PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effects of Novel Flash Glucose Monitoring System on Glycemic Control in Adult Patients with Type 1 Diabetes Mellitus: Protocol of a Multicenter Randomized Controlled Trial
AUTHORS	ZHOU, YONGWEN; Deng, Hongrong; Liu, Hongxia; Yang, Daizhi; Xu, Wen; Yao, Bin; Yan, Jinhua; Weng, Jianping

VERSION 1 – REVIEW

REVIEWER	Mohammed Aldawish, MD, FRCP, FACE, CCD Prince Sultan Military Medical City Kingdom of Saudi Arabia
REVIEW RETURNED	25-Apr-2020

GENERAL COMMENTS	page 8: use endpoints as mentioned in table 2(mainly TIR & TIT)

REVIEWER	R Ajjan University of Leeds, UK I declare that I have conducted studies with Freestyle Libre system before but do not feel this represents a conflict of interest.
REVIEW RETURNED	30-Apr-2020

GENERAL COMMENTS	This manuscript describes the study protocol of a multicentre open label randomised controlled trial comparing the effects of glucose monitoring using Freestyle Libre with self-monitoring of blood glucose (SMBG) in 76 individuals with type 1 diabetes with inadequate glycaemic control (HbA1c between 7-10%). The primary end point is difference in HbA1c between two study arms with a host of secondary end points including TIR, TIT, TAT and TBT as well as various measures of glucose variability. I have the following comments: 1. Exclusion criteria include end stage renal failure but in a different item the authors mention eGFR<45ml/min. Does this mean that even milder forms of renal dysfunction represent an exclusion criterion? Also, can the authors be more specific about retinopathy? Would proliferative stable disease successfully treated with photocoagulation represent an exclusion criterion? 2. What is the rationale for only including those with HbA1c between 7-10%? The authors may wish to provide an explanation. 3. Other than giving the participant a sensor, what kind of education will be provided? What do the authors mean by general diabetic education? More details are required. 4. Is the primary end point at week 14 or 26? One should be primary and the other secondary. 5. Sample size is a concern. I am not a statistician but according to my calculations, and using the data provided by the investigator,

	the study has less than 55% power to detect the specified change in HbA1c. I suggest the investigators seek professional statistical input and provide more details in the power calculation section (which benefit from a rewrite anyway). It would be helpful to provide reference(s) for the numbers used to determine SD of HbA1c. 6. There is a large number of glucose variability measures planned. It would perhaps be helpful to provide a brief explanation for the reasons that all these measures are required. 7. Will the investigator review sensor data during follow ups and advise on treatment adjustment? 8. According to the investigators, funding sources can be found on page 10 but I have been unable to find this. On page 12, investigators mention that the study is supported by the National Key R&D Program of China. If this the funder, the investigators need to make this more clear. Also, how were the sensors provided (purchased or provided for free by the company)? 9. It would have been nice for the investigators to include a brief discussion, highlighting the importance of the study, how it would add to the literature, how it may affect clinical management, strengths and weaknesses…etc… 10. Minor points a. There are various mistakes in the English language and the manuscript would benefit from a review by a native English speaker. b. Given number of centres and relatively small sample size, study recruitment period is relatively long. Can the investigators comment on this?

REVIEWER	Ben Wheeler University of Otago, New Zealand
REVIEW RETURNED	21-Jun-2020

GENERAL COMMENTS	Thank you for the opportunity to review this protocol pamanuscript discussing a proposed (although currently running) RCT of isCGM vs SMBG in sub-optimally controlled adults with Type 1 Diabetes. The study data will be of interest to the diabetes community and more studies on this topic are of value. I have the following comments regarding the manuscript: 1) Scientific English throughout needs editing prior to publication. I strongly recommend this is done by a native English speaker skilled in scientific english. 2)Dual report HbA1c in mmol/mol and % 3) Is there stratification for any variables? e.g. by study site (multi-site study) If not - why not? 4) Clarify which system glycaemic metrics will be taken from - iPro or isCGM system? (as both appear to be collected). Important not mixed. Also a discussion of why not using Libre Pro could be mentioned. 5) Discussion of cutaneous adverse events should reference recent systematic review in J. Diab Science and Tech, not just one industry funded short duration study.

	6) Sample size calculation - needs more details. no data referenced as to how this sample size was determined. S.D. of HbA1c within their proposed population, or taken from other data sets etc. Sample size appears too small to my eye to detect a 0.4% difference in HBA1c between groups. This needs to much more carefully discussed - and references explaining, and details given so this calculation could be checked externally. 7) I note the paper has no discussion at all. Ideally would have a few paragraphs highlighting the importance of the study, strenghts and limitations. 8) I note multiple papers have discussed CGM and isCGM use in sub-optimally controlled populations - most of these have not been referenced in this manuscript - I think bit more thorough interaction with the literature required. Speific comments: "professional" CGM mentioned multiple times - I assume this means blinded? TIR vs TIT - 3.9 - 7.4 is not the recomemnded metric for analysis . Usually 3.9-10 Page 4, line 15 - I would not recommend using the phrase "and so on" in a scietific paper. page 8 line 14 (and in other places) mentions "biomedical metrics" - what are these? Page 8 lines 50 - 55 - hypoglycaemic events need some clarification here. Tense - past/present/future needs checking carefully throughout. Page 10 - aspects on "professional" CGM - again this means blinde dI assume. Worth putting choice of this device in context with what was available at study commencement.

VERSION 1 – AUTHOR RESPONSE

Responses to the comments of Reviewer #1
Reviewer Name: Mohammed Aldawish, MD, FRCP, FACE, CCD
Institution and Country: Prince Sultan Military Medical City, Kingdom of Saudi Arabia

Comments:
page 8: use endpoints as mentioned in table 2(mainly TIR & TIT)
Response:
Thank you for your kind comments and for bringing this to our attention. We apologized for the wrong description of the endpoints. We had corrected the mistake and state all prespecified secondary outcomes in the revised manuscript (see Endpoints section, Page 6).

Responses to the comments of Reviewer #2
Reviewer Name: R Ajjan
Institution and Country: University of Leeds, UK
Please state any competing interests or state 'None declared': I declare that I have conducted studies with Freestyle Libre system before but do not feel this represents a conflict of interest.
Comments:
This manuscript describes the study protocol of a multicentre open-label randomised controlled trial comparing the effects of glucose monitoring using Freestyle Libre with self-monitoring of blood glucose (SMBG) in 76 individuals with type 1 diabetes with inadequate glycaemic control (HbA1c between 7-10%). The primary end point is difference in HbA1c between two study arms with a host of secondary end points including TIR, TIT, TAT and TBT as well as various measures of glucose variability. I have the following comments:

1. Exclusion criteria include end stage renal failure but in a different item the authors mention eGFR<45ml/min. Does this mean that even milder forms of renal dysfunction represent an exclusion criterion? Also, can the authors be more specific about retinopathy? Would proliferative stable disease successfully treated with photocoagulation represent an exclusion criterion?
Response:
Thank you for your comment and for raising this particularly important point on the exclusion criteria. As for your concern about the renal dysfunction, we admitted the mistakes leading to your confusion. Our trial aims to evaluate the effect of a new device for glycemic control in T1DM patients with suboptimal glycemic control. Therefore, we used the change of HbA1c level as our primary endpoint and the CGM-related metrics as the secondary endpoints. In this regard, any factor that influenced the HbA1c value should be excluded such as the serious diseases which might lead to acute metabolic disorders. However, for participants with mild forms of renal dysfunction or those with successful treatment on the proliferative stable disease as you mentioned, we admitted that they should not be excluded into recruitment especially those with stable and successful treatment. Therefore, considering the reasonability, we had removed the second item and used the eGFR<45ml/min/m^2 as the exclusive criteria (see Table 1. Inclusive and exclusive criteria section, Page 13)

2. What is the rationale for only including those with HbA1c between 7-10%? The authors may wish to provide an explanation.
Response:
Thank you for your question.
This trial aims to evaluate the effect of flash glucose monitoring system in adult patients with suboptimal glycemic control. Therefore, the enrolled adult participants in this trial should have poor glycemic control which is usually defined as HbA1c≥7% according to the guidelines of the American Diabetes Association and the Chinese Diabetes Society. The upper bound of the HbA1c level we chose refers to the previously published randomized clinical trials including the DIAMOND study (Beck RW, et al. JAMA. 2017;317(4):371-8) and the JDRF study (Tamborlane WV, et al. The New England journal of medicine. 2008;359(14):1464-76).

3. Other than giving the participant a sensor, what kind of education will be provided? What do the authors mean by general diabetic education? More details are required.
Response:
Thank you for your comment and we appreciate your attention to the general diabetic education. For each participant regardless of the group distribution, we will provide each patient with a Chinese version education brochure (English version is provided in uploaded SUPPLEMENT.2) about the general diabetic education including the self-management guideline when facing with hypoglycemia or hyperglycemia and therapy adjustment guideline. All of the guidelines in this brochure we made were according to the recommendations by the American Diabetes Association and the Chinese Diabetes Society.

4. Is the primary endpoint at week 14 or 26? One should be primary and the other secondary.
Response:
Thank you for your comment.
Following your comment, we had corrected the mistake in table 2 in the revised manuscript (see Table 2. Endpoints, Page 14). As we mentioned in the Endpoints Section (Page 6), the primary endpoint is the change in HbA1c level from baseline to week 26.

5. The sample size is a concern. I am not a statistician but according to my calculations, and using the data provided by the investigator, the study has less than 55% power to detect the specified change in HbA1c. I suggest the investigators seek professional statistical input and provide more details in the power calculation section (which benefit from a rewrite anyway). It would be helpful to provide reference(s) for the numbers used to determine SD of HbA1c.
Response:
Thank you for your comments and we appreciate your attention to the study's sample size. As this study is the first randomized clinical trial to evaluate the effect of FGM system among adult T1DM patients with sub-optimal HbA1c, the mean change and standard deviation of the HbA1c used for the sample size calculation is based on the summary of the published randomized clinical trials about continuous glucose (CGM) systems and the system reviews about the FGM systems. We have cited the references in the revised manuscript. The summary points are as follows:

- Randomized clinical trials about CGM systems:
  1. In the JDRF study (Tamborlane WV, et al. The New England journal of medicine. 2008;359(14):1464-76) which evaluated the effects of different kinds of CGM systems in patients with T1DM, the change of HbA1c from baseline among patients aged ≥25 years old in the intervention group is -0.5(0.56)%, with the significant mean difference in change of 0.53% (95% confidence interval [CI], −0.71 to −0.35) compared with the change from baseline in the control group(-0.02[0.45]%).
  2. In the Gold study (Lind M, et al. JAMA. 2017;317(4):379-87) which evaluated the effect of CGM in adults with T1DM treated with multiple daily insulin injections (aged≥18 yrs), the mean difference in the change of HbA1c from baseline between two groups is significant (mean difference, -0.43%,[95%CI,-0.57% to -0.29%]) and the respective change of HbA1c is from 8.35(0.9)% in baseline to 7.92(0.8)% in the invention group.
  3. The DIAMOND group (Beck RW, et al. JAMA. 2017;317(4):371-8) determined the effectiveness of CGM in adults with T1DM treated with insulin injections(Aged≥25 yrs), the mean reduction in HbA1c level from baseline was 1.0(0.8)% at 24 weeks in the CGM group and 0.4(0.7)% in the control group, respectively, and the adjusted treatment group difference in mean change in HbA1c level was –0.6%(95% CI, –0.8% to –0.3%; P < .001).
- System reviews about the change of HbA1c after using FGM and CGM systems:
  1. The meta-analysis of all identified studies in T1DM patients (34 studies comprising, 5466 participants) indicated the use of Flash GM for 2 to 24 months was associated with an

estimated HbA1c reduction from baseline of 0.4% (Gordon I, et al. Diabetes research and clinical practice. 2020:108158).

2. The meta-analysis of clinical trials and real-world observational studies reported that the overall mean change in HbA1c among adult patients with diabetes was -0.56% (95%CI -0.76, -0.36). （Evans M, et al. Diabetes therapy: research, treatment, and education of diabetes and related disorders. 2020;11(1):83-95)

Based on the literature reviews presented above, the mean difference change of HbA1c between groups is from -0.08 to -0.6% among patients aged ≥18 yrs, and the range of SD of HbA1c is from 0.61 to 0.9%. Considering the referenced articles above was mostly conducted on the participants aged more than 25 yrs and the younger inclusive age might lead to the lesser difference between groups as presented in the JDRF study, we chose the 0.4% as the mean difference and 0.8 (close to the summarized mean value) as the SD of HbA1c for the calculation of sample size. We acknowledged that it would be better to provide the references for the numbers used to determine SD of HbA1c and following your advice, we had added the references in the revised manuscript (see Sample size section, Page 8)

6. There is a large number of glucose variability measures planned. It would perhaps be helpful to provide a brief explanation for the reasons that all these measures are required.
Response:
Thank you for your question and kind suggestion.
Different from the HbA1c, the metrics of glucose variability can reflect the within- and between-day glucose variation, which can make the evaluation of the FGM system more comprehensive. The CGM metrics presented in the revised manuscript are the key metrics recommended by the International Consensus on Use of Continuous Glucose System (Danne T, et al. Diabetes Care. 2017;40(12):1631-40), which is necessary enough to present the overall profiles of glycemic control (see Table 2. Endpoints section, see Page 14).

7. Will the investigator review sensor data during follow-ups and advise on treatment adjustment?
Response:
Thank you for your comments.
We are sorry for not describing it clearly and we had made some corrections in the revised manuscript (see Follow-up visits (week 12-14 and week 24-26) section, Page 5). According to the design of our trial, on the day that participants finished their two-week sensor data collection in the respective period, the investigators will download and review the sensor data and give advice on treatment adjustment referring to the standard guidelines, usually at the end of week 2, week 14 and week 26. And during the gap period between two follow-ups, any advice on treatment and insulin adjustment from investigators was not encouraged unless there is an adverse event reported.

8. According to the investigators, funding sources can be found on page 10 but I have been unable to find this. On page 12, investigators mention that the study is supported by the National Key R&D Program of China. If this the funder, the investigators need to make this more clear. Also, how were the sensors provided (purchased or provided for free by the company)?
Response:
Thank you for your comment and your recommendation on providing more details about the funding sources.
This trial is funded by the National Key Research and Development Program of China(2017YFC1309600) and the sensors were purchased from this grant. We declared that the Bayer Company, Medtronic Company, and the Abbott Diabetes Care are not involved in carrying out the trial, data analysis, data management, and publication. We had made a clearer explanation in the revised manuscript (see FUNDING STATEMENT section, page 9).

9. It would have been nice for the investigators to include a brief discussion, highlighting the importance of the study, how it would add to the literature, how it may affect clinical management, strengths, and weaknesses…etc…
Response:
Thank you for your kind suggestion and we do appreciate your kind recommendation on the dimension of the discussion. We had added a few paragraphs discussing the importance of the study, strengths, and limitations (see DISCUSSIONS section, Page 8).

10. Minor points
a. There are various mistakes in the English language and the manuscript would benefit from a review by a native English speaker.
Response:
Thank you for the kind suggestion.
We have gone over our manuscript repeatedly manually and using the several spelling and grammar checking tools, and has modified some mistakes in the text. Regarding the fact that we are not native speakers, we appreciate your kindness to give more detail regarding grammar to help us to improve. we appreciate your kindness to give more detail regarding grammar to help us to improve.

b. Given number of centres and relatively small sample size, study recruitment period is relatively long. Can the investigators comment on this?
Response:
Thank you for your concern about the study recruitment duration.
The calculation of sample size is based on the results from the published randomized clinical trials on the use of CGM systems (please see the response to Comment #5). The involved centers in our trial are the participating centers in the project (2017YFC1309600) funded by the National Key Research and Development Program of China. Initially, according to our research experience, the recruitment duration we set was 12 months. However, the trial has been required to suspend for some time due to the COVID-19 isolation, which might lead to difficulties in recruitment and the potential loss of follow-up. Therefore, to make sure of the quality of the trial, we prolonged the recruitment time to the end of 2020.

Responses to the comments of Reviewer #3
Reviewer Name: Ben Wheeler
Institution and Country: University of Otago, New Zealand
Please state any competing interests or state ‘None declared’: None declared.

Comments
Thank you for the opportunity to review this protocol manuscript discussing a proposed (although currently running) RCT of isCGM vs SMBG in sub-optimally controlled adults with Type 1 Diabetes. The study data will be of interest to the diabetes community and more studies on this topic are of value. I have the following comments regarding the manuscript:
1) Scientific English throughout needs editing prior to publication. I strongly recommend this is done by a native English speaker skilled in scientific english.
Response:
Thank you for your kind suggestion.

We have gone over our manuscript repeatedly manually and using the several spelling and grammar checking tools, and has modified some mistakes in the revised manuscript. Regarding the fact that we are not native speakers, we appreciate your kindness to give more detail regarding grammar to help us to improve.

2)Dual report HbA1c in mmol/mol and %
Response:
Thank you for bringing it to my attention. We uniformly reported HbA1c in % in the revised manuscript.

3) Is there stratification for any variables? e.g. by study site (multi-site study) If not - why not?
Response:
Thank you for your question. In the randomization section, we did not use any stratification such as the study sites. The randomization method we used is simple randomization with the random order generated by the software and arranged into the sealed and opaque envelopes. There will be an independent researcher in charge of the envelope distribution (Details see Randomization section, Page 5). As the people with access to the envelopes are distinct from those recruiting investigators to the trial, this method is thought to be reasonable and be able to reduce and eliminate the bias risk (Clark L, et al. BMJ. 2016;355.). Therefore, considering the availability in reality, we decided to use this method.

4) Clarify which system glycaemic metrics will be taken from - iPro or isCGM system? (as both appear to be collected). Important not mixed. Also a discussion of why not using Libre Pro could be mentioned.
Response:
Thank you for your question. The glycemic metrics analyzed in the statistics for endpoints are derived from Ipro2® at baseline (0-2 weeks), 12-14 weeks, and 24-26 weeks. To realize the effect of the isCGM system, the glycemic metrics from the isCGM system will be also collected (methods see Follow-up visits section, Page 6). To avoid the confusion, we had added more details in the revised manuscript (see Endpoints section, Page 6, and the annotation in Table 2. Endpoints, Page 14).
As for not using Libre Pro in this trial, it is because there is no approval of the Libre Pro used in China until the study commencement and the protocol preparation. Therefore, only we could use for “blinded” data collection is the Ipro2® system.

5) Discussion of cutaneous adverse events should reference recent systematic review in J. Diab Science and Tech, not just one industry funded short duration study.
Response:
Thank you for your kind suggestion and your recommendation on this very recently published study. We had added more details in the Risks and adverse events (AEs) section (see Page 6).

6) Sample size calculation - needs more details. no data referenced as to how this sample size was determined. S.D. of HbA1c within their proposed population, or taken from other data sets etc. Sample size appears too small to my eye to detect a 0.4% difference in HBA1c between groups. This needs to much more carefully discussed - and references explaining, and details given so this calculation could be checked externally.
Response:
Thank you for your comments and we appreciate your concerns on the sample size in our study. This study is the first randomized clinical trial to evaluate the effect of FGM system among adult T1DM patients with sub-optimal HbA1c, the Mean change and SD of the HbA1c used for the sample size calculation is based on the summary of the published randomized clinical trials about continuous

glucose (CGM) systems and the system reviews about the FGM systems. We have cited the references in the revised manuscript. The summary points are as follows:

- Randomized clinical trials about CGM systems:
  1. In the JDRF study (Tamborlane WV, et al. The New England journal of medicine. 2008;359(14):1464-76) which evaluated the effects of different kinds of CGM systems in patients with T1DM, the change of HbA1c from baseline among patients aged ≥25 years old in the intervention group is -0.5(0.56)%, with the significant mean difference in change of 0.53% (95% confidence interval [CI], −0.71 to −0.35) compared with the change from baseline in the control group(-0.02[0.45]%).
  2. In the Gold study (Lind M, et al. JAMA. 2017;317(4):379-87) which evaluated the effect of CGM in adults with T1DM treated with multiple daily insulin injections (aged≥18 yrs), the mean difference in the change of HbA1c from baseline between two groups is significant (mean difference, -0.43%,[95%CI,-0.57% to -0.29%]) and the respective change of HbA1c is from 8.35(0.9)% in baseline to 7.92(0.8)% in the invention group.
  3. The DIAMOND group (Beck RW, et al. JAMA. 2017;317(4):371-8) determined the effectiveness of CGM in adults with T1DM treated with insulin injections(Aged≥25 yrs), the mean reduction in HbA1c level from baseline was 1.0(0.8)% at 24 weeks in the CGM group and 0.4(0.7)% in the control group, respectively, and the adjusted treatment group difference in mean change in HbA1c level was –0.6%(95% CI, –0.8% to –0.3%; P < .001).
- System reviews about the change of HbA1c after using FGM and CGM systems:
  1. The meta-analysis of all identified studies in T1DM patients (34 studies comprising, 5466 participants) indicated the use of Flash GM for 2 to 24 months was associated with an estimated HbA1c reduction from baseline of 0.4% (Gordon I, et al. Diabetes research and clinical practice. 2020:108158).
  2. The meta-analysis of clinical trials and real-world observational studies reported that the overall mean change in HbA1c among adult patients with diabetes was -0.56% (95%CI -0.76, -0.36). （Evans M, et al. Diabetes therapy: research, treatment, and education of diabetes and related disorders. 2020;11(1):83-95)

Based on the researches presented above, the mean difference change of HbA1c between groups is from -0.08 to -0.6% among patients aged ≥18 yrs, and the range of SD of HbA1c is from 0.61 to 0.9%. Considering the referenced articles above was mostly conducted on the participants aged more than 25 yrs and the younger inclusive age might lead to the lesser difference between groups as presented in the JDRF study, we chose the 0.4% as the mean difference and 0.8 (close to the summarized mean value) as the SD of HbA1c for the calculation of sample size. We acknowledged that it would be better to provide the references for the numbers used to determine SD of HbA1c and following your advice, we had added the references in the revised manuscript (see Sample size section, Page 8)

7) I note the paper has no discussion at all. Ideally would have a few paragraphs highlighting the importance of the study, strenghts and limitations.

Response:

Thank you for your kind suggestion and we appreciate your recommendation. We have added a few paragraphs discussing the importance of the study, strengths, and limitations (see DISCUSSIONS section, Page 8).

8) I note multiple papers have discussed CGM and isCGM use in sub-optimally controlled populations - most of these have not been referenced in this manuscript - I think bit more thorough interaction with the literature required.

Response:

Thank you for your kind suggestion and your kind recommendation on the updated literature. Regarding the fact that some updated literature were not referenced in this manuscript, we had

discussed the difference between CGM and isCGM use and some real-world evidence in the revised manuscript (see DISCUSSION section, Page 9).

Specific comments:
"professional" CGM mentioned multiple times - I assume this means blinded?
Response: Thank you for the question. We apologized for not describing it clearly. The “professional” CGM systems refer to the Ipro2® (Medtronic, USA), which is also called as the retrospective CGM systems. During the wearing time, the sensor data derived are not visible and only after the removal of the sensor and data download with retrospective SMBG data calibrations, the glycemic metrics and ambulatory glucose profile will be accessible to the patients and investigators. Therefore, it is a “blinded” CGM system. In order to present it clearer, we used the name “retrospective CGM” instead of “Professional CGM” and we had added more details in the SUPPLEMENT 1- 1.1 Retrospective CGM system section (see Page 1).

TIR vs TIT - 3.9 - 7.4 is not the recommended metric for analysis. Usually 3.9-10
Response: Thank you for bring it to our attention. We had corrected the mistake in the revised manuscript (see Endpoints section, Page 6).

Page 4, line 15 - I would not recommend using the phrase "and so on" in a scietific paper.
Response: Thank you for your kind suggestion. We had corrected the mistakes in the revised manuscript (see ABSTRACT, Methods, and analysis section, Page 2).

page 8 line 14 (and in other places) mentions "biomedical metrics" - what are these?
Response: Thank you for the question. We apologized for not describing it clearly. The biomedical metrics refer to biological data on HbA1c, lipid profiles, liver enzymes, renal function, thyroid function and antibodies, C-peptide and diabetic antibodies as described in the METHODS AND ANALYSIS, Run-in period (Baseline, week 0-2) section (Page 4, Line X) and Laboratory Analyses and Data management section (Page 7).

Page 8 lines 50 - 55 - hypoglycaemic events need some clarification here.
Response: Thank you for your kind suggestion. We had added more clarifications about the hypoglycemic events in the Risks and advent events (AEs) section (Page 6).

Tense - past/present/future needs checking carefully throughout.
Response: Thank you for your kind suggestion. We had gone over our manuscript repeatedly manually and had modified some mistakes in the revised manuscript. we appreciate your kindness to give more detail regarding grammar to help us to improve.

Page 10 - aspects on "professional" CGM - again this means blinded I assume. Worth putting the choice of this device in context with what was available at study commencement.
Response: Thank you for your kind suggestion. As mentioned in the response to your first specific comment, the Ipro2® is the “blinded” CGM system with the data not visible to anyone during the wearing period. Therefore, it is thought to be a perfect tool in the research with less interpretation. Meanwhile, at study commencement, it was the only accessible and updated “blinded” CGM system in China and the mean absolute relative difference of Ipro2® is relatively acceptable with only 9.9%. Therefore, we chose Ipro2® as the unified tool to collect the sensor data in both groups.

VERSION 2 – REVIEW

REVIEWER	R Ajjan University of Leeds, UK
REVIEW RETURNED	18-Aug-2020

GENERAL COMMENTS	Thank you for addressing my comments. I remain concerned about sample size calculations. If the SD of the variable is 0.8%, a sample size of 76 will not be enough to detect a difference of 0.4% in A1c. According to my calculations, the study has less than 55% power to detect the stated difference between two study arms. Can the authors confirm that they did seek expert statistical input into the study? If not, the work needs to be reviewed by a statistician to ensure the study is viable.

REVIEWER	B Wheeler University of Otago New Zealand
REVIEW RETURNED	16-Aug-2020

GENERAL COMMENTS	Thank you for the opportunity to re-review this manuscript. The topic is important and relevant. Current issues that should be dealt with before publication: 1) Boucher et al has published a very similar trial in adolescents and yound adults in diabetes care 2020. I note with interest that neither their pu lished protocol nor main data are referenced here. given the similarities this is a surprise. 2)The sample size calculation remaims a major issue. full details with all parameters and references needed to repeat this are needed. I do not think the sample size is large enough to detect a difference between groups of 0.4%, certainky this seems very ambitious from my knowledge of the literature. I would like to see this fully spelt our and the SD used in similar populations (the SD in well controlled CGM safety studies are not as relevant as data from high risk adult populations which are wider) stated for us to see to understand this. I am worried if 0.4% is the goal - then the study may be underpowered.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

Reviewer Name: R Ajjan
Institution and Country: University of Leeds, UK
Comments:

Thank you for addressing my comments. I remain concerned about sample size calculations. If the SD of the variable is 0.8%, a sample size of 76 will not be enough to detect a difference of 0.4% in A1c. According to my calculations, the study has less than 55% power to detect the stated difference between two study arms. Can the authors confirm that they did seek expert statistical input into the study? If not, the work needs to be reviewed by a statistician to ensure the study is viable.

Answer:

Firstly, sincerely thank you for raising this particularly important point on the sample size and we do appreciate for your efforts in addressing our manuscript and your valuable suggestions. Secondly, to your point, we had since retrospectively reviewed the method and we had also checked the power of initial size. The same as the result you had mentioned, the power was approximately 54% to detect the group difference of 0.4% with the SD of 0.8, which might lead to the increased margin of error.

Therefore, in this regard, we had consulted the statisticians and conducted additional analysis of our assumptions.

As the statisticians suggested, we had used the program of the inequality tests for two means in a repeated measures design set up in the PASS 11.0 software (Hintze, J. (2011). PASS 11. NCSS, LLC. Kaysville, Utah, USA. www.ncss.com.). This module calculates the power for testing the time-averaged difference between two means in a repeated measures design. The mixed model analysis of repeated measures data is also deemed to provide more flexibility in covariance specification and a greater degree of robustness in the presence of missing data. And with the correct covariance structure and an estimate of the within-patient correlation, the estimation was thought to be better on the power and the estimated sample size was thought to be necessary to achieve the objectives. Therefore, as our study collected data repeatedly and sustained for over 26 weeks, we believed that the method of calculation selected and recommended by the statisticians is appropriate for our research.

As for the parameters including the mean group difference in HbA1c change and the SD, we would like to use the value of 0.4% and 0.8%. Such a decision was referred to the method of the major randomized clinical trials and the summarized results of literatures among similar populations. As we could see in the Table 1 and Table 2, the range of the mean group difference the studies adopted was from 0.3 to 0.5% and the range from the results was from 0.18 to 1.1%, with the SD ranged from 0.45 to 0.8%, respectively. Therefore, we believed that 0.4% and 0.8% is reliable for sample size calculation in our study.

As for the Rho (Autocorrelation), we would like to use the value of 0.6. The reason might be as follows: 1) Rho means the correlation between observations on the same subject. Generally, the value of the Rho from 0.4 to 0.7 was deemed a significant moderate-to-large correlation (Page 431-1, Chapter 431, Guides of PASS 11.0). Previous study had presented that the correlation between baseline and follow-up HbA1c values was at least 0.5% (consistent with a scatterplot provided in the DIAMOND study [JAMA 2017;317(4):371-8]); 2) taking the parameter used in the other literature as reference, we could found that the correlation selected in the CONCEPTT study and the study addressing on the high-risk adolescents with T1DM was 0.4 and 0.7, respectively. Therefore, for conservation, we chose a value of 0.6. Therefore, with the assumption of the 10 % drop rate, the mean group difference in HbA1c change of 0.4%, the SD of 0.8% and the correlation of 0.6, the sample size should be expanded to 104. Figure 1 presented the detail information of the calculation. In this regard, we had corrected the sample size in the revised manuscript (see Line 1, Paragraph 2, Page 8, Sample size section) and will update the change in the clnicaitrials. gov.

Table 1. Summary of the parameters used for sample size calculation and the results of three main referred studies

Study	Mean	SD	Power	Drop rate
Parameters used for sample size calculation of the three main referred studies				
1. GOLD study (JAMA. 2017;317(4):379-87)	0.3	1.1	90%	10%
2. DIAMOND study (JAMA 2017;317(4):371-8)	0.4	0.7	90%	
3. JDRF study (NEJM. 2008;359(14):1464-76)	0.5	0.9	90%	10%

The results	Group	Baseline A1c (%)	Change(3-month)	Change(6-month)
1.GOLD study (JAMA. 2017;317(4):379-87)	DEXCOM G4	8.71(0.8)	/	/
	SMBG	8.70(0.9)	/	/
2.DIAMOND study (JAMA 2017;317(4):371-8)	DEXCOM G4	8.6(0.7)	-1.1(0.7)	-1.0(0.8)
	SMBG	8.6(0.6)	-0.5(0.7)	-0.4(0.7)
3.JDRF study (NEJM. 2008;359(14):1464-76)	CGMSs	≥25yrs: 7.6(0.5)	/	≥25yrs : -0.5(0.56)
		15-24 yrs: 8.0(0.7)		15-24 yrs: -0.18(0.65)
	SMBG	≥25yrs: 7.6(0.5)	/	≥25yrs: 0.02(0.45)
		15-24 yrs: 7.9(0.8)		15-24 yrs : -0.21(0.61)

Table 2. Summary of results of some studies about CGMSs and FGM

Study	Group	Baseline A1c (%)	Change(3-month)	Change(6-month)
The results among patients with well glycemic control (A1c <7.0%)				
1.JDRF study (Diabetes Care. 2009;32(8):1378-83)	CGMSs	6.4(0.5)	/	0.02(0.45)
	SMBG	6.5(0.3)	/	0.33(0.43)
2.Aleppo G, et al (Diabetes Care. 2017;40(4):538-45.)	CGM-only	7.1(0.7)	/	0.0(0.5)
	CGM+SMBG	7.0(0.6)	/	0.0(0.5)
The results among patients with sub-optimal control (A1c >7.0% or ≥7.5%)				
1.Campbell FM, Pediatric diabetes. 2018.	FGM	10.3(4.9)	-0.4(0.6) *	
2.GOLD study (JAMA. 2017;317(4):379-87)	DEXCOM G4	8.71(0.8)	/	/
	SMBG	8.70(0.9)	/	/
3.DIAMOND study (JAMA 2017;317(4):371-8)	DEXCOM G4	8.6(0.7)	-1.1(0.7)	-1.0(0.8)
	SMBG	8.6(0.6)	-0.5(0.7)	-0.4(0.7)
4.JDRF study (NEJM. 2008;359(14):1464-76)	CGMSs	≥25yrs: 7.6(0.5)	/	≥25yrs : -0.5(0.56)

		15-24 yrs: 8.0(0.7)		15-24 yrs: -0.18(0.65)
	SMBG	≥25yrs: 7.6(0.5) 15-24 yrs: 7.9(0.8)	/	≥25yrs: 0.02(0.45) 15-24 rs: -0.21(0.61)
5. Boucher et al. (Diabetes care vol.43,10 (2020): 2388-2395)	FGM	13-20yrs: 10.8(1.7)	13-20yrs: 11.2(1.6)	13-20yrs: 10.1(1.6)
	SMBG	13-20yrs: 11.2(1.6)	13-20yrs: 11.2(1.7)	13-20yrs: 10.7(1.5)

*Duration of the FGM wearing was eight weeks.

Figure 1. Report for the sample calculation made by PASS 11.0

Time-Averaged Difference (Normal Data) Power Analysis

Numeric Results

Two-Sided Test. Null Hypothesis: D = 0. Alternative Hypothesis: D ≠ 0.
Covariance Type = Compound Symmetry

Power	Group 1 Sample Size (N1)	Group 2 Sample Size (N2)	Sample Allocation Ratio (R)	Time Points (M)	Difference to be Detected (D1)	Standard Deviation (Sigma)	Auto-corr. (Rho)	Alpha	Beta
0.80798	47	47	1.000	3	0.400	0.800	0.600	0.050	0.19202

References

Brown, H. and Prescott, R., 2006. Applied Mixed Models in Medicine. 2nd ed. John Wiley & Sons Ltd. Chichester, West Sussex, England. Chapter 6.
Liu, H. and Wu, T., 2005. 'Sample Size Calculation and Power Analysis of Time-Averaged Difference.' Journal of Modern Applied Statistical Methods, Vol. 4, No. 2, pages 434-445.
Diggle, P.J., Liang, K.Y., and Zeger, S.L., 1994. Analysis of Longitudinal Data. Oxford University Press. New York, New York. Chapter 2.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 & N2 are the number of subjects in groups 1 and 2, respectively.
R is the ratio of the number of subjects in group 2 to the number in group 1 (R = N2/N1).
M is the number of time points (repeated measurements) at which each subject is observed.
D1 is the difference between the means of groups 1 and 2 under the alternative hypothesis.
Sigma is the standard deviation of a single observation. It is the same for both groups.
Rho is the correlation between observations on the same subject.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

Group sample sizes of 47 and 47 achieve 81% power to detect a difference of 0.400 in a design with 3 repeated measurements having a Compound Symmetry covariance structure when the standard deviation is 0.800, the correlation between observations on the same subject is 0.600, and the alpha level is 0.050.

Reviewer: 3
Reviewer Name: B Wheeler
Institution and Country: University of Otago, New Zealand
Comments:

Thank you for the opportunity to re-review this manuscript. The topic is important and relevant.Current issues that should be dealt with before publication:
1) Boucher et al has published a very similar trial in adolescents and yound adults in diabetes care 2020. I note with interest that neither their published protocol nor main data are referenced here.&xa0; given the similarities this is a surprise.

Answer:
Thank you for your comments and your recommendations on this very published study in Diabetes Care, in fact, came online after our manuscript submission (Boucher et al. Diabetes care vol. 43,10 (2020): 2388-2395). Even though the aim of the two studies was to evaluate the effect of flash glucose monitoring system (FGM) use, there were still several differences between the two studies. The major difference is the age of research population. Our study focused on the adult populations aged ≥18yrs, who were generally more mature than young adults (aged 13-20 yrs) and might better understand the method of glycemic control and device use. Whether the novel type of the device application could have better effect on glycemic control was worthy for a discussion. While we do admit that we did not cite any protocol researching on the FGM in the introduction part, which could not present the status quo of researches fully to the readers. Therefore, we had made revise on the manuscript (see Line 10, Paragraph 3, Page 3, INTRODUCTION section).

2)The sample size calculation remaims a major issue. full details with all parameters and references needed to repeat this are needed. I do not think the sample size is large enough to detect a difference between groups of 0.4%, certainky this seems very ambitious from my knowledge of the literature. I would like to see this fully spelt our and the SD used in similar populations (the SD in well controlled CGM safety studies are not as relevant as data from high risk adult populations which are wider) stated for us to see to understand this. I am worried if 0.4% is the goal - then the study may be underpowered.

Answer:

Thank you for your comments and your queries on the parameters and the method of sample size calculation. As concerns surrounding the underpower sample size and the reference of the parameter choice, please allow us to first of all take you through our rationale behind the parameters we used.

1. On setting up the value of 0.4% as the group difference in HbA1c change

The aim of our study was to study the efficacy of FGM in the management of diabetes, for adult patients with type 1 diabetes and inadequate glycemic control. In this regard, it is anticipated that improvement will be observed in the group with FGM use, which is over and above any improvement in control group. The mean group difference in HbA1c change from baseline we used initially based on the randomized clinical trials including the JDRF study (Tamborlane WV et al. NEJM. 2008;359(14):1464-76), the GOLD study (Lind M, et al. JAMA. 2017;317(4):379-87) and the DIAMOND study (Beck RW et al. JAMA. 2017;317(4):371-8) which evaluated the effect of DEXCOM G4 in adults (≥18 yrs and 25yrs, respectively) with T1DM. Details were presented in Table 1. The mean reduction in HbA1c level from baseline is 0.4 to0.6%. Therefore, we chose 0.4% as the value of mean group difference in HbA1c change for sample size calculation.

While in the recently published RCT that addressed the young patients with T1DM who had higher baseline HbA1c value than ours, the reduction of HbA1c was only approximately 0.6% (Boucher et al.

Diabetes care vol. 43,10 (2020): 2388-2395). The recent meta-analysis and reviews also concluded the estimated HbA1c reduction after FGM use (from 0.4 to 0.6%) while change in the control group was insignificant (Diabetes research and clinical practice. 2020:108158; Diabetes therapy. 2020;11(1):83-95). All in all, we believed that adopting the value of 0.4% as the mean group difference for sample size calculation is rational.

2. On setting up the value of 0.8% as standard deviation (SD)

As for the SD, as we could see in Table 1, on the one hand, the value of 0.8% referred to the method used in their studies (ranged from 0.7 to 1.1%). On the other hand, from the studies' results, the SD ranged from 0.45 to 0.80%. To the reviewer's point, we also acknowledged that the SD from the higher-risk adult populations should be broader than those with better control. As suggested, a detailed summarization was conducted (see Table 2) and we found the SD in the studies of similar populations ranged from 0.45 to 0.80%, which was in line with our previous assumption. Therefore, we believed that the value of 0.8 is reliable.

Novelty, in the study conducted on the youth with type 1 diabetes and high-risk(Boucher et al. Diabetes care vol. 43,10 (2020): 2388-2395), the SD was larger with the value of 14.4mmol/mol (equals to 1.7%) but we assumed that such a SD is too large for our study as the range of HbA1c set up for inclusion criteria in our study was smaller than that that in youths with high risk (HbA1c ≥9%). Furthermore, the inclusion criteria of HbA1c in our study were similar to those the other studies presented in Table 2. Therefore, we believed that the value of 0.8 is appropriate in our study.

3. On our calculation of sample size

Again, sincerely thank you for raising this particularly important point on the sample size and we do appreciate for your efforts in addressing our manuscript and your valuable suggestions. To your point, we had since retrospectively reviewed the method and we had also checked the power of initial size. The same as what you had mentioned, the power was approximately 54%, which might lead to the increased margin of error. Therefore, we had consulted the statisticians and conducted additional analysis of our assumptions.

As the statisticians suggested, we used the program of the inequality tests for two means in a repeated measures design set up in the PASS 11.0 software (Hintze, J. (2011). PASS 11. NCSS, LLC. Kaysville, Utah, USA. www.ncss.com.). This module calculates the power for testing the time-averaged difference between two means in a repeated measures design. The mixed model analysis of repeated measures data is also deemed to provide more flexibility in covariance specification and a greater degree of robustness in the presence of missing data. And with the correct covariance structure and an estimate of the within-patient correlation, the estimation was thought to be better on the power and the estimated sample size was thought to be necessary to achieve the objectives. Therefore, as our study collected data repeatedly and sustained for over 26 weeks, we believed that the method of calculation selected and recommended by the statisticians is appropriate for our research.

As for the metrics, we referred to the method of the major randomized clinical trials and the summarized results of literatures among similar populations. As we could see in the Table 1 and Table 2, the range of the mean group difference the studies used was from 0.3 to 0.5% and the range of from the results was from 0.18 to 1.1% with the SD ranged from 0.45 to 0.8%, respectively. Therefore, we believed that the value of the mean group difference of 0.4 and the SD of 0.8 is reliable for sample size calculation in our study.

As for the Rho (Autocorrelation), we would like to use the value of 0.6. The reason might be as follows: 1) Rho means the correlation between observations on the same subject. Generally, the value of the Rho from 0.4 to 0.7 was deemed a significant moderate-to-large correlation (Page 431-1,

Chapter 431, Guides of PASS 11.0). Previous study had presented that the correlation between baseline and follow-up HbA1c values was at least 0.5% (consistent with a scatterplot provided in the DIAMOND study [JAMA 2017;317(4):371-8]); 2) taking the parameter used in the other literatures as reference, we found that the correlation selected in the CONCEPTT study and the study addressing on the high-risk adolescents was 0.4 and 0.7, respectively. Therefore, for conservation, we chose a value of 0.6.

All in all, with the assumption of the 10% drop rate, the mean group difference in HbA1c change of 0.4%, the SD of 0.8% and the correlation of 0.6, the sample size should be expanded to 104 instead of 76. Figure 1 was the report of the calculation. In this regard, we had corrected the sample size in the revised manuscript (see Line 1, Paragraph 2, Page 8, Sample size section) and will update the change in the clnicaitrials. gov.

Table 1. Summary of the parameters used for sample size calculation and the results of three main referred studies

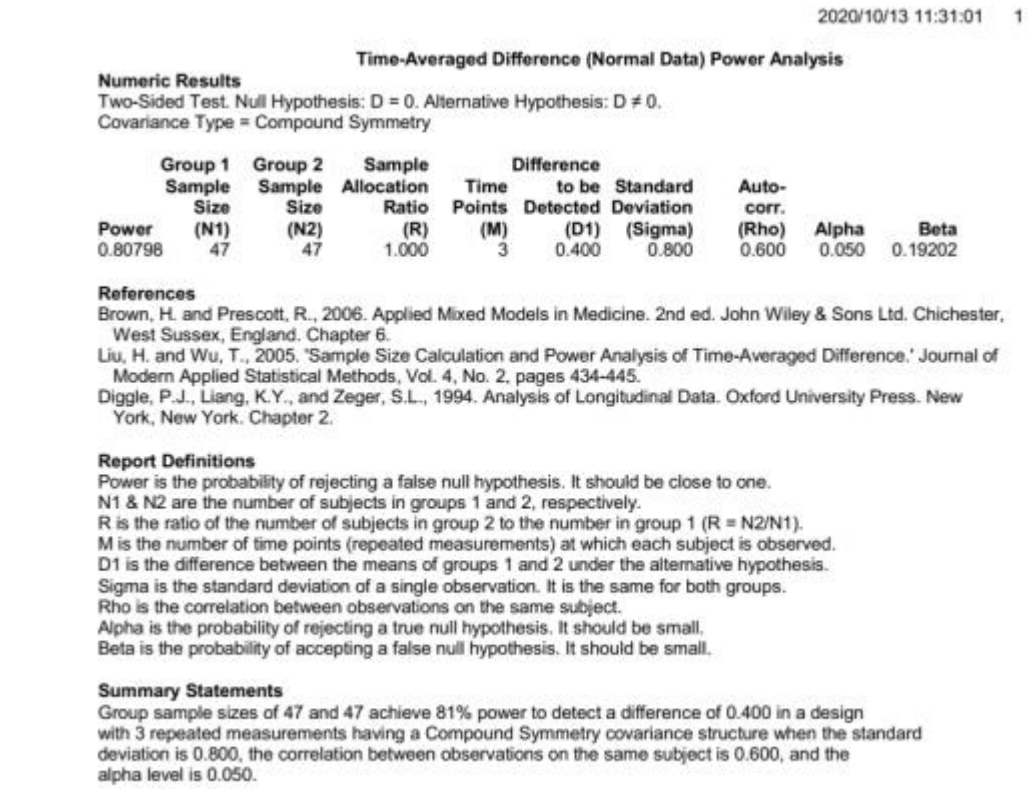
Study	Mean	SD	Power	Drop rate
Parameters used for sample size calculation of the three main referred studies				
1. GOLD study (JAMA. 2017;317(4):379-87)	0.3	1.1	90%	10%
2. DIAMOND study (JAMA 2017;317(4):371-8)	0.4	0.7	90%	
3. JDRF study (NEJM. 2008;359(14):1464-76)	0.5	0.9	90%	10%
The results	Group	Baseline A1c (%)	Change(3-month)	Change(6-month)
1.GOLD study (JAMA. 2017;317(4):379-87)	DEXCOM G4	8.71(0.8)	/	/
	SMBG	8.70(0.9)	/	/
2.DIAMOND study (JAMA 2017;317(4):371-8)	DEXCOM G4	8.6(0.7)	-1.1(0.7)	-1.0(0.8)
	SMBG	8.6(0.6)	-0.5(0.7)	-0.4(0.7)
3.JDRF study (NEJM. 2008;359(14):1464-76)	CGMSs	≥25yrs: 7.6(0.5)	/	≥25yrs : -0.5(0.56)
		15-24 yrs: 8.0(0.7)		15-24 yrs: -0.18(0.65)
	SMBG	≥25yrs: 7.6(0.5)	/	≥25yrs: 0.02(0.45)
		15-24 yrs: 7.9(0.8)		15-24 yrs : -0.21(0.61)

Table 2. Summary of results of some studies about CGMSs and FGM

Study	Group	Baseline A1c (%)	Change(3-month)	Change(6-month)
The results among patients with well glycemic control (A1c <7.0%)				
1.JDRF study (Diabetes Care. 2009;32(8):1378-83)	CGMSs	6.4(0.5)	/	0.02(0.45)
	SMBG	6.5(0.3)	/	0.33(0.43)
2.Aleppo G, et al (Diabetes Care. 2017;40(4):538-45.)	CGM-only	7.1(0.7)	/	0.0(0.5)
	CGM+SMBG	7.0(0.6)	/	0.0(0.5)
The results among patients with sub-optimal control (A1c >7.0% or ≥7.5%)				
1.Campbell FM, Pediatric diabetes. 2018.	FGM	10.3(4.9)	-0.4(0.6) *	
2.GOLD study (JAMA. 2017;317(4):379-87)	DEXCOM G4	8.71(0.8)	/	/
	SMBG	8.70(0.9)	/	/
3.DIAMOND study (JAMA 2017;317(4):371-8)	DEXCOM G4	8.6(0.7)	-1.1(0.7)	-1.0(0.8)
	SMBG	8.6(0.6)	-0.5(0.7)	-0.4(0.7)
4.JDRF study (NEJM. 2008;359(14):1464-76)	CGMSs	≥25yrs: 7.6(0.5) 15-24 yrs: 8.0(0.7)	/	≥25yrs : -0.5(0.56) 15-24 yrs: -0.18(0.65)
	SMBG	≥25yrs: 7.6(0.5) 15-24 yrs: 7.9(0.8)	/	≥25yrs: 0.02(0.45) 15-25 rs: -0.21(0.61)
5. Boucher et al. (Diabetes care vol.43,10 (2020): 2388-2395)	FGM	13-20yrs: 10.8(1.7)	13-20yrs: 11.2(1.6)	13-20yrs: 10.1(1.6)
	SMBG	13-20yrs: 11.2(1.6)	13-20yrs: 11.2(1.7)	13-20yrs: 10.7(1.5)

*Duration of the FGM wearing was eight weeks.

Figure 1. Report for the sample calculation made by PASS 11.0



2020/10/13 11:31:01 1

Time-Averaged Difference (Normal Data) Power Analysis

Numeric Results

Two-Sided Test. Null Hypothesis: D = 0. Alternative Hypothesis: D ≠ 0.
Covariance Type = Compound Symmetry

Power	Group 1 Sample Size (N1)	Group 2 Sample Size (N2)	Sample Allocation Ratio (R)	Time Points (M)	Difference to be Detected (D1)	Standard Deviation (Sigma)	Auto-corr. (Rho)	Alpha	Beta
0.80798	47	47	1.000	3	0.400	0.800	0.600	0.050	0.19202

References

Brown, H. and Prescott, R., 2006. Applied Mixed Models in Medicine. 2nd ed. John Wiley & Sons Ltd. Chichester, West Sussex, England. Chapter 6.
Liu, H. and Wu, T., 2005. 'Sample Size Calculation and Power Analysis of Time-Averaged Difference.' Journal of Modern Applied Statistical Methods, Vol. 4, No. 2, pages 434-445.
Diggle, P.J., Liang, K.Y., and Zeger, S.L., 1994. Analysis of Longitudinal Data. Oxford University Press. New York, New York. Chapter 2.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 & N2 are the number of subjects in groups 1 and 2, respectively.
R is the ratio of the number of subjects in group 2 to the number in group 1 (R = N2/N1).
M is the number of time points (repeated measurements) at which each subject is observed.
D1 is the difference between the means of groups 1 and 2 under the alternative hypothesis.
Sigma is the standard deviation of a single observation. It is the same for both groups.
Rho is the correlation between observations on the same subject.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

Group sample sizes of 47 and 47 achieve 81% power to detect a difference of 0.400 in a design with 3 repeated measurements having a Compound Symmetry covariance structure when the standard deviation is 0.800, the correlation between observations on the same subject is 0.600, and the alpha level is 0.050.

VERSION 3 – REVIEW

REVIEWER	Ben Wheeler University of Otago, New Zealand
REVIEW RETURNED	17-Oct-2020

GENERAL COMMENTS	The authors have been very responsive to feedback. I will look forward to seeing their results.