PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Investigating the day-to-day impact of hypoglycaemia in adults with type 1 or type 2 diabetes: design and validation protocol of the Hypo-METRICS application.
AUTHORS	Søholm, Uffe; Broadley, Melanie; Zaremba, Natalie; Divilly, Patrick; Nefs, Giesje; Mahmoudi, Zeinab; de Galan, B.; Pedersen-Bjergaard, Ulrik; Brennan, Alan; Pollard, Daniel; McCrimmon, Rory; Amiel, Stephanie; Hendrieckx, Christel; Speight, Jane; Choudhary, Pratik; Pouwer, Frans

VERSION 1 – REVIEW

REVIEWER	Katharine Barnard-Kelly Bournemouth University
REVIEW RETURNED	23-Aug-2021

GENERAL COMMENTS	Well-written article that is clear, concise and describes in detail the study.

REVIEWER	Thomas Kamarck University of Pittsburgh, Psychology
REVIEW RETURNED	26-Aug-2021

GENERAL COMMENTS	Review of BMJ Open manuscript ID bmjopen-2021-051651 The investigators describe the development of a self-report assessment tool to be used as part of a field assessment protocol ('Hypo-METRICS app') during daily life for investigating the symptoms and experiences associated with hypoglycemic episodes among insulin-dependent diabetics. The protocol described here involves three phases, with two of the phases already completed (the establishment of a working group to identify relevant areas of functioning, and the development of the 29 items for use as part of the app) and the third phase (examination of feasibility, reliability, and validity of the 'app' in the 'Hypo-METRICS clinical study') currently in process. The sample to be recruited for the clinical study is a group of 600 adults (18-85) with T1DM or insulin-treated T2DM to be recruited from eight specialist diabetes centres across five European countries. Recruited participants will be asked to complete the scales 3 times per day over the course of 10 weeks. The investigators propose to examine the factor structure of the items, to examine their convergent and discriminant validity in comparison with established questionnaires, to examine the feasibility or acceptability or the self-monitoring experience, and to examine the completion rates and the factors that may determine compliance with the self-monitoring protocol. This strikes me as an extremely valuable study, as hypoglycemia,

	and efforts to manage it, can be associated with significant functional and clinical problems in the setting of diabetes, yet our efforts to measure these effects to date (for example, with the use of global, retrospective questionnaires) are somewhat primitive. The use of EMA methods seems like a natural fit for this problem, insofar as these methods allow us to examine symptom changes in real time, over time, in the natural environment. The ability to map the dynamic within-person changes in symptoms against objective continuous measures of blood glucose is an especially exciting part of this proposal. The investigators have done a nice job on item development and initial user testing. The rationale for inclusion of each scale or 'module' is sound, and the examples that they have provided for some of the initial changes to item content based upon feedback from users are compelling. Because one of the goals of publishing study protocols is presumably to facilitate 'open science,' I would like to see a bit more detail about the hypotheses to be tested as part of the 'Hypo-METRICS clinical study' along with the specific predictions associated with each. 1. For example, there is little description of the ways in which the associations between self-reported symptoms/experiences and objective glucose monitoring will be examined. Will there be a cut-point established in the continuous glucose record (in terms of glucose level and duration of response) for identifying periods of 'objective' hypoglycemia? What is the prediction with respect to the association between 'objective' and 'subjective' hypoglycemia? Is this one of the ways that you plan to validate the Hypo-METRICS scale? What about individual differences in 'accuracy' (correspondence between objective and subjective measures of hypoglycemia)? Are there important demographic or clinical characteristics that distinguish 'accurate' and 'inaccurate' self-monitors? 2. There needs to be additional detail about how the construct validity of the Hypo-METRICS modules will be evaluated. First, how will you handle the confound between time of day and item administration? Will separate factor analyses be conducted for each of the three time points? Second, what are the 'validated self-report questionnaires' that will be administered as criteria against which the Hypo-METRICS scales will be validated? What are the constructs to be assessed here? Presumably, the construct validity plan will assist you in evaluating individual differences, but this will not assist you in validating the within-person changes in scale scores. How will the latter be conducted? 3. One of the primary goals of the Hypo-METRICS clinical study is to assess the feasibility and acceptability of the EMA monitoring experience. I am surprised that only 20 participants will be selected to participate in this part of the study. Will these participants (or, preferably, a larger group) be selected on the basis of important characteristics that may affect compliance or acceptability (age, sex, education, disease severity, etc.)? Some greater detail about these considerations would be useful. Similarly, some attention to the effects of time during the monitoring interval would be beneficial. Are there decreases in the reliability of the module measures over the course of the monitoring period, suggesting a failure to maintain

	consistent participant engagement over the 10-week period?. How will this be tested? 4. I have one last, very minor point. I suggest that you re-think the terminology used in the third paragraph under “Phase 3” on page 15. You refer to the “target population for this study” as 600 adults with T1DM. I believe that the correct way to phrase this would be to indicate that the target population involves European adults with T1DM, and that the “sample of participants” chosen to represent this population would consist of 600 adults age 18-85 from specialist clinics, etc.

REVIEWER	Jennifer Warnick Warren Alpert Medical School of Brown University, Psychiatry and Human Behavior
REVIEW RETURNED	27-Aug-2021

GENERAL COMMENTS	Thank you for the opportunity to review this protocol manuscript. The manuscript is well written, and the study is well designed. Below are a few very minor suggestions to improve clarity and potential impact. 1. In the abstract, consider adding that you will be collecting CGM data as this adds to the novelty and potential impact of the project. 2. Phase 2 Developing Items: I agree with the authors' choice to include branching logic for certain questions. However, participants may learn that answering certain questions will lead to more questions. I would suggest examining the accuracy of self-reported hypos to the CGM data to account for this potential issue. 3. Are participants able to self-initiate EMA prompts if they have a hypo-episode between fixed EMA prompts? 4. Phase 2 User Testing: Please clarify the user-testing process. It's not clear whether the participants met as a group and provided feedback on the questions and the potential of answering them 3x per day or whether participants actually tried using the app for a certain amount of time. I interpreted the former in the body of the text and the latter in the discussion section. 5. Phase 2 App Design: Could a participant complete an EMA prompt outside of the time-interval? For example, would it be possible for a participant to complete all 3 daily prompts in one sitting (in the evening)? If not, please specify that the prompts expire or close after the time interval ends. 6. Phase 3 Study Design: Why did the authors choose 10-weeks for the study duration? Please justify this decision in the manuscript. 7. Phase 3 Study Design: If a participant already owns/wears a CGM, will they be required to to wear a 2nd one for the study? 8. Phase 3 Study Design: Will participants be compensated and if so, will participants be compensated for each EMA prompt completed? 9. Will EMA adherence be monitored by study staff and will study staff intervene if participants are not completing prompts? 10. Please include any demographic and general diabetes management variables you plan to examine in your analyses. For example, will you be collecting data to identify participants who have pumps vs. engage in multiple daily injections and use a glucose meter vs. CGM daily? 11. Will the study be providing smartphones and/or data plans to participants who do not own personal devices or data plans?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Katharine Barnard-Kelly, Bournemouth University

Comment 1 from reviewer 1:
Well-written article that is clear, concise and describes in detail the study.

Response 1 to reviewer 1:
We thank the reviewer for taking the time to review our manuscript.

Reviewer: 2
Dr. Thomas Kamarck, University of Pittsburgh

Introductory comment from reviewer 2:
Review of BMJ Open manuscript ID bmjopen-2021-051651

The investigators describe the development of a self-report assessment tool to be used as part of a field assessment protocol ('Hypo-METRICS app') during daily life for investigating the symptoms and experiences associated with hypoglycemic episodes among insulin-dependent diabetics. The protocol described here involves three phases, with two of the phases already completed (the establishment of a working group to identify relevant areas of functioning, and the development of the 29 items for use as part of the app) and the third phase (examination of feasibility, reliability, and validity of the 'app' in the 'Hypo-METRICS clinical study') currently in process. The sample to be recruited for the clinical study is a group of 600 adults (18-85) with T1DM or insulin-treated T2DM to be recruited from eight specialist diabetes centres across five European countries. Recruited participants will be asked to complete the scales 3 times per day over the course of 10 weeks. The investigators propose to examine the factor structure of the items, to examine their convergent and discriminant validity in comparison with established questionnaires, to examine the feasibility or acceptability or the self-monitoring experience, and to examine the completion rates and the factors that may determine compliance with the self-monitoring protocol.

This strikes me as an extremely valuable study, as hypoglycemia, and efforts to manage it, can be associated with significant functional and clinical problems in the setting of diabetes, yet our efforts to measure these effects to date (for example, with the use of global, retrospective questionnaires) are somewhat primitive. The use of EMA methods seems like a natural fit for this problem, insofar as these methods allow us to examine symptom changes in real time, over time, in the natural environment. The ability to map the dynamic within-person changes in symptoms against objective continuous measures of blood glucose is an especially exciting part of this proposal.

The investigators have done a nice job on item development and initial user testing. The rationale for inclusion of each scale or 'module' is sound, and the examples that they have provided for some of the initial changes to item content based upon feedback from users are compelling.

Because one of the goals of publishing study protocols is presumably to facilitate 'open science,' I would like to see a bit more detail about the hypotheses to be tested as part of the 'Hypo-METRICS clinical study' along with the specific predictions associated with each.

Response to introductory comment from reviewer 2:
We thank the reviewer for his comments and kind words. Below we will address each comment point-by-point.

Comment 1 from reviewer 2:
1. For example, there is little description of the ways in which the associations between self-reported symptoms/experiences and objective glucose monitoring will be examined. Will there be a cut-point established in the continuous glucose record (in terms of glucose level and duration of response) for identifying periods of 'objective' hypoglycemia? What is the prediction with respect to the association between 'objective' and 'subjective' hypoglycemia? Is this one of the ways that you plan to validate the Hypo-METRICS scale? What about individual differences in 'accuracy' (correspondence between objective and subjective measures of hypoglycemia)? Are there important demographic or clinical

characteristics that distinguish 'accurate' and 'inaccurate' self-monitors?

Response 1 to reviewer 2:
We agree with the reviewer that this is an important point. In fact, the primary objective of the Hypo-METRICS study is to "*define the threshold and duration of sensor glucose that provides the optimal sensitivity and specificity for events that people living with diabetes experience as hypoglycaemia*." This text has been copied from the Hypo-METRICS study protocol paper that is currently in development. An important distinction to make here is between: 1) the "Hypo-METRICS app development" and 2) the "Hypo-METRICS clinical study protocol". The former is described in the present manuscript and the latter is described in a separate manuscript, which will directly address the questions raised by the reviewer.

To summarise, the purpose of the present manuscript is to inform the readers about the Hypo-METRICS app development process and planned psychometric analyses, and not to provide complete details on the Hypo-METRICS study protocol. To emphasize this distinction, we have added the following text to our manuscript on page 12:

"Further details on the Hypo-METRICS clinical study, including the full list of objectives, can be found here: https://www.clinicaltrials.gov/ct2/show/record/NCT04304963"

Comment 2.1 from reviewer 2:
2. There needs to be additional detail about how the construct validity of the Hypo-METRICS modules will be evaluated. First, how will you handle the confound between time of day and item administration? Will separate factor analyses be conducted for each of the three time points?

Response 2.1 to reviewer 2:
We will indeed conduct the factor analyses separately on each of the check-ins (morning, afternoon, evening), and have therefore added the following text to the manuscript on page 12:

"*A multilevel factor analysis will be conducted separately for each of the three check-ins to avoid violating assumptions of independency between the repeated measurements*"

Comment 2.2 from reviewer 2:
Second, what are the 'validated self-report questionnaires' that will be administered as criteria against which the Hypo-METRICS scales will be validated? What are the constructs to be assessed here?

Response 2.2 to reviewer 2:
Based on the reviewer's comment we have added the full list of questionnaires, including the construct assessed by each questionnaire to the Supplementary material (Table S1). This is now reflected in the following sentence on page 12 of the manuscript:

"Lastly, construct validity will be examined by analysing the correlations between the items or factor scores from the Hypo-METRICS app and validated self-report questionnaires (listed in Table S1 in supplementary material)"

Comment 2.3 from reviewer 2:
Presumably, the construct validity plan will assist you in evaluating individual differences, but this will not assist you in validating the within-person changes in scale scores. How will the latter be conducted?

Response 2.3 to reviewer 2:
We assume the reviewer is referring to an investigation of the within-person variability on the item/factor scores. This will be investigated with test-retest reliability analyses, to explore reliability over time of aggregated scores across a few weeks, and with intraclass correlations (ICC) and root mean square of successive differences (RMSSD), to explore between- vs within-person variability and day-to-day variability, respectively. This information has been added to the manuscript on page 12:

"To explore test-retest reliability, factor scores will be aggregated and compared (via correlation analysis) across two different weeks. To examine between- and within-person variability on an item level, intraclass correlations (ICC) [55] and root mean square of successive differences (RMSSD) will

be calculated [56]."

Comment 3.1 from reviewer 2:
3. One of the primary goals of the Hypo-METRICS clinical study is to assess the feasibility and acceptability of the EMA monitoring experience. I am surprised that only 20 participants will be selected to participate in this part of the study.

Response 3.1 to reviewer 2:
The feasibility and acceptability will indeed be explored by conducting in-depth semi-structured interviews with 20 participants who have used the Hypo-METRICS app for a 10-week study period (in the Hypo-METRICS clinical study). The results of this interview study will allow us to qualitatively assess: 1) participants' motivation and engagement, 2) content validity (including understandability, relevance, comprehensiveness etc.), 3) overall user experiences with the app (design, functionality etc.). A detailed interview guide has been developed for this purpose. Due to the qualitative method applied, we believe 20 participants is an appropriate sample size and is in accordance with COSMIN guidelines that suggest a sample size of >7 when conducting qualitative studies assessing content validity (https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf page 29).

The acceptability of the app will be further explored via the analysis of completion rates. Since this is quantitative analyses, it will be conducted for the full sample (n=600). To emphasize this, we have added the following text on page 13: *"An analysis of completion rates and patterns of missing data from the clinical study will be performed on the full sample (n=600)."*

Comment 3.2 from reviewer 2:
Will these participants (or, preferably, a larger group) be selected on the basis of important characteristics that may affect compliance or acceptability (age, sex, education, disease severity, etc.)? Some greater detail about these considerations would be useful.

Response 3.2 to reviewer 2:
We thank the reviewer for suggesting this addition. We have now described our purposive sampling strategy for the interviews in the manuscript on page 13: *"Participants will be purposively sampled to ensure diversity on the following characteristics: type of diabetes, sex, age and completion rate."*

Comment 3.3 from reviewer 2:
Similarly, some attention to the effects of time during the monitoring interval would be beneficial. Are there decreases in the reliability of the module measures over the course of the monitoring period, suggesting a failure to maintain consistent participant engagement over the 10-week period?. How will this be tested?

Response 3.3 to reviewer 2:
Participant engagement will be explored in two ways; first via the interviews (n=20), allowing the participants to elaborate on if and why their motivation changed over the course of the study, and how this could be changed in future studies, and secondly via the quantitative analysis of completion rates on the full sample size (n=600). The above-mentioned test-retest reliability analyses will also show if the app responses change considerably across a 4-week period.

Comment 4 from reviewer 2:
4. I have one last, very minor point. I suggest that you re-think the terminology used in the third paragraph under "Phase 3" on page 15. You refer to the "target population for this study" as 600 adults with T1DM. I believe that the correct way to phrase this would be to indicate that the target population involves European adults with T1DM, and that the "sample of participants" chosen to represent this population would consist of 600 adults age 18-85 from specialist clinics, etc.

Response 4 to reviewer 2:
We have addressed this point with the following text on page 12:
"The target population for this study is European adults with T1DM or insulin-treated T2DM, and the sample of participants chosen to represent this population will consist of 600 adults (aged 18-85 years) recruited from eight specialist diabetes centres across five countries (Austria, Denmark, France, The Netherlands, United Kingdom)."

Reviewer: 3
Dr. Jennifer Warnick, Warren Alpert Medical School of Brown University

Overall from reviewer 3:
Thank you for the opportunity to review this protocol manuscript. The manuscript is well written, and the study is well designed. Below are a few very minor suggestions to improve clarity and potential impact.

Overall response to reviewer 3:

We thank the reviewer for taking the time to review our manuscript and provide valuable feedback. A response to each comment is provided below.

Comment 1 from reviewer 3:
1. In the abstract, consider adding that you will be collecting CGM data as this adds to the novelty and potential impact of the project.

Response 1 to reviewer 3:
We agree with the reviewer that using the Hypo-METRICS app in conjunction with CGM is novel and have therefore added this information to the abstract (page 2) as follows: *"The first version was released mid-2020 for use (in conjunction with continuous glucose monitoring and activity tracking) in the Hypo-METRICS study; an international observational longitudinal study"*

Comment 2 from reviewer 3:
2. Phase 2 Developing Items: I agree with the authors' choice to include branching logic for certain questions. However, participants may learn that answering certain questions will lead to more questions. I would suggest examining the accuracy of self-reported hypos to the CGM data to account for this potential issue.

Response 2 to reviewer 3:
We agree this is an important point. Since the submission of this manuscript, we have developed a detailed interview guide for the interviews that we plan to conduct, and as part of this guide, we have specifically added questions regarding if the branching logic impacted on participants' use of the app and responses (i.e., if they specifically did not report certain hypos to avoid additional questions). Further, as part of the primary objective of the Hypo-METRICS study, which will be further detailed in the Hypo-METRICS study protocol (as described in **Response 1 to reviewer 2),** the accuracy of the self-reported hypos to the CGM detected hypos will be explored in detail.

Comment 3 from reviewer 3:
3. Are participants able to self-initiate EMA prompts if they have a hypo-episode between fixed EMA prompts?

Response 3 to reviewer 3:
Participants are able to open and respond to the EMA questions within the pre-defined time-intervals and once the time-interval has passed, the check-in on that day is no longer available. To emphasize this further in the manuscript, we have added the following text on page 12: "*Participants could self-initiate the check-ins but were not able to complete the individual check-ins outside these time-intervals.*"

Comment 4 from reviewer 3:
4. Phase 2 User Testing: Please clarify the user-testing process. It's not clear whether the participants met as a group and provided feedback on the questions and the potential of answering them 3x per day or whether participants actually tried using the app for a certain amount of time. I interpreted the former in the body of the text and the latter in the discussion section.

Response 4 to reviewer 3:
We agree this was lacking clarity in the original manuscript so have added the following on page 9:

"Participants met as two separate groups in two sessions to provide feedback on the app content; however participants only tested the questions as a paper-and-pencil version and not in the uMotif platform."

Comment 5 from reviewer 3:
5. Phase 2 App Design: Could a participant complete an EMA prompt outside of the time-interval? For example, would it be possible for a participant to complete all 3 daily prompts in one sitting (in the evening)? If not, please specify that the prompts expire or close after the time interval ends.

Response 5 to reviewer 3:
As indicated above (please see "Response 3 to reviewer 3") it is not possible for the participants to complete an EMA prompt outside the time-interval.

Comments 6-11 from reviewer 3:
6. Phase 3 Study Design: Why did the authors choose 10-weeks for the study duration? Please justify this decision in the manuscript.
7. Phase 3 Study Design: If a participant already owns/wears a CGM, will they be required to to wear a 2nd one for the study?
8. Phase 3 Study Design: Will participants be compensated and if so, will participants be compensated for each EMA prompt completed?
9. Will EMA adherence be monitored by study staff and will study staff intervene if participants are not completing prompts?
10. Please include any demographic and general diabetes management variables you plan to examine in your analyses. For example, will you be collecting data to identify participants who have pumps vs. engage in multiple daily injections and use a glucose meter vs. CGM daily?
11. Will the study be providing smartphones and/or data plans to participants who do not own personal devices or data plans?

Responses 6-11 to reviewer 3:

We thank the reviewer for this collection of very important and relevant questions. In line with "**Response 1 to reviewer 2"** above, there is an important distinction between 'The Hypo-METRICS clinical study' and 'Development of the Hypo-METRICS app'. The purpose of the present manuscript is to describe the development of the Hypo-METRICS app, rather than the details of the Hypo-METRICS protocol. For the reviewer's information, we have provided a response to each question below. The answers to these will be detailed in the Hypo-METRICS clinical study protocol (that will be submitted for publication later).

6: The 10 weeks were chosen due to power calculations for the primary objective.

7: Participants will be equipped with a blinded CGM for the purpose of the study and will therefore wear 2 CGMs if they already have one.

8: Participants will receive a voucher for finishing the 10-week participation in the clinical study, and will also be allowed to keep a FitBit tracker (used in the study). Each EMA prompt will not be compensated.

9: Participants will have weekly calls with study site staff who will be monitoring completion rates and any serious adverse events. The uMotif platform will further send additional reminders when the participant has been inactive for more than 2 days.

10: The analysis for the primary objective, as well as other objectives not outlined in this paper, will be further detailed in the Hypo-METRICS clinical study protocol.

11: Participants will borrow a smartphone for the duration of the study if they do not have one themselves.

VERSION 2 – REVIEW

REVIEWER	Thomas Kamarck University of Pittsburgh, Psychology
REVIEW RETURNED	10-Nov-2021

GENERAL COMMENTS	I may have missed this, but I didn't see any response to my previous comment about looking at the association between EMA responses and continuous glucose monitoring. It sounds like this is a major part of the validation study, but there is no description of how these analyses will be conducted.

REVIEWER	Jennifer Warnick Warren Alpert Medical School of Brown University, Psychiatry and Human Behavior
REVIEW RETURNED	11-Nov-2021

GENERAL COMMENTS	None. The authors have adequately addressed all comments.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Dr. Thomas Kamarck, University of Pittsburgh

Comment 1 from reviewer 2:
I may have missed this, but I didn't see any response to my previous comment about looking at the association between EMA responses and continuous glucose monitoring. It sounds like this is a major part of the validation study, but there is no description of how these analyses will be conducted.

Response 1 to reviewer 2:
We thank the reviewer for his additional comment. We apologise if the previously submitted comments to the reviewer did not make a clear and important distinction between 1) the "Hypo-METRICS app design and planned validation" and 2) the "Hypo-METRICS clinical study protocol". The associations between EMA responses and continuous glucose monitoring are not part of the planned validation described in this manuscript. However, we do plan to conduct these analyses as part of the Hypo-METRICS clinical study. Therefore, the planned analyses are described in the Hypo-METRICS clinical study protocol paper.

To further clarify this in the manuscript, we have added the following sentences (on page 12):
One of the key aims of the Hypo-METRICS clinical study is to explore associations between CGM data and Hypo-METRICS app responses. To avoid double reporting of results, these analyses will not be included in the current validation study.