PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Serious gaming during multidisciplinary rehabilitation for patients with complex chronic pain or fatigue complaints: study protocol for a controlled trial and process evaluation
AUTHORS	Vugts, Miel; Joosen, Margot; Mert, Agali; Zedlitz, Aglaia; Vrijhoef, Hubertus

VERSION 1 - REVIEW

REVIEWER	Dr Jacqui McKechnie Glasgow Caledonian University Cowcaddens Rd Glasgow G4 0BA
REVIEW RETURNED	07-Mar-2017

GENERAL COMMENTS	Review of Serious gaming during multidisciplinary rehabilitation for patients with complex chronic pain and fatigue complaints: study protocol for a controlled trial and process evaluation. This is an exciting study which should help to provide evidence for the efficacy of the use of serious games in this population. The protocol is well thought out and all aspects of the study have been considered thoroughly. However, it would be useful to have more information in the protocol about why the game LAKA is more effective compared to other games. I note that it is the first game which has been developed which promotes practice under adverse conditions, but more detail on this would be helpful. Also it would be useful to be clearer about the human computer interface to be used and what adaptions will be made to ensure that the game is accessible for those with extra needs in relation to mobility or vision etc. There are a few typos which I have highlighted below: Page 2 line 10: comma required after ‘circumstances’ Page 4 lines 10, 11 and 12: need to be reworded Page 5 lines 30 and 31: ‘Unfortunately, only small long-term improvements have been found in low back pain patients with severe and long-lasting pain and disability’. Needs an explanation about what is meant by ‘small’. Page 5 lines 33: comma required after ‘such’. Page 5 lines 35: could put ‘exergaming’ in quotation marks for consistency with the other terms used. Page 7 from line18 to 48: ‘proces’ should be ‘process’ Page 15 line 46: should be ‘in the development of LAKA’. I wish the authors success in the data collection and write-up to follow and will be interested to read the final study report when it is published.

REVIEWER	Köke, Albere University Maastricht Netherlands
REVIEW RETURNED	16-Mar-2017

GENERAL COMMENTS	This article describes a protocol for studying the effectiveness of serious gaming as a complementary modality during multidisciplinary rehabilitation for patients with complex chronic pain and fatigue complaints. The protocol includes a novel approach for chronic pain rehabilitation and the mixed method approach is impressive and well suited for the development of this type of treatment. Some minor issues to consider: Abstract Intro: change current intervention into multidisciplinary rehabilitation Methods; what is meant by 220 complete records. Do you aim at including patients until you have 220 complete records, or do you want to include 220 patients? In method section of the article it is not described as such. Discussion: I would not focus on effectiveness of this study, due to the chosen design but on aim 2 and 3 . Discuss this topic. Introduction In the title of the manuscript the target population were also patients with fatigue complaints. In the introduction fatigue is never mentioned. What is meant by fatigue complaints; patients with fatigue syndrome or patients with chronic pain and fatigue? Please make in the introduction clear what population you are aiming at. Perhaps title should be adapted (see also comment at section discussion) Also only the relative small effect of rehabilitation of low back pain is described. What about the effectiveness in other chronic pain problems? Is this a problem for all chronic pain populations or specific for low back pain The introduction is well written and understandable. I miss however reasons why there are only small long term improvements of rehabilitation and why serious gaming specifically would be able to improve these small effects. Please give a rationale or theoretical framework how serious gaming can fill in the gaps in traditional treatment. Now possible explanations are given how serious gaming could work. But what is the difference with more traditional methods? Are we not delivering the same treatment in a more fancy way? Why adding this to the treatment and why not provide it as a stand- alone treatment? The end of the introduction is about complexity of outcome evaluation. However for me it is not clear what the authors exactly mean. Specific the part of debriefing needs more explanation. The sentence: Indeed, more adequately …. (pag 5 sent 10): I don't understand this according to the previous text e. Of course adequately powered trials are always needed. What is the point? Study aims 1. Why so many primary outcome? What is the specific target of rehabilitation? In my opinion is daily functioning/participation the main outcome. Please explain why there are so many and so diverse outcomes. And why participation is not one of them. Perhaps focus on effectiveness should be less and two other goals are more

	important! 2. Please structure the information according to order as described in the aim; organization provider patient and innovation level. Give some examples of serious gaming features that is related to innovation levels. I don't understand the addition in brackets line 39 pag 5. Is this a methodological issue? 3. In aim 3 mechanisms are described of serious gaming; please elaborate on these in the introduction. What mechanisms are known and how can they contribute to or enhance the existing mechanisms in traditional treatment (see comment theoretical frame work earlier) Method and analysis Why is a quasi experimental design chosen? As your aim 1 is effectiveness then a RCT should be chosen. Recruitment What is meant by absence of disruptive events planned? Pag5 7 line 21 What is the role of local stakeholders of serious gaming in the study pag 7 line 25 Patients Why patients are invited as the completed half of their rehabilitation program. Why not at start of the treatment program? Page 7 line 28-29; Please explain this Interventions The intervention focuses on participation. However this is not part of your primary outcome measures. Explain this? The basic program is in accordance with individual care needs! How can you than evaluate effectiveness as content of the basic treatment is not uniform between patients? Please explain this, also this should be addressed in the discussion Mode of delivery. What is exact purpose of debriefing sessions. Please provide this information Programme theory This a very short description. Please describe this in more detail so the reader can understand the rational of the approach. Also combine this with the model behind pain rehabilitation and how you combine these models,. What is common and what is complementary? Or how interacts this with the basic program. Measures What is the exact starting point of the study. At page 7 it is stated that patient are invited after completion of half the program. Yet you already measure at baseline. On what data the effects of the treatment are evaluated? In table 2 baseline measurements seems included. Or have I interpreted the information on page 7 incorrectly. Again the main focus of rehabilitation, as you also stated, is participation. This is not measured. Why? How standardized are case mix variables measured and registered. How can you ensure adequate quality of these data? Give information about abbreviations in table 4 Qualitative data I don't understand the role of the interviewers towards more educational …. …What do you mean or what is the aim for this. For evaluation the interviewer must be neutral and open and explorative but not directional or educational in my view. So please explain what the exact role and aim is. Power calculation. On which primary outcome the power analysis was performed? Discussion Strengths and limitations are well discussed. However, the apparently diversities in the basic treatment (accordingly to patient's

	needs) is not discussed This also has effect on internal validity , as the quasi experimental design has . The aim of effectiveness should be less strong and focus on development and process evaluation are more important in my opinion based on the chosen design. Perhaps the title of the manuscript should reflect this more

VERSION 1 – AUTHOR RESPONSE

Our point-by-point response to comments of reviewer 1:

Thank you very much for your compliments on the thoroughness of our research protocol and your suggestions for improvement.

1.1 Reviewer 1 commented: However, it would be useful to have more information in the protocol about why the game LAKA is more effective compared to other games. I note that it is the first game which has been developed which promotes practice under adverse conditions, but more detail on this would be helpful.'
Response: Thank you for your suggestion to provide more detailed information about why the game LAKA may be more effective compared to other games. We did so by including the following phrase in the introduction (page 3, lines 32-34): 'Despite of promising results for various mono-disciplinary applications of gaming and simulation, no evident application seems to exist for supporting biopsychosocial adjustment processes in patients with CP or FSS.2 3 32-37

1.2 Reviewer 1 commented further: 'Also it would be useful to be clearer about the human computer interface to be used and what adaptions will be made to ensure that the game is accessible for those with extra needs in relation to mobility or vision etc.'

Response: Thank you for the suggestion to provide more clarity about the human computer interface and adaptations being made to ensure accessible of the game.
We added a sentence to the appendix, and a screenshot that provides additional illustration of the interface (supplementary file, p 2, lines 4-8): ´ User interface (accessibility): The human-computer interface is designed for being easy to use (i.e. there is no time pressure). It is controllable by individuals with low computer skill. It involves making decisions by taping on the screen (pre-selecting and confirmation). One of the casual mini-games involves usage of the tilting mechanism of the tablet pc, for steering an object. Progress is never dependent on gaming skills.'
Moreover, we made changes to the section 'mode of delivery' (page 9, lines 8-10), which now includes the following sentence: 'Staff members are available for consultation on accessing serious gaming (i.e. for technical issues and adaptation to special needs).'

1.3 Reviewer 1 commented:
'There are a few typos which I have highlighted below:
Page 2 line 10: comma required after 'circumstances'
Page 4 lines 10, 11 and 12: need to be reworded
Page 5 lines 30 and 31: 'Unfortunately, only small long-term improvements have been found in low back pain patients with severe and long-lasting pain and disability'. Needs an explanation about what is meant by 'small'.
Page 5 lines 33: comma required after 'such'.
Page 5 lines 35: could put 'exergaming' in quotation marks for consistency with the other terms used.
Page 7 from line18 to 48: 'process' should be 'process'
Page 15 line 46: should be 'in the development of LAKA'.

Response: Thank you for your diligent look at our language. The typos are corrected. Moreover, you

requested for an explanation about what is meant by ‘small’ in the following sentence: ‘Unfortunately, only small long-term improvements have been found in low back pain patients with severe and long-lasting pain and disability’. This sentences was changed (p. 3, lines 21-23): ‘Randomized controlled trials that compared health outcomes after multidisciplinary rehabilitation versus alternative treatments in patients with CP or FSS generally reported up to medium-sized differences’.

Our point-by-point response to comments of reviewer 2:

We are grateful for the compliments on the proposed mixed-method study, and for your hard work in making a list of minor issues. We think it helped a lot to improve the manuscript.

2.1 Reviewer 2 made the following suggestion on our abstract: (Intro) change current intervention into multidisciplinary rehabilitation.

Response: Thank you. We have adapted our terminology (p. 2, line 5).

2.2 Reviewer 2 what is meant by 220 complete records (methods)? Do you aim at including patients until you have 220 complete records, or do you want to include 220 patients? In method section of the article it is not described as such.

Response: Thank you for pointing at this potential source of confusion. Outcome data are received at once after all participants finished their rehabilitation programme. We calculated a required sample size of 212 patients (rounded up = 220), and that inclusion of 250 patients is necessary to have 220 full cases. We changed the phrase in our abstract (p. 2, line 12-14): ‘Multivariate mixed modelling analysis is planned for assessing how much variance in 250 patient records of routinely monitored pain intensity, pain coping and cognition, fatigue, and psychopathology outcomes is attributable to serious gaming.’

2.3 Reviewer 2 commented on the discussion within the abstract: I would not focus on effectiveness of this study, due to the chosen design but on aim 2 and 3. Discuss this topic.'
The suggestion was repeated later by the reviewer: ‘Perhaps focus on effectiveness should be less and two other goals are more important! The aim of effectiveness should be less strong and focus on development and process evaluation are more important in my opinion based on the chosen design. Perhaps the title of the manuscript should reflect this'

Response: Thank you, your suggestion is thought provoking and indeed worthwhile to discuss. With your suggestion, you are touching upon the essence of balancing interests in research and clinical practice. We as researchers have been prioritizing effectiveness, and had been seeking for optimal methods given actual practical limitations that emerged. In retrospect, increasing awareness of the impracticality of a randomized design, as well as the number of incontrollable factors, might have driven our focus more towards the questions of process. We understand (like no other) that the study will leave more uncertainty about effectiveness than needed under ideal conditions, but (in principle) we have not been perceiving this as an argument for adjusting our hierarchy of study goals and title of the manuscript. The editor requested us to delete the discussion part of the abstract, but this does not reduce the need for discussion on this point. In the discussion section, we added (p. 14, lines 20-24): ‘Still, conditional optimization of quasi-experimental methods is a legitimate strategy for obtaining evidence on the effectiveness of an intervention.79’, and ‘By the emergence of practical limitations, study strengths shift towards dealing with questions of process.’

2.4 Reviewer 2 commented: 'In the title of the manuscript the target population were also patients with fatigue complaints. In the introduction fatigue is never mentioned. What is meant by fatigue complaints; patients with fatigue syndrome or patients with chronic pain and fatigue? (a) Please make

in the introduction clear what population you are aiming at. (b) Perhaps title should be adapted (see also comment at section discussion). Also only the relative small effect of rehabilitation of low back pain is described. What about the effectiveness in other chronic pain problems? Is this a problem for all chronic pain populations or specific for low back pain?

Response: Thank you for addressing these irregularities in our manuscript. (a) Besides patients with various pain problems, the target population includes patients who are not in pain, but do have severe (CIS subjective fatigue > 36) chronic fatigue complaints. Target patients with chronic fatigue do not necessarily meet additional CDC-criteria for chronic fatigue syndrome. To answer your other questions and follow your suggestion, we elaborated the sentence in our introduction about effectiveness of multidisciplinary treatment for the target population (with additional references) (p. 3, lines 21-23): 'Randomized controlled trials that compared symptoms and functioning after multidisciplinary rehabilitation versus alternative treatments in patients with CP or chronic fatigue syndrome generally reported up to medium-sized differences.22-25' (b) In the title Instead 'chronic pain and chronic fatigue' is replaced by 'chronic pain or chronic fatigue' (p. 1, line 1).

2.5 Reviewer 2 commented: 'The introduction is well written and understandable. I miss however reasons why there are only small long term improvements of rehabilitation and why serious gaming specifically would be able to improve these small effects. (a) Please give a rationale or theoretical framework how serious gaming can fill in the gaps in traditional treatment. Now possible explanations are given how serious gaming could work. But what is the difference with more traditional methods? Are we not delivering the same treatment in a more fancy way?'
This point is repeated in the review by commenting on aim 3: 'mechanisms are described of serious gaming; please elaborate on these in the introduction. What mechanisms are known and how can they contribute to or enhance the existing mechanisms in traditional treatment (see comment theoretical frame work earlier)'.
This topic is also addressed in commenting on programme theory: 'This a very short description. (b) Please describe this in more detail so the reader can understand the rationale of the approach. Also combine this with the model behind pain rehabilitation and how you combine these models. What is common and what is complementary? Or how interacts this with the basic program?'

Response: Thank you for your request for more clarity throughout the article on our rationale for the addition of Serious Gaming for improving the effectiveness of multidisciplinary rehabilitation.
(a) Current literature provides limited insight into what distinctive features of serious gaming have to offer for rehabilitation beyond 'the same treatment in a more fancy way'. However, the literature does support the expectation that distinct qualities or degrees of patient engagement with treatment content could strengthen (moderate) the effect of (the same) intended treatment. First, we made the following change in the introduction to be clearer about existing areas of improvement (gaps) in traditional treatment (p.3, lines 23-26): 'Nonetheless, recent research addresses improvement of bio-psychosocial intervention models,26 27 'matching' and 'blending' therapeutic strategies and delivery modes,28 29 and promotion of patient engagement.30 As such, access, reach, adherence and effectiveness of bio-psychosocial interventions may be enhanced.'
Furthermore, we made a revision in the introduction to be clearer about what is known about distinctive mechanisms of serious gaming, and how they could 'fill' (some of the) current gaps or improvement opportunities (p. 3, lines 36-41): 'Features that distinguish serious games from traditional modes include covert learning techniques, interactivity, storytelling, sound effects, visuals, and 'debriefings'. They could offer relative benefits for behavioural change processes through distinctive attentional (presence), affective (enjoyment), and meta-cognitive processes.40-43 Further research into gaming mechanisms is needed,42 and may also inform about how biopsychosocial intervention mechanisms could be strengthened'.
(b) With regard to your suggestion to elaborate the 'programme theory' section, we want to stress that (at the point of drafting the protocol) it is premature to hypothesize about specific mechanisms and

interactions with context factors (i.e. other treatment components). Doing that is exactly the goal of qualitative data collection and analyses (before quantitative analyses) within the proposed study. To involve readers more in our preliminary thoughts on this issue, we made small revisions in the 'programme theory' section (p. 9, lines 24-27): ‘Besides by symptom categorization, serious gaming outcomes were interpreted by frameworks of rehabilitation mechanisms as self-improvements (see appendix).27 45 57 58 62 63 Two comprehensive implementation models are used for the classification of context factors, such as planning and compatibility relative to other treatment components.64 65’ Information is added about developer assumptions in the supplementary file: ‘Serious gaming sessions are planned after educational components (stress management and well-being, cognitive restructuring, and meditation) to enable complementary learning engagement and transfer’, and ‘Distinctive features of serious gaming strengthen (moderate) effects of behavioural change content on outcomes7’

2.6 Reviewer 2 proceeded comments about the rationale of serious gaming: ‘Why adding this to the treatment and why not provide it as a stand- alone treatment?

Response: Thank you for raising this understandable question. The true answer is simple: this choice was not for the authors to make. This is mentioned as a limitation on page 14, lines 6-7: ‘Instead, pragmatic considerations for the deployment of serious gaming during rehabilitation in two sites of a single Dutch centre led treatment allocation, recruitment, and data collection methods’. Serious gaming does not (and maybe cannot) contain all important elements of a stand-alone treatment (i.e. actual exposure) as needed by patients with an indication for multidisciplinary rehabilitation.

2.7 Reviewer 2 commented: 'The end of the introduction is about complexity of outcome evaluation. However for me it is not clear what the authors exactly mean. Specific the part of debriefing needs more explanation.

Response: We agree that clarity of end of the introduction could be enhanced. We did that as follows (p. 4, lines 8-14): ‘For example, some intervention studies show different outcomes of a computer delivered therapy when applied in different countries.10 This is also an important issue for serious gaming, because outcomes are clearly sensitive to context factors. 11 12 Therefore, ‘debriefings’ are suggested as a method for discussing and exploiting game-play experiences and strengthening learning outcomes.13 These studies leave uncertainties about how to effectively organize instructional support, i.e. via software or delivered by (trained) health care staff, via internet or face-to-face, in groups or individually.’

2.8 Reviewer 2 commented: 'The sentence: Indeed, more adequately …. (pag 5 sent 10): I don’t understand this according to the previous text. Of course adequately powered trials are always needed. What is the point?

Response: Thank you for alerting us that the point of this sentence is unclear. It was replaced with the following text that describes our point more accurately (p.4, lines 12-14): ‘There is strong consensus that adequately powered clinical trials are needed to determine the effectiveness of serious gaming.2 3 37’

2.9 Reviewer 2 commented on our study aims: Why so many primary outcomes? What is the specific target of rehabilitation? In my opinion is daily functioning/participation the main outcome. (a) Please explain why there are so many and so diverse outcomes. And why participation is not one of them. The same comment was repeated on interventions: 'The intervention focuses on participation. However this is not part of your primary outcome measures. (b) Explain this?'
Response: (a) In order to better address why we have chosen these (multiple) primary outcome measures, revisions are made in the ‘introduction’ (p. 4, lines 24-27): ‘Primarily, interdependent

outcome domains of pain, fatigue, and emotional functioning (pain intensity, pain coping and cognition, fatigue complaints, and psychological distress) are studied, because they are considered to be relevant and plausible for the intervention and population.27 45
(b) We choose not to mention participation as an outcome, because the validity of the only available measure is uncertain (a yet untested measure of participation was part of routine monitoring during the study), while the literature reached consensus on a set of outcomes in which participation has no prominent role (p. 14, lines 18-20). Therefore, we stated as a limitation: Different comparisons with serious gaming (i.e. usual care, waiting list, or text based computer-based intervention), more elaborate diagnostic assessment, and outcome measurements including role participation and long-term follow-up are precluded.'

2.10 Reviewer 2 requested: 'Please structure the information according to order as described in the aim; organization provider patient and innovation level. Give some examples of serious gaming features that is related to innovation levels. I don't understand the addition in brackets line 39 pag 5. Is this a methodological issue?'

Response: Thank you for these suggestions. We structured the information, added examples of innovation level factors, and the addition in brackets are removed, because it is not needed to illustrate the point here (p. 4, lines 29-32): 'Innovation level factors could be the quality of the serious game design, and its compatibility with user routines. Patient level facilitators or barriers could be demographic, health status and intervention history factors.'

2.11 Reviewer 2 asked a question about the Method and analysis section: 'Why is a quasi-experimental design chosen? As your aim 1 is effectiveness then a RCT should be chosen.'

Response: We are grateful for this question that enables to express how we think about this decision. After searching ways to implement an RCT in coordination with the clinic (e.g. a stepped-wedge design), we accepted that a quasi-experiment would be the optimal design under the given circumstances. This is deemed acceptable for an effectiveness study (but not for an efficacy study) (p. 14, lines 20-22): 'Still, conditional optimization of quasi-experimental methods is a legitimate strategy for obtaining evidence on the effectiveness of an intervention.79'

2.12 Reviewer 2 asked a question about recruitment: 'What is meant by absence of disruptive events planned?'

Response: Thank you. It was essentially the following (p. 7, line 7): 'For the recruitment of control subjects, two other sites (out of 18 sites as part of the same treatment centre) are selected based on similarity with regard to patient characteristics, facilities, protocols, history, personnel, location in or near a city in the southern Netherlands, and no other research projects planned during the intervention period.'

2.13 Reviewer 2 asked: 'What is the role of local stakeholders of serious gaming in the study'
Response: 'Providers', as the first group of stakeholders, are involved in;
- Intervention delivery as described in the manuscript under 'mode of delivery' (p. 9);
- Patient recruitment as described under 'Patients' (p. 7); and
- Focus group interviewing´ as described under 'Stakeholder (focus group) interviewing´ (p. 12). The latter section also describes that 'experts' and 'implementers' participated in focus groups before and after the experiment. A sentence was added for clarification about the role of local stakeholders during the experiment (p. 9, lines 13-15): 'For external validity, no specific roles were assigned to other local stakeholders for the delivery of serious gaming (i.e. to study 'natural' problem solving by implementers).

2.14 Reviewer 2 asked a question about patients: 'Why patients are invited as the completed half of their rehabilitation program. Why not at start of the treatment program?'

Response: Thank you for raising this valid question related to our pragmatic approach. This is possible without risking bias, because baseline data collected from patients by the clinic are foreclosed (see comment 2.17) and serious gaming can only affect treatment during the second half of the program. The reasons are: 1) We expected a higher response rate when recruiting before the second half of the program, because implications of participation in research may be more difficult to oversee for patients before the start of treatment given their unfamiliarity of program procedures, actual health conditions and priorities, and absence of a trust relationship with their caregivers at that time. 2) Recruiting clients during the second half of the program speeds up the study by two months, enabling to investigate the effects of serious gaming from the first time that clients receive it as part of the program. To conclude, we saw the chosen strategy as the most efficient way of recruitment.

2.15 Reviewer 2 asked a question about interventions: (a) The basic program is in accordance with individual care needs! How can you than evaluate effectiveness as content of the basic treatment is not uniform between patients? (b) Please explain this, also this should be addressed in the discussion'. In addition, comments were made about the discussion: 'Strengths and limitations are well discussed. However, the apparently diversities in the basic treatment (accordingly to patient's needs) is not discussed. This also has effect on internal validity, as the quasi experimental design has.

Response: You make a valid point. (a) First, we describe more precisely how the programme is tailored to individual care needs, we refined the sentence you refer to (p. 7, lines 31-32): 'The standardized 16-week programme consists of on average 95 hours of individual or group sessions that are organized in modules and centrally assigned to individual patients based on diagnostic findings.' Due to the centralized module allocation procedure, there is no reason to expect systematic differences in basic treatment between the study groups. We will be able to check this assumption with data about the basic treatment that is actually received. (b) We added to the discussion (p. 14, lines 22-23): 'Apparent confounding factors (i.e. differences in usual treatment received) should therefore be controlled for by appropriate methods.'

2.16 Reviewer 2 asked a question about mode of delivery. What is exact purpose of debriefing sessions? Please provide this information

Response: This information is added (p. 9, lines 12-13): 'The goal of the debriefings was to discuss experiences of game play, technology acceptance and learning, and facilitate learning transfer to daily life.'

2.17 Reviewer 2 asked a question about measures: 'What is the exact starting point of the study. At page 7 it is stated that patient are invited after completion of half the program. Yet you already measure at baseline. On what data the effects of the treatment are evaluated? In table 2 baseline measurements seems included. Or have I interpreted the information on page 7 incorrectly.

Response: This is a logical confusion, because there is a missing piece of information that crucial to clarify this. All outcome and case-mix data, including those collected at baseline, are used to evaluate the effects of treatment. However, we do not collect the data from patients ourselves, and have no access to data that are stored in patient records. Baseline data are useful for determining if there is a parallel trend in outcome improvement from baseline to intermediate. We made a revision for clarification (p. 10, lines 1-2): 'Outcome and case-mix variables are retrieved from routinely administered clinical patient records after all participants have completed their rehabilitation programme.'

2.18 Reviewer 2 asked a question about measures: 'How standardized are case mix variables measured and registered. How can you ensure adequate quality of these data?'

Response: We agree that more information about the measurement and registration of case-mix variables is useful. We revised the following lines (p. 10, lines 2-4): ‘All patient variables are collected by the clinic through a standardized and secured web-surveying procedure, including facilitation of patients without convenient computer access and promotion of follow-up completion.4 70’ Measurement and registration of case-mix variables is completely standardized. Moreover, we added information about measurement (between brackets) in table 3 (starting at p. 10, line 3). The process of data registration is audited on a yearly basis by independent auditors on stability, transparency, control and security following the internationally recognized ISAE 3402 protocol. Assuming that the data obtained by these simple measures are valid, quality of the extracted data was good in our previous feasibility study.

2.19 Reviewer 2 suggested that information should be given about abbreviations in table 4
Response: Of course, the following information about abbreviations is added below table 4 (p. 11, lines 15-16):
‘*1 Unified theory of acceptance and use of technology
*2 Positive and negative affect scale – short form’

2.20 Reviewer 2 asked a question about qualitative data: 'I don’t understand the role of the interviewers towards more educational …. …What do you mean or what is the aim for this. For evaluation the interviewer must be neutral and open and explorative but not directional or educational in my view. So please explain what the exact role and aim is.

Response: We understand that we have to present this point more clearly, because this recommended approach to interviewing in realist evaluation (see reference number 72) indeed differs from usual practice in qualitative research (p. 11, lines 23-25): ‘Accordingly, the interviewer starts with an open and explorative style, but may sometimes take an explanatory role to raise discussion about programme theory elements when CMO configurations become better delineated.’ With regard to the aim of it, the ‘process analysis’ section (p. 13, lines 4-13) describes that interview findings are part of the research phase of hypothesis generation and refinement rather than testing.

2.21 Reviewer 2 asked a question about power calculation: 'On which primary outcome the power analysis was performed?'

Response: Thank you. Your comment indicates that power calculations should be presented more accessibly. Power calculation was not performed on a particular primary outcome measure, because it was based on a multivariate test (MANOVA test of global effects)21. We based our effect-size estimate on a previous meta-analysis, because available pilot data were strongly affected by self-selection. Based on your question, we share the following information with our readers (p.13, lines 28-29): ‘By the same standards, it was checked if the determined sample size would also be sufficient for independent univariate tests of variance on each of the primary outcomes.’

VERSION 2 – REVIEW

REVIEWER	Dr Jacqui McKechnie Glasgow Caledonian University Scotland UK
REVIEW RETURNED	26-Apr-2017

GENERAL COMMENTS	I am happy that the authors have taken on board the comments from the first review and I am happy to recommend that this be accepted. I wish the authors all the best with their study and look forward to reading about it in due course.

REVIEWER	Albère Köke Maastricht University, the Netherlands
REVIEW RETURNED	19-Apr-2017

GENERAL COMMENTS	The authors have addressed the questions and comments on the first version in a sufficient way and therefore in my opinion the article can be published