe the patient path before randomization of the included RCTs. Please see our example table located at the bottom of this letter.

Reviewer 1:
Thank you for your reply. I agree that there certainly is a need for sham controlled and open RCTs for assessment of effectiveness of APM.

Conclusions for the second revision round by the Reviewer 1:
In a systematic review, the individual RCTs are the object of the study and the researchers' obligation is to document all the important descriptive characteristics. Without doing this, one is not able to judge how heterogeneous the original RCTs have been, and consequently has a meta-analysis been justified; and finally not be able to assess the generalizability of the results of the systematic review. The minimum set of description is: whether or not the selection of patients to the trials was described, and if yes how the patients were selected; whether or not the following characteristics were described and if they were a description of the findings: 1) inclusion and exclusion criteria, diagnostic characteristics of the patients included and severity of their indication; presence of absence of osteoarthrosis; behavioural factors related to exercise, work characteristics and socioeconomic characteristics; 2) What was the intended intervention contrast, what was the adherence to the intervention arms, what were the shares of patients crossing over to another treatment arm; what was the use of other health care services during the follow-up; 3) what were all the outcome measures in the study protocols; what were all the numerical values of all these outcomes as recorded in the published papers: 4) were the statistical analyses appropriately undertaken in the original RCTs under study.
A meta-analysis should be undertaken by including only those RCTs for which the authors can provide sound clinical reasoning that these RCTs are clinically sufficiently homogenous. The reader is able to make his/hers own judgements on this matter, when the descriptive data is provided in the paper.

Authors response to conclusions for second revision round:
We completely agree that the descriptive characteristics of the trials need to be included in the final IPDMA. This enables the reader to assess how comparable the RCTs are, as you describe. We will include this information per included RCT in our IPDMA, and will also include it if they are not described properly. We have already put the published data from four trials in a concept table below.
To inform the readers of the protocol that we will include this information in our final IPDMA we have added the following lines to our manuscript:

Authors actions:
In the manuscript, we added the following lines to the **"**Data check and risk of bias" paragraph.
Page 7, lines 180-185

"To enable to assessment of heterogeneity between the included trials, the following characteristics of the included RCTs will be described: 1) selection of participants, 2) inclusion- and exclusion criteria, 3) description of clinical path prior to inclusion, 4) number of participants that declined participation, 5) diagnostic characteristics, 6) work characteristics, 7) socioeconomic characteristics, 7) intervention and control treatment, 8) cross-over, 9) other health care services during follow-up, and 10) outcome measures."

Concept table 1: Main characteristics of the RCTs included in the IPDMA

	Østerås et al. (2012)	Kise et al. (2016) (OMEX trial)	Roos et al. (2018) (SLAMSHAM trial)	Van de Graaf et al. (2018) (ESCAPE trial)
Selection of participants				
Inclusion criteria	35-60 years; subjects with knee pain for more than 3 months and eligible for an arthroscopic partial meniscectomy and MRI showing a degenerative meniscus tear.	35–60 years; unilateral knee pain more than two months without a major trauma; medial degenerative meniscal tear by MRI; radiographic changes at most, grade 2 byKellgren-Lawrence[c]	35-55 years; knee pain for more than 2 months without significant trauma and an MRI-confirmed medial meniscus lesion. The patients must be eligible for outpatient surgery.	45 – 70 years; knee pain and a nonobstructive (no locking of the knee joint) meniscal tear confirmed by magnetic resonance imaging (MRI).
Exclusion criteria	CL rupture requiring acute trauma surgeries, osteoarthritis grade 3–4, haemarthroses, locking knee, symptomatic pain in contrary extremities, comorbidities excluding physical activities, not able to speak or read the language of interest.	Acute trauma, locked knee, ligament injury, and knee surgery in the index knee during the previous two years.	In need of acute surgery e.g. locking knees or high-energy trauma, grade 3 or 4 knee OA on the Kellgren & Lawrence classification, knee surgery within the previous 2 years, unable to speak language of interest and drug or alcohol abuse. Also, patients with trombophilia are excluded so as to prevent a high risk of deep venous thrombosis.	Locking of the knee, prior knee surgery, instability caused by an anterior or posterior cruciate ligament rupture, severe osteoarthritis (Kellgren-Lawrence score of 4) and a body mass index greater than 35.
Description of pathway prior to inclusion	Not described	Not described	Not described	Not described

Declined participation	6.9%	38%	27.1%	Not described
Pre-intervention therapy	Not described	Not described	Not described	Not described
Verification of diagnosis	MRI confirmed degenerative meniscus tear	Clinically and by MRI	MRI confirmed	MRI confirmed
Baseline characteristics				
Number of included patients per group (total)	9/8 (17)	70 (140)	22 (44)	159 / 162 (321)
Age (SD / range)	49.7 (9.3)	49.5 (35.7 – 59.9)	46.8 (5.7)	57.6 (6.5)
Gender (% male)	76.4%	61%	52%	49.3%
Comorbidity	No	No	No	No
Behavior / lifestyle	No	Yes (smoking)	Yes (overweight)	Yes (overweight)
Environmental factors	No	No	No	No
Socio-economic inequality	No	Yes (education)	No	Yes (education)
Interventions				
Treatment details	APM	APM	APM, postoperative care folder	APM, perioperative instructions and a home exercise program. Participants were only referred to PT after

				APM if they did not recover as anticipated
Control treatment	Exercise 3 times weekly, 3 months	Exercise 2-3 times weekly, 12 weeks	Skin-incision, knee manipulation to mimic real arthroscopy, postoperative care folder	Exercise 16 sessions of 30 minutes each conducted over 8 weeks
Attendance to exercise therapy	84% completed rehabilitation program	43/70 (61%) at least satisfactory	NA	145 / 162 (89.5%) completed PT protocol
Crossover to surgery	No	13/70 (19%)	36%	47 (29%)
Other healthcare services during follow-up	Not described	Not described	Not described	Not described
Outcomes				
Primary outcomes	Pain in the last week using a visual analogue scale	KOOS 4	KOOS 5	Change in patient-reported knee function on the Subjective Knee Form of the International Knee Documentation Committee (IKDC) from baseline over 24 months
Follow-up percentage treatment group	100%	64/70 (91%	22/22 (100%)	141/159 (89%)
Follow-up percentage control group	100%	62/70 (89%)	20/22 (96%)	141/162 (87%)
Reasons for dropout reported	NA	No	Yes	Yes

Reviewer: 2

Reviewer 2:
The “Ethics and dissemination ” in the “ABSTRACT” section need to be addressed more appropriately since it is an individual participant data meta-analysis.

Authors’ response:
Thank you for the response. To better reflect the ethics concerns of the IPDMA we have added additional lines to the abstract.

Authors actions:
Added lines to abstract (page 2, line 58 – 59)

All trial data will be anonymized before it is shared with the authors. The data will be encrypted and stored on a secure server located in the Netherlands. No major ethical concerns remain.

VERSION 3 – REVIEW

REVIEWER	Antti Malmivaara National Institute for Health and Welfare (THL)
REVIEW RETURNED	19-Dec-2019

GENERAL COMMENTS	Thank you for enriching the description of the RCTs in your manuscript. Please add the idea of using the descriptive data in assessing justification of meta-analysis and in assessing the generalizability of the findings not just in the methods section (please see the full appraisal) but also to the abstract and to the relevant sections of the manuscript itself. - The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.

VERSION 3 – AUTHOR RESPONSE

Reviewer(s)' Comments to Authors:

Reviewer 1:

Reviewer 1:
Thank you for adding the paragraph that you present above. As the actualization of the interventions in both study arms form the causal factors for the outcomes, please describe also 1) adherence to the interventions in both treatment arms, and 2) shares (and timing) of cross-overs from the index intervention to control intervention and vice versa.
Please separate the paragraph from the text dealing with publication bias; and please add to the paragraph that the description of the study characteristics will be used in assessing the clinical homogeneity/heterogeneity of the RCTs, and that the description will be used 1) in assessing which trials can be entered into a meta-analysis, and 2) in assessing the generalizability of the results of the meta-analysis.

Please add the idea of using the descriptive data in assessing justification of metaanalysis and in assessing the generalizability of the findings also to the abstract and to the relevant sections of the manuscript itself.
I would like to emphasize that the most important and fundamental question is more proximal than the meta-analysis itself: these RCT characteristics determine whether the trials are clinically homogenous enough to justify a meta-analysis. Thus only those studies for which the researchers can present an explicit rationale for clinical homogeneity can be put into the meta-analyses, and analysed as modifiers or confounders of outcome.

Authors' response:
Thank you for your response. As with every meta-analysis, it is important to check if a meta-analysis is justified and if pooling of studies is possible. As suggested we have added 1) adherence to the interventions in both treatment arms, and 2) cross-overs from the index intervention to control intervention and vice versa. Moreover, we have separated the paragraph and added that the characteristics will be used to assess which trials can enter the meta-analysis and to assess the generalizability of the final results.

Authors' actions:
Added text and separated the paragraph on page 7. Underlined is new.

"To enable to assessment of homogeneity/heterogeneity between the included trials, the following characteristics of the included RCTs will be compared and described in a table: 1) selection of participants, 2) inclusion- and exclusion criteria, 3) description of clinical path prior to inclusion, 4) number of participants that declined participation, 5) diagnostic characteristics, 6) work characteristics, 7) socioeconomic characteristics, 8) intervention and control treatment, 9) cross-over, 10) adherence to the intervention in both treatment arms, 11) other health care services during follow-up, and 12) outcome measures. These study characteristics will be used to assess which trials can enter the meta-analysis and to determine the generalizability of the results. "

Reviewer 1:
Thank you for the extension of description stated above. As stated above, please describe also: 1) adherence to interventions in both treatment arms, and 2) shares and timing of cross-overs from the index intervention to control intervention and vice versa. The most fundamental methodological issue here is the clinical homogeneity as the determinant of whether or a meta-analysis does or does not make sense.

Authors' response:
Thank you for your response. We have added both adherence to the intervention and two-sided cross-over to our manuscript.

Reviewer 1:
Again, as the actualization of the interventions in both study arms form the causal factors for outcome, please describe 1) adherence to interventions in both treatment arms, and 2) shares and timing of cross-overs from the index intervention to control intervention and vice versa. In addition, please take into account the degrees of adherences to the interventions and shares of cross-overs in both treatment arms, when

you make your conclusions. E.g. if adherence to APM is 35% in the control arm and nearly 100% in the index arm, you have to conclude that operating 35% instead of operating 100% will lead to the point estimates shown by the trial. In addition, if you will be able to pool results in meta-analyses, you have to take the adherences for the interventions into account when assessing the point estimates or at least when making inferences of the degree of effectiveness.
To conclude here, the Table must not be in isolation of the rest of the systematic review, but must be decisive for assessing the indication for a meta-analysis, and for assessing the generalizability of the results.

Authors' response:
Thank you for your response.
As described previously, we will include an adequate table explaining the characteristics of the included papers in the review and the heterogeneity between the studies will be assessed. After the assessment, we will determine if a fully combined meta-analysis is justified or if we have to stratify for different control treatments under comparison or other decisive factors. This table will not be isolated from the rest of the systematic review, instead it will be used to provide arguments for or against the decision to perform a meta-analysis as we have added to page 7 of the manuscript. However, we do not agree that this table should be located in an IPDMA protocol as data collection is still active and it is not clear which RCTs we can include.

Reviewer 1:
Thank you for adding the description of RCT characteristics in your paper, and including also the realization of the interventions. I think that it is also most important to differentiate between studies with sham controls and those with non-blinded controls, because in the latter the placebo effect is present. Therefore trials belonging to these two categories should not be entered into a same meta-analysis. I suggest that in item 5) diagnostic categories you would include information on traumatic vs non-traumatic mechanism, and presence/absence of osteoarthrosis; and in the item of selection (1), whether there was a previous conservative treatment before randomization and if there was, what was the length and intensity of this treatment.

Authors' response:
Thank you for your response and suggestions. We will add your suggestions to our manuscript.

Authors actions:
Added lines to the manuscript on page 7. (underlined is new)

"To enable to assessment of heterogeneity between the included trials, the following characteristics of the included RCTs will be compared and described in a table: 1) selection of participants, 2) previous (conservative) treatment(s) before randomization, 2) inclusion- and exclusion criteria, 3) description of clinical path prior to inclusion, 4) number of participants that declined participation, 5) diagnostic characteristics, including traumatic or non-traumatic injury and presence or absence of osteoarthrosis, 6) work characteristics, 7) socioeconomic characteristics, 8) intervention and control treatment, 9) cross-over, 10) adherence to the intervention in both treatment arms, 11) other health care services during follow-up, and 12) outcome measures."

Reviewer 1:

Thank you. Please note that including characteristics in the IPDMA is not enough. You have to consider the clinical homogeneity as a prerequisite for a meta-analysis.

Reviewer 1:
Thank you for including a much richer description of the actual PICO characteristics of the RCTs. However, using these characteristics only in a meta-analysis is grossly inadequate. I think that it is of fundament importance that this description is used in weighing whether meta-analysis is justified and between what studies. This justification must be made explicit in the discussion. The actual PICO characteristics must also be used when making interpretations of generalizability of the results.

Authors' response:
Thank you for your comment and we completely agree that these PICO characteristics serve multiple functions: to assess the justification for or against a complete meta-analysis, stratified meta-analysis or no meta-analysis at all, and secondly, for the interpretation of the generalizability of the results. However, an overview of these characteristics should in our opinion not be included in the an IPDMA protocol but instead in the final systematic review.

Reviewer 1:
Please use the comprehensive PICO description in weighing whether or not a meta-analysis is justified between some of the studies, and make these inferences explicit in the discussion.

Authors' response:
Thank you for your response. The discussion paragraph of an IPDMA protocol is typically used to describe the purpose and added value of an IPDMA, including opportunities and challenges. The justification for a meta-analysis does, in our opinion, not belong in the discussion paragraph of a protocol. We completely agree that the justification for a meta-analysis does need to be extensively discussed in the final systematic review (with or without meta-analysis) and we will surely include this in our IPDMA manuscript.

Reviewer 1:
When considering 1) the justification for a meta-analysis and 2) generalizability of results from individual RCTs, please take into consideration how well the actualized PICO was described. One is not able to make generalizations of RCTs which fail to describe appropriately the actual study material. We have to accept the current realities of having rather poor evidence overall. No meta-analysis, even an IPD one, can overcome this flaw.

Authors' response:
Thank you for your response. We will take the description of the PICO in mind when considering the justification for a meta-analysis and for the generalizability of our results.

Reviewer 1:
Thank you for improving the description of the PICO characteristics.
Please tabulate and describe all RCTs separately. And please include adherence to surgery, as well as other items suggested in this review.

Please add to the paragraph describing the study characteristics their use in assessing the clinical homogeneity/heterogeneity of the RCTs, and that this information will be used 1) in assessing which trials can be entered into a meta-analysis, and 2) in assessing the generalizability of the results of the meta-analysis.
Valid analytical research is dependent of a comprehensive description of all relevant RCT characteristics, and these characteristics must enter judgement based on substance understanding before a meta-analysis can be considered. A meta-analysis is justified only when the authors can provide sound clinical reasoning that RCTs included in a meta-analysis are clinically sufficiently homogenous. A comprehensive description of the recruited patients, adherences and cross-overs of the interventions, and the outcomes is mandatory for any systematic review. Also, the risk of bias of the RCTs must be evaluated and judgements of the current evidence must be based on RCTs with low risk of bias.

Authors' response:
Thank you for your response and review. We find it difficult to see the added value of the separate description and tabulation of the RCTs. A comprehensive table that includes all PICO related topics for all included RCTs will increase comparability and readability in our opinion and this will improve the discussion on the justification for or against a meta-analysis and the generalizability of the results. We will include such a table in our IPDMA manuscript, including the adequate justification for or against a meta-analysis.

VERSION 4 – REVIEW

REVIEWER	Antti Malmivaara National Institute for Health and Welfare
REVIEW RETURNED	06-Feb-2020

GENERAL COMMENTS	Dear authors, I am most pleased that you have added in the manuscript a comprehensive documentation of the characteristics of RCTs, and that these characteristics will be used for assessment of clinical homogeneity determining whether or not each RCT can be entered into a meta-analysis. And, that the characteristics of RCTs will be used when considering the generalizability of the findings. Please add a short sentence of the documentation of the RCTs, and using this when assessing justification for a meta-analysis and generalizability of the results also in the method section of the abstract, e.g. "Characteristics of RCTs will be used for assessment of clinical homogeneity and generalizability of the findings." - The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.