PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Study protocol for a randomised open label clinical trial examining the safety and efficacy of the Android Artificial Pancreas System (AAPS) with advanced bolus-free features in adults with type 1 diabetes: the 'CLOSE IT' (Closed Loop Open SourcE In Type 1 diabetes) trial
AUTHORS	Wilkinson, Tom; Tomic, Dunya; Boyle, Erin; Burren, David; Elghattis, Yasser; Jenkins, Alicia; Keesing, Celeste; Middleton, Sonia; Nanayakkara, Natalie; Williman, Jonathan; de Bock, Martin; Cohen, Neale

VERSION 1 – REVIEW

REVIEWER	Szypowska, Agnieszka Med Univ Warsaw, Department of Paediatrics Speaker fees from Dexcom, Medtronic, Abbott.
REVIEW RETURNED	16-Sep-2023

GENERAL COMMENTS	The study protocol for a randomised open label clinical trial comparing the safety and efficacy of the Android Artificial Pancreas System (AAPS) with the same system used as hybrid closed loop (HCL) in people with T1D is written according to the SPIRIT Statement. The protocol is well-written, the study objective is clearly defined, the research design and methodology are described sufficiently. Comment: According to the international consensus statement on CGM metrics in clinical trials the usage of concomitant medication that can affect CGM accuracy, e.g paracetamol, should be added to the exclusion criteria.

REVIEWER	Kudva, Yogish C. Mayo Clinic
REVIEW RETURNED	17-Sep-2023

GENERAL COMMENTS	The manuscript is the study protocol for a randomized open lable clinical trial examining the safety and efficacy of the Android Artificial pancreas System (AAPS) with advanced bolus-free features in adults with type 1 diabetes: the CLOSE IT trial. Subjects will complete a run in of 84 days to be competent in the use of the system and subsequently will be randomized 1:1 to FCL or HCL. Primary and secondary outcomes listed are appropriate. Non inferiorty definition was chosen prior to publication of definition by expert consensus. Subsequently, twenty subjects will use Fiasp and be randomized for 4 weeks to Novorapid versus Fiasp.

	Suggestions 1. The statistical analysis plan has been stated a priori. Why is the section starting with statistical analysis plan will be prepared by….? Also, it is important for all protocols to state statisical analysis plan a priori and not expand scope subsequently. This will go beyond the spirit of transparency stated in the publication of this manuscript. 2. Please state briefly the advanced bolus free features that permit FCL use in the Methods. 3. There are mult center studies using HCL being conducted with ultra rapid insulin Lyumjev in the outcomes publication and the iLET pivotal study had a subgroup of Fiasp users. Please mention this.

REVIEWER	Braune, Dr. Katarina Charité Universitätsmedizin Berlin, Institute of Medical Informatics
REVIEW RETURNED	19-Sep-2023

GENERAL COMMENTS	Thank you for providing me the opportunity to review this study protocol by Wilkinson et al. on a randomized clinical trial of the “AndroidAPS” open-source AID system. While the manuscript stands out for its rationale, robustness and clarity, addressing the following points will further strengthen its quality. I am optimistic that this research will make a significant contribution to the field. Major comments: 1. The intervention of this trial relies on the capability of the oref1 algorithm in operating the AID system in full-closed loop. It would be of interest to provide more details on why that might be specifically the case for oref1, and indicate whether any specific features will be enabled during the trial, such as dynamic Insulin Sensitivity Factor (ISF), unannounced meals (UAM), and supermicroboluses (SMBs). 2. In any AID algorithm like oref1, the accuracy of settings is paramount for their safety and efficacy. Given that these settings (e.g., target glycemia, ISF, maxIOB, basal multipliers) might differ between individuals and also within individuals when transitioning from HCL to FCL, how does the protocol propose to determine and assess these? 3. Considering potential scenarios where participants might neglect necessary interactions with the system during altered insulin sensitivity periods, how does the study intend to evaluate instances where participants have utilized temporary targets or made profile switches? For example, using the AID in FCL whilst physically active might result in hypos, impacting the person but also the study results. 4. If participants, who should be operating the system in FCL, decide to manually enter a bolus or input carbohydrate data intermittently, how will the study manage such deviations? Is there a predetermined protocol regarding which data would be considered outliers and excluded from the study results?

	5. The randomization process is well-described. However, do you have any stratification factors (e.g., based on gender, age, duration of T1D, or baseline HbA1c) to ensure balance between the groups? Minor: 6. It might be useful to explain why the run-in phase is specifically 84 days, rather than a more rounded number, and what research or reasoning supports this choice. 7. A brief explanation of why participants will switch from NovoRapid to Fiasp in the extension phase would provide context. What is the scientific or clinical reasoning behind this switch? 8. The calculation for the sample size is presented, but it might be helpful to give a brief rationale or justification for choosing a non-inferiority margin of 7% TIR, especially given the later mention of the 2023 international consensus recommending a 3% difference.

VERSION 1 – AUTHOR RESPONSE

Comments from Reviewer 1	
According to the international consensus statement on CGM metrics in clinical trials the usage of concomitant medication that can affect CGM accuracy, e.g paracetamol, should be added to the exclusion criteria.	Thank you – we have given some thought to excluding paracetamol use from analysis, however have decided against doing so. Although the international consensus statement on CGM metrics states that paracetamol use may affect the accuracy of the Dexcom G6, we are not aware of any studies clearly demonstrating this to be the case. Of note, the Dexcom G6 incorporates a permselective membrane to block the diffusion of paracetamol to the electrode surface (this was not present in the earlier G4 or G5 systems). One study suggested that repeated dosing of paracetamol at 4-hourly intervals may overcome this membrane and result in interference (reference below), however of note that study was sponsored by Abbott (manufacturer of the competing Freestyle Libre system) and concluded that the magnitude of any deviation would be unlikely to impact on insulin dosing. The product information for the Dexcom G6 states that paracetamol may be taken at standard doses of up to 1 gram every 6 hours without affecting the validity of sensor recordings (https://www.dexcom.com/en-us/interference) - this is consistent with standard dosing advice given to any patient taking paracetamol.

	Denham D. Effect of Repeated Doses of Acetaminophen on a Continuous Glucose Monitoring System with Permselective Membrane. *Journal of Diabetes Science and Technology*. 2021;15(2):517-518.
Comments from Reviewer 2	
The statistical analysis plan has been stated a priori. Why is the section starting with statistical analysis plan will be prepared by….? Also, it is important for all protocols to state statisical analysis plan a priori and not expand scope subsequently. This will go beyond the spirit of transparency stated in the publication of this manuscript.	The protocol contains the principal features of the planned statistical analysis. However, it is usual/recommended practice for a separate Statistical Analysis Plan document to also be prepared, which "contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol" (ICH E9). Please see doi: https://doi.org/10.1001/jama.2017.18556 for further information.
Please state briefly the advanced bolus free features that permit FCL use in the Methods.	These features include super micro boluses (SMBs) and an unannounced meals (UAM) feature. We have now specifically identified these as features that may permit FCL use.
There are multi center studies using HCL being conducted with ultra rapid insulin Lyumjev in the outcomes publication and the iLET pivotal study had a subgroup of Fiasp users. Please mention this.	We have provided more detail in the introduction regarding published trials that have assessed Fiasp and Lyumjev in closed loop systems, including the iLET pivotal study.
Comments from Reviewer 3	
Major	
The intervention of this trial relies on the capability of the oref1 algorithm in operating the AID system in full-closed loop. It would be of interest to provide more details on why that might be specifically the case for oref1, and indicate whether any specific features will be enabled during the trial, such as dynamic Insulin Sensitivity Factor (ISF), unannounced meals (UAM), and supermicroboluses (SMBs).	We will enable SMBs and UAM and have now included some detail regarding this. Dynamic ISF will not be used.
In any AID algorithm like oref1, the accuracy of settings is paramount for their safety and efficacy. Given that these settings (e.g., target glycemia, ISF, maxIOB, basal multipliers) might differ between individuals and also within individuals when transitioning from HCL to FCL, how does the protocol propose to determine and assess these?	Although the large number of settings allow for customizability, we acknowledge the complexity this adds. Settings will be adjusted on an individual basis and it is difficult to specify a strict protocol for this. We have included more detail about the process for regular review of settings by the study team.
Considering potential scenarios where participants might neglect necessary interactions with the system during altered insulin sensitivity periods, how does the study intend to evaluate instances where participants have utilized temporary targets or made profile switches? For example, using the AID in FCL	Participants will be able to switch profiles and utilize temporary targets to manage exercise through the trial, including in the FCL group. This is now specifically stated.

whilst physically active might result in hypos, impacting the person but also the study results.	
If participants, who should be operating the system in FCL, decide to manually enter a bolus or input carbohydrate data intermittently, how will the study manage such deviations? Is there a predetermined protocol regarding which data would be considered outliers and excluded from the study results?	Yes. The primary analysis will be conducted on individuals as randomized (intention-to-treat). However, a sensitivity (per protocol) analysis will also be conducted as described on page 11. "For the per protocol analysis CGM metrics will be considered to belong to the FCL group on days where the participant has delivered no manual boluses outside of protocol conditions, or to the HCL group on days where the participant has delivered ≥2manual boluses outside of protocol conditions." Days with only one manual bolus will be excluded from the per protocol analysis.
The randomization process is well-described. However, do you have any stratification factors (e.g., based on gender, age, duration of T1D, or baseline HbA1c) to ensure balance between the groups?	Thank you. Randomisation has been by stratified by the two study sites only. The strongest predictor of an outcome (e.g. TIR) is its value at baseline, which will be adjusted for in the analysis.
Minor	
It might be useful to explain why the run-in phase is specifically 84 days, rather than a more rounded number, and what research or reasoning supports this choice.	84 days represents 12 full weeks. We agree that this is an essentially arbitrary number. The reasoning behind a long run-in phase is to best allow optimization of settings prior to randomization and this is stated as a strength of the study in the article summary.
A brief explanation of why participants will switch from NovoRapid to Fiasp in the extension phase would provide context. What is the scientific or clinical reasoning behind this switch?	We have provided more context for Fiasp in the introduction, including reference to other studies of its use in the closed loop setting.
The calculation for the sample size is presented, but it might be helpful to give a brief rationale or justification for choosing a non-inferiority margin of 7% TIR, especially given the later mention of the 2023 international consensus recommending a 3% difference.	This study was planned prior to publication of the 2023 international consensus statement, hence the discrepancy in non-inferiority margins. We would require a sample size of 382 participants to be similarly powered to detect non-inferiority at a 3% margin, which would represent a considerable increase in required resource. Given the novel nature of this trial it will still generate useful information: if our study is able to show non-inferiority at 7% then a future larger trial might be conducted to show non-inferiority at 3%.