## PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to a another journal from Archives of Disease in Childhood but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Paediatrics Open. The paper was subsequently accepted for publication at BMJ Paediatrics Open.

## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Preterm children's developmental coordination disorder, cognition and quality of life - a prospective cohort study |
| **AUTHORS** | Uusitalo, Karoliina; Haataja, Leena; Nyman, Anna; Ripatti, Liisi; Huhtala, Mira; Rautava, Päivi; Lehtonen, Prof Liisa; Parkkola, Riitta; Lahti, Katri; Koivisto, Mari; Setänen, Sirkku |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Reviewer name: Spittle, Alicia<br>Institution and Country: University of Melbourne School of Health Services, Physiotherapy<br>Competing interests: I have no competing interests |
| **REVIEW RETURNED** | 21-Nov-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | This paper is examining motor outcomes including DCD and CP in a regional population of children in Finland at 11 years of age. I am familiar with the research of the team and this research question is now well defined and has appropriate methodology to answer the research question.<br><br>I have only a few minor comments.<br>1. Throughout the paper I would state "preterm children with and without DCD" to remind the reading we are referring to preterm children with DCD not children in generally with DCD.<br>2. How many multiples were included? I would suggest using generalising estimating equations or clustering for twins in your analysis depending on rates of multiple births. |

| | |
|---|---|
| **REVIEWER** | Reviewer name: Arichi, Tom<br>Institution and Country: King's College London<br>Competing interests: I have no conflicts of interest |
| **REVIEW RETURNED** | 02-Dec-2019 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is a well written and well carried out study – which contains potentially important longitudinal population information about the well characterised PIPARI preterm cohort in Finland. The results are interesting but are potentially confounded by the relatively small numbers of children with DCD and the usual bias associated with longitudinal studies of this type. That said, I can appreciate that this is valuable data and that little can be done about what is contained in already collected data – indeed, the authors acknowledge these limitations very well in the discussion. Therefore I would suggest that some careful rewording is needed in other parts to aid interpretation |

| | |
|---|---|
| | (and in particular to emphasise that the findings reflect their specific cohort rather than the preterm population as a whole).<br>Specific points:<br>1. I think caution is needed about overstating certain details in their results due to the numbers (the over-representation of boys) and follow-up bias (the HRQoL results). These results are certainly worth reporting but reflect findings within their own cohort and may not/do not reflect the preterm population as a whole – therefore I feel may not warrant a prominent place in the conclusions of the abstract and first paragraph of the discussion for instance. (I acknowledge that they have conceded these points appropriately in the discussion).<br>2. As there is a lack of a control population it is difficult to pick apart inter-related effects and therefore how to interpret the findings. If there was a control population (or it can be found in the literature in might help interpretation) as it may help to understand for example if poorer HRQoL is related to DCD alone in children with normal cognitive function. In this study, I can only see this relationship explored in the CP subgroup (of which there are only 9) in children with severe cognitive impairment. However, in the non CP group, DCD and cognition are clearly highly correlated (as are gestation at birth and a number of clinical risk factors) – which means that the authors cannot be confident that cognition alone does not explain the relationship between HRQoL and DCD?<br>3. The statistics state that "linear regression" was used to explore relationships – presumably this was multiple linear regression? Could the relationships have been explored using something like the mediation model used by Van Hus (DMCN 2014)?<br>4. In the introduction: the authors suggest that "etiology is unknown" and refer to a single review article to suggest that it may be associated with abnormalities in brain development and functioning. I think this underplays that there is a significant amount (although I concede not enormous literature) about this particularly using neuroimaging which – (reviewed nicely in Boitteau et al. 2016 Front Neurol) which certainly converges that there are (at the very least) definitely abnormalities in the brain underlying the condition. Whilst this does include alterations in the white matter, regional brain volume, and functional connectivity – including recent work (Dewey et al. 2019, Setanen 2016 (from the same cohort)) showing that abnormalities in neonatal imaging similarly can predict DCD in later childhood, which would support their role in the condition.<br>5. Although the authors state that there were no significant differences in the characteristics of the drop out cases – this refers to the demographic characteristics only (weight, gestational age etc) and so does not rule out the influence of this bias affecting the results. This drop-out bias for example may explain the higher HRQoL scores for the study population and therefore I think they should be cautious about overinterpreting this result.<br>6. Looking at Figure 3, it is also quite striking that the trends in answers are the same for the study group and population and that much of the difference seems to be explained entirely by very low scores in the "sleeping" and "discomfort" areas in the population – as I am unfamiliar with this data is there an explanation as to why these areas were particularly affected in Finnish normal children? |

**VERSION 1 – AUTHOR RESPONSE**

Reviewer 1:

This paper is examining motor outcomes including DCD and CP in a regional population of children in Finland at 11 years of age. I am familiar with the research of the team and this research question is now well defined and has appropriate methodology to answer the research question.

I have only a few minor comments.

1. Throughout the paper I would state "preterm children with and without DCD" to remind the reading we are referring to preterm children with DCD not children in generally with DCD.

Answer: Thank you for carefully revising our manuscript. We have made changes throughout the manuscript to meet this suggestion.

2. How many multiples were included? I would suggest using generalising estimating equations or clustering for twins in your analysis depending on rates of multiple births.

Answer: This is a good and important point. The rate of multiple births was similar in very preterm born children with and without DCD (33.8 % in children without DCD, 33.3 % in children with DCD). We can add this information to Table 1 (Background characteristics) if necessary.

Reviewer 2:

This is a well written and well carried out study – which contains potentially important longitudinal population information about the well characterised PIPARI preterm cohort in Finland. The results are interesting but are potentially confounded by the relatively small numbers of children with DCD and the usual bias associated with longitudinal studies of this type. That said, I can appreciate that this is valuable data and that little can be done about what is contained in already collected data – indeed, the authors acknowledge these limitations very well in the discussion. Therefore I would suggest that some careful rewording is needed in other parts to aid interpretation (and in particular to emphasise that the findings reflect their specific cohort rather than the preterm population as a whole).

Answer: Thank you for thorough revision of our manuscript. We have now made changes to clarify that these findings represent our cohort of very preterm born children.

Specific points:

1. I think caution is needed about overstating certain details in their results due to the numbers (the over-representation of boys) and follow-up bias (the HRQoL results). These results are certainly worth reporting but reflect findings within their own cohort and may not/do not reflect the preterm population as a whole – therefore I feel may not warrant a prominent place in the conclusions of the abstract and first paragraph of the discussion for instance. (I acknowledge that they have conceded these points appropriately in the discussion).

Answer: We thank you for the observations concerning the over-representation of boys and the possible follow-up bias of the HRQoL results. We have made corrections to the manuscript to avoid overemphasizing these findings.

2. As there is a lack of a control population it is difficult to pick apart inter-related effects and therefore how to interpret the findings. If there was a control population (or it can be found in the literature in might help interpretation) as it may help to understand for example if poorer HRQoL is related to DCD alone in children with normal cognitive function. In this study, I can only see this relationship explored in the CP subgroup (of which there are only 9) in children with severe cognitive impairment. However, in the non CP group, DCD and cognition are clearly highly correlated (as are gestation at birth and a number of clinical risk factors) – which means that the authors cannot be confident that cognition alone does not explain the relationship between HRQoL and DCD?

Answer: We appreciate your comment regarding the lack of a control population. We agree that a control population would have been very interesting and helpful when interpreting the data. In particular with the HRQoL results we feel that the literature and other studies did not offer solution to this since the reasonably large variation of the tools used to measure the quality of life. Therefore, we decided to use the normative of the 17-dimensional illustrated questionnaire as it was based on the results of the age and nationality although the information of the motor and cognitive development of the normative population was not available. We agree, that the small number of very preterm children with DCD and/or cognitive impairment limits the interpretation and have made changes to emphasize that this finding is reflecting our cohort of very preterm children.

3. The statistics state that “linear regression” was used to explore relationships – presumably this was multiple linear regression? Could the relationships have been explored using something like the mediation model used by Van Hus (DMCN 2014)?

Answer: Thank you for the observation. We have used multiple linear regression and clarified this to the statistics section of the manuscript. Unfortunately, we (including our statistician) are not familiar with the mediation model used by Van Hus in DMCN (2014).

4. In the introduction: the authors suggest that “etiology is unknown” and refer to a single review article to suggest that it may be associated with abnormalities in brain development and functioning. I think this underplays that there is a significant amount (although I concede not enormous literature) about this particularly using neuroimaging which – (reviewed nicely in Boitteau et al. 2016 Front Neurol) which certainly converges that there are (at the very least) definitely abnormalities in the brain underlying the condition. Whilst this does include alterations in the white matter, regional brain volume, and functional connectivity – including recent work (Dewey et al. 2019, Setanen 2016 (from the same cohort)) showing that abnormalities in neonatal imaging similarly can predict DCD in later childhood, which would support their role in the condition.

Answer: Thank you for your comment about the etiology of DCD. We have studied the review by Biottaeu et al. and the article of Dewey et al and other recent papers concerning the etiology, and especially brain imaging of the children/adolescents with DCD. We have now added changes to the manuscript according to the findings of these studies.

5. Although the authors state that there were no significant differences in the characteristics of the drop out cases – this refers to the demographic characteristics only (weight, gestational age etc) and so does not rule out the influence of this bias affecting the results. This drop-out bias for example may explain the higher HRQoL scores for the study population and therefore I think they should be cautious about overinterpreting this result.

Answer: We appreciate your observation. In our cohort of very preterm children we found no statistically significant differences between the background characteristics of the study children and the drop-outs. These background characteristics included mother’s education and father’s education (shown in Table 1).

6. Looking at Figure 3, it is also quite striking that the trends in answers are the same for the study group and population and that much of the difference seems to be explained entirely by very low scores in the “sleeping” and “discomfort” areas in the population – as I am unfamiliar with this data is there an explanation as to why these areas were particularly affected in Finnish normal children?

Answer: Thank you for the comment. After carefully revising the manuscript, we have tried to tone down the possible overinterpretation of the HRQoL data. Therefore, we decided to remove the Figure 3.

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Reviewer name: Amanda Kwong<br>Institution and Country: Murdoch Children's Research Institute<br>Competing interests: None to declare |
| **REVIEW RETURNED** | 09-Feb-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors present an excellent study of the relationships between preterm birth, motor impairment, cognitive function and quality of life. This is an important study that contributes to our knowledge of attributes of preterm infants for their quality of life related to motor/cognitive ability. This study is conducted with good scientific rigour and statistical analysis, with strengths including the diagnosis of CP and DCD according to the most up-to-date guidelines available and use of multiple imputation for missing data.<br><br>The main point that the authors should clarify is the reporting of the higher rate of DCD as a main finding, or a point of interest that contributes to the interpretation of the study's findings.<br><br>Some further points:<br>It would be useful to have a comparison of the characteristics of the included infants versus the drop-outs as a supplementary table.<br><br>The first line of the discussion does not appear to relate to any of the study aims posed within the final paragraph of the introduction. This statement is also not proved empirically within the methods or results of this paper. However, this might make an interesting discussion point later within the discussion section. Please also check the conclusion in the abstract and text.<br><br>It would be useful to give the geographical context of the other reports of rising trends of non-CP motor impairments in extremely preterm population.<br><br>The authors mention the comparison of other cohorts in reference to a control group. It would be useful to know if the authors felt that the absence of a control group would affect the results, and if so, how? It is mentioned briefly in the limitations but no robust discussion made.<br><br>Within the discussion, comment on the follow-up rate, difference (or lack of) in characteristics and how this may/may not have affected the study's results. |

| | |
|---|---|
| **REVIEWER** | Reviewer name: Kate L Cameron<br>Institution and Country: Murdoch Children's Research Institute, Melbourne, Australia<br>Department of Physiotherapy, University of Melbourne, Melbourne, Australia<br>Competing interests: None to declare |
| **REVIEW RETURNED** | 09-Feb-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | This paper examines motor and cognitive outcomes as well as |

| | |
|---|---|
| | health related quality of life for 11 year old children born very preterm and/or very low birthweight in Turku, Finland. This study addresses an important topic, and uses appropriate methodology.<br><br>General comments:<br>Throughout the paper, children are referred to as having or not having DCD based upon an MABC-2 score of ≤5th percentile and not having a neurological explanation for motor impairment (assessed using the Touwen neurological examination). Using the DSM-V criteria for DCD, criterions I, III and IV are adequately addressed, however criterion II (that motor difficulties adversely affect activities of daily living) does not seem to have been assessed. Although the authors have chosen a cut-off of ≤5th percentile on the MABC-2, perhaps it would be better to qualify the diagnosis, such as using the term probable DCD? I would also suggest discussing the above in either the methods or discussion section.<br><br>The authors often use the phrase 'very preterm children'. I would suggest changing to person-first language, ie 'children born very preterm'.<br><br>Introduction<br><br>Line 49.<br>Suggest changing to "such as DCD on HRQoL"<br><br>Line 49.<br>I was a little confused by the use of the word 'controversial' when referring to the articles by Dewey et. al. and Karras et. al.. From reading the referenced papers quickly, both studies found that children and adolescents with DCD experienced lower quality of life compared with their peers. Perhaps a quick explanation as to why these findings were controversial? Or a change in phrasing to reflect that the research to date supports a relationship between DCD and poor HRQoL?<br><br>Lines 53-60<br>In the aims, I suggest indicating that children were very preterm and/or very low birth weight.<br><br>Methods:<br><br>Participants<br>It appears this study includes two different cohorts with two different sets of inclusion criteria. If this is the case, I suggest making this clear and consistent in both the 'participants' paragraph (page 7) and figure 1. Currently, for children born between 2001 and 2003, figure 1 outlines children are included if they are born very preterm (≤32 weeks) and ≤1500g, while in the paragraph, the inclusion criteria is ≤1500g and <37 weeks.<br><br>Motor outcomes.<br>How did you assess for discrepancies between MABC-2 assessments? (referred to in the Motor Outcomes paragraph)<br><br>Statistical analysis<br>Line 35: an open bracket is needed before birth weight. |

| | |
|---|---|
| | Re the sentence; “Differences in the 17D scores between the groups were studies using Mann-Whitney U Test”. Are the groups referred to 1) children with vs without DCD, 2) with vs without CP and 3) with vs without severe cognitive impairment? Could this sentence be made a little clearer?<br><br>In the results section (health related quality of life), the authors report on correlations between MABC-2 and the 17D as well as full scale IQ and the 17D. Was this analysed using Pearson correlations? If this is the case, could this be added to the statistical analysis paragraph for clarity?<br><br>Results<br><br>Please clarify the 77.6% follow-up rate. Eg. 170 out of 219 children who were enrolled in the cohort study prior to the 11 year follow-up? (I’m unsure if I’ve interpreted this correctly)<br><br>I suggest using ‘children who withdrew’ (or something similar) rather than ‘drop-outs’ which is more colloquial.<br><br>Health related quality of life<br>How did you analyse the difference in QoL scores in your study compared with the general Finnish population? Please add a sentence to ‘Statistical Analysis’ to clarify which test was used.<br><br>Table 2<br><br>The heading of table 2 indicates that “cognitive outcome and health-related quality of life are shown in 11 year old….”, but only full scale IQ and each of the WISC-IV subscales are included in the table. Could the heading be adjusted to reflect the table?<br><br>Discussion<br><br>Line 8:<br>I suggest referencing “compared with population norms (5%)”<br><br>The Spittle et. al. paper cited in the discussion (reference 7) does not refer to children born preterm as having DCD, but rather non-CP motor impairment. I would suggest referring to motor impairment or non-CP motor impairment rather than DCD in this paragraph when citing this paper.<br><br>Paragraph 2 and 3. The authors use the word gender at the end of both paragraphs. I just wanted to clarify that the authors mean gender (social construct) rather than sex (biological)?<br><br>The WISC-IV has some items in the Processing Speed and the Perceptual Reasoning Index subtests that require fine motor control, such as holding a pen and drawing in a small space, and manipulating blocks. Is it possible that motor impairment may negatively affect a child’s performance in these subtests? Perhaps this could be included in the discussion, particularly as the paper concludes that there is a positive correlation between cognitive performance and motor skills.<br><br>Figure 1<br><br>Does ‘withdrew during the follow-up’ refer to the follow-up at 11 |

| | |
|---|---|
| | years? Were there any previous follow-ups between birth and 11 years? eg. withdrew at 5 year follow up? |

| | |
|---|---|
| **REVIEWER** | Reviewer name: Eirini Koutoumanou<br>Institution and Country: UCL, UK<br>Competing interests: None |
| **REVIEW RETURNED** | 13-Feb-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | Thank you for giving me the opportunity to review this interesting report. Please find below several comments which I think have the potential to contribute in an overall improved version of the current report.<br><br>- Does PIPARI stand for something (start of the Participants section)? Could a reference please be added about PIPARI study?<br><br>- Is the Turku University Hospital equivalent to any other hospital in Finland? Could some comments be added about this as it will assist in the generalisation of the presented results?<br><br>- What led to the expansion of the inclusion criteria in 2004?<br><br>- Is it important for the reader to know that one of three physicians evaluated the motor outcome of very preterm children at 11 years of age? I think probably not (but I might be wrong, in which case I'd appreciate clarification) – it'd be more interesting to know that this one physician was chosen randomly from a pool of physicians with similar knowledge and experience. So the important point to be made here is about the unbiasness of the one physician.<br><br>- This last comment also extends to the choice of the Finnish psychologist examining the cognitive development of the children. It is important to know that both the Finnish and Swedish psychologists were chosen randomly from a pool of psychologists with similar knowledge and experience. Currently we are told that there were two Finnish psychologists and only one was chosen – so unless I've misunderstood something, I can't see the added value from this comment.<br><br>- When the authors say that 'Touwen neurological examination was systematically used to confirm…no other neurological conditions' do they mean that children were examined in regular intervals throughout the follow up time? I am not sure about the use of the word systematic here. Could this be clarified please?<br><br>- Were comparisons of the video recordings done by the same people for all cases? Were they performed for all children or when there was an issue with the measurements or reason to suspect something had gone wrong?<br><br>- Which children completed the HRQoL questionnaire by themselves and which as part of an interview by a physician before their motor assessment? If this was decided by the children themselves, this might have led to a lot of bias in the results.<br><br>- The following URL http://www.15d-instrument.net/rmd should be replaced by http://www.15d-instrument.net/15d/replacing-missing-data/<br><br>- The correlation between the movement and various IQ scores were |

| | |
|---|---|
| | statistically significant but nevertheless not strong in magnitude. This result cannot be ignored simply because of low p-values and it needs more commenting. The scatterplot in figure 2 does not show a clear pattern between the two variables. Did the authors also try a non-parametric correlation coefficient to check the robustness of the Pearson's value?<br><br>- A lot of the comparisons made about HRQoL amongst the various groups of children were significant but in fact the actual differences between them (in terms of magnitude) were very small (as scores range from 0 to 1), i.e. 0.93 vs 0.96, 0.92 vs 0.98, 0.96 vs 0.99, etc. Could the authors please comment on whether these differences are actually clinically important? If we were given the actual values for the differences between healthy Finnish children and the very preterm cohort for sleep, discomfort and symptoms, depression, appearance, etc. we might had seen that the magnitude of the differences was about the same as the other dimensions of the HRQoL, but simply did not turn out statistically significant. The authors feedback on this might also necessitate changes to some of the overall statements about differences in HRQoL between the various groups of children throughout the paper.<br><br>- If not already given (as I might have missed it), can you please provide a reference for the 5% of DCD in the normal population (first sentence of the discussion section)?<br><br>- The authors mention sample size at their discussion section, but did they consider performing a sample size calculation at the start of the study to ensure that they only stop collecting data after a satisfactory number of children was observed? If not, what were the restrictions not allowing them to do so?<br><br>- Finally, it seems that the Movement ABC-2 score is used to classify children's motor abilities and put them in categories of DCD, no risk and some risk of motor problems. Has any research been done on the numerical format of the Movement ABC-2 scores as this might be more helpful that the categorisation strategy? This is a generic comment but one worth making as the authors might want to consider this option for future research in this field. |

**VERSION 2 – AUTHOR RESPONSE**

Reviewer: 1

The authors present an excellent study of the relationships between preterm birth, motor impairment, cognitive function and quality of life. This is an important study that contributes to our knowledge of attributes of preterm infants for their quality of life related to motor/cognitive ability. This study is conducted with good scientific rigour and statistical analysis, with strengths including the diagnosis of CP and DCD according to the most up-to-date guidelines available and use of multiple imputation for missing data.

Thank you for encouraging comments and the references.

1. The main point that the authors should clarify is the reporting of the higher rate of DCD as a main finding, or a point of interest that contributes to the interpretation of the study's findings.

Answer: Thank you for this comment. Our main findings are that DCD was still common at 11 years of age in children born very preterm in 2000s, and that DCD associated with adverse cognitive development and self-experienced HRQoL. We have revised the manuscript to clarify the main findings in the Abstract, Introduction, Results and Discussion sections.

Some further points:

2. It would be useful to have a comparison of the characteristics of the included infants versus the drop-outs as a supplementary table.

Answer: We have now added this information to the main Table 1 including a finding that the mothers of the drop-outs had lower educational level compared with mothers of the study children. We have also added this information to the Results section.

"The mothers of the children who withdrew had lower educational level compared with mothers of the study children (53% vs 36% with ≤12 years of education, p=0.04). No other differences in background characteristics shown in Table 1 were found between the study children and the children who withdrew."

3. The first line of the discussion does not appear to relate to any of the study aims posed within the final paragraph of the introduction. This statement is also not proved empirically within the methods or results of this paper. However, this might make an interesting discussion point later within the discussion section. Please also check the conclusion in the abstract and text.

Answer: We appreciate these comments and have made changes to the Abstract, Introduction, Results and Discussion accordingly.

4. It would be useful to give the geographical context of the other reports of rising trends of non-CP motor impairments in extremely preterm population.

Answer: This is an important comment regarding the geographical aspect of non-CP motor impairments. There are no population based norms for the Movement ABC-2 in Finland, but the test norms are widely in clinical use. The recent studies we referenced are from Sweden and Australia. We have revised the manuscript and added the following sentence to the discussion:

"The rate of DCD in children born extremely preterm of this PIPARI Study cohort was 18%, while two recent studies from Sweden and Australia have reported non-CP motor impairment rates of

26-37% in extremely preterm populations. The Swedish study reported that the rate of non-CP motor impairment was 37% when they used a cut-off based on their control group, but if the normative cut-offs had been used the rate would have been 12.5%.

5. The authors mention the comparison of other cohorts in reference to a control group. It would be useful to know if the authors felt that the absence of a control group would affect the results, and if so, how? It is mentioned briefly in the limitations but no robust discussion made.

Answer: Thank you for this question regarding the lack of control group. A control group of same age Finnish children would have been ideal for comparing the HRQL results. In addition, there was no possibility to compare the rate of DCD to peers born at term due to lack of a control group. Regarding the HRQoL results the Finnish normative of the same age population was available, although it was based on data collection before 1996, while this study was conducted from 2012 to 2017. We have improved the discussion of our manuscript according to this suggestion and added following sentences:

“There was no possibility to compare the rate of DCD to peers born at term due to lack of a control group. To assess the cognitive development we used WISC-IV which is a validated and widely used tool in Finland and the national cut-offs are precise and up-to-date. Regarding the HRQoL results the Finnish normative of the same age population was available, although it was based on data collection before 1996.”

6. Within the discussion, comment on the follow-up rate, difference (or lack of) in characteristics and how this may/may not have affected the study’s results.

Answer: We appreciate your comment regarding the follow-up rate and the difference in characteristics between study children and children who withdrew. We have added the descriptive statistics of the drop-outs with p-values to the Table 1 and revised the discussion accordingly.

“The strength of this study was a relatively high follow-up rate (78 %) from birth to 11 years of age”

“The mothers of the children who withdrew from the study had lower educational level compared with mothers of the study children. This may have influenced the results and may offer one explanation why the study cohort of children born very preterm reported their HRQoL better compared to the norms.”

Reviewer: 2

Comment to authors

This paper examines motor and cognitive outcomes as well as health related quality of life for 11 year old children born very preterm and/or very low birthweight in Turku, Finland. This study addresses an important topic, and uses appropriate methodology.

We thank the reviewer for very thorough review and comments.

General comments:

1. Throughout the paper, children are referred to as having or not having DCD based upon an MABC-2 score of ≤5th percentile and not having a neurological explanation for motor impairment (assessed using the Touwen neurological examination). Using the DSM-V criteria for DCD, criterions I, III and IV are adequately addressed, however criterion II (that motor difficulties adversely affect activities of daily living) does not seem to have been assessed. Although the authors have chosen a cut-off of ≤5th percentile on the MABC-2, perhaps it would be better to qualify the diagnosis, such as using the term probable DCD? I would also suggest discussing the above in either the methods or discussion section.

Answer: Thank you for this important comment. We agree that a clinical diagnosis for DCD requires confirmation that motor difficulties affect activities of daily living. Before the motor assessments the parents were interviewed according to the Developmental Coordination Disorder Questionnaire 2007 (DCDQ’07) to support the motor assessment.

2. The authors often use the phrase ‘very preterm children’. I would suggest changing to person-first language, ie ‘children born very preterm’.

Answer: We have made changes throughout the manuscript according to the suggestion.

Introduction

Line 49.

3. Suggest changing to “such as DCD on HRQoL”

Answer: We have revised the manuscript accordingly.

Line 49.

4. I was a little confused by the use of the word ‘controversial’ when referring to the articles by Dewey et. al. and Karras et. al.. From reading the referenced papers quickly, both studies found that children and adolescents with DCD experienced lower quality of life compared with their peers. Perhaps a quick explanation as to why these findings were controversial? Or a change in phrasing to reflect that the research to date supports a relationship between DCD and poor HRQoL?

Answer: Dewey et. al. found no differences in total HRQoL scores between children with or without DCD. However, they found that adolescents with DCD and AHDH had lower HRQoL on the mood and emotions and school environment. Their additional comparisons indicated that on both these subscales (the mood and emotions, and school environment) the adolescents with DCD and ADHD had significantly lower scores than adolescents with DCD only. Karras et. al. found that children with DCD reported significantly lower scores in four out of 10 domains: psychological well-being, moods and emotions, parent relations and home life, and school environment. Unlike Dewey et al they did not provide total scores. We have revised the Introduction accordingly.

“Dewey et. al. has found no differences on total HRQoL scores between adolescents with DCD and/or attention- deficit–hyperactivity disorder (ADHD), and typically developing adolescents. However, they found that adolescents with DCD and AHDH had lower HRQoL on the mood and emotions subscale, and school environment subscale. Their additional comparisons indicated that on both these subscales the adolescents with DCD and ADHD had significantly lower scores than adolescents with DCD only. Karras et. al. has found that children with DCD reported significantly lower scores in four out of 10 HRQoL subscales: psychological well-being, moods and emotions, parent relations and home life, and school environment.”

5. Lines 53-60

In the aims, I suggest indicating that children were very preterm and/or very low birth weight.

Answer: We have made corrections accordingly.

Methods:

6. Participants

It appears this study includes two different cohorts with two different sets of inclusion criteria. If this is the case, I suggest making this clear and consistent in both the ‘participants’ paragraph (page 7) and figure 1. Currently, for children born between 2001 and 2003, figure 1 outlines children are included if they are born very preterm (≤32 weeks) and ≤1500g, while in the paragraph, the inclusion criteria is ≤1500g and <37 weeks.

Answer: Thank you for this comment. The whole prospective study cohort from birth years 2001-2006 is from the same regional university hospital. When this study enrolled in 2001 the commonly used criteria to define the stage of prematurity was focused on the birth weight (≤1500 grams in very premature children). During the first years of the study the criteria for defining prematurity became more focused on the gestational age. Thus, the inclusion criteria were amended to include all infants born ≤32 weeks of gestational age irrespective of birth weight in 2004. This change added only a small amount of very preterm infants compared to the original definition.

Motor outcomes.

7. How did you assess for discrepancies between MABC-2 assessments? (referred to in the Motor Outcomes paragraph)

Answer: All the Movement ABC-2 assessments and Touwen neurological examinations were video recorded and if any hesitation regarding the assessments emerged the videos were reassessed by one experienced child neurologist. We have made changes to the methods section of the manuscript to better describe the assessment of discrepancies.

“All the Movement ABC-2 assessments and Touwen neurological examinations were video recorded. In case of any hesitation regarding the assessments, the videos were reassessed by one experienced child neurologist (LH).”

Statistical analysis

8. Line 35: an open bracket is needed before birth weight.

Answer: We have revised the text to include the missing open bracket.

9. Re the sentence; “Differences in the 17D scores between the groups were studies using Mann-Whitney U Test”. Are the groups referred to 1) children with vs without DCD, 2) with vs without CP and 3) with vs without severe cognitive impairment? Could this sentence be made a little clearer?

Answer: This is an important point. We have clarified the sentence so that the referred groups were unequivocally pointed out. The new sentence is:

“Differences in the 17D scores between the groups of 1) children with and without DCD, 2) with and without CP and 3) with and without severe cognitive impairment were studied using Mann-Whitney U Test.”

10. In the results section (health related quality of life), the authors report on correlations between MABC-2 and the 17D as well as full scale IQ and the 17D. Was this analysed using Pearson correlations? If this is the case, could this be added to the statistical analysis paragraph for clarity?

Answer: The correlations between the Movement ABC-2 and the 17D as well as full-scale IQ and the 17D were studied using Spearman's correlation as the 17D is not normally distributed. We have added this information to the statistical analysis paragraph.

Results

11. Please clarify the 77.6% follow-up rate. Eg. 170 out of 219 children who were enrolled in the cohort study prior to the 11 year follow-up? (I’m unsure if I’ve interpreted this correctly).

Answer: We thank for this careful observation. The follow-up rate was calculated using the cohort size of 170 children (i.e. children taking part in the study at the age of 11) out of 219 children born very preterm who filled the inclusion criteria and agreed to participate. We have made changes to the flow chart of the participants (Figure 1.) to clarify this issue.

12. I suggest using ‘children who withdrew’ (or something similar) rather than ‘drop-outs’ which is more colloquial.

Answer: We have made changes to the manuscript according to this suggestion.

Health related quality of life

13. How did you analyse the difference in QoL scores in your study compared with the general Finnish population? Please add a sentence to 'Statistical Analysis' to clarify which test was used.

Answer: The differences in 17D scores in study children born very preterm compared with Finnish population based normative results were studied using the Mann-Whitney U Test. We have added this information to the statistical analysis paragraph.

Table 2

14. The heading of table 2 indicates that "cognitive outcome and health-related quality of life are shown in 11 year old….", but only full scale IQ and each of the WISC-IV subscales are included in the table. Could the heading be adjusted to reflect the table?

Answer: Thank you for this observation. We have corrected the heading.

Discussion

Line 8:

15. I suggest referencing "compared with population norms (5%)"

Answer: Thank you for this suggestion. One of our main findings is that DCD was still common at 11 years of age in children born very preterm in 2000s. We have revised the manuscript to clarify the main findings in the Abstract, Introduction, Results and Discussion sections, and amended this item simultaneously.

16. The Spittle et. al. paper cited in the discussion (reference 7) does not refer to children born preterm as having DCD, but rather non-CP motor impairment. I would suggest referring to motor impairment or non-CP motor impairment rather than DCD in this paragraph when citing this paper.

Answer: This is an important comment. We have revised the Discussion of the manuscript according to this suggestion.

17. Paragraph 2 and 3. The authors use the word gender at the end of both paragraphs. I just wanted to clarify that the authors mean gender (social construct) rather than sex (biological)?

Answer: Thank you for pointing out this aspect. We have clarified the manuscript and changed the phrase "gender" to "sex" to prevent any misunderstandings.

18. The WISC-IV has some items in the Processing Speed and the Perceptual Reasoning Index subtests that require fine motor control, such as holding a pen and drawing in a small space, and manipulating blocks. Is it possible that motor impairment may negatively affect a child's performance in these subtests? Perhaps this could be included in the discussion, particularly as the paper concludes that there is a positive correlation between cognitive performance and motor skills.

Answer: We appreciate this observation. Since the possible negative effect of difficulties in fine motor skills to the child's performance in subtest mentioned cannot be ruled out we have revised the discussion to acknowledge this well-founded point:

"However, as WISC-IV has some items in the processing speed and the perceptual reasoning index subtests requiring fine motor control (e.g. holding a pen, drawing in a small space, and manipulating blocks) it is possible that motor impairment may have an effect on child's performance in these subtests."

Figure 1

19. Does ‘withdrew during the follow-up’ refer to the follow-up at 11 years? Were there any previous follow-ups between birth and 11 years? eg. withdrew at 5 year follow up?

Answer: Children born very preterm were classified as dropouts, if their families gave consent to participate the study at birth but then withdrew during the follow-up time of 11 years and the child did not take part in the study at 11 years of age.

The PIPARI Study protocol has included several follow-ups before the age of 11. Further information of our study group is found from our website: https://sites.utu.fi/pipari/en/ (in English). We have previously published a descriptive manuscript of the follow-up scheme in Finnish medical journal. We have added these references to our manuscript to meet this suggestion. Also, we are pleased to add a supplementary figure of the study protocol if that is allowed by the Editor.

Reviewer: 3

Thank you for giving me the opportunity to review this interesting report. Please find below several comments which I think have the potential to contribute in an overall improved version of the current report.

We thank the reviewer for the constructive criticism.

1. - Does PIPARI stand for something (start of the Participants section)? Could a reference please be added about PIPARI study?

Answer: We thank you for the interest in the name and background of the study group. PIPARI is an acronym for a Finnish project title “Pienipainoisten riskilasten käyttäytyminen ja toimintakyky imeväisiästä kouluikään” which translates to “The Development and Functioning of Very Low Birth Weight Infants from Infancy to School Age”. Further information of our study group including i.e. publications is found from our website: https://sites.utu.fi/pipari/en/ (in English). We have previously published a descriptive manuscript of the follow-up scheme in Finnish medical journal. We have added these references to our manuscript to meet this suggestion. Also, we are pleased to add a supplementary figure of the study protocol if that is allowed by the Editor.

2. - Is the Turku University Hospital equivalent to any other hospital in Finland? Could some comments be added about this as it will assist in the generalisation of the presented results?

Answer: This is an important question. Turku University Hospital is one of the five Finnish university hospitals, which are highest level (level III i.e. university) hospitals in Finland. We have revised the manuscript to clarify the level of our hospital as suggested.

“The participants were born to Finnish- or Swedish-speaking families from January 2001 to December 2006 in Turku University Hospital, Finland, which is one of the five level III hospitals in Finland.”

3. - What led to the expansion of the inclusion criteria in 2004?

Answer: Thank you for this question. The whole prospective study cohort from birth years 2001-2006 is from the same regional university hospital. When this study enrolled in 2001 the commonly used criteria to define the stage of prematurity was focused on the birth weight (≤1500 grams in very premature children). During the first years of the study the criteria for defining prematurity became more focused on the gestational age. Thus, the inclusion criteria were amended to include all infants born ≤32 weeks of gestational age irrespective of birth weight in 2004.

4. - Is it important for the reader to know that one of three physicians evaluated the motor outcome of very preterm children at 11 years of age? I think probably not (but I might be wrong, in which case I'd appreciate clarification) – it'd be more interesting to know that this one physician was chosen randomly from a pool of physicians with similar knowledge and experience. So the important point to be made here is about the unbiasness of the one physician.

Answer: We appreciate this question. The three physicians performing the motor and neurological assessments were PhD students in the PIPARI study group. We have made changes to the Methods considering motor outcome accordingly.

"The three physicians performing the motor and neurological assessments were PhD students of the PIPARI study group."

4.- This last comment also extends to the choice of the Finnish psychologist examining the cognitive development of the children. It is important to know that both the Finnish and Swedish psychologists were chosen randomly from a pool of psychologists with similar knowledge and experience. Currently we are told that there were two Finnish psychologists and only one was chosen – so unless I've misunderstood something, I can't see the added value from this comment.

Answer: The two psychologist performing the cognitive assessments were PhD students in the PIPARI study group. We have revised the Methods considering cognitive outcome accordingly.

"Finnish assessments were performed by one of the two psychologists, who were PhD students of the PIPARI study group."

5. - When the authors say that 'Touwen neurological examination was systematically used to confirm…no other neurological conditions' do they mean that children were examined in regular intervals throughout the follow up time? I am not sure about the use of the word systematic here. Could this be clarified please?

Answer: Thank you for this important question. Touwen neurological examination was performed to every child irrespective of the performance on Movement ABC-2. Also, before the assessments at 11 years of age there were several clinical follow-ups. We have revised the manuscript as suggested and added a previously published descriptive reference to clarify the follow-up routine.

"Touwen neurological examination was used to confirm that there were no other neurological conditions such as muscle diseases affecting the motor development."

6. - Were comparisons of the video recordings done by the same people for all cases? Were they performed for all children or when there was an issue with the measurements or reason to suspect something had gone wrong?

Answer: We appreciate this question regarding the methods. All the motor and neurological examinations were routinely videotaped and this was performed to provide reliability between the three physicians performing the examinations. In case of any hesitation regarding the assessments, the videos were reassessed by one experienced child neurologist (LH). No discrepancies were found. We have made changes to the methods section of the manuscript to better describe quality control of the assessment procedure.

"All the Movement ABC-2 assessments and Touwen neurological examinations were video recorded. In case of any hesitation regarding the assessments, the videos were reassessed by one experienced child neurologist (LH)."

7. - Which children completed the HRQoL questionnaire by themselves and which as part of an interview by a physician before their motor assessment? If this was decided by the children themselves, this might have led to a lot of bias in the results.

Answer: Thank you for this observation. All children except for one completed the HRQoL questionnaire themselves. Since one child was not able to read the questionnaire was completed by interview by the physician. Filling the questionnaire was done before the motor assessment. We have clarified this to the manuscript.

“The children completed the questionnaire by themselves before the motor assessment except for one child who was not able to read and was interviewed by the physician before the assessment.”

8. - The following URL http://www.15d-instrument.net/rmd should be replaced by http://www.15d-instrument.net/15d/replacing-missing-data/

Answer: Thank you for this important notice. We have replaced the URL.

9. - The correlation between the movement and various IQ scores were statistically significant but nevertheless not strong in magnitude. This result cannot be ignored simply because of low p-values and it needs more commenting. The scatterplot in figure 2 does not show a clear pattern between the two variables. Did the authors also try a non-parametric correlation coefficient to check the robustness of the Pearson’s value?

Answer: Thank you for this important comment. Even if the correlations were not strong in magnitude, the differences in the mean values of IQ scores between the groups of children born very preterm with and without DCD were 15 points for full-scale IQ, 8 points for verbal comprehension, 9 for perceptual reasoning, 16 for working memory, and 13 for processing speed, all in favour of children without DCD. These differences are considered of clinical significance. The Spearman’s correlation for non-parametric variables was also performed, but since the Movement ABC-2 and WISC-IV we normally distributed Pearson correlations was chosen.

“A positive correlation between motor outcome and cognitive development in children born very preterm was found even if the correlations were not strong in magnitude.”

“The differences were clinically significant in magnitude, i.e. 15 points for full-scale IQ, 8 points for verbal comprehension, 9 for perceptual reasoning, 16 for working memory, and 13 for processing speed, all in favour of children without DCD.”

10. - A lot of the comparisons made about HRQoL amongst the various groups of children were significant but in fact the actual differences between them (in terms of magnitude) were very small (as scores range from 0 to 1), i.e. 0.93 vs 0.96, 0.92 vs 0.98, 0.96 vs 0.99, etc. Could the authors please comment on whether these differences are actually clinically important? If we were given the actual values for the differences between healthy Finnish children and the very preterm cohort for sleep, discomfort and symptoms, depression, appearance, etc. we might had seen that the magnitude of the differences was about the same as the other dimensions of the HRQoL, but simply did not turn out statistically significant. The authors feedback on this might also necessitate changes to some of the overall statements about differences in HRQoL between the various groups of children throughout the paper.

Answer: We appreciate this important observation. It is true, that the numerical differences in HRQoL results are small as the scores range from zero to one. Although the differences between the groups of children were statistically significant the clinical differences may not be as significant. We have revised the manuscript and added following sentence to discussion:

“The absolute differences in HRQoL results between the groups were minor since the scoring system ranges from zero to one. Whether these statistically significant differences have clinical importance is not definite.”

11. - If not already given (as I might have missed it), can you please provide a reference for the 5% of DCD in the normal population (first sentence of the discussion section)?

Answer: Thank you for this observation. One of our main findings is that DCD was still common at 11 years of age in children born very preterm in 2000s. We have revised the manuscript to clarify the main findings in the Abstract, Introduction, Results and Discussion sections, and amended this item simultaneously.

12. - The authors mention sample size at their discussion section, but did they consider performing a sample size calculation at the start of the study to ensure that they only stop collecting data after a satisfactory number of children was observed? If not, what were the restrictions not allowing them to do so?

Answer: We acknowledge that this is an important issue. Whole study cohort of children born very preterm and fulfilling the inclusion criteria of the main study (PIPARI) who had not withdrawn before/at the age of 11 (years 2012-2017) was examined. Accordingly, sample size calculation was not performed at the start of this sub study.

13. - Finally, it seems that the Movement ABC-2 score is used to classify children’s motor abilities and put them in categories of DCD, no risk and some risk of motor problems. Has any research been done on the numerical format of the Movement ABC-2 scores as this might be more helpful that the categorisation strategy? This is a generic comment but one worth making as the authors might want to consider this option for future research in this field.

Answer: Thank you for this suggestion. The classification applied in this paper was done using the most recent EACD (European Academy of Childhood Disability) recommendations and the manual of the Movement ABC-2. However, in another recent publication we have used the MABC-2 standard scores as continuous variables (ref Lahti K, Saunavaara V, Munck P, Uusitalo K, Koivisto M, Parkkola R, Haataja L. Diffusion Tensor Imaging is associated with motor outcomes of very preterm born children at 11 years of age. Acta Paediatr. 2019 doi: 10.1111/apa.15004.)