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)	Return-to-work intervention versus usual care for sick-listed employees: health-economic investment appraisal alongside a cluster randomised trial
AUTHORS	Lokman, Suzanne; Volker, Danielle; Zijlstra-Vlasveld, Moniek; Brouwers, Evelien; Boon, Brigitte; Beekman, Aartjan; Smit, Filip; Van der Feltz-Cornelis, Christina

VERSION 1 – REVIEW

REVIEWER	Mark Pennington King's Health Economics King's College London UK
REVIEW RETURNED	20-Mar-2017

GENERAL COMMENTS	The paper provides a thorough economic analysis of an online module designed to help employees on sick leave with mental health problems return to work. In general the analysis appears to be robust. The exclusion of health benefits in the analysis from the health sector perspective is puzzling. Advanced statistical techniques have been applied to quantify the impact of uncertainty on the results. The results of this analysis might most usefully be conveyed through the production of Cost-effectiveness Acceptability Curves (CEACs) where the health benefits are included. The authors undertake analysis from four different perspectives: employee, employer, health care sector, and societal. Each is reported as a cost benefit analysis. However, the health benefits of the intervention are only included in the analysis from the employee's perspective. It's difficult to see why the health benefits would be excluded from analysis taking either a health sector perspective or a societal perspective. In my view both analyses should include the health benefits. Such analysis might be considered a cost-benefit analysis if the health benefits are monetised. In fact the health benefits are measured in QALYs and a conservative value of 20,000 Euros per QALY is assumed. In my view it would be more correct to consider this a cost-utility analysis. Simply monetising QALY gains at 20,000 Euros does not convey the impact of different values one might place on a QALY gain on the results of the analysis. It would be more useful to decision makers to provide the results as a CEAC in which the likelihood the intervention is cost-effective is conveyed as a function of the value placed on a unit gain in health (QALY). The analysis from the employer's perspective excludes health benefits which seems reasonable. I would regard the impact of presenteeism and absenteeism as a cost (averted) and not a benefit of the intervention.

	This is perhaps a semantic point but this renders this analysis a cost-minimisation analysis. Nothing wrong with that, but I don't think the analysis from the employer's perspective is a true CBA. The authors are to be commended in their use of robust techniques to deal with the missing data alongside bootstrapping to capture the sampling uncertainty. A little more methodological detail would be useful. Did the authors bootstrap prior to fitting the EM models? There was one aspect of the analysis that troubled me. The authors choose to use self-reported data for absence rather than administrative records. They acknowledge substantial differences between the two sources. Why make the self-report data the base case? Surely the administrative records are more accurate? The authors justify this on the grounds that the self-report figures are lower generating a more conservative analysis. However, the analysis using the administrative data generates greater uncertainty (higher p-values). Hence one might regard the analysis based on administrative data as more conservative. I think the study would benefit from the reporting of CEACs for the analyses from the perspective of the employee, the health sector, and from society. In case I would include the QALY gain. The CEACs can easily be generated by valuing the incremental QALY gain using the range of zero (or 20,000 Euros) to 80,000 Euros per QALY and plotting the proportion of bootstrap replicates generating a positive NMB for ECO across the range of values. Minor points The last sentence on P6 is unclear. I think you mean end of follow-up when you refer to drop-out in this context. Lines 31-32 on P7 should be '...to which all OPs were required to adhere.' Lines 46-48 on P8 are unclear. I think this refers to the value of time costs for the employee. Time was valued at 12.96 Euros for each hour of engagement by the employee in health care programmes which displaced domestic chores onto other members of the employee's family? Line 37 on P11. I think 'not statistically significant' is clearer terminology than 'statistically insignificant'. Line 7 on P14. You report that incremental net benefit from the employee's perspective is surrounded by uncertainty. In fact, the INB is highly significantly different from zero (p=0.001). Hence while there is uncertainty around the magnitude of the INB there is very little uncertainty regarding the cost-effective of the intervention from the employee's perspective. You report a number of results from various sensitivity analyses for the different perspectives. I think it would be very helpful to tabulate all the results in one table to allow ready comparison. You could tabule Net benefit with the 95% CIs in the same cell and columns for the different perspectives and a row for the base case and each of the sensitivity analyses. Line 31 on P15. 'Benefits largely stemmed from reduced absenteeism...' I would bullet the results from the societal persective in the principal findings section. Line 53 on P15. '...cost data are often non-normally distributed...' P16 You conclude as a limitation that there is much uncertainty in the results. I suspect that once the QALY gains are included and valued across the range of zero to 80,000 Euros there will be much less uncertainty.

	Again, to reiterate my earlier point I don't think your analysis based on self reported absenteeism is conservative since it shows less uncertainty with regard to the likelihood the INB is positive than the analysis based on administrative data. First line on P17. '...were sick-listed for between 10 weeks and 2 years.' Line 38 on P17. You might want to clarify that the return on investment in Noben et al. was 11 Euros per Euro spent. Line 40 on P17. You refer to your own study as a preventive intervention - is this really what you meant?

REVIEWER	Kim Sweeny Victoria Institute of Strategic Economic Studies Victoria University Australia
REVIEW RETURNED	22-Mar-2017

GENERAL COMMENTS	This is a well designed and executed piece of research in an area in which there are only a few other studies. It makes a solid contribution to the literature, especially as it presents estimates of returns on investment demonstrating the benefits of ECO under quite conservative assumptions. These estimates are in line with those for ROI in mental health and other areas of health. Can I suggest that there be a short description of the role of the occupational physician in the Dutch health system and the difference between an OP and a primary care physician. On a more substantive note there needs to be discussion of why the researchers obtained negative results for presenteeism in Tables 2 and 3, especially in the light of the improvement in quality of life which would suggest that employees might be happier under ECO than CAU. Also the literature suggests that presenteeism is a larger problem than absenteeism in terms of economic costs so the authors might like to comment on this. There are a few very minor grammatical mistakes which could be corrected in the final paper.

REVIEWER	William S. Shaw Liberty Mutual Research Institute for Safety
REVIEW RETURNED	27-Mar-2017

GENERAL COMMENTS	This manuscript describes a series of cost-benefit analyses for a cluster-randomized trial to improve return-to-work rates among sick-listed employees (n = 220) reporting depressive and anxious symptoms. The intervention consisted of a multi-module eHealth intervention (“Return@Work”) based on cognitive-behavioral principles combined with encouragement and support from an occupational physician who received regular automated updates from the on-line training program. The primary outcome data sources were self-reported health care use and days absent from work, the EuroQOL-5D-3L, and the TiC-P, which included a number of direct and indirect non-medical costs including estimates of productivity loss while back at work.

	An intention-to-treat analysis was used to compare intervention and usual care groups on total costs and cost savings. This was repeated using a variety of perspectives: employer, employee, health care payer, and overall societal perspective (considering all costs and cost savings). All outcomes were translated into Euros (including QALYs), and net benefits were estimated using bootstrapping techniques, and sensitivity analyses were conducted to assess the effect of missing data and outliers. The primary conclusion was a €3,187 advantage per employee from the employer perspective, a €3,537 advantage per employee at the societal level. However, confidence bands were large, and after trimming the 5% of cases with highest costs overall, the advantage from the societal perspective was €2,928 per employee. The authors conclude that the ECO intervention program offers good value for money for virtually all stakeholders, but wide confidence intervals require careful interpretation. Long-term disability for mental health disorders is a large and growing problem in many countries, and more methods of intervention are needed to help facilitate a sustainable return to work for employees who are on prolonged sickness absence due to mental health symptoms. This manuscript provides a detailed cost-effectiveness analyses for a cluster-randomized trial combining individual and provider-level approaches to improve coping and self-efficacy for returning to work. This is a low cost and highly feasible intervention strategy where even small individual gains may prove to be highly cost-effective when implemented on a large scale. Thus, this is an appropriate application of cost-benefit analyses that has potential relevance to a variety of stakeholders and policy-makers. Overall, the authors should be congratulated for a very clear and concise presentation of their methodology and findings. The analytic strategy also has a number of strengths: analyzing cost savings from a variety of stakeholder perspectives, imputation of missing data using a variety of methods, and sensitivity analyses to remove extreme outliers. This reviewer finds only minor criticisms and suggestions: 1. Abstract/Settings: Recommend rewording for clarity: “Occupational health care in the Netherlands. Occupational physicians (n = 62) clustered and randomized by region into an experimental and a control group”. 2. Abstract/Conclusions: The terminology “require careful interpretation” here in reference to confidence intervals is somewhat vague. What do these confidence intervals suggest about precision and generalizability of findings? 3. Methods/Participants (page 6): There should be some added rationale to explain why only employees from small and medium-sized companies were included and why the window of 4 to 26 weeks was chosen. 4. Methods/Participants (page 6): Some of the language in the manuscript implies that participating employees had a physician-diagnosed mental health “disorder, ” but this section states only that they self-reported a high level of depressive or anxious symptoms. It would be more accurate to use the terminology “common mental symptoms” rather than “common mental disorders” throughout the manuscript,

	as it's unclear whether or not employees actually met the DSM criteria for mood or somatization disorders. 5. Methods/Procedure (page 6, line 5): This line seems to suggest that consenting participants were again randomized to the 2 conditions, but actually it was the patient's provider who was randomized to one of the two groups. Please clarify wording. 6. Methods/Intervention (page 7, paragraph 2): The authors appropriately chose to follow an intent-to-treat analysis, but it would still be informative to know what level of dose was actually delivered for the eHealth intervention. To what extent did participants complete the modules? This has some relevance for interpretation of cost-savings. 7. Methods/Resource use and costing (page 8, paragraph 2): More details are needed to explain how questions on the TiC-P were translated into an estimate of productivity loss. 8. Discussion (page 18, paragraph 1): A more detailed discussion is needed to interpret the wide confidence intervals from the study. What does this mean in terms of generalizability of results, and how can this lack of precision be dealt with in future studies of this type?

VERSION 1 – AUTHOR RESPONSE

REVIEWER: 1

Reviewer Name: Mark Pennington

Institution and Country: King's Health Economics, King's College London, UK Competing Interests: None declared

The paper provides a thorough economic analysis of an online module designed to help employees on sick leave with mental health problems return to work. In general the analysis appears to be robust. The exclusion of health benefits in the analysis from the health sector perspective is puzzling. Advanced statistical techniques have been applied to quantify the impact of uncertainty on the results. The results of this analysis might most usefully be conveyed through the production of Cost-effectiveness Acceptability Curves (CEACs) where the health benefits are included.

The authors undertake analysis from four different perspectives: employee, employer, health care sector, and societal. Each is reported as a cost benefit analysis. However, the health benefits of the intervention are only included in the analysis from the employee's perspective. It's difficult to see why the health benefits would be excluded from analysis taking either a health sector perspective or a societal perspective. In my view both analyses should include the health benefits. Such analysis might be considered a cost-benefit analysis if the health benefits are monetised. In fact the health benefits are measured in QALYs and a conservative value of 20,000 Euros per QALY is assumed. In my view it would be more correct to consider this a cost-utility analysis. Simply monetising QALY gains at 20,000 Euros does not convey the impact of different values one might place on a QALY gain on the results of the analysis. It would be more useful to decision makers to provide the results as a CEAC in which the likelihood the intervention is cost-effective is conveyed as a function of the value placed on a unit gain in health (QALY).

The analysis from the employer's perspective excludes health benefits which seems reasonable. I would regard the impact of presenteeism and absenteeism as a cost (averted) and not a benefit of the intervention. This is perhaps a semantic point but this renders this analysis a cost-minimisation

analysis. Nothing wrong with that, but I don't think the analysis from the employer's perspective is a true CBA.
The authors are to be commended in their use of robust techniques to deal with the missing data alongside bootstrapping to capture the sampling uncertainty. A little more methodological detail would be useful. Did the authors bootstrap prior to fitting the EM models?
There was one aspect of the analysis that troubled me. The authors choose to use self-reported data for absence rather than administrative records. They acknowledge substantial differences between the two sources. Why make the self-report data the base case? Surely the administrative records are more accurate? The authors justify this on the grounds that the self-report figures are lower generating a more conservative analysis. However, the analysis using the administrative data generates greater uncertainty (higher p-values). Hence one might regard the analysis based on administrative data as more conservative.
I think the study would benefit from the reporting of CEACs for the analyses from the perspective of the employee, the health sector, and from society. In case I would include the QALY gain. The CEACs can easily be generated by valuing the incremental QALY gain using the range of zero (or 20,000 Euros) to 80,000 Euros per QALY and plotting the proportion of bootstrap replicates generating a positive NMB for ECO across the range of values.

Comment 1. It's difficult to see why the health benefits would be excluded from analysis taking either a health sector perspective or a societal perspective.

REPLY: We do understand that excluding health benefits (the economic value that could be placed on gaining one QALY) in a health-economic analysis is puzzling when taking the health care perspective. However, these analyses were conducted from a 'health care payer's perspective' (in the Dutch context: the perspective of health care insurance companies). Had we used the health care sector perspective then we'd have fully agreed with the reviewer that the inclusion of health benefits would have been appropriate. However, in this study we conducted an investment appraisal from different stakeholders' perspectives. In this context, we did not compute the net-benefits from a general health care sector perspective, but specifically from the perspective of the health care insurer. In our opinion QALY gains do not directly benefit a health care insurer. To drive this point home, we consistently used the phrase health care payer's perspective, which in the Dutch context is the same as saying the health care insurance perspective.

That said, we agree with the reviewer that it is consistent to include the QALY gains into the analysis that were conducted from the societal perspective, instead of merely focussing on hard outcomes such as tangible Euros. To that end, we conducted extra analyses and replaced the outcomes of the societal perspective with the new values that now include the monetary value of gaining QALYs.

Comment 2. In my view it would be more correct to consider this a cost-utility analysis. Simply monetising QALY gains at 20,000 Euros does not convey the impact of different values one might place on a QALY gain on the results of the analysis. It would be more useful to decision makers to provide the results as a CEAC in which the likelihood the intervention is cost-effective is conveyed as a function of the value placed on a unit gain in health (QALY).

REPLY: We thank the reviewer for his remark. In our analyses we made investment appraisals from the stakeholder's perspectives. We included QALY gains, but only in our analysis from the employee/patient perspective (and the all encompassing societal perspective), and not in other perspectives, such as the employer's perspective. Hence we cannot find a justification in calling our investment appraisal a cost-utility analysis.
We prefer to compute benefits under conservative assumptions. If we place too much value on the QALY gains (i.e. €50,000 or €80,000) then we would probably portray a too rosy picture, even to the point where the QALY gains will dominate the outcomes of the investment appraisals.

Comment 3. I would regard the impact of presenteeism and absenteeism as a cost (averted) and not a benefit of the intervention. This is perhaps a semantic point but this renders this analysis a cost-minimisation analysis

REPLY: We disagree politely. In a cost-minimisation analysis two interventions are compared that are (assumed to be) equally effective and one simply choses the intervention that is associated with the fewest costs. However, the ECO-intervention in our study is more effective than care as usual and our health-economic evaluation was conducted alongside a superiority trial (not a non-inferiority trial).

Comment 4. A little more methodological detail would be useful. Did the authors bootstrap prior to fitting the EM models?

REPLY: We conducted non-parametric bootstrapping after fitting the EM models. We clarified this in the 'analyses' section (page 10).

Comment 5. Why make the self-report data the base case? Surely the administrative records are more accurate? The authors justify this on the grounds that the self-report figures are lower generating a more conservative analysis. However, the analysis using the administrative data generates greater uncertainty (higher p-values). Hence one might regard the analysis based on administrative data as more conservative.

REPLY: We agree with the reviewer that administrative data about absenteeism are likely to be more accurate than self-reported data. However, we did not have access to administrative data for absenteeism 3, 6 and 9 months after baseline, only for the period of one calendar year. Please note, the administrative data that were based on one-year periods are not helpful for constructing fine graded cash flows (in one-month cycles) that were required for our return on investment analyses. An additional advantage of the self-reported data over the administrative data was that data about presenteeism were only available from self-reports and we preferred to base the analyses of absenteeism and presenteeism on the same data source.

MINOR POINTS

Comment 1. The last sentence on P6 is unclear. I think you mean end of follow-up when you refer to drop-out in this context.

REPLY: We thank the reviewer for pointing out the lack of clarity. We clarified the last sentence of the 'procedure' section of the revised manuscript (page 6, line 35-36).

Comment 2. Lines 31-32 on P7 should be '...to which all OPs were required to adhere.'

REPLY: We thank the reviewer for the comment. The sentence has been corrected in the revised manuscript accordingly (page 7, line 22).

Comment 3. Lines 46-48 on P8 are unclear. I think this refers to the value of time costs for the employee. Time was valued at 12.96 Euros for each hour of engagement by the employee in health care programmes which displaced domestic chores onto other members of the employee's family?

REPLY: These costs refer to the number of hours that (informal) caregivers spent on taking over cleaning activities and running domestic errands. We rephrased the lines in the manuscript to clarify the meaning of these costs (page 8, lines 31-33).

Comment 4. Line 37 on P11. I think 'not statistically significant' is clearer terminology than 'statistically insignificant'.

REPLY: We changed the sentence in the manuscript accordingly (page 11, line 7).

Comment 5. Line 7 on P14. You report that incremental net benefit from the employee's perspective is surrounded by uncertainty. In fact, the INB is highly significantly different from zero (p=0.001). Hence while there is uncertainty around the magnitude of the INB there is very little uncertainty regarding the cost-effectiveness of the intervention from the employee's perspective.

REPLY: The reviewer correctly noticed that there is no uncertainty about the significance of the INB from the employee’s perspective, but only about the magnitude of the INB. We rephrased this sentence accordingly (page 14).

Comment 6. You report a number of results from various sensitivity analyses for the different perspectives. I think it would be very helpful to tabulate all the results in one table to allow ready comparison. You could tabulate Net benefit with the 95% CIs in the same cell and columns for the different perspectives and a row for the base case and each of the sensitivity analyses.

REPLY: As suggested by the reviewer, we added a table presenting the results of the base case and the sensitivity analyses (page 15).

Comment 7. Line 31 on P15. 'Benefits largely stemmed from reduced absenteeism...'

REPLY: We thank the reviewer for pointing the grammatical error to us. We corrected it in the revised manuscript accordingly (page 15, line 28).

Comment 8. I would bullet the results from the societal perspective in the principal findings section.

REPLY: As suggested by the reviewer we bulleted the results from the societal perspective in the ‘principal findings’ section in the manuscript (page 16).

Comment 9. Line 53 on P15. '...cost data are often non-normally distributed...'

REPLY: We thank the reviewer and corrected the typo in the revised manuscript.

Comment 10. P16 You conclude as a limitation that there is much uncertainty in the results. I suspect that once the QALY gains are included and valued across the range of zero to 80,000 Euros there will be much less uncertainty.

REPLY: As indicated in an earlier remark, we did not place a higher monetary value than €20,000 on gaining one QALY. In doing so we adhere to the Dutch guideline for costing in health-economic evaluations and present results that were computed under conservative assumptions.

Comment 11. Again, to reiterate my earlier point I don't think your analysis based on self reported absenteeism is conservative since it shows less uncertainty with regard to the likelihood the INB is positive than the analysis based on administrative data.

REPLY: We agree with the reviewer that the uncertainty of the INB being positive is larger in the analysis based on the administrative data. What we tried to explain is that the analysis based on self-reported data shows a lower point estimate for the INB from a societal perspective than the analysis based on administrative data. We changed the sentence about uncertainty in the manuscript for greater clarity (page 14, line 31).

Comment 12. First line on P17. '...were sick-listed for between 10 weeks and 2 years.'

REPLY: We thank the reviewer and changed the sentence accordingly in the revised manuscript (page 17, line 18).

Comment 13. Line 38 on P17. You might want to clarify that the return on investment in Noben et al. was 11 Euros per Euro spent.

REPLY: We thank the reviewer and added ‘per Euro spent’ to the sentence (page 18, line 8).

Comment 14. Line 40 on P17. You refer to your own study as a preventive intervention - is this really what you meant?

REPLY: The reviewer correctly noticed that the term ‘preventive’ should not have been used. We deleted this adjective (page 18).

REVIEWER: 2
Reviewer Name: Kim Sweeny
Institution and Country: Victoria Institute of Strategic Economic Studies, Victoria University, Australia
Competing Interests: None declared

This is a well-designed and executed piece of research in an area in which there are only a few other studies. It makes a solid contribution to the literature, especially as it presents estimates of returns on investment demonstrating the benefits of ECO under quite conservative assumptions. These estimates are in line with those for ROI in mental health and other areas of health.

Comment 1. Can I suggest that there be a short description of the role of the occupational physician in the Dutch health system and the difference between an OP and a primary care physician

REPLY: We thank the reviewer the suggestion and we added a description of the OPs role in the Dutch health care system to the introduction of the revised manuscript (page 4).

Comment 2. On a more substantive note there needs to be discussion of why the researchers obtained negative results for presenteeism in Tables 2 and 3, especially in the light of the improvement in quality of life which would suggest that employees might be happier under ECO than CAU. Also the literature suggests that presenteeism is a larger problem than absenteeism in terms of economic costs so the authors might like to comment on this.

REPLY: The reviewer raises interesting points and we agree that attention should be paid in the discussion to why we obtained negative results for presenteeism (in the short-run, but not in the long-run) and why in our study absenteeism was a larger problem than presenteeism. We have added possible explanations to the Discussion:

“However, other than Noben and colleagues and several other studies we found negative results for presenteeism in the short run (first nine months), but these were alleviated in the longer run (at the end of the year). An explanation for the initially negative results on presenteeism might be that employees who returned to work early were not completely fit and as productive as normally. In other words there was an initial trade-off between reduced absenteeism and increased presenteeism. However, after the first nine months the additional costs caused by presenteeism ceased to exist and were reversed into benefits. This change is possibly driven by an improvement in quality of life when people work.”

“The literature suggests that in terms of economic costs presenteeism is a larger problem than absenteeism. Our results are not in line with these findings. This could be due to the Dutch system in which employees receive a substantial percentage of their wage during the first two years of their illness. In many other countries the fall in income is more acute when employees stay absent from their work, increasing the incentive to keep on working – even when work is then associated with greater levels of presenteeism.”

Comment 3. There are a few very minor grammatical mistakes which could be corrected in the final paper.

REPLY: We thank the reviewer for the remark. We corrected typos and grammatical mistakes in the manuscript.

REVIEWER: 3
Reviewer Name: William S. Shaw
Institution and Country: Liberty Mutual Research Institute for Safety, USA Competing Interests: None declared

This manuscript describes a series of cost-benefit analyses for a cluster-randomized trial to improve return-to-work rates among sick-listed employees (n = 220) reporting depressive and anxious symptoms. The intervention consisted of a multi-module eHealth intervention (“Return@Work”) based on cognitive-behavioral principles combined with encouragement and support from an occupational physician who received regular automated updates from the on-line training program. The primary outcome data sources were self-reported health care use and days absent from work, the EuroQOL-5D-3L, and the TiC-P, which included a number of direct and indirect non-medical costs including estimates of productivity loss while back at work. An intention-to-treat analysis was used to compare intervention and usual care groups on total costs and cost savings. This was repeated using a variety of perspectives: employer, employee, health care payer, and overall societal perspective (considering all costs and cost savings). All outcomes were translated into Euros (including QALYs), and net benefits were estimated using bootstrapping techniques, and sensitivity analyses were conducted to assess the effect of missing data and outliers. The primary conclusion was a €3,187 advantage per employee from the employer perspective, a €3,537 advantage per employee at the societal level. However, confidence bands were large, and after trimming the 5% of cases with highest costs overall, the advantage from the societal perspective was €2,928 per employee. The authors conclude that the ECO intervention program offers good value for money for virtually all stakeholders, but wide confidence intervals require careful interpretation.

Long-term disability for mental health disorders is a large and growing problem in many countries, and more methods of intervention are needed to help facilitate a sustainable return to work for employees who are on prolonged sickness absence due to mental health symptoms. This manuscript provides a detailed cost-effectiveness analyses for a cluster-randomized trial combining individual and provider-level approaches to improve coping and self-efficacy for returning to work.

This is a low cost and highly feasible intervention strategy where even small individual gains may prove to be highly cost-effective when implemented on a large scale. Thus, this is an appropriate application of cost-benefit analyses that has potential relevance to a variety of stakeholders and policy-makers.

Overall, the authors should be congratulated for a very clear and concise presentation of their methodology and findings. The analytic strategy also has a number of strengths: analyzing cost savings from a variety of stakeholder perspectives, imputation of missing data using a variety of methods, and sensitivity analyses to remove extreme outliers. This reviewer finds only minor criticisms and suggestions:

Comment 1. Abstract/Settings: Recommend rewording for clarity: “Occupational health care in the Netherlands. Occupational physicians (n = 62) clustered and randomized by region into an experimental and a control group”.

REPLY: We thank the reviewer for the recommendation and clarified the sentence in the manuscript as suggested (page 3).

Comment 2. Abstract/Conclusions: The terminology “require careful interpretation” here in reference to confidence intervals is somewhat vague. What do these confidence intervals suggest about precision and generalizability of findings?

REPLY: We rephrased the sentence and now say: “the sometimes wide 95% confidence intervals suggest that the costs and benefits are not always very precise estimates and real benefits could vary considerably.” (page 3).

Comment 3. Methods/Participants (page 6): There should be some added rationale to explain why only employees from small and medium-sized companies were included and why the window of 4 to 26 weeks was chosen.

REPLY: There was a minimum of 4 weeks of sickness absence to avoid including patients with ‘spontaneous’ recovery. A maximum of 26 weeks of sickness absence duration was chosen because the probability of return to work reduces in case of longer absence (Henderson, 2005). The choice for small and medium-sized companies was made for practical reasons. The study was conducted in collaboration with Arbo Vitale and GGzBreburg. Arbo Vitale is a large occupational health service, offering employers insurance for the costs of sick leave and sickness guidance. Arbo Vitale observed a lot of costs due to absenteeism among employees of small-sized to medium-sized companies, which was the rationale to focus on this target group in the study.

Comment 4. Methods/Participants (page 6): Some of the language in the manuscript implies that participating employees had a physician-diagnosed mental health “disorder, ” but this section states only that they self-reported a high level of depressive or anxious symptoms. It would be more accurate to use the terminology “common mental symptoms” rather than “common mental disorders” throughout the manuscript, as it’s unclear whether or not employees actually met the DSM criteria for mood or somatization disorders.

REPLY: We thank the reviewer for raising this issue. We do understand that the use of the term disorder can be confusing as there was no clinical interview at baseline to assess diagnostic status. However, to be eligible for participation the employees did have to score ≥10 on either the depression or the somatization scale of the Patient Health Questionnaire (PHQ-9) or the Generalized Anxiety Disorder questionnaire (GAD-7). These cut-off scores usually indicate the presence of a disorder.

Hence, it is most likely that the majority of the participants did have a disorder, but the reviewer is right that we cannot be 100% sure if they were meeting the diagnostic criteria for DSM disorders. Therefore, we changed ‘common mental disorder’ into ‘(symptoms of) common mental disorder’.

Comment 5. Methods/Procedure (page 6, line 5): This line seems to suggest that consenting participants were again randomized to the 2 conditions, but actually it was the patient’s provider who was randomized to one of the two groups. Please clarify wording.

REPLY: We thank the reviewer for pointing this out. We rephrased the sentence in the ‘procedure’ section on page 6 of the manuscript to clarify.

Comment 6. Methods/Intervention (page 7, paragraph 2): The authors appropriately chose to follow an intent-to-treat analysis, but it would still be informative to know what level of dose was actually delivered for the eHealth intervention. To what extent did participants complete the modules? This has some relevance for interpretation of cost-savings.

REPLY: We added a description about the intervention adherence of the ECO intervention by the employees to the revised manuscript (page 11).

Comment 7. Methods/Resource use and costing (page 8, paragraph 2): More details are needed to explain how questions on the TiC-P were translated into an estimate of productivity loss.

REPLY: As indicated in the manuscript, days not worked owing to absenteeism were costed using the average gross gender- and age-specific daily wages (as per the Dutch guideline for costing in health-economic evaluations). For presenteeism we took a similar approach, but days not worked were now estimated by multiplying a self-reported inefficiency score (ranging between 0 and 1) with the number of days worked inefficiently. To illustrate, if an employee reports and inefficiency score of 0.50 for 7 working days then we assume that 3.5 working days have been lost due to inefficiency. We added this illustration to the manuscript (page 9, lines 9-10).

Comment 8. Discussion (page 18, paragraph 1): A more detailed discussion is needed to interpret the wide confidence intervals from the study. What does this mean in terms of generalizability of results, and how can this lack of precision be dealt with in future studies of this type?

REPLY: As noted, some 95% confidence intervals of the net-benefits are wide. By implication, one should not rely too much on the point estimates of net-benefits, return on investment ratios, break-even points, because they lack precision. In other words, our estimates, although conservative, have some degree of uncertainty and are therefore no substitute for one’s own business judgement.

REFERENCES
Salomon JA, Vos T, Hogan DR, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. Lancet. 2012;380:2129–43.

VERSION 2 – REVIEW

REVIEWER	Mark Pennington Senior lecturer health economics, King's Health Economics King's College London UK
REVIEW RETURNED	23-May-2017

GENERAL COMMENTS	The authors have addressed some although not all of my criticisms. Most notably the analysis from the health service perspective remains a cost-minimisation analysis with health benefits excluded, and the valuation of health benefits is undertaken with non-standard methodology. My view remains that the study would have been strengthened by the inclusion of Cost-Effectiveness Acceptability Curves to capture the uncertainty in cost-effectiveness of the intervention across a range of values for a QALY. Such analysis would have been relevant for the patient, health care and societal perspectives. That said, the methodology applied is valid and clearly reported. I have a number of minor points that the authors should probably address: First bullet of 'strengths and limitations' - drop the word 'only' Last paragraph of Intro - in my view the analyses from the health sector and employers perspectives are cost-minimisation rather then cost-benefit analyses as benefits of the intervention have not been included. I would regard the impact on presenteeism and absenteeism as productivity costs. This crops up again in the second paragraph under 'analyses' - duration of sick leave, as valued in your analysis, is a productivity cost and not a monetised benefit (as the QALY gain is). First para of methods - randomisation is at the area level not the level of the occupational physician Loss to follow-up in results - you mention that 40% of those assigned to the intervention completed half of the modules. How many completed all of them? Results from the health sector perspective - you conclude that ECO does not save the health sector money. It's worth noting that if the health sector values patient health outcomes (which it almost certainly does!) then the intervention may be cost-effective at a very modest willingness to pay for a QALY gain. Limitations, first bullet - you note the wide confidence intervals and conclude that a larger study might be needed. In fact, your study failed to recruit the planned 360 participants. Is a larger study feasible? I wonder if it is really necessary - it may well be that there is a high likelihood the intervention is cost-effective at a value of 50,000 euros per QALY. Perhaps all that is needed is a full cost-utility analysis of the data you have already collected. Third bullet - it seems highly likely that adjusting for baseline differences would have enhanced the cost-effectiveness of ECO, but I don't think you can conclude that with certainty. Finally, the last line of the manuscript jarred with me. Have the courage of your convictions! Your analysis is subject to considerable uncertainty, and you explore this in some depth. However, I find your analysis far more convincing than simply relying on 'one's own business judgement.'

	There was no author rebuttal in the manuscript or supplementary files as far as I could tell. Has this been erroneously excluded? I regard it as primarily the editor's job to assess whether the rebuttal addresses criticisms not addressed in the revised manuscript and I will leave it to the editor to assess that.

REVIEWER	Kim Sweeny Victoria University Australia
REVIEW RETURNED	05-Jun-2017

GENERAL COMMENTS	The authors have addressed my concerns adequately.

REVIEWER	William Shaw Liberty Mutual Research Institute for Safety, USA
REVIEW RETURNED	16-May-2017

GENERAL COMMENTS	All of this reviewer's concerns have been adequately addressed or incorporated. Thank you for making revisions and clarifications to the manuscript.

VERSION 2 – AUTHOR RESPONSE

REVIEWER: 1
Reviewer Name: Mark Pennington
Institution and Country: King's Health Economics, King's College London, UK Competing Interests: None declared

Comment 1. The authors have addressed some although not all of my criticisms. Most notably the analysis from the health service perspective remains a cost-minimisation analysis with health benefits excluded, and the valuation of health benefits is undertaken with non-standard methodology. My view remains that the study would have been strengthened by the inclusion of Cost-Effectiveness Acceptability Curves to capture the uncertainty in cost-effectiveness of the intervention across a range of values for a QALY. Such analysis would have been relevant for the patient, health care and societal perspectives. That said, the methodology applied is valid and clearly reported.

REPLY: We can understand that the reviewer addresses his former points again, because our former reply was not sent to him. In our former reply we addressed all of the criticisms, including this first point of the reviewer. We copied and pasted a part of the reply here. The main point here is that we made a conscious choice not to conduct a (standard) cost-utility analysis (as we have done many times before), but on this occasion we wanted to carry out a trial-based investment appraisal. Presenting an investment appraisal / business case for implementing a return-to-work intervention was deemed to be more appealing to e.g. employers and other stakeholders. It is worth noting that investment appraisals rely on a methodology which is somewhat different from the standard cost-utility analysis. Now we are happy to see that Reviewer 1 finds the (investment appraisal) methodology that we applied both valid and clearly reported.

Comment: It's difficult to see why the health benefits would be excluded from analysis taking either a health sector perspective or a societal perspective.

REPLY: We do understand that excluding health benefits (the economic value that could be placed on gaining one QALY) in a health-economic analysis is puzzling when taking the health care perspective. However, these analyses were conducted from a 'health care payer's perspective' (in the Dutch context: the perspective of health care insurance companies). Had we used the health care sector perspective then we'd have fully agreed with the reviewer that the inclusion of health benefits would have been appropriate. However, in this study we conducted an investment appraisal from different stakeholders' perspectives. In this context, we did not compute the net-benefits from a general health care sector perspective, but specifically from the perspective of the health care insurer (payer). In our opinion QALY gains do not directly benefit a health care insurer. To drive this point home, we consistently used the phrase health care payer's perspective, which in the Dutch context is the same as the health care insurance perspective.

That said, we agree with the reviewer that it is consistent to include the QALY gains into the analysis that were conducted from the societal perspective, instead of merely focussing on hard outcomes such as tangible Euros. To that end, we conducted a new analyses and replaced the outcomes of the societal perspective with the new values that now include the monetary value of gaining QALYs.

Comment: In my view it would be more correct to consider this a cost-utility analysis. Simply monetising QALY gains at 20,000 Euros does not convey the impact of different values one might place on a QALY gain on the results of the analysis. It would be more useful to decision makers to provide the results as a CEAC in which the likelihood the intervention is cost-effective is conveyed as a function of the value placed on a unit gain in health (QALY).

REPLY: We thank the reviewer for his remark. In our analyses we made investment appraisals from the stakeholder's perspectives. We included QALY gains, but only in our analysis from the employee/patient perspective (and the all-encompassing societal perspective), and not in other perspectives, such as the employer's perspective. Hence, no cost-utility analysis (with QALYs included) for some of the perspectives.

Comment: I would regard the impact of presenteeism and absenteeism as a cost (averted) and not a benefit of the intervention. This is perhaps a semantic point but this renders this analysis a cost-minimisation analysis

REPLY: We disagree politely. In a cost-minimisation analysis two interventions are compared that are (assumed to be) equally effective and one simply choses the intervention that is associated with the fewest costs. However, the ECO-intervention in our study is more effective than care as usual and our health-economic evaluation was conducted alongside a superiority trial (not a non-inferiority trial).

Comment 2. First bullet of 'strengths and limitations' - drop the word 'only'

REPLY: We thank the reviewer for the comment. The sentence has been corrected in the revised manuscript accordingly.

Comment 3. Last paragraph of Intro - in my view the analyses from the health sector and employers perspectives are cost-minimisation rather then cost-benefit analyses as benefits of the intervention have not been included. I would regard the impact on presenteeism and absenteeism as productivity costs. This crops up again in the second paragraph under 'analyses' - duration of sick leave, as valued in your analysis, is a productivity cost and not a monetised benefit (as the QALY gain is).

REPLY: We addressed this point in our former reply, as copied and pasted in the reply to the first comment of the reviewer.

Comment 4. First para of methods - randomisation is at the area level not the level of the occupational physician.

REPLY: Good point. We changed the sentence by adding the term ‘area’ in the manuscript.

Comment 5. Loss to follow-up in results - you mention that 40% of those assigned to the intervention completed half of the modules. How many completed all of them?

REPLY: We added an additional sentence about the percentage of the participants that completed at least 70% of the prescribed sessions.

Comment 6. Results from the health sector perspective - you conclude that ECO does not save the health sector money. It's worth noting that if the health sector values patient health outcomes (which it almost certainly does!) then the intervention may be cost-effective at a very modest willingness to pay for a QALY gain.

REPLY: We addressed this point in our former reply. The main point here is that we are not looking at a “health care” perspective, but a “health care payer” (insurance) perspective.

Comment 7. Limitations, first bullet - you note the wide confidence intervals and conclude that a larger study might be needed. In fact, your study failed to recruit the planned 360 participants. Is a larger study feasible? I wonder if it is really necessary - it may well be that there is a high likelihood the intervention is cost-effective at a value of 50,000 euros per QALY. Perhaps all that is needed is a full cost-utility analysis of the data you have already collected.

REPLY: We thank the reviewer for his remark. The trial did succeed to recruit the required number of participants. The correct number of planned participants was 200, as mentioned in the design article of Volker and colleagues (2013). The former number of 360 in the research protocol was based on separate analyses of participants with different complaints (i.e.: depression, anxiety and somatization).

With regard to the valuation of the QALY, we addressed this point in our former reply to the revised manuscript. However, this was not included for the reviewer, so we will paste our reply:

We prefer to compute benefits under conservative assumptions. If we place too much value on the QALY gains (i.e. €50,000 or €80,000) then we would probably portray a too rosy picture, even to the point where the QALY gains will dominate the outcomes of the investment appraisals.

Comment 8. Third bullet - it seems highly likely that adjusting for baseline differences would have enhanced the cost-effectiveness of ECO, but I don't think you can conclude that with certainty.

REPLY: We adapted the sentence by adding ‘most likely’ to the sentence.

Comment 9. Finally, the last line of the manuscript jarred with me. Have the courage of your convictions! Your analysis is subject to considerable uncertainty, and you explore this in some depth. However, I find your analysis far more convincing than simply relying on one's own business judgement.
REPLY: We agree with the reviewer and deleted the part of the sentence about relying on one's own business judgement.