

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)	The In-Motion-App for remote General Movement Assessment: A multi-site observational study.
AUTHORS	Adde, Lars; Brown, Annemette; van den Broeck, Christine; DeCoen, Kris; Eriksen, Beate; Fjørtoft, Toril; Groos, Daniel; Ihlen, Espen; Osland, Siril; Pascal, Aurelie; Paulsen, Henriette; Skog, Ole; Sivertsen, Wiebke; Støen, Ragnhild

VERSION 1 – REVIEW

REVIEWER	Maria Luisa Scattoni Istituto Superiore di Sanità, Italy
REVIEW RETURNED	15-Oct-2020

GENERAL COMMENTS	<p>Thank you for the opportunity to review the manuscript “The In-Motion-App for remote General Movement Assessment: A multi-site observational study.” Examining whether videos taken by parents with a hand-held camera are in accordance with required standards and could be remotely scored by a trained General Movement Assessment (GMA) observer encourages the clinical feasible computer-based movement analysis. The use of smartphone technology could support an equal access to GMA evaluation without geographical constraints and an early identification of Cerebral Palsy or other neurodevelopmental disorders. According to the authors, this is the first research in this field. The research method used is appropriate, but there are some concerns about this manuscript, which are described in detail below.</p> <p>Introduction:</p> <p>-Page 1, lines 15-16: The authors did not mention the writhing type of general movements. I suggest explaining the reasons. Do they not have a high predictive validity for later CP and, thus, the authors focused on fidgety movements?</p> <p>-Page 3, lines 34-37: The authors declare that they have developed a machine-learning model which predicted CP with high accuracy comparable to observational GMA. I suggest specifying the level of accuracy and the video capture methods on which they based the machine-learning model (i.e.: video acquired by clinicians or parents, stationary or hand-held camera).</p> <p>Participants:</p> <p>- “Parents of infants admitted to one of five participating level III-IV Neonatal Infant Care Units (NICU) in Norway, Denmark and</p>
-------------------------	--

Belgium from 2017-2018 were consecutively recruited at referral to the hospital follow-up program before discharge from the NICU. Families were recruited based on willingness to participate and the infant being evaluated as at high-risk of CP". Please, specify the number of centres and families recruited in the current study.

Data collection procedure:

- Page 6, lines 21-23: how participants were assisted to download and install the In-Motion-App? For example: by phone, by email, during a clinical assessment

- Page 6, lines 25-27: are the weeks post term age (PTA) at corrected age for infants recruited in Norway, Denmark and Belgium? I suggest specifying it

- Page 6, lines 28-29: How the local coordinator contacted the family to ask the reason why they had not uploaded any videos? For example: by phone, by email, during a clinical assessment

The In-Motion-App and instructional video:

- Who developed In-Motion-App and the instructional video?

- Page 4, lines 58-60: please, specify the instruction given to the parents. For example, how clothing of the infants should be, how the surface/underlay, ect.

Assessment of General Movements:

-All videos classified as "GMA scorable" were consecutively assessed by one certified and experienced. This should be stated as limitation of the study.

In-Motion body point tracking:

- Why did the authors use only a subset of videos recorded by the In-Motion-App for automatic

infant motion tracking, instead of the total number of videos?

-Why did the authors chosen the displacement of 7-body points and set the size of the circular area to 10% of the infant head size? Is there any reference? I suggest including more details

Result:

-Did the authors carry on the analysis considering the two different time windows (12+1 - 13 +6 and 14+1 - 17+6 weeks post term age)?

-How did the authors documented whether the hand-held video had optimal stability, events of abrupt displacement, was predominantly unstable, whether an adequate underlay was used and whether overall video image quality was sufficient? Did the authors consider these factors in the analysis? If yes, how the authors consider them.

	<p>-It might be interesting to see if results change depending on the country where the app was used</p> <p>Discussion:</p> <ul style="list-style-type: none"> - More discussion should be provided on the survey results <p>Tables:</p> <p>Table 3, first line: substitute the comma with point for the decimals.</p> <p>Minor changes:</p> <ul style="list-style-type: none"> - check for clarity the fourth point of the list “Strengths and limitations of this study” - check the terms “homed-based” and “onsie” through the manuscript - check Page 5 Line 58 “parental worries, were a statement was made”
--	--

REVIEWER	Amanda Kwong Murdoch Children's Research Institute
REVIEW RETURNED	18-Oct-2020

GENERAL COMMENTS	<p>This manuscript details a study of the In-Motion smartphone app to record infant general movements on parents' phones for both human and potential automated general movements assessment. The app was successfully used by parents to capture videos which were compliant with conditions for auto-detection of key body points needed for further computer processing of movements. The methods and results are articulated well. I have only the following minor concerns to be addressed:</p> <ul style="list-style-type: none"> * The authors state that recruitment occurred between 2017 and 2018. Please clarify the length of recruitment. If data are available for excluded/declined consent versus number approached, please provide this in the supplementary information. * The study uses a single human assessor to score infant movements, where most studies advocate for cases to be scored by at least two assessors. Please justify why only one assessor was used. * The authors state that a subset of 66 videos were used for automatic infant motion tracking. Please describe how these 66 videos were selected. * There appears to be some confusion as to whether the 66 videos were automatically tracked by the app, but later within the same paragraph (line 40) states that all videos used in the study were manually annotated. * Two filming windows were set, yet participants were able to submit videos before, and after this time, or submit more than two
-------------------------	---

	videos. Please justify why these two time points were used if families were then able to submit videos outside these parameters.
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

Introduction:

*Page 1, lines 15-16: The authors did not mention the writhing type of general movements. I suggest explaining the reasons. Do they not have a high predictive validity for later CP and, thus, the authors focused on fidgety movements?

Reply: Yes, we agree that the writhing type of general movements do not have as high predictive validity for later CP than the fidgety type of general movements and this is the reason for us to focus on fidgety movements. To clarify this, we have made the following change to the sentence in the introduction (second paragraph, page 3): “The fidgety type of general movements (GMs) observed at 9-20 weeks’ corrected age has shown the highest predictive validity for later CP, compared to the writhing type of general movements observed before 9 weeks corrected age (5-8)”.

*Page 3, lines 34-37: The authors declare that they have developed a machine-learning model which predicted CP with high accuracy comparable to observational GMA. I suggest specifying the level of accuracy and the video capture methods on which they based the machine-learning model (i.e.: video acquired by clinicians or parents, stationary or hand-held camera).

Reply: We have specified sensitivity, specificity, video acquired by clinicians and use of a stationary camera and changed the sentence as follows (paragraph four, page 3) : “Our research group has recently presented a machine-learning model which predicted CP with high accuracy (sensitivity of 92%, specificity of 81%) performed by clinician using a stationary camera, comparable to observational GMA (15).”

Participants:

**“Parents of infants admitted to one of five participating level III-IV Neonatal Infant Care Units (NICU) in Norway, Denmark and Belgium from 2017-2018 were consecutively recruited at referral to the hospital follow-up program before discharge from the NICU. Families were recruited based on willingness to participate and the infant being evaluated as at high-risk of CP”. Please, specify the number of centres and families recruited in the current study.

Reply: We have specified the number of hospitals and families recruited in the study, corrected the years for inclusion and changed the sentence as follows (page 4): “Parents of infants admitted to one of five participating level III-IV Neonatal Infant Care Units (NICU) in Norway (three hospitals including 13, 11 and 4 families, respectively), Denmark (one hospital including 43 families) and Belgium (one hospital including 15 families) from 2018-2019 (12 months recruitment period) were consecutively recruited at referral to the hospital follow-up program before discharge from the NICU.”

Data collection procedure:

*Page 6, lines 21-23: how participants were assisted to download and install the In-Motion-App? For example: by phone, by email, during a clinical assessment

Reply: We agree that the sentence needs some specification and have changed the sentence as follows (page 4): “Included participants were assisted by a research physiotherapist/pediatrician at the

time of inclusion to download and install the In-Motion-App by smartphone, containing the instructional video from Google Play or iTunes.”

*Page 6, lines 25-27: are the weeks post term age (PTA) at corrected age for infants recruited in Norway, Denmark and Belgium? I suggest specifying it

Reply: We have specified and changed the sentence (page 4): “They got information about the time window for performing two separate video recordings for their infant between 12+1 - 13 +6 and 14+1 - 17+6 weeks post term age (PTA).

*Page 6, lines 28-29: How the local coordinator contacted the family to ask the reason why they had not uploaded any videos? For example: by phone, by email, during a clinical assessment

Reply: We have specified as follows (page 4): “If no videos were returned from the families before 17+6 weeks PTA, the local study coordinator contacted the family by phone to ask the reason why they had not uploaded any videos.”

The In-Motion-App and instructional video:

*Who developed In-Motion-App and the instructional video?

Reply: We acknowledge the need for this information and have adjusted the following sentence (first paragraph): “The In-Motion-App and instructional video was designed by GMA trained personnel (LA, TF, RS, SO) at St. Olavs Hospital in Trondheim, Norway, for parents to give basic insight into recording....”

*Page 4, lines 58-60: please, specify the instruction given to the parents. For example, how clothing of the infants should be, how the surface/underlay, ect.

Reply: We have specified instructions given. The sentence is now as follows (second paragraph, page 5):“The main themes aimed at ensuring quality standards for remote GMA included: a) clothing of infant (just a diaper or a onsie), b) surface/underlay for infant (single-color blanket or rug), c) lighting (enough light avoiding sidelight that can cause shadows), d) state of infant (awake, alert, content, not disturbing baby, no pacifier), e) positioning (baby on floor- stand next to the baby’s feet, whole body must be visible) and f) length of video (3 minutes).

Assessment of General Movements:

*All videos classified as “GMA scorable” were consecutively assessed by one certified and experienced. This should be stated as limitation of the study.

Reply: We acknowledge this limitation and have made an addition to a sentence concerning limitations of the study in the discussion (fifth paragraph, page 10): “There is a risk that more videos might have been classified as non-scorable by another GMA observer with less training and experience or if there had been several GMA experts observing the same videos.”

In-Motion body point tracking:

*Why did the authors use only a subset of videos recorded by the In-Motion-App for automatic infant motion tracking, instead of the total number of videos?

Reply: All 7 body points in 5493 video frames were manually annotated and used as ground truth for automated body point position evaluation in this study. This process was labor demanding and consequently only performed on the subset containing 66 videos received before the chosen date

19th of September 2018. We have changed the following sentence to clarify this issue (page 5): “A subset of 66 videos from 36 infants recorded by the In-Motion-App by September 19th 2018 was used for automatic infant motion tracking....”

*Why did the authors chosen the displacement of 7-body points and set the size of the circular area to 10% of the infant head size? Is there any reference? I suggest including more details

Reply: The trained body point tracker was presented in another study from our research group (please see reference in text). The main rationale using 7-body points in that study was the aim to track body movements representing different body segments (like upper and lower extremities, trunk and head) and not specific kinematic details about different joints.

The displacement threshold of 10% of head size is established as a common evaluation metric in human pose estimation for assessing the ability of body part tracking with high precision. We have specified as follows (page 5) in the text and added an reference in the reference list: “In accordance with the established metric for evaluating pose-estimation (17), size of the circular area was set to 10% of the infant head size and was normalized to adjust for different scaling (i.e., video zoom) (Figure 3)”.

Result:

*Did the authors carry on the analysis considering the two different time windows (12+1 - 13 +6 and 14+1 - 17+6 weeks post term age)?

Reply:

The effect of practicing could have been evaluated by comparing videos taken during the first with videos taken during the second time window. This was beyond the scope of the study and has not been analyzed.

*How did the authors documented whether the hand-held video had optimal stability, events of abrupt displacement, was predominantly unstable, whether an adequate underlay was used and whether overall video image quality was sufficient? Did the authors consider these factors in the analysis? If yes, how the authors consider them.

Reply: Hand-held video stability, events of abrupt displacement, predominantly unstable videos, adequate underlay and whether overall video image quality was sufficient was documented into yes/no categories by video observations by the GMA expert. These factors were only described as overall results demonstrating robustness of the tracker used on hand-held smartphone recordings. No detailed considerations were made in the analysis. We have clarified the issue in the following adjustment of the sentence in the manuscript (in the Assessment of video quality for remote GMA paragraph under Methods, page 5): “In addition, all videos were observed by the same GMA expert who also categorized yes/no whether the hand-held video had optimal stability,.....”

*It might be interesting to see if results change depending on the country where the app was used

Reply:

The potential impact of country or cultural context could have been evaluated by comparing videos taken in different countries. However, this was beyond the scope of the study, and additionally, number of included families from each country are relatively low and unevenly distributed. Such considerations could/should be included in future studies.

Discussion:

*More discussion should be provided on the survey results

Reply: We have reviewed the discussion carefully and added some sentences in the third paragraph in page 10 in the discussion. We hope this will complement the discussion with respect to the survey results.

Tables:

*Table 3, first line: substitute the comma with point for the decimals.

Reply: Thank you for the comment. We have substituted the comma with point for decimals in first line in Table 3.

Minor changes:

*check for clarity the fourth point of the list “Strengths and limitations of this study”

Reply: We have made changes to clarify the fourth point:

“Study did not assess socio-demographic factors as reasons for families not to record or return videos.”

*check the terms “homed-based” and “onsie” through the manuscript

Reply: The terms “homed-based” and “onsie” are checked throughout the whole manuscript.

*check Page 5 Line 58 “parental worries, were a statement was made”

Reply: Page 5 line 58 checked for correct spelling and meaning.

Reviewer #2

* The authors state that recruitment occurred between 2017 and 2018. Please clarify the length of recruitment. If data are available for excluded/declined consent versus number approached, please provide this in the supplementary information.

Reply: We have made a quality check of our inclusion period and found correct time period to be in 2018-2019, lasting for a 12 months period. Consequently, we have corrected the sentence, clarified the length of recruitment period and adjusted according to comments from reviewer 1 (page 4): “Parents of infants admitted to one of five participating level III-IV Neonatal Infant Care Units (NICU) in Norway (three hospitals including 13, 11 and 4 families, respectively), Denmark (one hospital including 43 families) and Belgium (one hospital including 15 families) from 2018-2019 (12 months recruitment period) were consecutively recruited at referral to the hospital follow-up program before discharge from the NICU.”

We acknowledge the relevance of supplementary data concerning excluded/declined consent versus number approached if data were available. However, after checking the issue it seems that such data is not available for all sites and can consequently not be provided in the supplementary information.

* The study uses a single human assessor to score infant movements, where most studies advocate for cases to be scored by at least two assessors. Please justify why only one assessor was used.

Reply: We appreciate and understand the question. This study did not focus on GMA for the prediction of CP and aimed at evaluating whether quality of recordings fulfilled standards making a GMA observation possible or not. The focus was not on the GMA classifications and associations with later outcome. Hence, we considered one certified and experienced GMA observer as adequate answering the main research question in our study. In a clinical or research context predicting CP by GMA we fully agree the strength using two observers. We have added a sentence in the “Assessment of General Movements” paragraph in page 5: “The use of one observer was chosen due to the study design not focusing on GMA and prediction of outcome.”

In addition, a sentence in the discussion concerning limitations of the study are changed due to correlated comment from reviewer #1 (fifth paragraph, page 10): “There is a risk that more videos might have been classified as non-scorable by another GMA observer with less training and experience or if there had been several GMA experts observing the same videos.”

*The authors state that a subset of 66 videos were used for automatic infant motion tracking. Please describe how these 66 videos were selected.

Reply: Please also see comment to reviewer #1. All 7 body points in 5493 video frames were manually annotated and used as ground truth for automated body point position evaluation in this study. This process was labor demanding and consequently only performed on the subset containing 66 videos received before the chosen date 19th of September 2018. We have described the issue and changed the following sentence to clarify this issue (page 5): “A subset of 66 videos from 36 infants recorded by the In-Motion-App by September 19th 2018 was used for automatic infant motion tracking....”

* There appears to be some confusion as to whether the 66 videos were automatically tracked by the app, but later within the same paragraph (line 40) states that all videos used in the study were manually annotated.

Reply: We appreciate the comment and the opportunity to clarify and be a bit more specific. We have specified as follows (page 5): “All 7 body points in the 5493 video test frames used in this study were manually annotated and used as ground truth, comparing them with the automatic body point positions for evaluation.”

* Two filming windows were set, yet participants were able to submit videos before, and after this time, or submit more than two videos. Please justify why these two time points were used if families were then able to submit videos outside these parameters.

Reply: We acknowledge the relevant question. We wanted to evaluate whether submitted videos were in accordance to the communicated filming windows in the app for the individual infant/family without programming parameters for these time windows into the app. This is now discussed in the third paragraph on page 10 in the discussion and we have changed the sentence to emphasize the issue more with respect to limitations: “These findings indicate a limitation in the design of the app reminders and lack of programmed filming windows parameters. These app functionalities need to be improved in a process involving the users”.

VERSION 2 – REVIEW

REVIEWER	Maria Luisa Scattoni Istituto Superiore di Sanità, Italy
REVIEW RETURNED	21-Dec-2020

GENERAL COMMENTS	The authors adequately addressed my concerns
-------------------------	--

REVIEWER	Amanda Kwong Murdoch Children's Research Institute, Australia
REVIEW RETURNED	22-Dec-2020

GENERAL COMMENTS	<p>Thank you, authors for revising your original manuscript. The changes satisfy queries raised from the first version. However, clarity in the process for manual and automatic annotation currently remains unclear.</p> <p>Page 5 line 49 – Eighty and 20% out of 5493 frames - this sentence is not clear</p> <p>Page 5 “In motion body point tracking” - It is still not clear if the 5493 frames were from the subset of 66 videos that were manually annotated. Can I please clarify – the first step is you have manually annotated each of the 5493 frames (none are automatically annotated) to pick out the 7 body points, as if in preparation for automatic infant motion tracking? I had to read page 5 under “In-Motion body point tracking” several times to interpret this, though I still don’t think I have the meaning correct. Later in the paragraph, it points out that automatic body point positions were compared – so were these 5493 frames both manually and automatically annotated for 7 body points and then compared for accuracy? It is still not clear what the purpose of the 80/20% of 5493 frames is.</p> <p>Distinct similarities exist between Kwong et al. 2018 and the current study (e.g. reporting of return rates and survey engagement, flow diagram Figure 3, time frames set for the two filming windows, setting in-app reminders for filming and parent survey) with only the parent survey referenced back to Kwong et al. 2018.</p>
-------------------------	---

VERSION 2 – AUTHOR RESPONSE

We appreciate that reviewer #1 found that we adequately addressed the raised concerns in the first review process. We also appreciate that reviewer #2 found the queries raised from the first version satisfactorily changed and give us the possibility of further clarifications with respect to the clarity of the motion tracker paragraph in page 5 of the main document.

Below you will find our point-by point comments to reviewer #2 concerns. All changes of the manuscript according to these points have been marked using “Track Changes” function in Microsoft Word in the revised version of the manuscript.

Reviewer #2

*However, clarity in the process for manual and automatic annotation currently remains unclear. Page 5 line 49 – Eighty and 20% out of 5493 frames - this sentence is not clear Page 5 “In motion body point tracking” - It is still not clear if the 5493 frames were from the subset of 66 videos that were manually annotated. Can I please clarify – the first step is you have manually annotated each of the 5493 frames (none are automatically annotated) to pick out the 7 body points, as if in preparation for automatic infant motion tracking? I had to read page 5 under “In-Motion body point tracking” several times to interpret this, though I still don’t think I have the meaning correct. Later in the paragraph, it points out that automatic body point positions were compared – so were these 5493 frames both

manually and automatically annotated for 7 body points and then compared for accuracy? It is still not clear what the purpose of the 80/20% of 5493 frames is.

Reply: We totally agree that the described process remains unclear and acknowledge that the description in the “motion tracker” paragraph describing manual and automatic annotations needs more clarity. Hence, we appreciate the reviewer comments and contribution, making it possible for us to improve descriptions. Our clarification according to reviewer #2 comments is as follows:

Training of the infant motion tracker:

The 7-body points infant motion tracker algorithm has previously been trained on high-risk infants that had participated in a previous study (8). The reader is referred to Groos et al (16) for details about the training procedures of the infant motion tracker. The following two sentences from the middle part of page 5 “In-Motion body point tracking” paragraph is now changed and moved to the beginning of the paragraph for clarity of the previously performed training and stepwise process: “The infant motion tracker algorithm consists of a convolutional neural network trained on 7-body points on 14900 video frames on high-risk infants that had participated in another study from our group (8). For further technical details of the previous trained convolutional neural net, the reader is referred to Groos et al (16)”.

Subset of videos and selected frames for the evaluation of the infant motion tracker:

For evaluation of the infant motion tracker algorithm, 5493 video frames were selected from the subset of 66 videos and used for automatic infant motion tracking in this study using the previous trained neural net (16). This step is clarified by changing the sentence in the paragraph as follows: “To evaluate the infant motion tracker, 5493 video frames was selected from a subset of 66 videos from 36 infants, recorded by the In-Motion-App by September 19th 2018”

Purpose of 80/20% of 5493 frames as a part of the evaluation:

The purpose of selection of 80/20% of 5493 frames is to specify that 80% were selected by random and that 20% were specifically selected in order to include difficult body part occlusions that can occur in young infants. The sentence is moved and changed as follows for clarity: “Eighty percent out of the selected 5493 frames were selected by random. The other 20% were selected manually in order to include body part occlusions (for example right wrist occluded behind left wrist) that may be challenging to track.”

Comparison for accuracy/evaluation of the infant motion tracker:

All 7 body points in the 5493 frames were manually annotated for evaluation purposes only in this study. These manually annotations were used as ground truth to evaluate the correctness of the automatic tracker and had nothing to do with the training of the automatic tracker that were performed in another previous study (16). To clarify the issue, we have changed wording in the paragraph as follows: “The performance of the infant motion tracker was assessed and reported by the following

three steps: First, the automatic motion tracking was performed to detect the position of 7-body points (nose, thorax (center between shoulders), wrists, pelvis, and ankles) in each of the 5493 video frames. Secondly, all 7 body points in the 5493 selected video frames was manually annotated. These manually annotations are the ground truth for the evaluation of the infant motion tracker. Thirdly, the performance of the infant motion tracker is reported as percentage of points within a circular area centered at the manually annotated body point for the 5493 frames.”

In summary, we have changed and added wording and some of the sequence in the “In-Motion body point tracking” paragraph at page 5 and 6 and hope clarity of the stepwise procedure have improved according to reviewer #2 comments.

*Distinct similarities exist between Kwong et al. 2018 and the current study (e.g. reporting of return rates and survey engagement, flow diagram Figure 3, time frames set for the two filming windows, setting in-app reminders for filming and parent survey) with only the parent survey referenced back to Kwong et al. 2018.

Reply: We totally agree that there exist similarities between Kwong et al. 2018 and the current study and that reference back to Kwong et al. 2018 according to parent survey and topic about return rates in the discussion are uncomplete. Hence, we have added reference to Kwong et al. 2018 (11) at the following places in the manuscript:

- Page 4, paragraph about data collection procedure, description of filming windows.
- Page 4, last paragraph after description of filming windows
- Page 4, last sentence about app reminders

In conclusion, we hope you will find that we have amended the paper in line with all comments from reviewer #2, and that you will find the revised manuscript worth of being published in BMJ Open. If further clarifications are needed according to the motion tracker paragraph, we hope to get the opportunity to respond to those, even though this is the second review process.

VERSION 3 – REVIEW

REVIEWER	Amanda Kwong Murdoch Children's Research Institute
REVIEW RETURNED	11-Feb-2021
GENERAL COMMENTS	Thank you to the authors for addressing concerns around the clarity of analysis processes. I have no further comments to add.