PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Quality of life and added value of a tailored palliative care intervention in patients with soft tissue sarcoma undergoing treatment with trabectedin: A multicenter, cluster-randomized trial within the Germany Interdisciplinary Sarcoma Group (GISG)
AUTHORS	Hentschel, Leopold; Richter, Stephan; Kopp, Hans-Georg; Kasper, Bernd; Kunitz, Annegret; Grünwald, Viktor; Kessler, Torsten; Chemnitz, Jens Marcus; Pelzer, Uwe; Schuler, Ulrich; Freitag, Janet; Schilling, Andrea; Hornemann, Beate; Arndt, Karin; Bornhaeuser, Martin; Schuler, Markus Kajo

VERSION 1 – REVIEW

REVIEWER	Robin L Jones and Eugenie Younger Royal Marsden and Institute of Cancer Research Robin L Jones has been a consultant for: Adaptimmune Blueprint Clinigen Eisai Epizyme Daichii Deciphera Immunedesign Johnson and Johnson Lilly Merck Pfizer Pharmamar Upto Date
REVIEW RETURNED	11-Dec-2019

GENERAL COMMENTS	I reviewed this manuscript with Dr Eugenie Younger. This is an important study worthy of publication. The following suggestions/ comments will it is hoped improve this excellent manuscript. This manuscript provides results of the randomized 'Yonlife' study, which evaluated the effect of a tailored palliative care intervention on QOL in patients with advanced soft tissue sarcoma (STS) treated with Trabectedin. Although this was the primary objective, the manuscript also focused on the feasibility of collecting QOL data using ePROs, and impact of the palliative care intervention on other PRO measures such as anxiety and depression and pain severity. This study is one of the first that has used electronic patient reported outcomes (ePROs) to collect quality of life (QOL) data in patients with advanced STS treated with chemotherapy. As the authors state,

	patients with advanced STS have limited effective therapeutic options, and treatment decisions involve balancing efficacy, impact on QOL and potential side effects. These data therefore add important knowledge to aid understanding of QOL in patients with sarcoma, particularly as research in this area lags behind studies in patients with common cancers. Although the data are interesting and worth publishing, a major revision is required to address the following issues. Major concerns Comment 1 I would recommend that the authors ask a native English speaker, who is proficient in proof reading scientific manuscripts, to edit the manuscript for English content. Comment 2 Table 1 give details of the full analysis set (n=78), however the authors have said that analysis of efficacy endpoints will only be performed on the per-protocol set (PPS). How many patients were in the PPS? What were the demographic and clinical details of the patients in the PPS. It is important to know which patients were remaining after drop out (PRO drop out and chemotherapy drop out). Figure 1 legend shows at time point one (v1), there were only 19 patients in the IC and 14 patients in the CC. There is no explanation as to why so few patients provide PRO data at baseline. Which patients dropped out? Why was the response rate low? The abstract does not give the details of the number of patients in the study. All of the tables should include the number of patients evaluable at each time point as the interpretation is based on how many patients completed the PRO questionnaires at each time point. Comment 3 The method section states that patients were recruited just after starting Trabectedin or currently under treatment with Trabectedin. The table shows that some patients had received up to 17 cycles of Trabectedin before starting the study. Could the authors provide rationale for including patients starting Trabectedin, with those who have been on treatment for many months? Comment 4 Table 2 shows a beneficial interventional trend on FACT-G total score and most FACT-G functional domains, but a potentially adverse trend with regard to anxiety, pain and symptom severity. The authors have stated that these are not significant referring to the p-values, but have not discussed the possible reasons for these findings. The p-values for the FACT-G scores were not significant either (looking at the p-values), however the authors have focused on the possible beneficial effect of the intervention using Cohen's d-test, while not discussing these other findings e.g. perhaps the questionnaires were not sensitive or specific for this population in this group. It is also surprising that palliative care intervention led to an adverse effect on pain severity scores. Please could the authors comment on these results.

	Comment 5 The abstract conclusion is focused on the feasibility of ePRO collection in a multicentre study, however this is not a new finding. This does not appear to be the main finding of the study. I would recommend revising the conclusion to include an opinion related to the primary objective i.e. what was the value of a palliative care intervention on QOL? Comment 6 The authors should put in context the development of trabectedin and cite the previous randomized Phase 3 trial leading to FDA approval, and previously published patient reported outcome data (Cancer 2019;125:4435-4441). Minor comments Comment 1 Page 2, Line 19. Please review the term 'supportive intervention'. As the reader, I feel that this is too vague and am left wondering what the intervention involves. Given that the assessment of the intervention on QOL is the main goal of the study, more detail is needed. Comment 2 Page 2, Line 24. Please provide more detail on participants of interest and how many will be recruited (if not described in the results) Comment 3 Page 2, Line 28. The authors have said that the intervention for the IC involves ePRO with four-step assessment for palliative care. This does not tell the reader what the intervention is, only that more assessments will take place in the IC. Please define the intervention here. Comment 4 Page 2, Line 40. The authors have not described the population at the start of the results section. How many patients took part? How many were in each treatment arm? What was the median age of patients? What line of treatment was Trabectedin for most patients? Although some of these data are in the table, it would be helpful to have a summary of the group in the text. Comment 5 Page 2, Line 47. Survival (in days) was almost double in the intervention cluster compared to the control group, however the p-value is not significant. This is surprising – please could the authors comment. Comment 6 Page 2, Line 51-53. The authors have described the influence of several factors (such as age, gender, depression and anxiety) on

	QOL, however this objective was not mentioned in the abstract. It is unusual to describe results without mentioning that this was going to be assessed. Outcomes which were described (such as impact of intervention on pain intensity, anxiety and depression) are not mentioned in the results section. Comment 7 Page 3, Line 6. The main conclusion should consider whether the supportive care/palliative care intervention had a beneficial impact on QOL in the intervention cluster (i.e. added 'value' as per the title). Although the trial does add knowledge about QOL in this patient group, that was not the main principal aim of the study. The authors have concluded that PROs can be used in a multi-centre study but this is not new knowledge. The authors should focus their conclusion on the principal findings of their study. Comment 8 Page 4, line 12. ePRO assessment is feasible in a multicentre study – this is not a key strength of this study as this has been shown before in multiple other studies (albeit in different tumour groups). Comment 9 Page 4, line 14-16. Please review this sentence as it is not clear. What is meant by "important outcomes of palliative care"? Do the authors mean that the intervention seems to improve certain quality of life outcomes? Comment 10 Page 4, line 19. The authors state that the sample size is a major limitation, however the number of patients had not yet been described in the abstract. Comment 11 Page 5, line 5. Please review the first sentence of the introduction. What do the authors mean by 'evolved in the past'? Should this say 'evolved over recent years'? It is not clear whether the authors think that toxicity and impact on QOL has increased over the last few years? Additionally, there is no reference for this sentence. Please rewrite this sentence so that it is more clear. Comment 12 Page 5, Lines 19-22. 'Mental problems' should be replaced with 'mental health problems'. Comment 13 Page 5, line 22. Paramount is not the correct English word to use in this context. Please review. Comment 14 Page 5, line 28. The authors say that drug have 'specific safety profiles'. Safety tends to refer to phase 1 drugs. Please review this sentence. Do the authors mean to refer to side effect profiles?

	Comment 15 Page 4, line 33. Please give an example where convenience can affect treatment decisions, such as inpatient versus outpatient regimen. Comment 16 Page 5, line 49. What do the authors mean by 'time to deterioration'? Please be more specific do you mean deterioration in symptoms or disease progression on imaging or declining performance status? Comment 17 Page 5, line 54. Please add more details about the beneficial outcomes, the settings and the entities. For example describe what beneficial outcomes have been achieved using PRO, in which tumour groups and using what interventions. This is important to put this research paper into context of existing research in this field. Comment 18 Page 6, line 12. A multi-professional expert panel is not an intervention. Please describe what the expert panel are doing i.e. the intervention that they are making. Comment 19 Page 7, line 8. Please define what is meant by 'just started' (e.g. just received the first cycle?). Comment 20 Page 7, line 10. It is important to state at which line of therapy these patients are receiving Trabectedin to understand more about the target population. Are these patients having Trabectedin as second line treatment (e.g after Doxorubicin), or are some of the patients having Trabectedin as first line treatment of metastatic disease? Comment 21 Page 7, line 20. The incorrect symbol has been used – should be ≤2. Comment 22 Page 7, lines 27-32. Please use the same tense throughout the methods section (past at beginning, then present tense in these sentences). Comment 23 Page 7, lines 40-45. Could the authors explain why patients and public were not involved in the design of the research study? Comment 24 Page 8, line 42. Why was this time point chosen for the primary

	outcome? Is this related to the timing of scans? Is this related to the number of cycles that were expected to be received by patients? Please provide rationale for the time point. Comment 25 Page 8, line 49. Please specify which functional domains the FACT-G measures rather than using the word 'several' – this is too vague and the functional domains are important measures in this study. Comment 26 Page 9, line 40. Please add detail on the exploratory outcome shown in the results section of the abstract regarding the impact of variables such as age and gender on QOL. This is mentioned later in the statistical considerations but important to include this as an outcome of the study. Comment 27 Page 11. Results. Please provide more details on the patient group. For example, how many patients had locally-advanced disease versus metastatic disease, how many previous lines of treatment had patients received, what prior treatments had patients received (e.g. anthracycline based)? Comment 28 (also major comment) Page 11. Results. The authors have provided a table for the full analysis set, however there is no detail on the per protocol set (PPS). How many patients were in the PPS and how many in each group? This is important if the analysis is only based on patients who completed V1 and V4 assessment and had treatment with Trabectedin throughout. Comment 29 (also major comment) Page 11. Table 1. It appears that at least one patient in the study had already received 17 cycles of trabectedin. This contradicts what is stated in the methods, i.e. patients were recruited when they had only just started treatment. Please could the authors clarify why this patient was included? Is it possible to compare the QOL of patients starting trabectedin with those who have been on treatment for such a long period of time? Comment 30 Page 11, Table 1. The median values in the row ' number of previous cycles of another chemotherapy' are surprising. Patients normally receive at least two cycles of a chemotherapy drug before having imaging to assess response. Please could the authors comment on why this is? Comment 31 (also major comment) Page 11, line 55. Does the primary outcome refer only to the PPS? If so, please could the authors describe the number of patients in the PPS to which these data refer. Comment 32 (also major comment)

	Page 12, Table 2. Please could the authors comments on why they think that the intervention may have had an adverse effect on anxiety, pain severity and interference, symptom severity and interference compared to the control cluster? The authors have focused on the potentially beneficial impact of the intervention but not discussed these adverse findings. Comment 33 (also major comment) Page 14. Lines 3-5. The authors state that there were no significant differences between the IC and CC with regard to other domains of PRO, but the p-values were not significant for any of the FACT-G domains. These show a possible adverse trend in of the intervention and therefore this should be discussed. Comment 34 Page 14. Table 4. The p-values are presented and are not significant, however it would be interesting to also define the effect size for these PRO domains as with the FACT-G domains. Comment 35 Page 14. Table 4. It is interesting that palliative care intervention did not make any difference to scores on MD Anderson symptom inventory – could the authors comment on why that might be? Could this be because this inventory is not sensitive enough to detect a difference? Comment 36 Page 16. Discussion. Please could the authors review the first paragraph – several sentences appear to be repetition of each other. Comment 37 Page 16 line 40. Limitations are usually described towards the end of the discussion. Please could the authors focus on the importance of their findings for patients and for clinical practice. Perhaps the FACT-G scores could be compared to scores in other studies looking at similar populations of patients with advanced cancer. Comment 38 Page 17, line 12. Please could the authors be more specific when they refer to 'several reports with different designs'. This is a very vague statement and examples could be provided to enhance the discussion. Comment 39 Page 17, lines 26-28. Please rewrite this sentence as the meaning is not clear. Comment 40 Page 17, conclusion. The conclusion is well written and important.

REVIEWER	Karin Ribi International Breast Cancer Study Group, Switzerland
REVIEW RETURNED	10-Jan-2020

GENERAL COMMENTS	This paper presents a cluster-randomized trial investigating the added value of a tailored palliative care intervention in patient population that has been studied less often compared to other cancer entities. The results are therefore of interest. However, some major information are lacking, which makes it difficult to appraise the methodological rigor. 1. Abstract There is no clear objective formulated in the abstract. The abstract states that patients were randomized, this is contradictory to the methods section p.8) stating that centers were randomized. The number of patients included in this study is missing. 2. Introduction The first two sentences need to be supported by references. The intervention is a palliative care intervention that can be applied across cancer types and treatments. A rational why this intervention is relevant for sarcoma patients and why the authors chose to only include patients receiving a specific treatment is lacking. Prediction of Qol is not mentioned as an objective. Please provide also a rational why the design of a cluster-randomized trial was chosen. 3. Methods: There are some doubts, if this study was a true cluster-randomized study: The level of randomization is not clear, patient or centers? Please clarify the randomization procedure. If centers were randomized the cluster size is 11 not 33. Also the authors used STROBE for cohort study checklist instead of the CONSORT for cluster randomised trial. Procedures: The case vignettes were created based on PRO data at baseline only. It is not clear at which time points the recommendations were discussed, just once in the beginning? PROs can change and recommendation may be adapted accordingly, how was this taken into account? Some more details on the intervention and assessment time points are needed, as readers may not go back to the publication of the protocol. E.g. Time points of PRO assessments during study phase and follow-up, when started the follow-up assessment? After completing therapy? Ref 21 seems not to be correct. There is a specific publication on Cohen's d by Cohen 1988. 4- Results A flow chart is missing. There is a Legend for Figure 1 at the end of the document for the number of patients at each assessment, but Figure 1 shows changes in QoL scores. Please use the CONOSRT for Cluster randomized trials detailing enrollment, allocation, follow-up and analysis. Patient characteristics: If randomization was done on center level, were there any imbalances between patients in the IC vs CC? The number of previous cycles has a quite lager range. Assuming that a certain accumulation of symptoms can occur over time the v1 assessment may vary with regard to PROs. How did the authors account for this? There are no information on cluster characteristics. Were the clusters similar, in size, type, number of patients treated per year, etc.? It is important to know, if cluster characteristics may contribute to the observed results. Page 12: For the median TUD, there is two times IC mentioned. What was the purpose of including patients in the reference center? Was the data gathered from these patients analysed?

	Univariate regression models are questionable, as the predictor, variables may be interdependent and controlling for this in a multivariate regression model would have been preferable. 5-Discussion The authors state “mechanism about how a supportive care intervention has to be composed and how it has to be implement is barely understood”. Some reflection by the authors on study design, intervention and implementation applied in this study and its implications for the study results should be added.

REVIEWER	Kwan Yu Heng Duke-NUS Medical School, Singapore
REVIEW RETURNED	14-Feb-2020

GENERAL COMMENTS	I am performing the statistical review of this paper attempting to assess a tailored set of intervention on various domains of QoL and to explore effect sizes using different PROM in patients with advanced STS undergoing treatment with trabectedin. 1. I felt since there is multiple measures used, a linear mixed method is more appropriate compared to regression in view that autocorrelation is very likely to occur between PROMs at various time point. The linear mixed model will adjust for this but regression does not. 2. Table 5 should be presented as beta, 95% CI and p instead of just p and R2. R2 is generally irrelevant unless we do model building and prediction. Therefore, the authors should re-analyse the data according to the above before useful interpretation can be derived. Thank you for the opportunity to review.

VERSION 1 – AUTHOR RESPONSE

Comments from Reviewer #1:

Reviewer Name: Robin L Jones and Eugenie Younger; Institution and Country: Royal Marsden and Institute of Cancer Research, UK. Robin L Jones has been a consultant for: Adaptimmune, Blueprint, Clinigen, Eisai, Epizyme, Daichii, Deciphera, Immunedesign, Johnson and Johnson, Lilly, Merck, Pfizer, Pharmamar.

I reviewed this manuscript with Dr Eugenie Younger. This is an important study worthy of publication. The following suggestions/ comments will it is hoped improve this excellent manuscript.

This manuscript provides results of the randomized ‘Yonlife’ study, which evaluated the effect of a tailored palliative care intervention on QOL in patients with advanced soft tissue sarcoma (STS) treated with Trabectedin. Although this was the primary objective, the manuscript also focused on the feasibility of collecting QOL data using ePROs, and impact of the palliative care intervention on other PRO measures such as anxiety and depression and pain severity.

This study is one of the first that has used electronic patient reported outcomes (ePROs) to collect quality of life (QOL) data in patients with advanced STS treated with chemotherapy. As the authors state, patients with advanced STS have limited effective therapeutic options, and treatment decisions involve balancing efficacy, impact on QOL and potential side effects. These data therefore add

important knowledge to aid understanding of QOL in patients with sarcoma, particularly as research in this area lags behind studies in patients with common cancers. Although the data are interesting and worth publishing, a major revision is required to address the following issues.

MAJOR CONCERNS

Reviewer #1; Question 1: I would recommend that the authors ask a native English speaker, who is proficient in proof reading scientific manuscripts, to edit the manuscript for English content.
Author response**:** We agree with the reviewer's observation. As already stated this has been done accordingly.

Reviewer #1; Question 2: Table 1 give details of the full analysis set (n=78), however the authors have said that analysis of efficacy endpoints will only be performed on the per-protocol set (PPS). How many patients were in the PPS?
What were the demographic and clinical details of the patients in the PPS. It is important to know which patients were remaining after drop out (PRO drop out and chemotherapy drop out).
Figure 1 legend shows at time point one (v1), there were only 19 patients in the IC and 14 patients in the CC. There is no explanation as to why so few patients provide PRO data at baseline. Which patients dropped out? Why was the response rate low? The abstract does not give the details of the number of patients in the study. All of the tables should include the number of patients evaluable at each time point as the interpretation is based on how many patients completed the PRO questionnaires at each time point.
Author response**:** We included 41 patients in the PPS. We now added the new Table 2 including characteristics of patients in the PPS. Concerning Figure 1 (now Figure 2), we included only patients that provided data for all sub-dimensions of FACT-G at each of the three time points (V1, V4, V7). This in fact reduces total numbers. We now include the number of patients already in the abstract. We amended all tables with the respective numbers of assessable patients according to your request.

Reviewer #1; Question 3: The method section states that patients were recruited just after starting Trabectedin or currently under treatment with Trabectedin. The table shows that some patients had received up to 17 cycles of Trabectedin before starting the study. Could the authors provide rationale for including patients starting Trabectedin, with those who have been on treatment for many months?
Author response**:** As YonLife has been designed to be an explorative study in the field of PRO, we aimed to include a broad range of patients with advanced disease. When designing the study we were convinced that a PRO-based intervention should be beneficial at any time point during treatment. Also, to the best of our knowledge no data contradicting this assumption were available. For clarification we'd also like to point out that in the methods section we now stated that patients had to have received at least one dose which includes patients receiving first dose just recently as well as patients that are under ongoing treatment.

Reviewer #1; Question 4: Table 2 shows a beneficial interventional trend on FACT-G total score and most FACT-G functional domains, but a potentially adverse trend with regard to anxiety, pain and symptom severity. The authors have stated that these are not significant referring to the p-values, but have not discussed the possible reasons for these findings. The p-values for the FACT-G scores were not significant either (looking at the p-values), however the authors have focused on the possible beneficial effect of the intervention using Cohen's d-test, while not discussing these other findings e.g. perhaps the questionnaires were not sensitive or specific for this population in this group. It is also surprising that palliative care intervention led to an adverse effect on pain severity scores. Please could the authors comment on these results.
Author response**:** Following the reviewers suggestion we now have added the following part in the Discussion section: *"As the intervention yields beneficial effects on QoL it seemed adverse on symptom domains such as average pain, as well as anxiety and depressivity. For the former, the applied intervention might not have been timely enough as adequate pain management needs immediate action instead of recommendation that take several days. Complex syndromes such as*

anxiety and depressivity need ongoint treatment, either psychooncological or pharmaceutical, which take more time to be effective".
Concerning the remarks about the non-existing sarcoma-specific module, we added the following statement to the already existing part in the discussion: "*Furthermore, sarcoma-specific QoL or symptom-measures are still missing, while the FACT-G and MDASI are generic instruments which might not cover syndromes and aspects specific for sarcoma patients.*"

Reviewer #1; Question 5: The abstract conclusion is focused on the feasibility of ePRO collection in a multicentre study, however this is not a new finding. This does not appear to be the main finding of the study. I would recommend revising the conclusion to include an opinion related to the primary objective i.e. what was the value of a palliative care intervention on QOL?
Author response**:** Following the reviewer suggestion we have rewritten the abstract conclusions as follows: "*Our data suggest that a tailored intervention based on ePRO may improve global QoL and subscales, while it did not have a beneficial impact on single symptoms*".

Reviewer #1; Question 6: The authors should put in context the development of trabectedin and cite the previous randomized Phase 3 trial leading to FDA approval, and previously published patient reported outcome data (Cancer 2019;125:4435-4441 PubMed).
Author response**:** Thank you for proposing this important work, which unfortunately was not yet published by the time of submission. Following the reviewer's suggestions the additional information regarding the regulatory status of trabectedin as well as the results of the phase III study and its subanalysis has been added in the Introduction section as follows: "*Trabectedin (Yondelis®) is a semi-synthetic drug originally isolated from the sea squirt Ecteinascidia turbinata with a complex multimodal mechanism of action [11,12]. Trabectedin was the first marine-derived antineoplastic drug approved in 2007 in the European Union and in over 70 countries across the globe for the treatment of patients with advanced STS after failure of anthracyclines and ifosfamide, or who are unsuited to receive these agents [13]. In 2015, trabectedin was also approved in the United States based on a pivotal phase III trial, which demonstrated that trabectedin had a significantly longer PFS compared with dacarbazine in patients with advanced liposarcoma or leiomyosarcoma after failure of prior chemotherapy [14]. Noteworthy, an ad hoc analysis of the phase III trial, which compared inpatient with outpatient infusion of trabectedin, showed that safety, efficacy and PROs outcomes were comparable between both treatment settings. In addition, an analysis of the MD Anderson Symptom Inventory (MDASI) PRO scores reported no clinically meaningful differences among patients reporting severe symptoms (MDASI score ≥7) who were treated with trabectedin either in an inpatient or outpatient treatment setting [15]."*

Additional Major Concerns

Reviewer #1; Question 7: Page 11. Results. The authors have provided a table for the full analysis set, however there is no detail on the per protocol set (PPS). How many patients were in the PPS and how many in each group? This is important if the analysis is only based on patients who completed V1 and V4 assessment and had treatment with Trabectedin throughout.
Author response**:** We have added the new Table 2 which included the requested characteristics.

Reviewer #1; Question 8: Page 11. Table 1. It appears that at least one patient in the study had already received 17 cycles of trabectedin. This contradicts what is stated in the methods, i.e. patients were recruited when they had only just started treatment. Please could the authors clarify why this patient was included? Is it possible to compare the QOL of patients starting trabectedin with those who have been on treatment for such a long period of time?
Author response**:** We have deleted the misleading terms "just started". Indeed, the study included patients with a broad range regarding the number of previous cycles. As toxicity and symptoms are prevalent throughout the course of the treatment, we decided to include patients regardless of their number of previous cycles.

Reviewer #1; Question 9: Page 11, line 55. Does the primary outcome refer only to the PPS? If so, please could the authors describe the number of patients in the PPS to which these data refer?

Author response**:** We have added the number of evaluable patients for each outcome parameter to Table 3.

Reviewer #1; Question 10: Page 12, Table 2. Please could the authors comments on why they think that the intervention may have had an adverse effect on anxiety, pain severity and interference, symptom severity and interference compared to the control cluster? The authors have focused on the potentially beneficial impact of the intervention but not discussed these adverse findings.
Author response**:** We have included an additional statement as described in the response to your question 4.

Reviewer #1; Question 11: Page 14. Lines 3-5. The authors state that there were no significant differences between the IC and CC with regard to other domains of PRO, but the p-values were not significant for any of the FACT-G domains. These show a possible adverse trend in of the intervention and therefore this should be discussed.
Author response**:** We absolutely agree with the reviewer's observations. We now changed this sentence as follows "*Overall, there were non-significant, adverse trends in other domains of PRO (MDASI, FAACT, HADS and BPI scales) (Table 3 and Supplementary Table 2)".* Furthermore, we added a paragraph to the discussion section regarding the adverse trends as proposed by your Major Concern #4.

MINOR CONCERNS

Reviewer #1; Question 1: Page 2, Line 19. Please review the term 'supportive intervention'. As the reader, I feel that this is too vague and am left wondering what the intervention involves. Given that the assessment of the intervention on QOL is the main goal of the study, more detail is needed.
Author response**:** Thank you for this hint. We now use ""tailored multi step intervention" in order to be more specific. As we include more details to answer your question 3 in the method section and due to limited word count in the abstract, we refrain from giving more details already here.

Reviewer #1; Question 2: Page 2, Line 24. Please provide more detail on participants of interest and how many will be recruited (if not described in the results)
Author response**:** Due to limited word count, we only include number of cluster per center as well as number of patients randomized. Further details can be found in the CONSORT-flowchart.

Reviewer #1; Question 3: Page 2, Line 28. The authors have said that the intervention for the IC involves ePRO with four-step assessment for palliative care. This does not tell the reader what the intervention is, only that more assessments will take place in the IC. Please define the intervention here.
Author response**:** We apologize as our wording might have been inaccurate. We now replace the misleading term "assessment" and sketch the intervention already in the abstract. Due to the strict word limit in the abstract we refrain from elaborating on the intervention already here and add further details in the Intervention-Section.

Reviewer #1; Question 4: Page 2, Line 40. The authors have not described the population at the start of the results section. How many patients took part? How many were in each treatment arm? What was the median age of patients? What line of treatment was Trabectedin for most patients? Although some of these data are in the table, it would be helpful to have a summary of the group in the text.
Author response**:** Accordingly, we have added the number of patients in the participants section of the abstract upon request of the second reviewer. As we agree that more details would be fruitful, due to strict word limit in the abstract, we would like to refrain from including more details already in the abstract.

Reviewer #1; Question 5: Page 2, Line 47. Survival (in days) was almost double in the intervention cluster compared to the control group, however the p-value is not significant. This is surprising – please could the authors comment.

Author response**:** Due to the small sample size and few events of deaths the statistical power of this test is quite limited.

Reviewer #1; Question 6: Page 2, Line 51-53. The authors have described the influence of several factors (such as age, gender, depression and anxiety) on QOL, however this objective was not mentioned in the abstract. It is unusual to describe results without mentioning that this was going to be assessed.
Author response: we now added the following to the objective-section of the Abstract: "…as well as identifying predictors of QoL."

Outcomes which were described (such as impact of intervention on pain intensity, anxiety and depression) are not mentioned in the results section.
Author response**:** We have now added the following sentence: "Smaller adverse trends between arms were observed in MDASI, FAACT, HADS and BPI scales"

Reviewer #1; Question 7: Page 3, Line 6. The main conclusion should consider whether the supportive care/palliative care intervention had a beneficial impact on QOL in the intervention cluster (i.e. added 'value' as per the title). Although the trial does add knowledge about QOL in this patient group, that was not the main principal aim of the study. The authors have concluded that PROs can be used in a multi-centre study but this is not new knowledge. The authors should focus their conclusion on the principal findings of their study.
Author response**:** We agree that we should focus more on the interventional effect. Upon your major concern #5 we change the conclusion as follows "*Our data suggest that a tailored intervention based on ePRO may improve global QoL and subscales, while it did not have a beneficial impact on single symptoms".*

Reviewer #1; Question 8: Page 4, line 12. ePRO assessment is feasible in a multicentre study – this is not a key strength of this study as this has been shown before in multiple other studies (albeit in different tumour groups).
Author response**:** We have deleted this sentence.

Reviewer #1; Question 9: Page 4, line 14-16. Please review this sentence as it is not clear. What is meant by "important outcomes of palliative care"? Do the authors mean that the intervention seems to improve certain quality of life outcomes?
Author response**:** We have deleted this sentence.

Reviewer #1; Question 10: Page 4, line 19. The authors state that the sample size is a major limitation, however the number of patients had not yet been described in the abstract.
Author response**:** We now describe the number of patients in the abstract.

Reviewer #1; Question 11: Page 5, line 5. Please review the first sentence of the introduction. What do the authors mean by 'evolved in the past'? Should this say 'evolved over recent years'? It is not clear whether the authors think that toxicity and impact on QOL has increased over the last few years? Additionally, there is no reference for this sentence. Please rewrite this sentence so that it is more clear.
Author response**:** Thank you for your request. We now changed the first two sentences as follows "*The armamentarium of systemic treatment in advanced soft tissue sarcoma (STS) has evolved over the past decade. Yet, the burden of disease remains high and drug related adverse events are frequent [1–3], even in patients who experience long lasting clinical benefit.*".

Reviewer #1; Question 12: Page 5, Lines 19-22. 'Mental problems' should be replaced with 'mental health problems'.
Author response**:** It was changed as proposed.

Reviewer #1; Question 13: Page 5, line 22. Paramount is not the correct English word to use in this context. Please review.
Author response**:** We rephrase this sentence as following: "*Mental health problems such as distress, depression and anxiety are as frequent as in other cancer patients*."

Reviewer #1; Question 14: Page 5, line 28. The authors say that drug have 'specific safety profiles'. Safety tends to refer to phase 1 drugs. Please review this sentence. Do the authors mean to refer to side effect profiles?
Author response**:** We completely agree with your request as our statement was incorrect. Accordingly, we rephrased the sentence as follows: "*... there are distinct and drug-specific side effects*").

Reviewer #1; Question 15: Page 4, line 33. Please give an example where convenience can affect treatment decisions, such as inpatient versus outpatient regimen.
Author response**:** We decided to delete the misleading term "convenience", as it is not particularly relevant for this manuscript.

Reviewer #1; Question 16: Page 5, line 49. What do the authors mean by 'time to deterioration'? Please be more specific do you mean deterioration in symptoms or disease progression on imaging or declining performance status?
Author response**:** We now change to "time to deterioration of QoL". We described the operationalization of time to deterioration in the method section as follows: *"Additionally, the time until QoL deterioration (TUD) was also assessed as a change of at least 3.3 points between V1 and V4 as defined by King et al. [27]."*.

Reviewer #1; Question 17: Page 5, line 54. Please add more details about the beneficial outcomes, the settings and the entities. For example describe what beneficial outcomes have been achieved using PRO, in which tumour groups and using what interventions. This is important to put this research paper into context of existing research in this field.
Author response**:** We added details as follows: "*For instance, Basch et al found benefits of their STAR (Symptom Tracking and Reporting) intervention in prolonging time on chemotherapy, less unexpected admission and longer quality-adjusted survival [17]. In brief, they randomized 766 patients from a single institution under chemotherapy for solid tumors to either usual care or STAR. The intervention consisted of 12 different symptoms collected remotely, providing treating physicians with graphical representations of results and alerting nurses when a preset cut-off of worsening condition was met. Another randomized multi-center trial evaluated the effect of a web-based, self-report assessment and educational intervention on symptom distress during cancer therapy in 752 ambulatory patients from different entities and with various diagnoses [18]. In this multicenter sample of participants they reported that Web-based patients-rated symptoms and communication coaching reduced symptom distress after active cancer treatment, particularly in those aged >50 years*."

Reviewer #1; Question 18: Page 6, line 12. A multi-professional expert panel is not an intervention. Please describe what the expert panel are doing i.e. the intervention that they are making.
Author response**:** We sketch the job of the expert panel as follows: "[…] *a multi-professional expert panel that discusses PRO-results and derive treatment recommendations*". As we are describing the role of our expert panel in the intervention section in the Methods as well as giving a more detailed description in our previous publication (Schuler et al., 2017), we would like to refrain from giving deeper details in this introduction section.

Reviewer #1; Question 19: Page 7, line 8. Please define what is meant by 'just started' (e.g. just received the first cycle?).
Author response**:** We have changed this paragraph upon your request (#8) and deleted "just started".

Reviewer #1; Question 20: Page 7, line 10. It is important to state at which line of therapy these patients are receiving Trabectedin to understand more about the target population. Are these patients having Trabectedin as second line treatment (e.g after Doxorubicin), or are some of the patients having Trabectedin as first line treatment of metastatic disease?
Author response**:** Patients received trabectedin in an in-label prescription. We now added number of previous lines of another chemotherapy to Table 2.

Reviewer #1; Question 21: Page 7, line 20. The incorrect symbol has been used – should be ≤2.
Author response**:** This has been changed accordingly.

Reviewer #1; Question 22: Page 7, lines 27-32. Please use the same tense throughout the methods section (past at beginning, then present tense in these sentences).
Author response**:** We made the changes accordingly.

Reviewer #1; Question 23: Page 7, lines 40-45. Could the authors explain why patients and public were not involved in the design of the research study?
Author response**:** Although patients were not involved in study design, they've been deeply involved in study conduction as a representative of the national sarcoma patient advocacy organization ("Das Lebenshaus") took part in the expert panel discussions. We changed the "Patient and public involvement" section as follows: "*We are grateful to all patients that participated in the YonLife trial. A member of the national sarcoma patient advocacy group "Das Lebenhaus" took part in the expert panel discussion."*

Reviewer #1; Question 24: Page 8, line 42. Why was this time point chosen for the primary outcome? Is this related to the timing of scans? Is this related to the number of cycles that were expected to be received by patients? Please provide rationale for the time point.
Author response**:** Yes, this is correct. In your review you pointed out already two of the reasons for the rational of selecting 9 weeks as time point of primary outcome. Moreover, we expected 9 weeks to be appropriate in order to take actions after receiving the case vignette. We made according changes in the manuscript as follows: "*Nine weeks was set as time for primary outcome assessment since this period provides enough time to take action concerning interventional proposals*".

Reviewer #1; Question 25: Page 8, line 49. Please specify which functional domains the FACT-G measures rather than using the word 'several' – this is too vague and the functional domains are important measures in this study.
Author response**:** We have specified this point and now it includes the four domains in brackets. The sentence now reads "*The FACT-G is a PRO measure used to assess health-related QoL in patients undergoing cancer therapy as a total sum score (ranging from 0 to 108) comprising four subscales of QoL (physical, social, emotional, functional well-being) [26]."*

Reviewer #1; Question 26: Page 9, line 40. Please add detail on the exploratory outcome shown in the results section of the abstract regarding the impact of variables such as age and gender on QOL. This is mentioned later in the statistical considerations but important to include this as an outcome of the study.
Author response**:** We have added the following sentence "*We assessed the predictive value of the following varibles at V1 for QoL: gender, age, performance status (ECOG), tumor stage (UICC-classification), symptom severity (MDASI), symptom interference (MDASI), depressivity (HADS), anxiety (HADS), patients satisfaction (IN-PATSAT32) [34], anorexia/cachexia (FAACT)."*

Reviewer #1; Question 27: Page 11. Results. Please provide more details on the patient group. For example, how many patients had locally-advanced disease versus metastatic disease, how many previous lines of treatment had patients received, what prior treatments had patients received (e.g. anthracycline based)?
Author response**:** We have added the information about metastatic-disease to Table 2. Range and median of previous lines of another chemotherapy have been added to Table 2. Concerning the type of previous treatment, analysis are currently ongoing. We assume that the specific kind of previous chemotherapy does not influence our intervention.

Reviewer #1; Question 28: Page 11, Table 1. The median values in the row 'number of previous cycles of another chemotherapy' are surprising. Patients normally receive at least two cycles of a chemotherapy drug before having imaging to assess response. Please could the authors comment on why this is?
Author response**:** We agree that this is surprising, yet our patients were treated with trabectedin in an in-label prescription. We are currently analyzing potential reasons that might have prevented the use of usual front-line chemotherapy such as ifosfamide or anthracycline-based regimen.

Reviewer #1; Question 29: Page 14. Table 4. The p-values are presented and are not significant, however it would be interesting to also define the effect size for these PRO domains as with the FACT-G domains.
Author response**:** Thank you for this important advice. We now included the effect sizes in Table 3.

Reviewer #1; Question 30: Page 14. Table 4. It is interesting that palliative care intervention did not make any difference to scores on MD Anderson symptom inventory – could the authors comment on why that might be? Could this be because this inventory is not sensitive enough to detect a difference?
Author response**:** We agree with your suggestion that there is no measurable effect because of low sensitivity. As stated in the discussion section there is no sarcoma-specific module available in most PRO-measures, including the MDASI. Therefore, as other measures, the MDASI might lack content validity in a sarcoma population. We have amended the discussion accordingly.

Reviewer #1; Question 31: Page 16. Discussion. Please could the authors review the first paragraph – several sentences appear to be repetition of each other.
Author response**:** Thank you for your hint – we have rewritten the first paragraph in the Discussion section.

Reviewer #1; Question 32: Page 16 line 40. Limitations are usually described towards the end of the discussion. Please could the authors focus on the importance of their findings for patients and for clinical practice. Perhaps the FACT-G scores could be compared to scores in other studies looking at similar populations of patients with advanced cancer.
Author response**:** We have included additional references in order to compare our findings to increase interpretability. This reads as follows: "*This change is well in line with findings reported by Franciosi and colleagues [37] who reported in a multi-center randomized trial. Patients either received early palliative care or standard oncologic care. They reported a comparable decline in FACT-G score of ~2 in 281 patients suffering from advanced cancer (gastric, pancreatic, NSCLC, biliary tract), while their total FACT-G scores after twelve weeks (70.1 and 69.6) is comparable to the score found in IA (73.9) and CA (69.4) after nine week in the YonLife-trial. FACT-G score (76.4) was also comparable to the YonLife baseline score (74.2) in a sample of 42 patients suffering from different sarcoma entitites in a single center, cross-sectional description by Chan and colleagues [38].*

Furthermore, we moved the part describing limitations to the end of the manuscript.

Reviewer #1; Question 34: Page 17, lines 26-28. Please rewrite this sentence as the meaning is not clear.
Author response**:** We delete this sentence as it yields no essential benefit to the manuscript.

Reviewer #1; Question 35: Page 17, conclusion. The conclusion is well written and important.
Author response**:** Thank you for your important remarks and critical requests.

Comments from Reviewer #2:

Reviewer Name: Karin Ribi
Institution and Country: International Breast Cancer Study Group, Switzerland Please state any competing interests or state 'None declared': None declared

This paper presents a cluster-randomized trial investigating the added value of a tailored palliative care intervention in patient population that has been studied less often compared to other cancer entities. The results are therefore of interest. However, some major information are lacking, which makes it difficult to appraise the methodological rigor.

Reviewer #2; Question 1: The Abstract: There is no clear objective formulated in the abstract. The abstract states that patients were randomized, this is contradictory to the methods section p.8) stating that centers were randomized. The number of patients included in this study is missing.

Author response**:**

- We now formulated the objective as follows: "*The YonLife trial aimed to explore the effect of a tailored multi-step intervention on QoL, symptoms and survival in patients with advanced STS undergoing treatment with trabectedin as well as identifying predictors of QoL.*"
- Our previous wording was misleading: We randomized centers, not patients. We have rephrased the participants section of the abstract to clarify and added the number of patients.

Reviewer #2; Question 2: Introduction: The first two sentences need to be supported by references.
Author response: We have rephrased the first two sentences upon request from editor #1 and provided references as well. The new sentences read as follows "*The armamentarium of systemic treatment in advanced soft tissue sarcoma (STS) has evolved over the past decade. Yet, the burden of disease remains high and drug related adverse events are frequent [1–3], even in patients who experience long lasting clinical benefit."* We added references as well."

The intervention is a palliative care intervention that can be applied across cancer types and treatments. A rational why this intervention is relevant for sarcoma patients and why the authors chose to only include patients receiving a specific treatment is lacking.
Author response: As other authors (Denis et al., 2019; Klinkhammer-Schalke et al., 2012) we aimed at a single tumor entity. This seems reasonable, because there are specific issues in distinct cancer types. Moreover, patients receiving trabectedin usually have advanced disease and suffer from a broad range of symptoms which can be very severe. Especially those patients could eventually benefit from a comprehensive supportive intervention. Of course it would have been possible to include patients receiving other anticancer treatments, but our rationale was to facilitate recruitment and analysis when keeping the patient group quite homogenous regarding treatment.

Prediction of QoI is not mentioned as an objective.
Author response**:** We now mention this in the Abstract as described above.

Please provide also a rational why the design of a cluster-randomized trial was chosen.
Author response**:** We have added a rational to the Randomization section.

Reviewer #2; Question 3: Methods: There are some doubts, if this study was a true cluster-randomized study: The level of randomization is not clear, patient or centers? Please clarify the randomization procedure.
Author response**:** Thank you for your advice.

- Our previous wording was misleading as YonLife randomized centers, not patients into the control arm and the interventional arm (containing 3 clusters each) as well as the reference center.
- We rewrite the randomization procedure and created a new subheading to clarify and improve readability. The new part reads as follows: "*Six German centers were cluster-randomized in a 1:1 ratio in an IA (3 centers) and a CA (3 centers). This trial was designed as a cluster-randomized trials to avoid contamination that might result in a type 2 error. If randomized on patient level, contamination might have been occurred as patients talked to each other about the recommendations or the treating physician transferred recommendations from one patient to another."*
- *The seventh center where the supportive care recommendations were created served as a reference center (RC). Patients treated at the RC received the same intervention as in the IA but were analyzed separately. The RC was invented in order to avoid bias from a dual role of participating clinicians as being part of treatment staff in the center and taking part in the expert panel at the same time. Furthermore, we initiated the RC as the first center in order to get to know and solve any technical or logistical barriers in a mono-center setting before spreading it procedures to a multi-center-setting.* "
- We changed the wording throughout the manuscript accordingly from "IC" to "IA" and "CC" to "CA".

If centers were randomized the cluster size is 11 not 33.
Author response**:** Thank you for your advice. We now changed this issue according your hint in the "statistical considerations" section as well as throughout the manuscript.

Also the authors used STROBE for cohort study checklist instead of the CONSORT for cluster randomized trial.
Author response**:** We added the CONSORT flowchart as new Figure 1 for cluster randomized trials as requested by you and the editorial team.

Procedures: The case vignettes were created based on PRO data at baseline only. It is not clear at which time points the recommendations were discussed, just once in the beginning? PROs can change and recommendation may be adapted accordingly, how was this taken into account?
Author response**:**

- The PRO-results were discussed after V1 and before V2 by the expert panel. Recommendations were sent to treating physician before V2. This is now added to the Intervention paragraph in the Methods.
- Thank you for bringing up the important question whether we took into account possible changes of PRO over time and how to react regarding our recommendations. We were discussing this issue before study conduction but refrained from sending recommendations at several time points because of the following reasons:
    - We assessed symptoms which are more stable over time (i.e. QoL, depressivity, anxiety) and others that might change quicker (symptoms such as pain, diarrhea…). Even when sending recommendations at every visit, our reaction would still be slower than the clinician on site for those domains changing rapidly.
    - Due to the aforementioned point, our recommendation were more likely to affect the more stable dimensions. As most treatments (e.g. psycho-oncological intervention or – if necessary – anxiolytic medication) for more complex domains need time to be initiated and to be effective, adaption of once-given recommendation might have been before treatment effects occurred.
    - Our intervention gave insight into symptoms that might have been missed by the treatment team. Like a comprehensive geriatric assessment, it encompasses domains not regularly assessed in daily routine with validated instruments and aimed to generate treatment recommendations for a broader range of areas.

Some more details on the intervention and assessment time points are needed, as readers may not go back to the publication of the protocol. E.g. Time points of PRO assessments during study phase and follow-up, when started the follow-up assessment? After completing therapy?
Author response**:** Thank you for sharing your concerns about changing PRO and requesting more details on study-design.

- We now added assessment time points to the "trial design and objectives" part in the Methods as follows: "*Outcomes were assessed at baseline (i.e. visit [V] 1), after 3 (V2), 6 (V3), 9 (V4) weeks. Follow-up was conducted 21 (V5), 35 (V6) and 61 (V7) weeks after baseline.*"
- Following a request from editorial team, we added new Table 1 which included time points of respective PRO measures throughout study phase.

Ref 21 seems not to be correct. There is a specific publication on Cohen's d by Cohen 1988.
Author response**:** We have changed this reference as proposed.

Reviewer #2; Question 4: Results: A flow chart is missing.
Author response**:** We added the CONSORT flowchart as new figure 1 for cluster randomized trials as requested by you and the editorial team.

There is a Legend for Figure 1 at the end of the document for the number of patients at each assessment, but Figure 1 shows changes in QoL scores. Please use the CONOSRT for Cluster randomized trials detailing enrollment, allocation, follow-up and analysis.
Author response**:** We have added these number in the CONSORT-Flow-Chart as well. We would like to refrain from deleting the numbers in the figure as readers do not need to take a look on both, the CONSORT and the figure as the same time.

Patient characteristics: If randomization was done on center level, were there any imbalances between patients in the IC vs CC?
Author response**:** We now added details to the RESULTS section as follows "*At baseline, the IA included 38 patients (19 of whom included in PPS), while CA consists of 29 patients (14 of whom included in PPS). No difference concerning age, gender and the number of previous cycles of trabectedin was observed between the arms. In the CA more patients had a higher tumor stage (P=0.083) and less patients suffer from leiomyosarcoma (Table 2).*

The number of previous cycles has a quite lager range. Assuming that a certain accumulation of symptoms can occur over time the v1 assessment may vary with regard to PROs. How did the authors account for this?
Author response: We aimed to include a broad range of patients with advanced disease.

There are no information on cluster characteristics. Were the clusters similar, in size, type, number of patients treated per year, etc.? It is important to know, if cluster characteristics may contribute to the observed results.
Author response**:** All seven participating hospitals are tertiary referral centers with a university affiliation and are members of the German Interdisciplinary Sarcoma Group. They all provide care according to national and international sarcoma care guidelines. The expected number of patients fulfilling inclusion criteria was expected to range from 10 to 50 with a median of 20 patients.

Page 12: For the median TUD, there is two times IC mentioned.
Author response**:** We apologize for this was mistakenly written. The correct sentence reads as follows: "*The median TUD differed slightly between IA (25 days, 95% CI: 6.2-43.8) and CA (22 days, 95% CI: 16.5-27.5; P=0.927)."*

What was the purpose of including patients in the reference center? Was the data gathered from these patients analyzed?
Author response: We now added the following paragraph in order to clarify on the requested issue: "*The seventh center where the supportive care recommendations were created served as a reference center (RC). Patients treated at the RC received the same intervention as in the IA but were analyzed separately. The RC was invented in order to avoid bias from a dual role of participating clinicians as being part of treatment staff in the center and taking part in the expert panel at the same time."*

Univariate regression models are questionable, as the predictor, variables may be interdependent and controlling for this in a multivariate regression model would have been preferable.
Author response**:** We totally agree with your concern. We added the results of the multivariable regression in new Table 4 and added a corresponding sentence to the Results section.

Reviewer #2; Question 5: Discussion: The authors state "mechanism about how a supportive care intervention has to be composed and how it has to be implement is barely understood". Some reflection by the authors on study design, intervention and implementation applied in this study and its implications for the study results should be added.
Author response**:** We agree with the reviewer's observations. We added a new paragraph on this topic entitles as "*YonLife-intervention - unanswered questions and future research".*

Comments from Reviewer #3:

Reviewer Name: Kwan Yu Heng

Institution and Country: Duke-NUS Medical School, Singapore Please state any competing interests or state 'None declared': None declared

I am performing the statistical review of this paper attempting to assess a tailored set of intervention on various domains of QoL and to explore effect sizes using different PROM in patients with advanced STS undergoing treatment with trabectedin.

Reviewer #3; Question 1: I felt since there is multiple measures used, a linear mixed method is more appropriate compared to regression in view that autocorrelation is very likely to occur between PROMs at various time point. The linear mixed model will adjust for this but regression does not.
Author response**:** We apologize if our heading of Table 5 (now Table 4) might have been misleading. We changed it accordingly as follows: "*Univariate regression of FACT-G total score after nine weeks (V4) on parameters measured at baseline (V1) over all groups*". The regression analyses aimed to examine effects of baseline parameters on the FACT-G total score on visit 4. No multiple measures per patient were analyzed in the model, so we believe a linear mixed model is not appropriate here.

Reviewer #3; Question 1: Table 5 should be presented as beta, 95% CI and p instead of just p and R2. R2 is generally irrelevant unless we do model building and prediction. Therefore, the authors should re-analyses the data according to the above before useful interpretation can be derived.
Author response**:** Thank you for this important advice. We now added the 95% CI to Table 4.

Thank you for the opportunity to review.

VERSION 2 – REVIEW

REVIEWER	Robin Jones and Eugenie Younger Royal Marsden and Institute of Cancer Research, UK Robin L Jones: Receipt of grants/research support: MSD GSK Receipt of consultation fees: Adaptimmune Athenex Blueprint Clinigen Eisai Epizyme Daichii Deciphera Immunedesign Lilly Merck Pharmamar Upto Date
REVIEW RETURNED	02-Apr-2020

GENERAL COMMENTS	The only remaining concern is that the abstract states that 79 patients were included. However for the per protocol analysis (on which results/conclusions are based) there were 19 patients in the intervention arm and 14 patients in the control arm. We could not see an explanation of why patients dropped out before completing

	the ePROs and were not available for analysis. we think the abstract should state the true number of patients who took part in the PRO assessments and were in the intervention arm and control arm.

REVIEWER	Karin Ribi International Breast Cancer Study Group
REVIEW RETURNED	19-Apr-2020

GENERAL COMMENTS	The authors have addressed my comments and questions in a sufficient manner. An English revision for the revised paragraphs would further improve the manuscript.

VERSION 2 – AUTHOR RESPONSE

Comments from Reviewer #1:

Reviewer Name: Robin Jones and Eugenie Younger Institution and Country: Royal Marsden and Institute of Cancer Research, UK Please state any competing interests or state 'None declared': Robin L Jones: Receipt of grants/research support: MSD GSK. Receipt of consultation fees:Adaptimmune, Athenex, Blueprint, Clinigen, Eisai, Epizyme, Daichii, Deciphera, Immunedesign, Lilly, Merck, Pharmamar, Upto Date

Reviewer #1; Question 1: The only remaining concern is that the abstract states that 79 patients were included. However, for the per protocol analysis (on which results/conclusions are based) there were 19 patients in the intervention arm and 14 patients in the control arm. We could not see an explanation of why patients dropped out before completing the ePROs and were not available for analysis. We think the abstract should state the true number of patients who took part in the PRO assessments and were in the intervention arm and control arm.

Author response**:** We agree with the reviewer's observation and added the number of analyzed patients in the abstract. The explanation why patients weren't analyzed is provided within the CONSORT-Flowchart. Further information on why patients did not provide PRO-data were not reported by the participating centers.

Comments from Reviewer #2:

Reviewer Name: Karin Ribi
Institution and Country: International Breast Cancer Study Group, Switzerland Please state any competing interests or state 'None declared': None declared

The authors have addressed my comments and questions in a sufficient manner. An English revision for the revised paragraphs would further improve the manuscript.

Author response: Following the editors' suggestion we have asked a native English speaker to edit the manuscript for English content.

We greatly appreciate the time and efforts of the editorial staff and reviewers and we believe that the manuscript has been revised to satisfy all the concerns and recommendations.