We appreciate the opportunity to revise and improve the manuscript. Editors' and reviewers' comments are listed in italics, and our response is listed immediately thereafter. Changes made, or parts of the document relevant to our responses are copied, with reference page numbers, paragraphs and line numbers.

Editors' comments:

1. *Might you add a few words on what this adds to the recent SR (Weisman A et al. Effect of artificial pancreas systems on glycaemic control in patients with type 1 diabetes: a systematic review and meta-analysis of outpatient randomised controlled trials. Lancet Diabetes Endocrinol. 2017;5:501-512).*

Authors' response:

In our discussion we introduced an additional paragraph titled Relation to previous studies, where we moved existing relevant text but also added more details describing the main advantages of our work compared to the systematic review of Weisman et al, which are the following:

- Inclusion of a considerably larger pool of trials (n=41) and patients (n=1042), some of which were identified from grey literature sources (hence minimizing the risk for publication bias), as opposed to Weisman et al (24 trials and 585 patients).

- Assessment of more outcomes.

- By conducting separate analyses based on all four combinations in terms of duration of outcome measurement (24h or overnight) and time period that the intervention was used (throughout 24h or solely overnight), our analysis provides answers to the following clinical relevant questions:
  - how efficacious is AP throughout the whole day when used throughout the whole day?
  - how efficacious is AP throughout the whole day when used only overnight?
  - how efficacious is AP throughout the night when used throughout the whole day?
  - how efficacious is AP throughout the night when used only overnight?

  On the other hand, Weisman et al, analysed only 24h outcomes for studies that used the intervention for 24h and only overnight outcomes for studies using the intervention overnight, even when individual trials provided data for both periods. On this account, their systematic review addresses only two of the questions mentioned earlier:
  - how efficacious is AP throughout the whole day when used throughout the whole day?
  - how efficacious is AP throughout the night when used only overnight?

  Hence, our analysis addresses the topic in a more comprehensive and detailed manner by providing answers to two additional clinical questions, which we believe are equally important in order to draw well-rounded conclusions regarding choice of intervention (overnight only or throughout the day) for every individual patient.

- Finally, there are certain methodological considerations regarding handling of cross-over trials and median values by Weisman et al (summarised in a relevant reference in the main text), which have been addressed differently in our analysis.

2. *Could you update the search to help differentiate this from the Lancet review? As it is, the search is one year old now.*

Authors' response:

Following the editors' recommendation, we have updated our search which has identified 7 additional eligible trials (250 additional patients). As a result, we have now incorporated data from 41 studies (1042 participants), which constitutes a significantly larger pool of evidence compared to the meta-analysis by Weisman et al (24 studies, 585 participants). On this ground, we believe that, in addition to its clinical relevance and importance, our work contributes significantly to preexisting knowledge on artificial pancreas systems.

3. *Is 'Closed-loop insulin therapy' the best term to use? Others use the term 'artificial pancreas'.*

Authors' response:

Both terms, as well as the term "bionic pancreas", are used to describe the same devices. Considering the editors' comment we believe that readers would indeed be more familiar to the term "artificial pancreas" rather than "closed-loop" (which is more of a technical term). Therefore, we updated our manuscript (title, text, figures and appendix), and now use consistently the term "artificial pancreas".

4. *The paper only looks at glucose levels, rather than complications, clinical outcomes and doesn't tell us much about safety.*

Authors' response:

We focused on practical outcomes related to glycaemic control and glucose levels, based on current guidance that advocates use of such outcomes in trials evaluating artificial pancreas systems. We did not assess additional clinical and safety endpoints (such as quality of life or ketoacidosis) or outcomes related to technical issues, based on our protocol, relevant guidance and inconsistent reporting in individual reports. We do consider this as a limitation of our systematic review and report these concerns in the relevant section in the discussion. However, with regard to safety outcomes, we evaluated incidence of severe hypoglycaemia, which is considered a clinically important outcome. Moreover, in our "Implications" section we highlight the need for further research evaluating the impact of artificial pancreas systems on patients' quality of life.

5. *One could read this entire paper without having any idea of exactly how much time (in hours or minutes) people with these systems spend in the desired glucose range. Table 2 makes it look like this is a very small proportion of time.*

Authors' response:

We agree that converting time percentages to actual hours (and minutes) would facilitate overall interpretation and add informative value to our findings. Thus, in our Results we now also report weighted mean differences (for artificial pancreas against control) in hours and minutes for the three key outcomes (time spent in normoglycaemia, hyperglycaemia and hypoglycaemia throughout the day). In addition, we have slightly modified the first paragraph in the Discussion to incorporate these values when summarising our findings. Of note, numerical values in table 2 may seem low, but this is because they represent mean differences in % of time between artificial pancreas (AP) (single- or double-hormone) and comparator, rather than mean values for each individual arm. To avoid any misinterpretations, we have updated headings in table 2 to reflect that values presented are WMD vs control arm. It should be underlined that we present mean differences between AP and control rather than mean values for each arm (AP or control), based on the Cochrane Collaboration recommendations

and on additional relevant guidance (1) for handling of data from crossover trials (see also our reply to Reviewer 4, comment 3 in our previous submission).

6. *All of the included studies seem to compare these systems to another type of pump - are there no studies comparing these systems to daily injections? Wouldn't that be a relevant comparison? Editors would like to know how much value is added (in addition to convenience) with this technology.*

Authors' response:

Multiple daily injections (MDI) would indeed be a relevant comparator to artificial pancreas. Therefore, MDIs were included among eligible comparators in our systematic review, as we describe in our Methods (Search strategy and selection criteria) and in our protocol. However, we did not identify any RCT comparing an artificial pancreas system with MDI.

Yet, lack of such trials is not surprising, given that prior RCTs have already compared MDI with continuous subcutaneous insulin infusion (CSII)/insulin pump therapy, which represents the early stage in the development of sensor augmented pump (SAP) therapy and subsequently of artificial pancreas. In fact, meta-analyses and individual RCTs suggest a favourable effect of insulin pump (CSII) compared to MDI in terms of glycaemic control, quality of life and treatment satisfaction. In addition, artificial pancreas can further reduce burden for patients compared to insulin pump or SAP, by automatically adjusting insulin infusion rate based on sensor glucose values. On this ground, it seems reasonable that available RCTs assess the effectiveness of artificial pancreas in comparison to either CSII/insulin pump or SAP and not versus MDI (which had already been assessed against insulin pump therapy).

Based on the above, we have added the following phrases (and respective references) in the introduction:

- "Moreover, a meta-analysis of 19 trials concluded that CSII has a favourable effect on glycaemic control in adults with type 1 diabetes compared to MDI, while a recent cluster randomised controlled trial demonstrated that patients with type 1 diabetes that used CSII instead of MDI reported additional benefits in quality of life and greater treatment satisfaction."
- "Therefore, compared to insulin pump or SAP therapy, the artificial pancreas can reduce burden for patients by automatically adjusting the amount of insulin entering the body based on sensor glucose levels."

Additionally, we thought that readers would be interested to know that SAP was cost-effective compared to insulin pump in a cost-effectiveness analysis in the U.K. Thus, we also cite this study in our introduction alongside a short phrase reading: "Interestingly, based on a cost-effectiveness analysis published in 2016, the use of SAP plus LGS was found to be cost-effective compared to CSII plus self-monitoring of blood glucose for patients with type 1 diabetes in the United Kingdom."

Finally, with regards to convenience and ease-of-use of artificial pancreas, we now highlight in our discussion its favourable effect based on our sensitivity analysis of trials under free-living conditions without remote monitoring. In particular, we modified previous text to read: "Results were robust both for single- and dual-hormone systems, and were consistent in all sensitivity analyses performed, including an analysis restricted to trials under free-living conditions without remote monitoring, supporting the convenience and ease-of-use of artificial pancreas systems".

Reviewer Comments:

Reviewer #1:

Comments:

I am satisfied that my previous comments have now been addressed.

Authors' response:

We are pleased that we have responded adequately to reviewer's comments and thank her for her contribution in improving our manuscript.

Reviewer #2:

The authors have responded very well to my comments. I still find the main outcomes about percentage of time in a particular range to be sub-optimal, but this is a limitation of the original studies (I think it should be noted more by the authors though as a limitation in this review, and other approaches in new studies like a repeated measures analysis of the actual value over time would be preferred). So, it difficult to address otherwise without the raw data. I only have minor additional things to say, as the paper is very well written and the methods are well described and chosen:

Authors' response:

We are happy that we have adequately addressed the reviewer's previous comments and thank him for his practical and thoughtful suggestions.

We agree that open access to raw data from primary research would allow for additional analysis approaches and facilitate optimization of more clinically relevant outcome measures in the research field of artificial pancreas. We believe that it would be more suitable to address this issue in the section related to the implications for future research, rather than the limitations section. Therefore, the relevant text in our discussion now reads: "To maximise yield of information and to facilitate analysis and synthesis of the totality of evidence, it may be important to agree on the use of a common individual patient data repository. Such repositories would facilitate free dissemination of raw trial data, allowing for replication of previous research findings utilising various analysis approaches (for example a repeated measures analysis) of clinically relevant outcomes".

Minor

1. *Move the results about small study effects to the end of the results (currently it comes before the main meta-analysis results, yet discusses how the meta-analysis results would change if 'missing' studies were included)*

Authors' response:

Indeed, it is more reasonable to report results about small study effects after our outcome findings. We thank the reviewer for pointing this out and have moved the relevant section at the end of the results.

2. *Change "Of note, 95% prediction intervals were statistically significant when closed-loop was used overnight (3.97 to 11.62) suggesting that closed-loop will be beneficial in at least 95% of the individual study settings when applied overnight, but not when applied throughout 24h (-6.14 to 27.06)."*

   To "Of note, 95% prediction intervals were entirely above zero when closed-loop was used overnight (3.97 to 11.62), suggesting that closed-loop will be beneficial in at least 95% of the individual study settings when applied overnight; however, the prediction interval contained negative values when applied throughout 24h (-6.14 to 27.06), and so in some settings may not be beneficial"

Authors' response:

We have rephrased this section according to reviewer's suggestion. Of note, numerical values have been updated to reflect our results based on the updated literature search.

3. *The numbers in the brackets are often given without any explanation. E.g. "…was decreased by 8.32% (5.10 to 11.53, 84%, 36.43, 17 studies)" – what does 84%, 36.43, etc mean? Please address throughout.*

Authors' response:

Following the reviewer's recommendation, we now provide explanations for numbers presented in brackets more consistently in our Results. More specifically, in a similar manner to a recent meta-analysis published in the BMJ(2), explanations are given at the start of each separate paragraph.

4. *Be consistent in using negative values when something is reduced. For example, in the following the sign keeps changing even though the effect is (I think) in the same direction: "Use of closed-loop had a favourable effect on time spent in hyperglycaemia (> 10 mmol/L) during the whole day which was decreased by 8.32% (5.10 to 11.53, 84%, 36.43, 17 studies) compared to control, both in trials where closed-loop was used only overnight (−6.51%, −9.42 to −3.60, 0%, 0.0, two studies), and in trials using closed loop throughout 24h (−8.62%, −12.41 to −4.84, 86%, 45.87, 15 studies)"*

Authors' response:

In order to be consistent when reporting negative values, we have rephrased the sentence to read: *"Use of artificial pancreas had a favourable effect on time spent in hyperglycaemia (> 10 mmol/L) during the whole day which was decreased by approximately two hours (overall WMD -8.52% 95% CI -11.14 to -5.9 , I^2 80%, Tau2 28.98, 22 studies) compared to control, both in trials where artificial pancreas was used only overnight (WMD -6.0%, -8.4 to -3.6 , 0%, 0.0, three studies), and in trials using artificial pancreas throughout 24h (-9.08%, -12.23 to -5.93, 83%, 37.53, 19 studies)"*.

5. *Appendix 8, the contour enhanced funnel plot's regions (e.g. the white region) looks very odd indeed. Is this actually correct? I think there is an error.*

Authors' response:

We appreciate the reviewer for bringing this to our attention. Indeed, in our previous submission the contour funnel plot figure was distorted, even though the original figure generated by Stata was correct. We assume that this distortion occurred when we converted the appendix from .doc to .pdf format. We now provide a rectified and updated contour funnel plot in appendix 19.

References

1. Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. International journal of epidemiology. 2002;31(1):140-9.
2. Lopez-Lopez JA, Sterne JAC, Thom HHZ, Higgins JPT, Hingorani AD, Okoli GN, et al. Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis. Bmj. 2017;359:j5058.