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)	A digital epidemiological and citizen science methodology to capture prospective physical activity in free-living conditions: A SMART platform study
AUTHORS	Katapally , Tarun; Chu, Luan

VERSION 1 – REVIEW

REVIEWER	Rubén López-Bueno University of Zaragoza, Spain
REVIEW RETURNED	28-Feb-2020

GENERAL COMMENTS	Dear authors, The study subject is a hot research topic and is important for the sports science field. However, there are some issues related to aims, participants and readability than need to be addressed. 1.- Is there any possibility to offer values in the abstract? 2.- An important limitation of the study is that PA is self-reported 3.- Lines 105-108 The objective of the study needs to be more succinct and clear.Besides, Have you thought of using more punctuation marks along the text (i.e. periods and semi-colons). 4.- How does EMA categorise domains? Is the user who set it? 5.- There is more to tell about the data sample. At least a table with characteristics and a brief description. 6.- There are many missing participants. Any cause imputable?

REVIEWER	Nesrin Nazlieva Uppsala University, Sweden
REVIEW RETURNED	05-Mar-2020

GENERAL COMMENTS	Abstract: The abstract was clear and well-defined. I would suggest mentioning the number (N=89) of included respondents, perhaps at the beginning of the sentence at line 46. I understand that you might hesitate due to the small number of respondents but it remains essential. Results: At line 215, you mention the number of females; however, not males'. I understand we should assume the rest might have identified themselves as males but this should be explicitly stated as there is the possibility of them deciding not to identify themselves. My suggestion is making it more explicit. Strengths and Limitations: I would suggest elaborating on the importance of "the development of novel and replicable methodology to capture prospective PA comprehensively from large populations using citizen-owned devices." (lines 308-309).

REVIEWER	Eric T Hyde Centers for Disease Control and Prevention, USA
REVIEW RETURNED	14-Apr-2020

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. This study describes a novel citizen-science methodology approach to using ecological momentary assessments (EMAs) for measuring prospective physical activity across different social and physical contexts, and comparing these measures to a validated retrospective self-reported measure of physical activity (IPAQ). While the research question is interesting and important for improving the use of EMAs for measuring physical activity, there are fundamental issues of study design and analysis that outweigh the strengths of this study. First, the different methodologies of physical activity measurement did not occur simultaneously. Rather, each measured physical activity across different time frames, therefore hampering the ability to directly compare the EMA approach to the IPAQ. The IPAQ was administered at the beginning of the study, and asked about physical activities over the previous 7 days that were of at least 10 consecutive minutes in duration. Conversely, the EMAs measured physical activity each day for 8 consecutive days. This is a limitation of the study that is not mentioned by the authors. Next, the authors main findings are not supported by the data. Based on the results from Tables 1 and 2, overall, light-, and moderate-to-vigorous PA did not differ by measurement method (p-values: 0.331, 0.322, 0.995, respectively). The authors claim in lines 223-225 in the Results section: "These findings show that although there are no significant differences between activities intensities reported via EMAs and IPAQ, citizen scientists consistently overestimated their PA using IPAQ." First, this is an interpretation of results, which goes beyond the scope of the Results section and and should be removed. Next, it is inappropriate to assert that participants over-reported their PA using IPAQ when the standard deviations for both methods are large and there's no gold standard comparison group. Over-reported compared to what? It is also possible that PA was under-reported using EMAs. This assertion is made throughout the manuscript, though it is misleading. In addition, Tables 4 and 5 were not included in the manuscript. Lastly, there are several other corrections that would need to be made before this could be considered for publication. For example, Abstract: 1. Participants: From the initial sample of 538 citizen scientists, only 89 participants were included in the analysis. It is misleading to report 538 citizen scientists here. 2. Results: None of the results from the significance testing are reported here. Lines 54-56 ("Daily time-triggered EMAs were able to capture not only prospective light and moderate-to-vigorous PA, but also enabled PA reporting across varied physical and social contexts") are not results, but rather methods of measurement. Methods

	Why was "answering the daily EMA on at least 3 days" an inclusion criteria? Is this based off criteria used in other studies? The authors list this criteria as a "key advancement of this study" in the Discussion, yet there is no scientific justification reported for it's use. Results As mentioned previously, Tables 4 and 5 are missing. Discussion Line 290 - Tables/Figures should not be cited in the discussion. Additionally, these findings are not particularly novel, as we have long known that most PA accumulation occurs at home, through active transport, at the gym, and in public spaces (parks), and that most PA is performed alone. This should be expanded upon or removed altogether. Line 297 - "Current evidence clearly indicates that there is no gold standard in assessing prospective PA using mobile EMAs" this statement needs references. Line 308 - If the primary strength of this study is the novel and replicable methodology, then these features should be expanded upon further in the discussion. While this study has merit, it needs further revision before it is suitable for publication.

VERSION 1 – AUTHOR RESPONSE

Reviewer # 1:
Comment #1: Is there any possibility to offer values in the abstract?

Response #1: We have included values to the revised abstract.

Comment #2: An important limitation of the study is that PA is self-reported.

Response #2: Yes, we have now added this limitation to strengths and limitations bullet points (after the abstract) (line 76), as well as the strengths and limitations subsection in the Discussion (line 326).

Comment #3: Lines 105-108: The objective of the study needs to be more succinct and clearer. Besides, have you thought of using more punctuation marks along the text (i.e. periods and semi-colons).

Response #3: We have made objectives sentence more succinct and clear. The text (line 111-115) now reads: “The objective of this study is to address current deficiencies in active living EMA approaches by developing a replicable citizen science methodology of smartphone-based EMAs to capture prospective PA within free-living social and physical contexts. This objective will be achieved by leveraging citizen-owned smartphones running on both Android and iOS systems, and by comparing EMA measures with traditional self-report measures of PA within the same cohort”

Comment #4: How does EMA categorise domains? Is the user who set it?

Response #4: The EMA domains were described in the “measures” and “Derived variables – Intensities and Volume of PA” subsections of the Methods section. Users did not set the domains as that would not standardize measures. The EMAs were designed so that users could respond to

triggers and report data, which are categorized to physical and social contexts. Thereafter, the research team derived these data into comparable domains across both EMAs and IPAQ.

Comment #5: There is more to tell about the data sample. At least a table with characteristics and a brief description.

Response #5: We have now included a table (Table 1) showing participants' characteristics and summarized demographics at the beginning of the results section.

Comment #6: There are many missing participants. Any cause imputable?

Response #6: Since the purpose of this study was to develop a methodology, rather than modeling to infer associations, we did not employ data imputation techniques. EMAs are currently novel methods that need standardization and missing data in EMAs studies are not uncommon. More importantly, we applied a strict inclusion criterion, where we included only participants with valid PA information on at least 3 days in the final analysis, which resulted in exclusion of most participants. This rigorous inclusion criterion is an essential step in standardizing EMA measures, and obtaining valid and reliable data.

Reviewer # 2

Comment #1: Abstract: The abstract was clear and well-defined. I would suggest mentioning the number (N=89) of included respondents, perhaps at the beginning of the sentence at line 46. I understand that you might hesitate due to the small number of respondents but it remains essential.

Response #1: We have now included the final sample size of 89 in the abstract: "538 citizen scientists (≥18 years) provided PA data during 8 consecutive days using a custom-built smartphone app, and after applying a rigid inclusion criteria, 89 were included in the final analysis".

Comment #2: Results: At line 215, you mention the number of females; however, not males'. I understand we should assume the rest might have identified themselves as males but this should be explicitly stated as there is the possibility of them deciding not to identify themselves. My suggestion is making it more explicit.

Response #2: We have now explicitly stated the number of males: (line 215-216). We also provided full sample characteristics in Table 1.

Comment #3: Strengths and Limitations: I would suggest elaborating on the importance of "the development of novel and replicable methodology to capture prospective PA comprehensively from large populations using citizen-owned devices." (lines 308-309).

Response #3: Thank you for this suggestion. We have elaborated the sentence: "The primary strength of the study is the development of novel and replicable methodology to capture prospective PA comprehensively from large populations using citizen-owned devices. This citizen science approach, if replicated appropriately, can transform surveillance of physical PA among large populations by leveraging citizen owned-devices. Implementing such innovative approaches of PA surveillance will be critical to develop appropriate interventions to address global physical inactivity". (Lines: 327-331)

Reviewer: 3

Comment #1: First, the different methodologies of physical activity measurement did not occur simultaneously. Rather, each measured physical activity across different time frames, therefore hampering the ability to directly compare the EMA approach to the IPAQ. The IPAQ was administered at the beginning of the study, and asked about physical activities over the previous 7 days that were of at least 10 consecutive minutes in duration. Conversely, the EMAs measured physical activity each day for 8 consecutive days. This is a limitation of the study that is not mentioned by the authors.

Response #1: This is a good point and we appreciate you bringing it up. It is true that IPAQ and EMAs measured PA in different timeframes. As IPAQ captures data retrospectively and EMAs capture data prospectively, they cannot be issued simultaneously. We realize that IPAQ could have been issued on day 8, but we did not do that based on evidence from our pilots, which showed that compliance to burdensome traditional recall surveys such as IPAQ is much higher when it is issued as close to participant enrolment in the study as possible i.e., day 1 of citizen scientists downloading the app. We have now mentioned this in the limitations section: “Another limitation is that IPAQ and EMAs measured PA in different timeframes. As IPAQ captures data retrospectively and EMAs capture data prospectively, they cannot be issued simultaneously. However, IPAQ could have been issued on day 8, but we refrained from such late deployment based on the evidence from our pilots, which showed that compliance to burdensome traditional recall surveys such as IPAQ is much higher when it is issued as close to participant enrolment in the study as possible”. (Lines 334-339)

Comment #2: Next, the authors main findings are not supported by the data. Based on the results from Tables 1 and 2, overall, light-, and moderate-to-vigorous PA did not differ by measurement method (p-values: 0.331, 0.322, 0.995, respectively). The authors claim in lines 223-225 in the Results section: " First, this is an interpretation of results, which goes beyond the scope of the Results section and and should be removed. Next, it is inappropriate to assert that participants over-reported their PA using IPAQ when the standard deviations for both methods are large and there's no gold standard comparison group. Over-reported compared to what? It is also possible that PA was under-reported using EMAs. This assertion is made throughout the manuscript, though it is misleading. In addition, Tables 4 and 5 were not included in the manuscript.

Response #2: It is correct that citizen scientists could have under-reported PA using EMA, hence the assertion that they over-reported using IPAQ has been modified by clearly stating that PA was over-reported using IPAQ, in comparison with EMAs. Yes, the results are not significant across intensities of PA and we acknowledge that, however the results are significant across different physical contexts as depicted in Table 4. Moreover, the interpretation of results is unbiased because we also did correlational analysis between IPAQ and EMA and showed moderate correlation between both measures. In the end, the objective of this study is to move towards standardizing EMAs and comparing with IPAQ is a start. The lack of significant results when comparing intensities of PA between IPAQ and EMAs is probably due to the low sample size, which is a trade-off for choosing a rigorous inclusion criteria (explained below in comment #5) – again an effort towards standardization. We have now rectified the error of missing tables, which seem to have been deleted erroneously during the formatting process.

Lastly, there are several other corrections that would need to be made before this could be considered for publication. For example,

Comment #3: Abstract:
Participants: From the initial sample of 538 citizen scientists, only 89 participants were included in the analysis. It is misleading to report 538 citizen scientists here.

Response #3: We have now included the final sample size of 89 in the abstract.

Comment #4: Abstract
Results: None of the results from the significance testing are reported here. Lines 54-56 ("Daily time-triggered EMAs were able to capture not only prospective light and moderate-to-vigorous PA, but also enabled PA reporting across varied physical and social contexts") are not results, but rather methods of measurement.

Response #4: The abstract has been revised to report statistical values from results.

Comment #5: Abstract
Methods: Why was "answering the daily EMA on at least 3 days" an inclusion criteria? Is this based off criteria used in other studies? The authors list this criteria as a "key advancement of this study" in the Discussion, yet there is no scientific justification reported for it's use.

Response #5: EMAs are currently novel methods that need standardization. We applied a strict inclusion criterion, where we included only participants with PA data on at least 3 days in the final analysis, which resulted in exclusion of most participants. We did this even at the risk of reducing our sample size because this rigorous inclusion criterion is an essential step in standardizing EMA measures, and obtaining valid and reliable data. This is not very different from accelerometry standardization methods, where data are considered valid if participants wear accelerometers for at least several hours (e.g. 10 hours) on at least 2-3 days in a one-week study period. We have expanded this explanation in the discussion as this is an important step towards standardization of prospective PA EMA data collection (Lines: 281-291).

Comment #6: Results. As mentioned previously, Tables 4 and 5 are missing.

Response #6We have now rectified the error of missing tables, which seem to have been deleted erroneously during the formatting process.

Comment #7: Discussion
Line 290 - Tables/Figures should not be cited in the discussion. Additionally, these findings are not particularly novel, as we have long known that most PA accumulation occurs at home, through active transport, at the gym, and in public spaces (parks), and that most PA is performed alone. This should be expanded upon or removed altogether.

Response #7: Tables/figures have been removed from the discussion. We did not state that these findings are novel. These findings are merely a result of our novel EMA design, which allow for capturing physical and social contexts along with volume and intensity of PA prospectively using a single time-triggered EMA per day. We have now clarified this in the discussion (Lines 300-302).

Comment #8: Discussion. Line 297 - "Current evidence clearly indicates that there is no gold standard in assessing prospective PA using mobile EMAs" this statement needs references.

Response #8: We have now added references to support this statement.

Comment #9: Discussion. Line 308 - If the primary strength of this study is the novel and replicable methodology, then these features should be expanded upon further in the discussion.

Response #9: Thank you for this suggestion. We have further expanded discussion to emphasize this strength (Lines: 312-323; 326-330).

VERSION 2 – REVIEW

REVIEWER	Rubén López-Bueno University of Zaragoza, Spain
REVIEW RETURNED	27-Apr-2020

GENERAL COMMENTS	Dear authors, The manuscript has substantially improved regarding the initial version.

REVIEWER	Nesrin Nazlieva Uppsala University, Sweden
REVIEW RETURNED	28-Apr-2020

GENERAL COMMENTS	The authors have addressed the editor's and reviewers' comments and suggestions.

REVIEWER	Eric Hyde USA
REVIEW RETURNED	05-May-2020

GENERAL COMMENTS	The authors adequately addressed the comments of previous reviewers and the paper is much improved.