PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	COST-EFFECTIVENESS OF AN 8-WEEK SUPERVISED EDUCATION AND EXERCISE THERAPY PROGRAM FOR KNEE AND HIP OSTEOARTHRITIS: A PRE-POST ANALYSIS OF 16,255 PATIENTS PARTICIPATING IN GOOD LIFE WITH OSTEOARTHRITIS IN DENMARK (GLA:D®)
AUTHORS	Grønne, Dorte; Roos, Ewa; Ibsen, Rikke; Kjellberg, Jakob; Skou, Søren

VERSION 1 – REVIEW

REVIEWER	Collins, J. E. Harvard Medical School, Orthopedic Surgery
REVIEW RETURNED	19-Mar-2021

GENERAL COMMENTS	bmjopen-2021-049541: Cost-effectiveness of an 8-week supervised education and exercise therapy program for knee and hip osteoarthritis: a pre-post analysis of 16,255 patients participating in good life with osteoarthritis in denmark (GLA:D®) This is an interesting and well written manuscript describing a cost-effectiveness analysis of an 8-week supervised exercise therapy program for osteoarthritis. The manuscript is generally clear and sufficiently detailed, and adheres to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS), and the authors acknowledge the main limitation which is the pre- post-nature of this study design (i.e., no true control group). There are a few places where more detail/clarification could be provided. 1) Lines 216-217, 'area under the curve' to determine post-intervention EQ-5D. I read this multiple times and do not understand how QOL the year after intervention was determined from 3 and 12 month EQ-5D. 2) Line 218, repeated measures model – EQ-5D was collected pre-intervention (baseline) and at 3 and 12 months. It's unclear when/how costs were obtained and why a repeated measures model is necessary here. Similarly on line 208, "gamma regression model for repeated measures" – do the repeated measures here refer to pre- and post-intervention period? 3) Line 220, lack of model convergence – this should be noted as a limitation, since all models could not adjust for the same set of covariates. It is unclear, with such a large sample size, why model convergence would be a problem. 4) Line 261, "predicted health care costs" – how were costs predicted? Lines 150-182 seem to indicate that costs were directly measured through the Danish National Health Insurance Service Registry.

REVIEWER	Marsh, Jacquelyn

	The University of Western Ontario, Faculty of Health Science
REVIEW RETURNED	10-May-2021

GENERAL COMMENTS	COST-EFFECTIVENESS OF AN 8-WEEK SUPERVISED EDUCATION AND EXERCISE THERAPY PROGRAM FOR KNEE AND HIP OSTEOARTHRITIS: A PRE-POST ANALYSIS OF 16,255 PATIENTS PARTICIPATING IN GOOD LIFE WITH OSTEOARTHRITIS IN DENMARK (GLA:D®) Summary This is a registry-based pre-post study design where the objective was to evaluate the cost-effectiveness of an 8-week supervised education and exercise therapy program (GLA:D®) for patients with symptomatic knee or hip OA, using patient level data from the GLA:D® registry and national registries in Denmark. The authors conclude that the structured 8-week supervised education and exercise therapy program was cost-effective at one year in patients with knee or hip OA supporting large scale implementation in clinical practice. ABSTRACT • P.2, line 32 The time horizons are unclear to me. In the abstract you state it's a one-year time horizon, however in the methods you also discuss a three-year time horizon. Please clarify. • P.2, line 32 Please specify that the change in healthcare costs refers to a change from baseline. • P. 2, line 33 Please specify what measures you were assessing with the GEE. Was it an actual incremental cost-effectiveness ratio? • P.2, line 39 Please continue using (95%CI…) throughout. It is a little confusing with all the numbers jumbled without context. Also, please replace the / with the word "and" to make things a little clearer that it's looking at different values between knee and hip. • P.2, line 42 Please consider revising "lower change" to "a reduction" here and throughout. • P.3, line 44 Although the mean values of your evaluations were below the WTP thresholds, the CIs for the hip include the NICE guideline threshold and therefore it should be acknowledged here. INTRO • P.4, line 72 Please consider revising to "… pressure health care services are facing around the world." METHODS • P.5, line 97 Please consider revising to "registry" rather than register here, and elsewhere in the manuscript. • P.5, line 100 The perspective used for the analysis needs to be reported here (eg, healthcare payer perspective). • P.5, line 101 The three-year horizon is mentioned here and several other areas throughout the manuscript (including tables), however its significance is never really discussed in the results nor discussion towards the end of the manuscript. Please provide a rationale for the inclusion of the 3-year data. Were there EQ-5D

data available for patients at the 3-year mark as well that could be analyzed also?

• P.5, line 102 Please clarify - the outcome you are comparing is the quality-adjusted life year in your ratio, not quality of life specifically.
• P.6, line 109 Earlier you state that the program is an 8-week program, however here you state it consists of 12 one-hour sessions, delivered twice weekly. Please clarify.
• P.6, line 111 Please provide a statement with who typically delivers the 2-day GLA:D course and their training.
• P.6, line 114 Please revise municipals to “municipalities”.
• P.6, line 116 Given only 40% of the fees associated with private physiotherapy are covered through public reimbursement, the 60% cost to the patient that comes out-of-pocket needs to be addressed as well. I understand the study design does not allow you to account for these types of costs, however, it should be acknowledged and discussed as a limitation of the study in the discussion. Similarly, it is unclear how services are therefore accounted for with municipal centers as patients are not charged. Later in your methods you mention interventions in municipal settings were not available and not included. Did you account for physiotherapy-related costs in these patients in any capacity?
• P.6, line 116 Please specify your criteria for “clinical diagnosis” of knee and/or hip OA. Is it the American College of Rheumatology criteria?
• P.7, line 142 Can you provide more description on how the public transfer payments system works in Denmark and its aims?
• P.7, line 147 Please clarify a few weeks following the intervention or approximately a month after the intervention if the program is 8 weeks. If completed a few weeks after the program, the effects on the EQ-5D may be different than if completed immediately after completing GLA:D.
• P.7, line 149 How did you define and collect compliance?
• P.8, line 157 Again, please clarify the significance of the 3-year horizon if the data were not reported on in the results or included in the discussion.
• P.8, line 162 Please provide a short description of how the Danish Case Mix System works and its purpose.
• P.8, line 170 Please revise contacts to “contexts”.
• P.8, line 173 What about medications that are not covered by public funds. Were these accounted for in the analysis in any way? An overall discussion point about some of the important costs that were not captured given the nature of the study design as mentioned above would be important.
• P.9, line 177 This section was a little unclear to me. Does Statistics Denmark, provide individual costs per person for personal care and practical help or does it rather provide more of a summary cost?
• P.9,Lines 188-189: “QALYs combine time lived and QOL into a single index number where ‘1’corresponds to one year of full health and ‘0’ corresponds to being dead.” Above you say -0.64.?
• P.9, line 192 Please clarify what you mean by “western” or “not western” as categorical variables for ethnic background.
• P.10, line 206 Please describe the link between the weeks receiving public transfer and the ultimate outcomes from the study. How was the information used in your analysis?
• P.10, line 214 A limitation that will need to be discussed as well is the assumption that the EQ-5D score collected at baseline corresponds to a quality of life throughout the entire year prior to intervention. You were unable to do the area under the curve like

you did in the intervention phase. Please include a short description in the discussion.

• P.10, line 217 Please clarify the exact predictor and outcome of the analysis. It is still a little unclear to me here whether the outcome was the ratio between cost and EQ-5D scores or whether the change scores of the costs and the EQ-5D were analyzed separately. I am guessing time was included as a predictor variable in the model as well, but please confirm. Also, why was EQ-5D included in this model rather than the QALY outcome that was discussed earlier in the methods section? Provide a clear description of your model in full, along with specific model parameters you used to fit your data.

• P.11, line 239 Again, only 40% of the costs associated with private PT are captured by your system, whereas the 60% unaccounted for is out of pocket for the patient. The physiotherapy costs were also not captured for the municipal rehabilitation centers. I am not sure you can reliably assess cost-effectiveness between these groups when not all physiotherapy costs are accounted for in each group. This should be discussed in the limitations.

RESULTS

• P12; line 251; What do you mean by ‘most likely’? Is there a known clinically important difference?

• P.12, line 255 Please specify the types of surgeries that were reported. Were the surgeries primarily total joint replacement or were some arthroscopies as well? The difference in cost between replacement and arthroscopic surgeries are likely sizeable.

• P.13, line 270 Labeling a “one-year horizon” when talking about a year horizon for two separate time periods (2 years) makes it a little confusing. I think it would make it less confusing if you specifically stated “one year prior” and “the intervention year” or something along those lines.

• P.13, line 272 It is still unclear to me the exact timeframe for the three-year horizon. The sentence here makes it appear like you are talking about the 3rd year costs only, not the entire accumulated costs from the time of intervention to year three. Can you please clarify?

• P.13, line 277 Please report the mean costs associated with the surgeries themselves here.

• P.13, line 278 As mentioned earlier in the manuscript, some of the patients had a TJR prior to the 1-year mark where the EQ-5D would have been completed. How many patients underwent TJR prior to year 1? This would likely impact the EQ-5D score that was reported if it was complete soon after joint replacement surgery.

• P.14, line 289 Again, it is still not clear to me how your regression was modeled and what the outcome was. Here it seems like it’s the incremental cost-effectiveness ratio, however, please clarify in your methods.

• P.14, line 290 Please provide some context as to what the one-year adjusted cost/QALY gained was “lower” than.

• P.14, line 305 Did you attribute a cost associated with death in the patients who died?

DISCUSSION

• P.15, line 310 As discussed above, it will be important to discuss costs that were not included in your analysis and its implications. For example, individual costs to the patient, any

potential caregiver time associated with the disease, time away from employment not captured by system costs or for those who are self-employed, etc.

• P.15, line 310 As discussed above, it becomes a little confusing when saying a “one-year horizon” when the time horizons are staggered within the same subjects. Please consider rewording here and throughout to make it clearer that this analysis was within-subject with staggered time intervals.

• P15l Lines 313-314: Our study demonstrated that an 8-week supervised patient education and exercise therapy program for knee or hip OA implemented in primary care is cost-effective in a one-year horizon with health care costs of 8,497€ per QALY for knee patients and 22,568€ for hip patients

• P.15, line 316 Please specify the costs associated with hip and knee surgeries here to highlight the relative contributions for each and the proportion of patients this applied to.

• P. 15, line 329 I think it’s important to acknowledge here that the present analysis didn't technically compare between groups, therefore the results are not directly comparable to previous analyses that compare two groups over the same timeframe.

• P. 16, line 347 I think it’s also important to consider there is a lot of variation in how individuals are referred for and opt to have knee or hip surgery. For example, several sociodemographic factors may influence this, and sex and gender may play a role as well. For example, previous studies have shown that females are less likely to be offered surgery due to gender biases. As your sample was predominantly female, this may have an effect on the number of individuals who went on to total joint replacement.

• P. 16, line 350 Please specify where the cost savings apply. This relates back to the perspective of the analysis.

• P. 16, line 351-352 “….supervised education and exercise therapy is cost-effective.” Cost-effective compared to no program? As this is a pre-post design, there was no comparison to an alternative intervention. Is pre/baseline assumed to be ‘no Glad’?

• P. 17, line 355 Please specify your criteria for “compliant” here again.

• P. 17, line 363 Please specify what you mean by “uncertain compliance”.

• P. 18, line 393 It would also be important to mention that the changes observed in health outcomes may also be related to placebo effects that were not captured in the analysis.

• P. 18, line 397 The significance of the three-year horizon is still unclear to me since no health outcomes were related for this time point. Please provide further clarification here.

• P.19, line 408 There are also limitations with such a large volume loss to follow-up (ie, 39%) at one year. It could be that these patients abandoned the program because they weren't doing well and didn't think it was worth the time any longer. I think this should be mentioned when discussing potential for data not missing at random later in the discussion.

• P.19, line 412 Please expand more on the implications of selection bias in this context.

• P.19, line 414 Please consider revising to “…compared to those who provided this information”.

• P.20, line 425 Given the WTP thresholds (for the UK) cross the CI for hip OA, I think it’s important to acknowledge here.

TABLES

	• Table 3 Please consider revising the title of the last column to more clearly state it is the calculated QALY over the entire year. Post-period QALY to me makes it sound like it was the QALY measured specifically at 12 months only. • Table 4 Please specify how the confidence intervals were generated for the MI. Are you saying you used the MI data with a different set of CI?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Cost-effectiveness of an 8-week supervised education and exercise therapy program for knee and hip osteoarthritis: a pre-post analysis of 16,255 patients participating in good life with osteoarthritis in denmark (GLA:D®) This is an interesting and well written manuscript describing a cost-effectiveness analysis of an 8-week supervised exercise therapy program for osteoarthritis. The manuscript is generally clear and sufficiently detailed, and adheres to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS), and the authors acknowledge the main limitation which is the pre-post- nature of this study design (i.e., no true control group). There are a few places where more detail/clarification could be provided

Response:

We would like to thank the reviewer for the thoughtful comments and efforts towards improving our manuscript. In the following, we highlight the concerns of the reviewer and our response to address these concerns.

Comment: Lines 216-217, 'area under the curve' to determine post-intervention EQ-5D. I read this multiple times and do not understand how QOL the year after intervention was determined from 3 and 12 month EQ-5D.

Response: Contrary to using the 3m EQ5D measurement representing the QOL in the first three months post index date and the 12 m EQ5D measurement representing the QOL month four to twelve, the change from the index date to 3m and further to 12m was gradually taken into account as illustrated in the figure added to the supplementary appendix.

Author Action: Figure S1 added to the supplementary appendix and at line xx text added: *"(Figure S1, Supplementary Appendix*)".

Comment: Line 218, repeated measures model – EQ-5D was collected pre-intervention (baseline) and at 3 and 12 months. It's unclear when/how costs were obtained and why a repeated measures model is necessary here. Similarly on line 208, "gamma regression model for repeated measures" – do the repeated measures here refer to pre- and post-intervention period?

Response: In the description of variables (starting at line 155) it is described that all costs were obtained on an actual individual level i.e. costs both the year prior to and the year following index date are actual costs covering the relevant period. To further clarify this approach the term 'actual' was added at various places.

A model of repeated measures was used as the same patients are followed in the pre and the post period. This argument has been added to the text.

Author Action: '*Actual'* was added at line 217, 220, 227. '*A model for repeated measures was applied as the same patients were included in the pre and post period.*' was added at line 222.

Comment: Line 220, lack of model convergence – this should be noted as a limitation, since all models could not adjust for the same set of covariates. It is unclear, with such a large sample size, why model convergence would be a problem.

Response: Thank you for addressing this issue. In the main analysis evaluating the change in costs to change in QALY only minor differences occurred and a comment regarding this has been added to the discussion. Even if the sample size is large some cells are small and consequently lack of model convergence can be an issue.

Author Action: Added line 426: *'As a consequence of lack of model convergence marital status and ethnicity was omitted as covariates in the adjusted model evaluating the costs for home care estimating change in costs per QALY gained in a one-year horizon. As costs related to home care comprises a rather small proportion of the total costs it is not considered to affect the main result.'*

Comment: Line 261, "predicted health care costs" – how were costs predicted? Lines 150-182 seem to indicate that costs were directly measured through the Danish National Health Insurance Service Registry.

Response: It is correct that all costs were actual individual level costs extracted from the national registries. The term 'predicted' was used through the manuscript to specify that the adjusted results were estimated and not exact. To avoid confusion the term 'predicted' was changed to 'adjusted' through the manuscript.

Author Action: 'Predicted' changed to 'adjusted' or 'estimated' at line 217, 218,221, 282, 283, 284, 294, Table 2, S3 and Figure 2.

Reviewer 2

Summary. This is a registry-based pre-post study design where the objective was to evaluate the cost-effectiveness of an 8-week supervised education and exercise therapy program (GLA:D®) for patients with symptomatic knee or hip OA, using patient level data from the GLA:D® registry and national registries in Denmark. The authors conclude that the structured 8-week supervised education and exercise therapy program was cost-effective at one year in patients with knee or hip OA supporting large scale implementation in clinical practice.
Response:

We would like to thank the reviewer for many relevant comments that helped us improve the manuscript. In the following, we addressed these concerns.

Comment: P.2, line 32 � The time horizons are unclear to me. In the abstract you state it's a one-year time horizon, however in the methods you also discuss a three-year time horizon. Please clarify.

Response: The main analysis evaluating the health care costs per QALY is carried out in a one-year horizon and only the main analysis is referred in the abstract being the most relevant analysis. In addition, as a secondary analysis costs were also reported in a three-year horizon. To clarify the text has been changed.

Author Action: Text changed at line 104: *'In the primary analysis, we reported health care costs per QALY gained in a one-year horizon calculated as the ratio of change in health care costs to change in QOL. In addition, as a secondary analysis, mean actual health care costs and costs to home care and public transfer payments were reported in a three-year horizon to assess how costs develop over time in this population of patients with a chronic condition.'*

Comment: P.2, line 32 ⮚ Please specify that the change in healthcare costs refers to a change from baseline.
Response: Specification that the change refers to change from baseline added to the abstract.

Author Action: Word added at line 33: '*from baseline'*

Comment: P. 2, line 33 ⮚ Please specify what measures you were assessing with the GEE. Was it an actual incremental cost-effectiveness ratio?

Response: The GEE model is used to assess adjusted costs and adjusted EQ5D at the different time points. In the abstract it is described which measures were used (change in health care costs and change in EQ5D) and that the adjusted measures were estimated using GEE. Text added to clarify.

Author Action: Line 34 added *'All'.*

Comment: P.2, line 39 ⮚ Please continue using (95%CI…) throughout. It is a little confusing with all the numbers jumbled without context. Also, please replace the / with the word "and" to make things a little clearer that it's looking at different values between knee and hip.

Response: '95% CI' added in all parenthesis describing confidence interval both in the abstract and in the text and '/' replaced with 'and'. Further adaptations of the abstract were made to keep the word count below the limit.

Author Action: **'***95% CI*' added and '/' replaced with '*and'* several places in the abstract.

Comment: P.2, line 42 ⮚ Please consider revising "lower change" to "a reduction" here and throughout.
Response: Thank you for your suggestion. As the change is not reducing but is lower compared to other change measures, we prefer to keep the existing wording.

Author Action: None

Comment: P.3, line 44 ⮚ Although the mean values of your evaluations were below the WTP thresholds, the CIs for the hip include the NICE guideline threshold and therefore it should be acknowledged here.
Response: Information added in abstract

Author Action: Added line 47: **'***except the upper limit of the 95% CI for hip patients which was in between the two thresholds'*.

Comment: P.4, line 72 ▯ Please consider revising to "… pressure health care services are facing around the world."
Response: Thank you for your suggestion improving the phrasing. Revised as suggested.

Author Action: Text changed at line 77: *'pressure health care services are facing around the world'*

Comment: P.5, line 97 ▯ Please consider revising to "registry" rather than register here, and elsewhere in the manuscript.
Response: Thank you for your suggestion improving the phrasing. Revised as suggested.
Author Action: Word changed at line xx and line xx to *'registry'*.

Comment: P.5, line 100 ▯ The perspective used for the analysis needs to be reported here (eg, healthcare payer perspective).

Response: Thank you for pointing this out, perspective added in the text.

Author Action: Added line 101 and line 105: '*in a healthcare payer perspective'*

Comment: P.5, line 101 ▯ The three-year horizon is mentioned here and several other areas throughout the manuscript (including tables), however its significance is never really discussed in the results nor discussion towards the end of the manuscript. Please provide a rationale for the inclusion of the 3-year data. Were there EQ-5D data available for patients at the 3-year mark as well that could be analyzed also?

Response: Reporting a three-year horizon on costs was included as secondary analysis to the main analysis evaluating the cost effectiveness of the intervention and we find it relevant to provide this additional information about how costs develop over time in this population. Looking at the results further justify inclusion of this analysis. Unfortunately, EQ-5D was not available at 3 years, and evaluating the cost effectiveness in a longer horizon than 1 year would require repetition of the survey.

The development in costs in the three-year horizon is already referred in the results section from line 294 to line 301. The discussion refers to the three-year horizon at line 367 and 433, but the main analysis is in focus in the discussion. As it is a secondary analysis the results are not included in the conclusion.

The text describing the study design is changed and we have highlighted that this was a secondary analysis and why it was relevant.

Author Action: At line 104 text changed to: *'In the primary analysis, we reported health care costs per QALY gained in a one-year horizon calculated as the ratio of change in health care costs to change in QOL. In addition, as a secondary analysis, mean actual health care costs and costs to home care and public transfer payments were reported in a three-year horizon to assess how costs develop over time in this population of patients with a chronic condition*

Comment: P.5, line 102 Please clarify - the outcome you are comparing is the quality-adjusted life year in your ratio, not quality of life specifically.
Response: The outcome is '*health care costs […] per QALY gained in a one-year horizon calculated as the ratio of change in health care costs to change in QOL'*. As the time horizon is 1 year there is actually no difference between the QALYs gained and change in QOL used as a denominator in the ratio. As the existing text describes how the QALY gained were calculated no changes were made.

Author Action: None

Comment: P.6, line 109 Earlier you state that the program is an 8-week program, however here you state it consists of 12 one-hour sessions, delivered twice weekly. Please clarify.
Response: The 8 week period refers to the whole program also including first visit at the clinic, performance of the functional tests and attending patient education sessions. This is why the length of the program is described as 8 weeks and not only 6 weeks. Details added to the text.

Author Action: Added at line 116: **'***delivered over approximately 8 weeks'*

Comment: P.6, line 111 Please provide a statement with who typically delivers the 2-day GLA:D course and their training.
Response: The course certifying clinicians is delivered at the University of Southern Denmark by researchers, clinicians and a former patient. This information is added to the text.

Author Action: Added at line 120: '*at the University of Southern Denmark delivered by researchers, clinicians, and a former patient'*.

Comment: P.6, line 114 Please revise municipals to "municipalities".

Response: Thank you for correcting the spelling

Author Action: At line 125 word changed to: '*municipalities'*.

Comment: P.6, line 116 Given only 40% of the fees associated with private physiotherapy are covered through public reimbursement, the 60% cost to the patient that comes out-of-pocket needs to be addressed as well. I understand the study design does not allow you to account for these types of costs, however, it should be acknowledged and discussed as a limitation of the study in the discussion. Similarly, it is unclear how services are therefore accounted for with municipal centers as patients are not charged. Later in your methods you mention interventions in municipal settings were not available and not included. Did you account for physiotherapy-related costs in these patients in any capacity?

Response: Thank you for addressing this topic. Neither the patient's out-of-pocket costs nor the costs covering expenses to the programs in municipal settings were taken into account in the analyses as the information was not available. As the increase in costs in primary health care sector the first year following index date only constitute a very low proportion of the increased costs in total, this limitation is not considered to have a significant effect on the overall results. Further, since costs for health care services delivered in municipal settings were not available, all analyses were repeated stratified for patients attending GLA:D® in private physiotherapy clinics vs. in municipal rehabilitation centers. Paragraph added to the discussion.

Author Action: Paragraph added at line 439: '*Only around 60% of the costs covering the program for most patients attending GLA:D® in private physiotherapy clinics were taken into account in the analyses i.e. patients out-of-pocket costs and costs covering the program in municipal settings as well as medications bought over the counter were not included. As the increase in costs in the primary health care sector and in costs covering medications the first year following index date only constitute a very low proportion of the increased costs in total, this limitation is not considered to substantially affect the overall results.*'

Comment: P.6, line 116 ⃞ Please specify your criteria for "clinical diagnosis" of knee and/or hip OA. Is it the American College of Rheumatology criteria?

Response: All treating therapists are educated as physiotherapists and are certified in delivering GLA:D®. At the course they were instructed in how to diagnose osteoarthritis as well as differential diagnosis. This information is added to the text. Since also patients with early stage of osteoarthritis and patients at a younger age can have osteoarthritis and can benefit from the intervention no definite set of clinical classification criteria was used. Taking the objective of this manuscript into account it is too detailed to elaborate on this topic, but if you want further description it can be found in the following two publications: DOI: 10.1016/j.joca.2019.09.003 and DOI: 10.1016/j.ocarto.2020.100111

Author Action: Paragraph added at line 121: '*All therapists were instructed in how to diagnose osteoarthritis and informed about differential diagnosis*'

Comment: P.7, line 142⃞ Can you provide more description on how the public transfer payments system works in Denmark and its aims?

Response: Thank you for your interest in the Danish welfare system and acknowledging that the context has implications for the results. Given that the manuscript already exceeds the recommended 4000 words, we do not think that adding information about the danish system in the text is feasible, but we have added a reference.

Author Action: Sentence added at line 154: "Please find more information about the Danish health care system elsewhere, [22]."

Comment: P.7, line 147 ⃞ Please clarify a few weeks following the intervention or approximately a month after the intervention if the program is 8 weeks. If completed a few weeks after the program, the effects on the EQ-5D may be different than if completed immediately after completing GLA:D.

Response: As the program is implemented in primary care and are not a strictly regulated trial some variation in both delivery of the program and follow up time occurs. The content of the program covers 8 weeks (6 weeks of exercise + patient education), but the calendar time spent is often a bit more to fit into patients' everyday life with e.g. vacations and other reasons to not being able to attend two classes a week. Also, some waiting time from first visit at the clinic and baseline measurement to start intervention may occur. This is why the follow up measurement is approximately 3 months after first visit at the clinic.

Author Action: Text added at line 158 *'… as the program is implemented in primary care and some variation in follow up time occurs'*. Added at line 116: **'***delivered over approximately 8 weeks'*

Comment: P.7, line 149 How did you define and collect compliance?
Response: In line 209 it is described that *'information on compliance were therapist-reported and high compliance was defined as patients attending at least 10 supervised exercise sessions'.*

Author Action: None

Comment: P.8, line 157 Again, please clarify the significance of the 3-year horizon if the data were not reported on in the results or included in the discussion.
Response: Please see answer to reviewer's comment for P.5, line 101.

Author Action: At line 104 text changed to: '*In the primary analysis, we reported health care costs per QALY gained in a one-year horizon calculated as the ratio of change in health care costs to change in QOL. In addition, as a secondary analysis, mean actual health care costs and costs to home care and public transfer payments were reported in a three-year horizon to assess how costs develop over time in this population of patients with a chronic condition.'*

Comment: P.8, line 162 Please provide a short description of how the Danish Case Mix System works and its purpose.

Response: Short description added to the text.

Author Action: Text added at line 178: **'***which organize patients with similar diseases and similar expenses into groups that each have annually adjusted tariffs that reflects practice'*.

Comment: P.8, line 170 Please revise contacts to "contexts".

Response: 'Contacts' is the correct word as it refers to hospital contacts described a few lines above as '*all inpatient admissions and outpatient activities'*.

Author Action: None

Comment: P.8, line 173 What about medications that are not covered by public funds. Were these accounted for in the analysis in any way? An overall discussion point about some of the important costs that were not captured given the nature of the study design as mentioned above would be important.

Response: Neither medications bought over the counter were included in the study. Paragraph added to the discussion.

Author Action: Paragraph added to the discussion at line 439: **'***Only around 60% of the costs covering the program for most patients attending GLA:D® in private physiotherapy clinics were taken into account in the analyses i.e. patients out-of-pocket costs and costs covering the program in municipal settings as well as medications bought over the counter were not included. As the increase in costs in the primary health care sector and in costs covering medications the first year following index date only constitute a very low proportion of the increased costs in total, this limitation is not considered to substantially affect the overall results.'*

Comment: P.9, line 177 This section was a little unclear to me. Does Statistics Denmark, provide individual costs per person for personal care and practical help or does it rather provide more of a summary cost?

Response: As described in line 161, patients were linked to national registries via their Civil Registration number and all utilisation of the different services were retrieved on an actual individual level.

Author Action: Added at line 188: '*Individual level'*

Comment: P.9,Lines 188-189: "QALYs combine time lived and QOL into a single index number where '1'corresponds to one year of full health and '0' corresponds to being dead." Above you say -0.64.?

Response: The range of the scale using the Danish cross walk value set is -0.624 to 1. Describing that scoring '0' corresponds to being dead means that scoring below '0' you will have a quality of life that is worse than being dead. We find the existing description precise and sufficient.

Author Action: None

Comment: P.9, line 192 Please clarify what you mean by "western" or "not western" as categorical variables for ethnic background.

Response: Based on that people immigrating to Denmark from different countries overall can be grouped regarding demographic and socioeconomic parameters, the categorization divides countries in EU, associated countries and the four Anglo-Saxon countries into the western category and all other countries into non-western countries. This categorization is used by Statistics Denmark and similar categorizations are used elsewhere. Details added to the text.

Author Action: Text added at line 205-206: '… *ethnic background (western [countries in EU, associated countries and the four Anglo-Saxon countries] or not western [other countries])'.*

Comment: P.10, line 206 Please describe the link between the weeks receiving public transfer and the ultimate outcomes from the study. How was the information used in your analysis?

Response: Reporting public transfer payment is a part of the secondary analysis and therefor not the main focus in the study as a consequence neither are the results discussed or a part of the conclusion. The results are presented in the results section in the section starting at line 282 and we also think that the results add valuable information describing the three year course of cost in these patient groups.

Author Action: None

Comment: P.10, line 214 A limitation that will need to be discussed as well is the assumption that the EQ-5D score collected at baseline corresponds to a quality of life throughout the entire year prior to intervention. You were unable to do the area under the curve like you did in the intervention phase. Please include a short description in the discussion.

Response: We do agree with you that it can be questioned whether the EQ5D score at baseline is valid to use as measure describing the QOL the year pre intervention. Possible bias related to regression to the mean was already is discussed in the text and further paragraph was added.

Author Action: Paragraph added at line 422: *'In the analysis EQ-5D measured at baseline represented the QOL the year prior to the intervention, but there is a risk that the change in QOL were overestimated as patients often seek treatment at time of worsening of symptoms'.*

Comment: P.10, line 217 Please clarify the exact predictor and outcome of the analysis. It is still a little unclear to me here whether the outcome was the ratio between cost and EQ-5D scores or whether the change scores of the costs and the EQ-5D were analysed separately. I am guessing time was included as a predictor variable in the model as well, but please confirm. Also, why was EQ-5D included in this model rather than the QALY outcome that was discussed earlier in the methods section? Provide a clear description of your model in full, along with specific model parameters you used to fit your data.

Response:

As described in line 227 we calculated the ratio of change in healthcare costs to change in EQ5D: "We estimated health care costs per QALY gained as the ratio of change in actual total health care costs to change in QOL." Change in costs were estimated using a GEE model (as described in line 220-223) and, similarly, change in QOL was also estimated using a GEE model (as described in line 233-235). I.e. the change scores were estimated separately and afterwards the ratio was calculated. Both models were adjusted for the same factors (gender, age, marital status, ethnicity, educational level and administrative region, i.e. time was not included in the model).

Clarification added to the text.

Author Action: Added at line 236: "*In the first step change in health care costs and change in QOL were estimated in two different models, where* both raw and adjusted analyses were conducted, including gender, age, marital status, ethnicity, educational level and region as covariates. In case of no convergence in the model, selected covariates were omitted. *In the second step the ratio of change in health care cost to change in QOL were calculated*."

Comment: P.11, line 239 Again, only 40% of the costs associated with private PT are captured by your system, whereas the 60% unaccounted for is out of pocket for the patient. The physiotherapy costs were also not captured for the municipal rehabilitation centers. I am not sure you can reliably assess cost-effectiveness between these groups when not all physiotherapy costs are accounted for in each group. This should be discussed in the limitations.
Response: Please see response to comment P.6, line 116

Author Action: Paragraph added at line 439: *'Only around 60% of the costs covering the program for most patients attending GLA:D® in private physiotherapy clinics were taken into account in the analyses i.e. patients out-of-pocket costs and costs covering the program in municipal settings as well as medications bought over the counter were not included. As the increase in costs in the primary health care sector and in costs covering medications the first year following index date only constitute a very low proportion of the increased costs in total, this limitation is not considered to substantially affect the overall results.'*

Comment: P12; line 251; What do you mean by 'most likely'? Is there a known clinically important difference?

Response: Determining at what level a difference is clinically important is somewhat arbitrary and is highly dependent on the context. Based on our knowledge we deem the difference not to be clinically important but to reflect that there is a doubt and that there is no clear guidance the sentence is somewhat vague phrased to express some uncertainty.

Author Action: None

Comment: P.12, line 255 Please specify the types of surgeries that were reported. Were the surgeries primarily total joint replacement or were some arthroscopies as well? The difference in cost between replacement and arthroscopic surgeries are likely sizeable.

Response: As described at line 183 use of resources in contacts where surgery in the knee or hip occurred were reported separately i.e. this covers both joint replacements, arthroscopies and other surgeries. Unfortunately, the type of surgeries was not specified in the data available for the analysis.

Author Action: None

Comment: P.13, line 270 Labeling a “one-year horizon” when talking about a year horizon for two separate time periods (2 years) makes it a little confusing. I think it would make it less confusing if you specifically stated “one year prior” and “the intervention year” or something along those lines.

Response: We agree with you that this may seem confusing, but we prefer to keep the existing wording as this help distinguish between the one-year horizon and the three-year horizon as defined at line 104-109.

Author Action: None

Comment: P.13, line 272 It is still unclear to me the exact timeframe for the three-year horizon. The sentence here makes it appear like you are talking about the 3rd year costs only, not the entire accumulated costs from the time of intervention to year three. Can you please clarify?
Response: As described from line 147 and forward it is not the same patients that are included in analyses reporting the costs in the one-year horizon and the three-year horizon, respectively. Those included in the analysis reporting costs in a three-year horizon are restricted to be included in GLAD before December 2014 to allow for three years of follow up time. Also costs in the one-year horizon and the three-year horizon are reported in different measures (mean costs per month or year respectively) as described in line 168-170. Therefore, results from the two different analyses are reported separately. Costs reported in the three-year horizon are reported as mean yearly costs and it is correctly that the description at line 274 refers costs in the third year and not accumulated costs. This also corresponds to Figure 2.

Author Action: None

Comment: P.13, line 277 Please report the mean costs associated with the surgeries themselves here.

Response: We thank the reviewer for the interest in details in the results. All details are given in Table S2 (raw) and Table S3 (adjusted) as well as illustrated in figure 2. We find it appropriate not to go more into details in the text an only report the main message as in the existing text.

Author Action: None

Comment: P.13, line 278 □ As mentioned earlier in the manuscript, some of the patients had a TJR prior to the 1-year mark where the EQ-5D would have been completed. How many patients underwent TJR prior to year 1? This would likely impact the EQ-5D score that was reported if it was complete soon after joint replacement surgery.
Response: We agree with you that undergoing TJR within the first year after entering the intervention probably would have an impact on the EQ-5D. Also this would indeed have an impact on the costs as hospital admissions and surgeries are relative expensive. This study evaluates the GLA:D intervention regardless of whether the patient receives a surgery i.e. for all patients enrolled in GLA:D in real life where the program is implemented in primary care. Proportion of patients receiving a joint replacement surgery added to text.

Author Action: Text added at line 276: *'Seven percent and 17% of knee and hip patients, respectively, reported to have had a joint replacement surgery between start intervention and the 12 m follow up measurement.'*

Comment: P.14, line 289 □ Again, it is still not clear to me how your regression was modeled and what the outcome was. Here it seems like it's the incremental cost-effectiveness ratio, however, please clarify in your methods.

Response: Please see answer to reviewer's previous comment on the manuscript P.10, line 217.

Author Action: Added at line 236: "*In the first step change in health care costs and change in QOL were estimated in two different models, where* both raw and adjusted analyses were conducted, including gender, age, marital status, ethnicity, educational level and region as covariates. In case of no convergence in the model, selected covariates were omitted. *In the second step the ratio of change in health care cost to change in QOL were calculated*."

Comment: P.14, line 290 □ Please provide some context as to what the one-year adjusted cost/QALY gained was "lower" than.
Response: Text added for clarification.

Author Action: Text added at line xx: '*compared to all patients'*.

Comment: P.14, line 305 □ Did you attribute a cost associated with death in the patients who died?

Response: All costs related to the health care system are included in the analyses. Costs held by other parts of the well-fare system or out of pocket expenses held by relatives were not included in the analyses. As only a minor number of patients (n:53) died during the study period this is not elaborated further in the manuscript.

Author Action: None

Comment: P.15, line 310 □ As discussed above, it will be important to discuss costs that were not included in your analysis and its implications. For example, individual costs to the patient, any

potential caregiver time associated with the disease, time away from employment not captured by system costs or for those who are self-employed, etc.
Response: We do agree with the reviewer, that this is important. As this is a health economic evaluation with a healthcare payer perspective and not a societal perspective, it was not included, also information on these costs were not available. We have, however, included it in the discussion as proposed. Please also see response to reviewer's comment to P.6, line 116.

Author Action: Paragraph added at line 439: *'Only around 60% of the costs covering the program for most patients attending GLA:D® in private physiotherapy clinics were taken into account in the analyses i.e. patients out-of-pocket costs and costs covering the program in municipal settings as well as medications bought over the counter were not included. As the increase in costs in the primary health care sector and in costs covering medications the first year following index date only constitute a very low proportion of the increased costs in total, this limitation is not considered to substantially affect the overall results.'*

Comment: P.15, line 310 As discussed above, it becomes a little confusing when saying a "one-year horizon" when the time horizons are staggered within the same subjects. Please consider rewording here and throughout to make it clearer that this analysis was within-subject with staggered time intervals.

Response: In the title it is described that the study is a pre-post analysis. This is also described e.g. in the design at line 101 as well as that we reported the ratio of change in health care costs to change in QOL. In the section describing the statistical analyses at line 217 and through the results section the results are described as 'change'. I.e. we think that it is clearly described that the analyses is carried out as a pre-post analysis evaluating change within-subjects.

Text added that the outcome is calculated as the ratio of change in health care costs to change in QOL in the same patients.

Author Action: Text added in line 104: *'In the primary analysis, we reported health care costs in a healthcare payer perspective per QALY gained in a one-year horizon calculated as the ratio of change in health care costs to change in QOL in the same patients.'*

Comment: P15l Lines 313-314: Our study demonstrated that an 8-week supervised patient education and exercise therapy program for knee or hip OA implemented in primary care is cost-effective in a one-year horizon with health care costs of 8,497€ per QALY for knee patients and 22,568€ for hip patients

Response: We are unsure, whether any question or comment from the reviewer is missing here.

Comment: P.15, line 316 Please specify the costs associated with hip and knee surgeries here to highlight the relative contributions for each and the proportion of patients this applied to.
Response: Details added in the results section and in the discussion section.

Author Action: Text added in line 298: *'The increase in mean health care costs was mainly due to costs related to surgeries in the knee or hip which the first year after index date in the adjusted analysis accounted for 46€/month of an increase in costs of 68€/month in knee patients and*

130.8€/month of an increase in costs of 162.8€/month in hip patients (Table S3, Supplementary Appendix).' and text added in line 341: *'(accounting for 70 and 80% of the increased costs, respectively)'*

Comment: P. 15, line 329 I think it's important to acknowledge here that the present analysis didn't technically compare between groups, therefore the results are not directly comparable to previous analyses that compare two groups over the same timeframe.

Response: Thank you for providing focus on this issue. Text is added to specify.

Author Action: Text added at line 355: '*Even though our study is a pre-post study and therefore not directly comparable'*

Comment: P. 16, line 347 I think it's also important to consider there is a lot of variation in how individuals are referred for and opt to have knee or hip surgery. For example, several sociodemographic factors may influence this, and sex and gender may play a role as well. For example, previous studies have shown that females are less likely to be offered surgery due to gender biases. As your sample was predominantly female, this may have an effect on the number of individuals who went on to total joint replacement.

Response: We completely agree with the reviewer that several factors influence the decision whether a patient undergo a joint replacement surgery or not including factors mentioned. Though we find it not to be within the scope of the current study to go into discussions about this subject. Also, at line 464 we already stated that: '*patients attending GLA:D® are a preselected group of patients … which might limit the generalizability'*.

Author Action: None

Comment: P. 16, line 350 Please specify where the cost savings apply. This relates back to the perspective of the analysis.
Response: Thank you for pointing out that this information is useful to add.

Author Action: Added in line 377: '*in the Australian health care system'.*

Comment: P. 16, line 351-352 "….supervised education and exercise therapy is cost-effective." Cost-effective compared to no program? As this is a pre-post design, there was no comparison to an alternative intervention. Is pre/baseline assumed to be 'no Glad'?

Response: As the reviewer put forward the intervention is not compared to an alternative care path and no comparisons are made in the text. Pre baseline is not assumed to be neither no GLA:D or other intervention – but from the study we can conclude that in those who signed up for GLA:D the intervention is cost-effective.

Author Action: Additional text added at line 339: *'Our study demonstrated that an 8-week supervised patient education and exercise therapy program for knee or hip OA implemented in primary care is cost-effective in a one-year horizon with health care costs of 8,497€ per QALY for knee patients and 22,568€ for hip patients <u>who signed up for the intervention</u>'.*

Comment: P. 17, line 355 ⍰ Please specify your criteria for "compliant" here again.
Response: The categorization of patients into compliant group is described in the method section line 210. Detail added for clarification.

Author Action: Added at line 383 *'(i.e. attending at least 10 supervised exercise sessions)'*

Comment: P. 17, line 363 □ Please specify what you mean by "uncertain compliance".

Response: Uncertain compliance refers to studies were minimum one criterion of three possible (intensity, duration, frequency) was not reported, thus the classification into American College of Sports Medicine's recommendations was not possible. Clarification added to the text.

Author Action: Added at line 392 *'(studies where compliance was not possible to categorize according to recommendations)'*

Comment: P. 18, line 393 ⍰ It would also be important to mention that the changes observed in health outcomes may also be related to placebo effects that were not captured in the analysis.
Response: We agree with you that the observed change in outcomes can be affected by placebo effects as well as nocebo effects, regression to the mean, residual confounding etc. In regard to the pre-post design the following was already mentioned: '*Without a proper control group, it cannot be ruled out that the observed change in EQ-5D is related to other factors than the treatment such as regression to the mean'* and placebo is added to this sentence.

Author Action: Text added at line 421: '*placebo or'*.

Comment: P. 18, line 397 ⍰ The significance of the three-year horizon is still unclear to me since no health outcomes were related for this time point. Please provide further clarification here.
Response: Please see response to reviewer comment to P.5, line 101. Text added to further describe the background for conducting the analyses. No further description is added to the discussion as the main focus in the discussion is the primary analysis (cost effectiveness in one-year horizon). We have highlighted that this was a secondary analysis and why it was relevant.

Author Action: Text changed at line 104: *'In the primary analysis, we reported health care costs per QALY gained in a one-year horizon calculated as the ratio of change in health care costs to change in QOL. In addition, as a secondary analysis, mean actual health care costs and costs to home care and public transfer payments were reported in a three-year horizon to assess how costs develop over time in this population of patients with a chronic condition.'*

Comment: P.19, line 408 ⍰ There are also limitations with such a large volume loss to follow-up (ie, 39%) at one year. It could be that these patients abandoned the program because they weren't doing well and didn't think it was worth the time any longer. I think this should be mentioned when discussing potential for data not missing at random later in the discussion.
Response: We do agree with the reviewer that there is a risk of selective loss to follow up e.g. due to outcome from the GLA:D program i.e. either bad or good outcome and that the assumption of data (EQ-5D) being missing at random can be questioned. The existing text already covers this in line 448: **'***…indicating a risk of selective loss to follow up in the GLA:D® registry'* and line 452 '…*However, there is a risk that loss to follow up was related to unobserved factors not available for the analysis'*. As we could speculate on a range of differences between those who provided follow up measures and those who did not we find it most appropriate to use general terms as in the existing text. However, to meet

the reviewer's comments we added the example of outcome from GLA:D being a possible reason for not providing follow up data. As described in the existing text at line 446: *'conducting a sensitivity analysis restricted to patients with complete information revealed that they had less mean change in health care costs than all included patients'* this might indicate that a good outcome could be one reason for drop out.

Author Action: Added at line 453: *'(e.g. good or bad outcome from the GLA:D® program)'*.

Comment: P.19, line 412 Please expand more on the implications of selection bias in this context.
Response: The main message is that there is a risk of selection bias in the sub analysis stratifying for compliance meaning that the results could be affected by potentially systematically differences between those providing and not providing information on compliance. Describing text added.

Author Action: Added at line 456: *'i.e. that the lower change in health care costs could be due to systematically differences in the use of health care services between those providing and not providing information about compliance rather than due to the intervention'*.

Comment: P.19, line 414 Please consider revising to "…compared to those who provided this information".

Response: Thank you for suggesting a wording improving the readability of the sentence.

Author Action: Text changed at line 459 to: '*who provided'*.

Comment: P.20, line 425 Given the WTP thresholds (for the UK) cross the CI for hip OA, I think it's important to acknowledge here.

Response: Based on the methods used, results and discussions, we overall conclude that the intervention is cost effective. Added to text that the health care costs per QALY in hip patients were in between the two thresholds.

Author Action: Text added at line 472: ***'**except the upper limit of the 95% CI for hip patients which was in between two thresholds'.*

Comment: Table 3 Please consider revising the title of the last column to more clearly state it is the calculated QALY over the entire year. Post-period QALY to me makes it sound like it was the QALY measured specifically at 12 months only.

Response: Clarification added

Author Action: Added in the title of the last column in table 3: '*Composite'*

Comment: Table 4 Please specify how the confidence intervals were generated for the MI. Are you saying you used the MI data with a different set of CI?
Response: The method used to impute missing data on EQ5D at follow up does not create CI. Therefore, neither the raw analysis for change in EQ5D or the ratio do not have CIs.
In table 4, a note already describes that the CIs are not created for the MI.
Author Action: None