PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Early Moves: A protocol for a population-based prospective cohort study to establish General Movements as an early biomarker of cognitive impairment in infants.
AUTHORS	Elliott, Catherine; Alexander, Caroline; Salt, Alison; Spittle, Alicia; Boyd, Roslyn; Badawi, Nadia; Morgan, Catherine; Silva, Desiree; Geelhoed, Elizabeth; Ware, Robert; Ali, Alishum; McKenzie, Anne; Bloom, David; Sharp, Mary; Ward, Roslyn; Bora, Samudragupta; Prescott, Susan; Woolfenden, Susan; Le, Vuong; Davidson, Sue-Anne; Thornton, Ashleigh; Finlay-Jones, Amy; Jensen, Lynn; Amery, Natasha; Valentine, Jane

VERSION 1 – REVIEW

REVIEWER	Dr Monica Juneja Maulana Azad Medical College and Associated Lok Nayak Hospital New Delhi-110002 India
REVIEW RETURNED	13-Jul-2020

GENERAL COMMENTS	It is a very important and well designed study which has the potential to change the way early intervention services will be provided world over; however some minor modifications are being suggested and few clarifications are being sought. Title : At line 5 - It may be changed to ‘early biomarkers ‘ or instead add ‘GM movements as an early biomarker of cognitive impairment’ Abstract (page 6) : Methodology : The details of study setting (primary care or Tertiary care level , single centre from western Australia etc) and study population (Preterms / Extreme Preterm/ Term babies etc) need to be added . Article Summary (page 7) It is a single centre study may be added as another limitation . Introduction (page 8): Data related to prevalence of cognitive delay in the population should also be provided in the introduction. Secondary aim of phase -1 • The secondary aim of phase -1 has been given as ‘to develop automated scoring of the GMA through applying machine based learning to the data’. However , last paragraph of page 14 describes the already developed advanced machine learning methods for classification of fidgety GMs with an accuracy of 62.9% . Please clarify whether this study plans to further fine tune automated scoring of Fidgety GMs ; or develop automated scoring for writhing GMs primarily . • The hypothesis (page 10, line 51) states that the automated scoring of GMA will have >90% specificity and >85% specificity to detect GM abnormalities, with lower accuracy for optimality scores. Apart from correcting the typographical error of >90% specificity as

	‘sensitivity ‘ ; please clarify the basis for this because as per the literature (last para of page 9 ,line 50-51) automated readings of fidgety movements have a sensitivity and specificity of 79-85% and 63-71% respectively . Methods Setting (page 11) : • Since details of “ORIGINS” study are not available as an open resource(ref 33) , the relevant details of study setting and study methodology need to be provided. • Study population : The birth cohort has been described as majority ‘Low risk population ‘ on page 10 line 44; at page 11 line 57 it has been described as ‘representative population ‘- please clarify likely proportion of preterm and extreme preterm babies as well as other high risk population . • Method of sample selection - whether these 2000 babies from active participants and 1000 from non active participants be selected randomly? • The details of data collection from active participants which is not a part of routine follow up for example ‘Biological sampling ‘ needs to be provided also . • Exclusion Criteria (page 12) : Please clarify whether all high risk babies -for example Down Syndrome , HIE-3 or IVH grade 4 be included in the intervention arm and thus get excluded or only a small proportion of such children will get excluded. What about infants with syndromes or problems known to be definitely associated with cognitive impairment? Since there is no need for screening of children definitely at high risk for Cognitive impairment, do you plan to exclude these infants from the study? Why children with absent Fidgety are being included even though they are being referred for EI services ? • Outcomes : Primary out come of phase -1 being BSID-4 . Please clarify whether babies with delay in only one of the cognition or language domain will be classified as cognitively delayed or not . Sample size estimation : • The study design for sample size calculation has been given as single masked study, whereas it is a double masked study . • What is the basis for assuming 90% sensitivity and 94% specificity for GM movements in predicting cognitive impairment? • It is not clear what is the primary outcome which has been used for sample size estimation (the association of GMs with cognitive delay has been mentioned as secondary outcome). Apart from these , minor modifications in language are also suggested ; for example motor impairments on page 8, line 7 may be changed to motor skills ; videoing may be modified to video-taping .

REVIEWER	Dr. Britta Huening University Hospital Essen Department of Pediatrics I Neonatology, Pediatric Intensive Care and Neuropediatrics Germany
REVIEW RETURNED	29-Jul-2020

GENERAL COMMENTS	**Comments to the Author** Thank you to Dr. Elliott and colleagues for their submission of “Early Moves: A protocol for a population-based prospective cohort study to identify early biomarkers of cognitive impairment in infants” to the

	BMJ Open. They propose a study protocol on the evaluation of an early motor assessment (General Movements Assessment – GMA) as an early biomarker for cognitive outcome at two years of age. The authors aim to develop an automated scoring of GMA by applying machine based learning in phase 1. Phase 2 aims to develop screening algorithms for the identification of children at risk of cognitive impairment additionally including social and environmental profile data. Phase 3 comprises cost and economic evaluations. The abstract was concise, informative, accurate and structured within journal guidelines except for the requirement to draw on the dissemination strategy. The authors state in the introduction that "*without timely targeted intervention, children at risk of cognitive impairments have poorer health care outcomes and do not reach their full life potential*." The reviewer considers this statement to be a too far-reaching interpretation that is less scientifically proven. Authors may want to give evidence for this statement in the introduction section of the manuscript. The reviewer is not sure if the authors imply the developmental domain of speech within their term of cognitive impairments as the Bayley – 4 reports cognitive scales and speech independently. The article summary encapsulates all study projects well. However, the feasibility of the smart phone app was, as the authors state on page 10, already determined. This should be removed from this section. The introduction is well written citing predominantly relevant literature. However, for readers not familiar with the GMA it might be difficult to understand what writhing or fidgety movements look like. The reviewer suggests including one or two sentences to make that clear to the reader. Furthermore, abnormal GMs should be defined at both ages (writhing and fidgety). The authors cite a study protocol for the time period of the occurrence of fidgety movements. The reviewer recommends citing Ferrari F, et al. Early Hum Dev, 2016 or the book by Einspieler et al on Prechtl's Method, University Press, 2004, instead. The authors repeatedly highlight the need for early identification of children at risk of cognitive impairment before the age of 2 years. However, they do not provide the reader with any evidence for effective and timely interventions. The reviewer likes to make minor and linguistic remarks: • P 5, first sentence – reference missing • P 5, second paragraph, second sentence – verb missing • P 5, ref 7 – considerable delay… - aren't children regularly seen by a pediatrician?

	• P 5, third paragraph, missing space before ref 9 • P 5, third paragraph, " health of baby's brain function – consider revising to "integrity of infant's brain function • P 5, fourth paragraph – consider using consistent wording for GM ages e.g. weeks or month • P 6, second paragraph – "every movement criterion" – explain criteria in more detail • P 6, second paragraph – ref 19 – this sentence is confusing, consider using two sentences to structure the results according to normal/abnormal/absent fidgety movements and then to describe the results of the MOS. • P 6, third paragraph – "videoing" – consider using video recording instead • P 7, Citation – consider introducing "WA" as abbreviation before first use The aims and hypothesis section is well structured and well written in 3 phases. Phase 1 – primary aim: does "abnormal GM" apply for both ages? The study design, masking, statistical analysis and public/patient involvement paragraphs of the Methods section is well written in an understandable way. The data collection and GM Assessment is well designed and thoughtful. The use of the Baby moves app and the automated reading of GMA is highly innovative. The setting, however, is confusing. The reviewer had to visit the webpage of the ORIGINS project to understand the composition of the cohort. The authors should state that the recruitment of the study had already started. There is a delineation necessary between biological profiling, biobank information and biological sampling. Please specify what kind of information, type of biological material is collected and for what purpose. The authors write on page 8 that participants are recruited into one of two arms. Are the groups set by the time of recruitment (antenatally versus prior to discharge) or are participants randomized? The Early moves study will recruit 3000 participants who are enrolled in the ORIGINS Project. Please provide information to the reader how the sample size was estimated and what the selection criteria are, e.g. Do the first 3000 participants meet the inclusion criteria? How do the inclusion criteriae of the ORIGIN project and the Early Moves Project differ? How many infants will be preterm and how many term born infants. What rate of cognitive impairment do the authors expect in preterm/term infants. The authors mention the participation in an ORIGINS intervention study as an exclusion criterion but have not previously described this

	intervention in the paper. Authors may consider adding that these interventions aim to promote cognitive and language outcome. The authors plan to administer a developmental screening tool for children lost to follow up. The reviewer recommends using the term "drop out" instead. The reviewer suggests including the inhomogeneity of the cohorts as a potential bias (genetic disorders, asphyxia, etc.). A further bias may be any kind of intervention such as physiotherapy that infants receive within the study period. On page 10, second paragraph the authors should include that babies need to be in supine position. On page ten, description of time periods should be correctly defined as 1+*0* to 2+6 weeks. Page 12, below the table – the authors state that the Bayley-4 is the most frequently used test in infant development. As the Bayley-4 was first available in September 2019 this statement can only be made for former editions. On page 13, third paragraph – The sentence: "Health-care resource-use data collected …" does not make sense – consider revising. The sample size estimation section it does not explain how the authors defined the sample size. Page 14, first paragraph – the term "attrition" is confusing. What are the authors trying to explain? The term "images" might be changed to "video". Check this paragraph for the positioning of spaces. The figure is confusing. The reviewer suggests to extract the titles of each box e.g. "Recruitment" etc in order to have the down arrow in the centre of the two remaining boxes. Why did the authors include "normal" and "abnormal"? The box "refer to EI network" should be further explained in the text. The discussion section is interesting to read and covers recent literature. The reviewer recommends to revise the first sentence into active speech. Furthermore, it is of the opinion the reviewer that it is not the early intervention itself that is hindered but its referral. In the second paragraph: authors should be explicit in that not all participants will be recruited antenatally.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

It is a very important and well-designed study which has the potential to change the way early intervention services will be provided world over; however some minor modifications are being suggested and few clarifications are being sought.

Thank you for reviewing this paper, we are so pleased you see the value in this work. Thank you for your comments, we found these to be very useful and have responded to each one below. We hope the responses and changes made in the manuscript meet your expectations.

Title : At line 5 - It may be changed to 'early biomarkers ' or instead add 'GM movements as an early biomarker of cognitive impairment'

Thank you for this suggestion, we have incorporated it into the title

Abstract (page 6) :

Methodology : The details of study setting (primary care or Tertiary care level , single centre from western Australia etc) and study population (Preterms / Extreme Preterm/ Term babies etc) need to be added .

We have updated the study setting and population details in the abstract.

Article Summary (page 7)

It is a single centre study may be added as another limitation .

Since submission of the original manuscript, the expansion of Early Moves to include additional sites has been progressed. The manuscript has been updated to reflect this.

Introduction (page 8): Data related to prevalence of cognitive delay in the population should also be provided in the introduction.

We appreciate the lack of prevalence data being identified, and have now included this in the introduction.

Secondary aim of phase -1

The secondary aim of phase -1 has been given as 'to develop automated scoring of the GMA through applying machine based learning to the data'. However , last paragraph of page 14 describes the already developed advanced machine learning methods for classification of fidgety GMs with an accuracy of 62.9% . Please clarify whether this study plans to further fine tune automated scoring of Fidgety GMs ; or develop automated scoring for writhing GMs primarily .

The secondary aim is to generate accurate scoring of global assessment and optimality scoring for both periods. For the global assessment of the fidgety period, this will involve further development on existing models that are available to the researchers. We hope the change to the wording in the manuscript more accurately reflects this aim.

The hypothesis (page 10, line 51) states that the automated scoring of GMA will have >90% specificity and >85% specificity to detect GM abnormalities, with lower accuracy for optimality scores. Apart from correcting the typographical error of >90% specificity as 'sensitivity ' ; please clarify the basis for this because as per the literature (last para of page 9 ,line 50-51) automated

readings of fidgety movements have a sensitivity and specificity of 79-85% and 63-71% respectively .
Thank you for noting the typographical error, we have corrected this. We have now included more detail in the introduction to justify the basis for our hypothesised improvement in specificity and sensitivity. The two main reasons we anticipate improved accuracy are the adoption of more advanced machine learning technique, and the large training dataset that the Early Moves study offers.

Methods

Setting (page 11) :

Since details of "ORIGINS" study are not available as an open resource(ref 33) , the relevant details of study setting and study methodology need to be provided.

Thank you for this. There is now an open resource available that we have cited in text (Silva et al., 2020). We have also updated the details around the ORIGINS study to provide further clarification around this setting. Early Moves is progressing plans to expand to one additional site, and the manuscript has been updated to reflect this.

Study population: The birth cohort has been described as majority 'Low risk population ' on page 10 line 44; at page 11 line 57 it has been described as 'representative population '- please clarify likely proportion of preterm and extreme preterm babies as well as other high risk population .We have clarified this and ensured that the sample is referred to as representative population throughout the manuscript. As all mothers birthing at each site are invited to participate in the study we anticipate the rates of preterm and extreme preterm to be similar to that found in the general population at 8.2% and 1.7% (Cheong & Doyle, 2012)

Method of sample selection - whether these 2000 babies from active participants and 1000 from non active participants be selected randomly?

The aims of this project can all be achieved with the data available for participants regardless of what arm of ORIGINS they are enrolled in, as it relies only on routinely collected data. Therefore we have removed the reference to the ORIGINS arms in this protocol as we appreciate the confusion caused.

The details of data collection from active participants which is not a part of routine follow up for example 'Biological sampling ' needs to be provided also .

As per above, the biological sampling may be used at some point to investigate relationships between biological profiles and GM/Bayley outcomes, but this is not a part of the main research questions as presented in the manuscript. We have therefore opted to remove reference to additional data collection from the manuscript.

Exclusion Criteria (page 12) : Please clarify whether all high risk babies -for example Down Syndrome , HIE-3 or IVH grade 4 be included in the intervention arm and thus get excluded or only a small proportion of such children will get excluded.

This is a valid and interesting point, thank you. We anticipate only a small portion of children will get enrol in additional interventions that meet the criteria for exclusion, however, we have now included in

the protocol review of this data to fully understand the profile of children being excluded on the grounds of involvement in a research study with a targeted cognitive intervention.

What about infants with syndromes or problems known to be definitely associated with cognitive impairment? Since there is no need for screening of children definitely at high risk for Cognitive impairment, do you plan to exclude these infants from the study?
For a comprehensive understanding of the relationship of general movements and developmental outcomes, no children will be excluded on the basis of their risk for cognitive impairment. Though we know babies at high risk, for example those with Trisomy 21 are highly likely to show cognitive impairment at the 2 year assessment; there is little data on the general movement profiles for some of these populations. Therefore, to fully understand the potential for general movements to act as a biomarker for cognitive impairment, all babies will be included.
We understand that children who are at high risk may be likely to receive cognitive interventions or support as part of evidence based best practice care. As this intervention is being delivered as part of their standard clinical care, these children will not be excluded on these grounds. We will however seek to accurately describe any care received, through access to the Medicare Benefits Scheme and health resource use questionnaires, and will perform sensitivity analysis to understand the impact including these children may have on the final results. We have also updated the statistical analysis section to control for known high- risk factors as appropriate.

Why children with absent Fidgety are being included even though they are being referred for EI services ?
As per the above examples with other cases of children with known risk factors, children with absent Fidgety movements will not be excluded from the project as intervention received is through standard clinical care. Medicare benefits scheme and health resource use questionnaires will be used to identify any infants in the cohort who receive interventions as part of their clinical care and sensitivity analysis will be performed to assess the impact of their inclusion.

Outcomes : Primary out come of phase -1 being BSID-4 . Please clarify whether babies with delay in only one of the cognition or language domain will be classified as cognitively delayed or not.
A combined cognition *and* language score, computed as the average of the cognitive score and the language score will be used to provide a relevant and practical composite outcome for nonverbal cognitive and language development (Johnson, Moore, & Marlow, 2014; Moore, Johnson, Haider, Hennessy, & Marlow, 2012; Yi, Sung, & Yuk, 2018).
Secondary analysis will also be conducted looking at the cognitive score alone. This has been clarified in text.

Sample size estimation :
The study design for sample size calculation has been given as single masked study, whereas it is a double masked study .
Apologies for this typographical error, we have corrected this to double masked.

What is the basis for assuming 90% sensitivity and 94% specificity for GM movements in predicting cognitive impairment?
As this is the first study to look at the predictive utility of GM movements for cognitive impairment in a representative sample of infants there is no available data to base a sensitivity and specificity estimate from. We accept the diagnostic statistics used to inform our calculation may be considered ambitious. We have reviewed the calculations for sample size with lower sensitivity and specificity assumptions, and have revised the text appropriately.

It is not clear what is the primary outcome which has been used for sample size estimation (the association of GMs with cognitive delay has been mentioned as secondary outcome).
Sample size estimation is based on the primary aim of the study, to determine the diagnostic value of abnormal GMs for cognitive delay at two-years. Manuscript text has been updated for clarity.

Apart from these , minor modifications in language are also suggested ; for example motor impairments on page 8, line 7 may be changed to motor skills ; videoing may be modified to video-taping .

Reviewer: 2
Thank you to Dr. Elliott and colleagues for their submission of “Early Moves: A protocol for a population-based prospective cohort study to identify early biomarkers of cognitive impairment in infants” to the BMJ Open.
Thank you for your review and commentary on the submission. We found your comments and insights very useful and constructive. We have responded to each of your comments below, and made changes in the manuscript as appropriate. We trust these changes are satisfactory.
They propose a study protocol on the evaluation of an early motor assessment (General Movements Assessment – GMA) as an early biomarker for cognitive outcome at two years of age. The authors aim to develop an automated scoring of GMA by applying machine based learning in phase 1. Phase 2 aims to develop screening algorithms for the identification of children at risk of cognitive impairment additionally including social and environmental profile data. Phase 3 comprises cost and economic evaluations.
The abstract was concise, informative, accurate and structured within journal guidelines except for the requirement to draw on the dissemination strategy. The authors state in the introduction that “*without timely targeted intervention, children at risk of cognitive impairments have poorer health care outcomes and do not reach their full life potential*.” The reviewer considers this statement to be a too far-reaching interpretation that is less scientifically proven. Authors may want to give evidence for this statement in the introduction section of the manuscript.
We appreciate the feedback regarding this statement and have made adjustments to tone it down to be a more appropriate interpretation of current research.

The reviewer is not sure if the authors imply the developmental domain of speech within their term of cognitive impairments as the Bayley – 4 reports cognitive scales and speech independently.

The article summary encapsulates all study projects well. However, the feasibility of the smart phone app was, as the authors state on page 10, already determined. This should be removed from this section.

The feasibility for the app was established for fidgety age videos in babies who were at high risk of cerebral palsy, therefore likely to have high levels of motivation to engage with the app and personal investment in the results. This is the first study to use the app in a predominantly low risk population, and the first to use the app also for the writhing period. We have therefore revised the wording in the article summary to better reflect this.

The introduction is well written citing predominantly relevant literature. However, for readers not familiar with the GMA it might be difficult to understand what writhing or fidgety movements look like. The reviewer suggests including one or two sentences to make that clear to the reader. Furthermore, abnormal GMs should be defined at both ages (writhing and fidgety).

Thank you for this valuable suggestion, we have incorporated this in the manuscript.

The authors cite a study protocol for the time period of the occurrence of fidgety movements. The reviewer recommends citing Ferrari F, et al. Early
Hum Dev, 2016 or the book by Einspieler et al on Prechtl's Method, University Press, 2004, instead.

Thank you for this recommendation. We have updated the reference for this.

The authors repeatedly highlight the need for early identification of children at risk of cognitive impairment before the age of 2 years. However, they do not provide the reader with any evidence for effective and timely interventions.

To date there has been no reliable way to identify children at risk of cognitive impairment before the age of 2 years who have no established risk factors, such as very preterm or very low birth weight. Without capacity to detect risk of cognitive intervention in otherwise healthy infants, with no known risk factors, we have no way to assess the impact of early interventions on this cohort

There is however, research in looking at early interventions in higher risk groups, which we have now included in the manuscript. In this population, evidence is mixed on the long term benefits of early intervention (Orton, Spittle, Doyle, Anderson, & Boyd, 2009; Spittle, Orton, Anderson, Boyd, & Doyle, 2015). However, research suggests that factors such as birth weight and maternal education can impact the long term effectiveness of early interventions (Brooks-Gunn, Goss, Kraemer, Spiker, & Shapiro, 1992) with the studies suggesting benefit of early intervention on cognitive outcomes is sustained for the higher-low birth weight babies up until 18 years of age; though it is not sustained for lighter-low birth weight babies (McCormick et al., 2006). This finding suggests to the authors that limited evidence for long term benefit of early intervention in preterm and low birthweight infants cannot be extrapolated to otherwise healthy term babies, who may have biological and environmental profiles that are more primed to benefit from early interventions. It is our hope that through identifying a very early biomarker for cognitive impairment, we may facilitate the development and testing of effective and timely interventions for this population.

The reviewer likes to make minor and linguistic remarks:

P 5, first sentence – reference missing

Thank you, we have referenced this now.

P 5, second paragraph, second sentence – verb missing

P 5, ref 7 – considerable delay… - aren't children regularly seen by a pediatrician?

In Australia, children have access to routine health and development checks with community child health nurses, and see their general practitioner for routine vaccinations. Children may be referred to specialists if they meet certain criteria on standardised assessments such as the Ages and Stages questionnaire. Waiting times for specialist appointments in the public health system can lead to considerable delay between concerns being raised and receiving specialist care and a potential diagnosis. Further, due to the natural variability in child development, a "wait and see" approach is often taken, adding valuable time between raising of concerns and the initial specialist referral.

P 5, third paragraph, missing space before ref 9

A space has been added here

P 5, third paragraph, " health of baby's brain function – consider revising to "integrity of infant's brain function

We have accepted this revision

P 5, fourth paragraph – consider using consistent wording for GM ages e.g. weeks or Month

Thank you for pointing this out, we have ensured wording is consistent

P 6, second paragraph – "every movement criterion" – explain criteria in more detail

We believe that full explanation of the movement criteria is beyond the scope of this article, though we appreciate readers may wish for more detail. We have therefore directed readers to the relevant references to attain this detail.

P 6, second paragraph – ref 19 – this sentence is confusing, consider using two sentences to structure the results according to normal/abnormal/absent fidgety movements and then to describe the results of the MOS.

We appreciate this feedback, and have revised the wording to make it more clear.

P 6, third paragraph – "videoing" – consider using video recording instead

Updated, thank you.

P 7, Citation – consider introducing "WA" as abbreviation before first use

We have included the abbreviation earlier in the introduction for clarity.

The aims and hypothesis section is well structured and well written in 3 phases.

Phase 1 – primary aim: does “abnormal GM” apply for both ages?
Abnormal GM does apply for both ages. Research in preterm infants has indicated that both abnormal writhing and abnormal fidgety outcomes are correlated with later cognitive outcomes. To date no longitudinal research has been conducted in term infants assessing writhing and fidgety periods, and cognitive outcomes. This study aims to understand the if the GM outcomes at each time period, as well as the trajectories of GMs across the time periods, is predictive of cognitive impairment.

The study design, masking, statistical analysis and public/patient involvement paragraphs of the Methods section is well written in an understandable way. The data collection and GM Assessment is well designed and thoughtful. The use of the Baby moves app and the automated reading of GMA is highly innovative.
The setting, however, is confusing. The reviewer had to visit the webpage of the ORIGINS project to understand the composition of the cohort.
The authors should state that the recruitment of the study had already started. There is a delineation necessary between biological profiling, biobank information and biological sampling. Please specify what kind of information, type of biological material is collected and for what purpose.
We apologise for the confusion caused by reference to the biological sampling and biobank information being collected through the ORIGINS project. As this information is not being used directly for the purpose of achieving the aims stated in this protocol, we have decided to remove mention of this in the protocol manuscript. Future works being planned to integrate biological profiling and the Early Moves data will be outlined in a separate paper, with appropriate research aims.

The authors write on page 8 that participants are recruited into one of two arms. Are the groups set by the time of recruitment (antenatally versus prior to discharge) or are participants randomized?
Again, thank you for pointing out the confusion here. Recruitment in to the arms is chosen by the participant on the basis of how much they are willing to commit to the project. Participants who wish to be involved in research, but do not wish to fill in lengthy surveys or provide biological samples, for example, may opt to enrol in the non-active arm.
As per our response to your previous comment, as the information required to achieve the aims set out in this project is available from all participants, regardless of their ORIGINS involvement, we have removed reference to the different arms from the protocol manuscript. We trust this makes participant recruitment clearer.

The Early moves study will recruit 3000 participants who are enrolled in the ORIGINS Project. Please provide information to the reader how the sample size was estimated and what the selection criteria are, e.g. Do the first 3000 participants meet the inclusion criteria? How do the inclusion criteriae of the ORIGIN project and the Early Moves Project differ?
We have updated the manuscript to clarify the recruitment process. As the study aims to recruit a representative sample of Australian infants, the inclusion criteria is very broad with all mothers birthing at the study site being eligible for the study. The only exclusion criteria is involvement in a interventional study targeting cognitive or language outcomes. The first 3000 eligible participants who provide consent will be enrolled in the study.

How many infants will be preterm and how many term born infants. What rate of cognitive impairment do the authors expect in preterm/term infants.
As all mothers birthing at each site are invited to participate in the study we anticipate the rates of preterm and extreme preterm to be similar to that found in the general population at 8.2% and 1.7% (Cheong & Doyle, 2012).

The authors mention the participation in an ORIGINS intervention study as an exclusion criterion but have not previously described this intervention in the paper. Authors may consider adding that these interventions aim to promote cognitive and language outcome.
Thank you for this recommendation, we have incorporated this in the manuscript.

The authors plan to administer a developmental screening tool for children lost to follow up.
The reviewer recommends using the term “drop out” instead. The reviewer suggests including the inhomogeneity of the cohorts as a potential bias (genetic disorders, asphyxia, etc.). A further bias may be any kind of intervention such as physiotherapy that infants receive within the study period.
These recommendations have been noted, and the manuscript updated.

On page 10, second paragraph the authors should include that babies need to be in supine position.
We have included this, thank you.

On page ten, description of time periods should be correctly defined as 1+*0* to 2+6 weeks
This has been corrected.

Page 12, below the table – the authors state that the Bayley-4 is the most frequently used test in infant development. As the Bayley-4 was first available in September 2019 this statement can only be made for former editions.
Thank you for pointing out the misleading structure of this sentence, we have now corrected it.

On page 13, third paragraph – The sentence: “Health-care resource-use data collected …” does not make sense – consider revising.
We trust our revisions to the sentence are appropriate.

The sample size estimation section it does not explain how the authors defined the sample size. Page 14, first paragraph – the term “attrition” is confusing. What are the authors trying to explain? The term “images” might be changed to “video”. Check this paragraph for the positioning of spaces.
We have changed the term “attrition” to “drop out”, which refers to drop out at the 2-year follow up, as we have now described in text. Images has also been changed to video as suggested.

The figure is confusing. The reviewer suggests to extract the titles of each box e.g. “Recruitment” etc in order to have the down arrow in the centre of the two remaining boxes. Why did the authors include “normal” and “abnormal”? The box “refer to EI network” should be further explained in the text.

We hope the changes made have clarified the figure.

The discussion section is interesting to read and covers recent literature.
The reviewer recommends to revise the first sentence into active speech. Furthermore, it is of the opinion the reviewer that it is not the early intervention itself that is hindered but its referral.

Thank you for this valuable recommendation, we agree with you that it is the referral is being hindered by the lack of early biomarkers and have updated the manuscript to reflect this.

In the second paragraph: authors should be explicit in that not all participants will be recruited antenatally.

This has been amended in the manuscript, thank you.

References

Brooks-Gunn, J., Gross, R. T., Kraemer, H. C., Spiker, D., & Shapiro, S. (1992). Enhancing the Cognitive Outcomes of Low Birth Weight, Premature Infants: For Whom Is the Intervention Most Effective? *Pediatrics, 89*(6), 1209-1215. Retrieved from https://pediatrics.aappublications.org/content/pediatrics/89/6/1209.full.pdf

Cheong, J. L., & Doyle, L. W. (2012). Increasing rates of prematurity and epidemiology of late preterm birth. *Journal of Paediatrics and Child Health, 48*(9), 784-788.

Johnson, S., Moore, T., & Marlow, N. (2014). Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatric research, 75*(5), 670-674. doi:10.1038/pr.2014.10

McCormick, M. C., Brooks-Gunn, J., Buka, S. L., Goldman, J., Yu, J., Salganik, M., . . . Bernbaum, J. C. (2006). Early intervention in low birth weight premature infants: results at 18 years of age for the Infant Health and Development Program. *Pediatrics, 117*(3), 771-780.

Moore, T., Johnson, S., Haider, S., Hennessy, E., & Marlow, N. (2012). Relationship between Test Scores Using the Second and Third Editions of the Bayley Scales in Extremely Preterm Children. *The Journal of Pediatrics, 160*(4), 553-558. doi:https://doi.org/10.1016/j.jpeds.2011.09.047

Orton, J., Spittle, A., Doyle, L., Anderson, P., & Boyd, R. (2009). Do early intervention programmes improve cognitive and motor outcomes for preterm infants after discharge? A systematic review. *Developmental Medicine & Child Neurology, 51*(11), 851-859.

Silva, D. T., Hagemann, E., Davis, J. A., Gibson, L. Y., Srinivasjois, R., Palmer, D. J., . . . Prescott, S. L. (2020). Introducing the ORIGINS project: a community-based interventional birth cohort. *Rev Environ Health*. doi:10.1515/reveh-2020-0057

Spittle, A., Orton, J., Anderson, P. J., Boyd, R., & Doyle, L. W. (2015). Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive

impairment in preterm infants. *Cochrane Database of Systematic Reviews*(11). doi:10.1002/14651858.CD005495.pub4

Yi, Y. G., Sung, I. Y., & Yuk, J. S. (2018). Comparison of Second and Third Editions of the Bayley Scales in Children With Suspected Developmental Delay. *Annals of rehabilitation medicine, 42*(2), 313-320. doi:10.5535/arm.2018.42.2.313

VERSION 2 – REVIEW

REVIEWER	Dr. Monica Juneja Maulana Azad Medical College and associated L.N. Hospital New Delhi India
REVIEW RETURNED	13-Oct-2020

GENERAL COMMENTS	It is a very important and well designed study which has the potential to change the way early screening and early intervention services will be provided the world over. The methodology has been described lucidly .