PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A Longitudinal Cohort Study Investigating Neurodevelopmental and Socio-emotional Outcomes in School-Entry aged Children after Open Heart Surgery in Australia and New Zealand: The NITRIC Follow-up Study Protocol
AUTHORS	Long, Debbie; Anderson, Vicki; Crossley, Louise; Sood, Nikita; Charles, Karina; MacDonald, Anna; Bora, Samudragupta; Pestell, Carmela; Murrell, Kathryn; Pride, Natalie; Anderson, Peter; Badawi, Nadia; Rose, Brian; Baillie, Heidi; Masterson, Kate; Chumbes Flores, Jenipher; Sherring, Claire; Raman, Sainath; Beca, John; Erickson, Simon; Festa, Marino; Anderson, Benjamin; Venugopal, Prem; Yim, Deane; Andrews, David; Cheung, Michael; Brizard, Christian; Gentles, Thomas; Iyengar, Ajay; Nicholson, Ian; Ayer, Julian; Butt, Warwick; Schlapbach, Luregn; Gibbons, Kristen

VERSION 1 – REVIEW

REVIEWER	Flavia Wehrle University Children’s Hospital Zurich, University Children's Hospital Zurich
REVIEW RETURNED	30-May-2023

GENERAL COMMENTS	Thank you for the opportunity to review this study protocol. It describes the NITRIC Follow-up Study that aims to investigate the neurodevelopmental trajectories of children with CHD after open heart surgery between 2 and 5 years of age. The combination of annual online screenings and a face-to-face assessment at 5-years is very elegant as it might provide clinically-relevant insight into what screenings are helpful to identify those most in need for support at school entry. Also, the study follows a large cohort of children, thus, allowing relatively complex statistical modeling to identify potential risk and protective factors for neurodevelopmental outcome. Below, some issues are mentioned that should be addressed before the manuscript is ready for publishing. Abstract - Methods and analysis: some more details about the planned statistical models could be added if word limit allows Introduction Overall, the introduction provides a concise overview of the topic of the study. However, while I fully agree that historically much research has focused on moderate/severe neurodevelopmental impairments in at risk populations, I believe that in its current state, the manuscript does not acknowledge enough the variety of studies that have been reporting on milder and more subtle forms of neurodevelopmental problems, both in the preterm population but

	also in the CHD population. For example, recently, several meta-analyses have been published that provide strong evidence that executive functions are particularly affected in patients with CHD (e.g., Feldmann et al., Pediatrics, 2021; Jackson et al., Cardiology in the Young, 2021). The substantial number of original studies that were included in these meta-analyses reflects the considerable body of research that has been investigating “subtle” neurodevelopmental problems such as executive function deficits in patients with CHD. This body of research should be acknowledged in more detail in the introduction. Throughout the introduction (and the manuscript as a whole): The phrasing/terminology that is used seems to claim that long-term follow-up needs to continue well-beyond early childhood/preschool age, e.g., - P. 10: EPICure: “role of epidemiological studies in advancing understanding of life-course consequences of extreme prematurity” - P. 10: “it remains unclear which tools, at which specific time points, have the best performance to predict child outcomes at school age” However: the current study aims to follow patients up until 5 years of age - I think the introduction should state more clearly why a time-frame of 5 years for a follow-up after open heart surgery is relevant and provides a sensible time frame and important information on the (long-term) neurodevelopmental outcome of patients with CHD. - I am currently not convinced that the terminology “long-term” is adequate for the study design with a 5-year time-frame. For example, in the preterm population, many studies have looked at outcomes well-beyond this time-frame (e.g., into adolescence and even adulthood). In light of the literature that is referenced in the current manuscript (examples provided above), follow-ups at least into school-age seem more adequate to be termed “long-term”. As the repeated assessments at high temporal resolution (i.e., annually) are a clear strength of the current study, I suggest to focus on this part of the design when choosing the terminology (e.g., longitudinal, trajectories,…) - The aims are phrased in a very concise way and I think this type of terminology should be used throughout the manuscript o “early identification” o “map trajectories from 2 to 5 years” o “explore phenotypes at 5 years” o “whether screening predicts outcomes for children with CHD once they reach school age” Methods and Analyses - overall: very clear description of design Participants: - it would be helpful to include the number of eligible participants here; n=1371 is mentioned in the introduction but likely, this number has changed due to death after the initial outcome assessment etc. This is mentioned later on in the Sample Size paragraph but I think it should already be mentioned here. - Are children with confirmed or suspected genetic syndromes included? Follow-up assessments Annual online screening o are any attempts made to receive reports from 2 caregivers per

	child (e.g., the mother and the father)? Reports on parent emotional well-being and parenting stress may be quite different between caregivers, so both reports are potentially of interest. Particularly, as the literature on paternal well-being is scarce to date, reports of fathers are of interest. o If only one parental report is collected per child, (how) do you confirm that the same parent reports on child and self-outcome at each time-point? Particularly to investigate trajectories of parental well-being, it is essential to have the report of the same parent available at each time-point. But likely, also the report on the child (e.g.., SDQ) should be completed by the same caregiver across time-points to allow the investigation of within-child trajectories o Healthcare utilization: it is not quite clear whether the utilization refers to the child or to the parent (Table 1 listed under parent-focused questionnaires) Face-to-face assessment o Language: from the Table it seems that this is a test that is designed for students ("appropriate for the student's age"). Why was this task chosen if children are for the most part of the longitudinal study not yet students (i.e., aged 2 to 5 years)? o Parent-reported questionnaires: it is not clear why the fatigue questionnaire (but none of the others, e.g., the BRIEF) is administered during the face-to-face-assessment again as it is just before assessed as part of the 5-year screening. - P. 10/11: You mention the CNOC research agenda. There has also been a recent publication that summarizes the CNOC recommendations for neurodevelopmental evaluations from birth through 5 years (Ware et al., Cardiology in the Young, 2020). Could you comment on whether and how the test battery that was designed for the NITRIC Follow-up Study is in alignment with these recommendations or not? Transcriptomics/Biomarkers - There is very little information on what kind of biomarkers will be assessed (paragraph 'Prediciton models': "…several layers of biomarkers on host response to CPB (transcriptomics, metabolomics, proteomics"). Please add some more details on what is assessed and why these biomarkers were chosen. Is a whole genome/exome sequencing performed? Data analysis Derivation of socioemotional phenotypes - The paragraph on the derivation of socioemotional phenotypes is quite difficult to understand in its current form for non-statistical experts. I believe the paragraph would benefit from some restructuring, particularly by adding an introductory sentence stating the overall aim of these analyses before going into (more technical) details. - The data-driven approach including a derivation and validation sample appears relevant and feasible considering the large sample. The sentence "The appropriate clustering method will be chosen following review of the data structure." is quite vague and would benefit from adding some example of possible approaches. - It is not straightforward to understand how/why candidate variables that are drawn from the language, attention, EF and memory domains are expected to contribute to the socioemotional phenotype: Are these cognitive outcomes expected to contribute to the socioemotional phenotype and if so, in the same way/to the same extent as the outcomes from the social behavioral and

	functional domains? - How is the optimal number of socioemotional phenotypes determined (e.g., through model fit indices?). - How are “highly correlated variables” defined (e.g., r>.90 or some other threshold)? Prediction models - It is not quite clear whether the prediction models also aim to predict the profile types and the trajectory types that are identified via the analyses described in the previous 2 paragraphs (in addition to predicting “outcomes in the neurodevelopmental and socioemotional domains”). This could provide interesting insight into distinct risk and protective factors for (homogeneous) subgroups of patients with CHD - Please specify in more detail what kind of models will be used, e.g., multiple regression models, mixed-effect models,…(for example to make the sentence “The model will allow for instance where the child does not have a full set of questionnaires" better understandable; is this a mixed-effect model with random effects (intercept and slope))? Missing data - Please provide some information on who missing data imputation will be approached. This is particularly relevant as for SEM approaches, complete data is needed. Feasibility and Engagement Providing feedback to the participants and there families seems highly adequate, not only as an appreciation for their efforts but also to ensure continued engagement. However, I am somewhat concerned about the strategy to provide written feedback and an encouragement to contact primary healthcare providers in case of identified domains. There is a risk that parents with higher education/better English skills/more awareness of the healthcare system are more likely to actually take action if prompted in a written form. Personal contact in case of problems might reduce this risk, e.g., discussion of concerning findings on the phone with a study team member and specific discussion/instruction of who to contact etc. I think this is relevant for the study outcome because a potential bias in the amount of therapy that children receive might impact the findings on the longitudinal trajectories of neurodevelopmental outcome. The authors mention the potential impact of feedback on service utilization in the limitation section. Assessing healthcare utilization is definitely an appropriate measure. However, I suggest that it should be discussed how to prevent a potential bias in who accesses therapies as a reaction to the study feedback. Discussion It is mentioned as a strength that the cohort is based on a large high-quality pragmatic trial with broad inclusion criteria. This is difficult to judge based on the information on the original trial that is provided in the manuscript. If the word limit allows, a few additional information on the original trial (e.g., what were the broad inclusion criteria) would be very helpful to the reader to better understand the overall characteristics of the cohort. Minor comments: - P. 10, line 21-24: it is not clear why the distinction from the preterm population is highlighted here.

	- P. 8 (and throughout the manuscript): “consumers” sounds a bit strange in this context, at least for non-native English speakers; is this a standard term to use in this context? Patients and families? - P. 11/P. 40: “associations of outcomes with the host response to CPB” sounds a bit strange. Are the hosts the children with CHD? - It is not quite clear why the chapters “Measures” and “Follow-up Assessments” are presented separately because the instruments presented in the chapter “Follow-up Assessments” are measures, so why aren’t they presented in the respective chapter? Consider merging these two chapters and grouping the respective information on the annual questionnaires and the face-to-face assessment. - Table 1: o Column “comments”: the number of age-appropriate questionnaires is mentioned (e.g., 21 for the ASQ). Does “questionnaires” refer to items? Or scales? I don’t think “questionnaires is the correct term? o I may have missed it but I could not find the “^” and the “§” in the Table but only in the footnote - Table 2 o Typo : Fatigue: “.5-point Likert scale” o “^” missing in the Table? - P. 36 : what do “procedures for escalating efforts to reach participants” entails? - P. 39: What is meant by “exploring social determinant interactions”?

REVIEWER	Thiviya Selvanathan The Hospital for Sick Children, Pediatrics
REVIEW RETURNED	23-Jun-2023

GENERAL COMMENTS	The authors describe the study protocol for longitudinal neurodevelopmental follow-up of a multi-center prospective cohort study of children with congenital heart disease who underwent open heart surgery.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1.

Thank you for the opportunity to review this study protocol. It describes the NITRIC Follow-up Study that aims to investigate the neurodevelopmental trajectories of children with CHD after open heart surgery between 2 and 5 years of age. The combination of annual online screenings and a face-to-face assessment at 5-years is very elegant as it might provide clinically-relevant insight into what screenings are helpful to identify those most in need for support at school entry. Also, the study follows a large cohort of children, thus, allowing relatively complex statistical modeling to identify potential risk and protective factors for neurodevelopmental outcome.

Many thanks for your kind words and review.

Introduction

Overall, the introduction provides a concise overview of the topic of the study. However, while I fully agree that historically much research has focused on moderate/severe neurodevelopmental impairments in at risk populations, I believe that in its current state, the manuscript does not acknowledge enough the variety of studies that have been reporting on milder and more subtle forms of neurodevelopmental problems, both in the preterm population but also in the CHD population. For example, recently, several meta-analyses have been published that provide strong evidence that

executive functions are particularly affected in patients with CHD (e.g., Feldmann et al., Pediatrics, 2021; Jackson et al., Cardiology in the Young, 2021). The substantial number of original studies that were included in these meta-analyses reflects the considerable body of research that has been investigating “subtle” neurodevelopmental problems such as executive function deficits in patients with CHD. This body of research should be acknowledged in more detail in the introduction.

Many thanks for raising this important point. We have reworded the introduction to acknowledge the work undertaken in this area, including the following on pg 10:

In particular, two recent systematic reviews have demonstrated consistent evidence for executive function impairment in school-aged children with CHD, underscoring the need for follow-up (16, 17). Despite the median age at follow-up in these papers being closer to high school age, the American Heart Association guidelines recommend starting screening for executive function at 6 years of age (18). Moreover, problems may present prior to formal schooling, therefore earlier screening may be beneficial. Executive functions begin to emerge during infancy and are core skills critical for the life-course, including success in school and in life.

Throughout the introduction (and the manuscript as a whole): The phrasing/terminology that is used seems to claim that long-term follow-up needs to continue well-beyond early childhood/preschool age, e.g.,

P. 10: EPICure: “role of epidemiological studies in advancing understanding of life-course consequences of extreme prematurity”

P. 10: “it remains unclear which tools, at which specific time points, have the best performance to predict child outcomes at school age”

However: the current study aims to follow patients up until 5 years of age

I think the introduction should state more clearly why a time-frame of 5 years for a follow-up after open heart surgery is relevant and provides a sensible time frame and important information on the (long-term) neurodevelopmental outcome of patients with CHD.

Many thanks for this advice. We have made changes to the introduction, including the following commencing on page 10:

These impairments in children with CHD are important indicators of school readiness, with increasing awareness of the need to obtain an adequate developmental assessment before school entry so that education, family and child supports can be put into place to optimise outcomes.

I am currently not convinced that the terminology “long-term” is adequate for the study design with a 5-year time-frame. For example, in the preterm population, many studies have looked at outcomes well-beyond this time-frame (e.g., into adolescence and even adulthood). In light of the literature that is referenced in the current manuscript (examples provided above), follow-ups at least into school-age seem more adequate to be termed “long-term”. As the repeated assessments at high temporal resolution (i.e., annually) are a clear strength of the current study, I suggest to focus on this part of the design when choosing the terminology (e.g., longitudinal, trajectories,…)

We agree that other populations have looked at outcomes well beyond 5 years, however the use of ‘long-term’ is more reflective of outcomes beyond the acute hospitalisation period. We have elected to keep the term ‘long-term’ where appropriate.

The aims are phrased in a very concise way and I think this type of terminology should be used throughout the manuscript

- “early identification”
- “map trajectories from 2 to 5 years”
- “explore phenotypes at 5 years”
- “whether screening predicts outcomes for children with CHD once they reach school age”

Thank you for this feedback. We have reviewed the manuscript and amended terminology where appropriate.

Methods and Analyses

Overall: very clear description of design

Thank you.

No changes made to the manuscript.

Participants:

It would be helpful to include the number of eligible participants here; n=1371 is mentioned in the introduction but likely, this number has changed due to death after the initial outcome assessment etc. This is mentioned later on in the Sample Size paragraph but I think it should already be mentioned here.

Many thanks for this feedback. We have incorporated the following information in the participant section on page 13:

In the NITRIC Follow-Up Study, we anticipate that 1150 surviving children from Australian and New Zealand sites will be eligible to participate.

Are children with confirmed or suspected genetic syndromes included?

Yes. Confirmed genetic syndromes at time of surgery were collected during the NITRIC RCT and will be included as covariates in the models.

No changes made to the manuscript.

Follow-up assessments

Annual online screening

Are any attempts made to receive reports from 2 caregivers per child (e.g., the mother and the father)? Reports on parent emotional well-being and parenting stress may be quite different between caregivers, so both reports are potentially of interest. Particularly, as the literature on paternal well-being is scarce to date, reports of fathers are of interest.

Thank you for raising this important question. We agree that this is an area in need of further research, however, unfortunately we are only requesting information from the primary caregiver.

No changes made to the manuscript.

If only one parental report is collected per child, (how) do you confirm that the same parent reports on child and self-outcome at each time-point? Particularly to investigate trajectories of parental well-being, it is essential to have the report of the same parent available at each time-point. But likely, also the report on the child (e.g.., SDQ) should be completed by the same caregiver across time-points to allow the investigation of within-child trajectories.

Thank you for this question. While we don't confirm that it is the same parent/carer, we do collect, at each survey timepoint, who is completing the survey (eg. mother, father, grandparent, other). Our previous experiences indicate that the surveys are most likely done by the same caregiver as this is the primary point of contact for each annual assessment. We will report on the parent/carer completing the surveys in the results manuscript.

No changes made to the manuscript.

Healthcare utilization: it is not quite clear whether the utilization refers to the child or to the parent (Table 1 listed under parent-focused questionnaires)

Apologies for this confusion. We are assessing healthcare utilisation of the child. We have moved this information to sit within the child-focussed questionnaires in Supplemental Table S1.

Face-to-face assessment

Language: from the Table it seems that this is a test that is designed for students ("appropriate for the student's age"). Why was this task chosen if children are for the most part of the longitudinal study not yet students (i.e., aged 2 to 5 years)?

Many thanks requesting this clarification. This face-to-face assessments are only undertaken at 5 years of age. The CELF-5 ANZ Screener is often used at school entry, hence the term student (taken from manual). In Australia and New Zealand, most children enter formal schooling in the year they turn 5, so many children are 4 years of age at school entry. We have also changed the term 'student' to 'child'.

Parent-reported questionnaires: it is not clear why the fatigue questionnaire (but none of the others, e.g., the BRIEF) is administered during the face-to-face-assessment again as it is just before assessed as part of the 5-year screening.

Only general fatigue subscale of the PedsQL multidimensional fatigue scale is measured at each annual screening. The whole fatigue scale is assessed at the face-to-face assessment.

No changes made to the manuscript.

P. 10/11: You mention the CNOC research agenda. There has also been a recent publication that summarizes the CNOC recommendations for neurodevelopmental evaluations from birth through 5 years (Ware et al., Cardiology in the Young, 2020). Could you comment on whether and how the test

battery that was designed for the NITRIC Follow-up Study is in alignment with these recommendations or not?

Thank you for this comment. Yes, we are aware of Ware et al's recent recommendations. The NITRIC FU assessments align with the domains outlined by Ware et al, however assessments are undertaken at additional 24, and 48 month timepoints. The named assessments don't fully align with these recommendations based on local understanding, familiarity and availability of the tools.

No changes made to the manuscript.

Transcriptomics/Biomarkers

There is very little information on what kind of biomarkers will be assessed (paragraph 'Predicton models': "…several layers of biomarkers on host response to CPB (transcriptomics, metabolomics, proteomics"). Please add some more details on what is assessed and why these biomarkers were chosen. Is a whole genome/exome sequencing performed?

Thank you for raising this. Whole genome/exome sequencing is not being performed. We have updated this paragraph to provide additional information on how, and which, biomarkers are being assessed. However, as the purpose of this protocol publication is to describe the neurodevelopmental outcome assessments, and we plan to publish separately detailed information on omics analyses, we have only provided brief detail in this manuscript on page 39.

Transcriptomics data will be generated on the full cohort with matched pre- and post-surgery samples and metabolomics data and proteomics data will be generated on subset of cohort. We will use forward selection algorithms to identify variables from each data set to discover novel biomarkers to predict patient outcomes after CPB. We will also combine these datasets to derive a combination biomarker (including gene expression, metabolites and proteins) to predict short-term and long-term patient outcomes.

Data analysis

Derivation of socioemotional phenotypes

The paragraph on the derivation of socioemotional phenotypes is quite difficult to understand in its current form for non-statistical experts. I believe the paragraph would benefit from some restructuring, particularly by adding an introductory sentence stating the overall aim of these analyses before going into (more technical) details.

Thank you for this feedback. We have reviewed the paragraph on page 37 and revised to aid readability.

The data-driven approach including a derivation and validation sample appears relevant and feasible considering the large sample. The sentence "The appropriate clustering method will be chosen following review of the data structure." is quite vague and would benefit from adding some example of possible approaches.

Thank you for this comment. For brevity, and to meet the journal word count, we initially did not elaborate on these techniques. We have amended the paragraph on page 37 based on the reviewer's feedback to provide more detail related to the process that will be undertaken during clustering.

To derive neurodevelopmental and socioemotional phenotypes at 5 years of age, the cohort will firstly be split into derivation and validation subsets (65:35 using a temporal split). We will ensure the subsets are balanced for the original intervention in the NITRIC trial to avoid bias by intervention, as well as the original NITRIC trial stratification variables (age group and cardiac pathophysiology). Outcomes from the assessments undertaken at 5 years of age (listed in Supplemental Table S2) will be used to derive neurodevelopmental and socioemotional phenotypes. These will include the language, attention, executive functioning, and memory, and social behavior and functioning domains. As such, the cohort will be restricted to children who have completed at least one assessment at the 5 years face-to-face visit. Where children have not completed the full assessment, multiple imputation will be used to impute missing outcome data. Descriptive analysis will firstly be performed to assess missingness, correlation and distribution, and to identify highly correlated outcomes. If two outcomes are highly correlated (r >0.8), only one will be retained in the clustering analysis to avoid redundancy. Due to the potential for missing outcome data, multiple imputed datasets will be generated, and k-means clustering undertaken on each to assess stability. Standard indices will be used to identify the optimal number of phenotypes (e.g., Silhouette index, Gap index, Dunn index), and one set of phenotypes from the multiple imputed datasets used for the remaining analyses. Graphical methods will be used to describe and visualize the composition of the phenotypes. Latent class analysis will then be used to assess the reproducibility of the phenotypes within the entire dataset.

It is not straightforward to understand how/why candidate variables that are drawn from the language, attention, EF and memory domains are expected to contribute to the socioemotional phenotype: Are these cognitive outcomes expected to contribute to the socioemotional phenotype and if so, in the same way/to the same extent as the outcomes from the social behavioral and functional domains?

We apologise for this error. We meant to state neurodevelopmental and socioemotional phenotypes. The outcomes for the neurodevelopmental phenotype will be drawn from the language, attention, executive functioning and memory domains. The outcomes for the socioemotional phenotype will be drawn from the social behavior and functioning domains. This has been updated, commencing page 37.

Outcomes from the assessments undertaken at 5 years of age (listed in Supplemental Table S2) will be used to derive neurodevelopmental and socioemotional phenotypes. These will include the language, attention, executive functioning, and memory, and social behavior and functioning domains, respectively.

How is the optimal number of socioemotional phenotypes determined (e.g., through model fit indices?).

Thank you for this question. We have added a number of standard indices that are commonly used to identify the optimal number of clusters on page 38.

Standard indices will be used to identify the optimal number of phenotypes (e.g., Silhouette index, Gap index, Dunn index).

How are “highly correlated variables” defined (e.g., r>.90 or some other threshold)?

Many thanks for this question. We have added the following to the descriptive analysis steps on page 37.

Descriptive analysis will firstly be performed to assess missingness, correlation and distribution, and to identify highly correlated outcomes. If two outcomes are highly correlated (r >0.8), only one will be retained in the clustering analysis to avoid redundancy.

Prediction models

It is not quite clear whether the prediction models also aim to predict the profile types and the trajectory types that are identified via the analyses described in the previous 2 paragraphs (in addition to predicting “outcomes in the neurodevelopmental and socioemotional domains”). This could provide interesting insight into distinct risk and protective factors for (homogeneous) subgroups of patients with CHD.

Thank you for this comment. We have revised this section on page 38 to ensure it is clear that we will be developing models to predict both the phenotypes, and the individual outcomes.

Mixed-effects models will be developed to investigate which individual, parent, surgical, PICU treatment and sociodemographic factors known at the time of surgery are associated with both neurodevelopmental and socioemotional outcomes, and derived phenotypes. The models will account for risk factors for cognitive delays (identified through existing literature and clinical judgement), the original NITRIC trial intervention and stratification variables (as fixed effects), and study site (random effect).

Please specify in more detail what kind of models will be used, e.g., multiple regression models, mixed-effect models,…(for example to make the sentence “The model will allow for instance where the child does not have a full set of questionnaires" better understandable; is this a mixed-effect model with random effects (intercept and slope))?

Many thanks. We have provided more detail, commencing page 38, as requested.

Mixed-effects models will be developed to investigate which individual, parent, surgical, PICU treatment and sociodemographic factors known at the time of surgery are associated with both neurodevelopmental and socioemotional outcomes, and derived phenotypes. The models will account for risk factors for cognitive delays (identified through existing literature and clinical judgement), the original NITRIC trial intervention and stratification variables (as fixed effects), and study site (random effect).

Missing data

Please provide some information on who missing data imputation will be approached. This is particularly relevant as for SEM approaches, complete data is needed.

Thank you for this comment. Throughout the various sections in the “Data Analysis” section, commencing page 36, we have included information relating to missing data and how it will be handled.

Developmental trajectories: Previous experience has demonstrated that variables from the NITRIC RCT have minimal missing data, however when missing data is evident, multiple imputation methods will be used for covariate data.

Derivation of Neurodevelopmental and Socioemotional Phenotypes: … As such, the cohort will be restricted to children who have completed at least one assessment at the 5 years face-to-face visit. Where children have not completed the full assessment, multiple imputation will be used to impute missing outcome data. … Due to the potential for missing outcome data, multiple imputed datasets will be generated, and k-means clustering undertaken on each to assess stability.

Feasibility and Engagement

Providing feedback to the participants and their families seems highly adequate, not only as an appreciation for their efforts but also to ensure continued engagement. However, I am somewhat concerned about the strategy to provide written feedback and an encouragement to contact primary healthcare providers in case of identified domains. There is a risk that parents with higher education/better English skills/more awareness of the healthcare system are more likely to actually take action if prompted in a written form. Personal contact in case of problems might reduce this risk, e.g., discussion of concerning findings on the phone with a study team member and specific discussion/instruction of who to contact etc. I think this is relevant for the study outcome because a potential bias in the amount of therapy that children receive might impact the findings on the longitudinal trajectories of neurodevelopmental outcome. The authors mention the potential impact of feedback on service utilization in the limitation section. Assessing healthcare utilization is definitely an appropriate measure. However, I suggest that it should be discussed how to prevent a potential bias in who accesses therapies as a reaction to the study feedback.

Many thanks for this important question. We acknowledge that this is a scientific limitation, however as you comment this also provides important feedback for families. In weighing up the risk and benefits to this scenario, we decided that it was in the best interests of the children and their families to provide the reports. Parents are also given the opportunity to discuss their concerns with a study team member. We plan report on whether parents have shared their child's report with any healthcare providers.

No changes made to the manuscript.

Discussion

It is mentioned as a strength that the cohort is based on a large high-quality pragmatic trial with broad inclusion criteria. This is difficult to judge based on the information on the original trial that is provided in the manuscript. If the word limit allows, a few additional information on the original trial (e.g., what were the broad inclusion criteria) would be very helpful to the reader to better understand the overall characteristics of the cohort.

Many thanks for this comment. Given the discussion section has been removed, we have provided reference to the Statistical Analysis Plan and the NITRIC RCT protocol and Main Results paper in the introduction on page 12.

Minor comments:

P. 10, line 21-24: it is not clear why the distinction from the preterm population is highlighted here.

Many thanks. We have removed the reference to the preterm population on page 10.

P. 8 (and throughout the manuscript): “consumers” sounds a bit strange in this context, at least for non-native English speakers; is this a standard term to use in this context? Patients and families?

Consumers is a term commonly used in Australia and New Zealand, however we have changed consumer to families throughout the text.

P. 11/P. 40: “associations of outcomes with the host response to CPB” sounds a bit strange. Are the hosts the children with CHD?

Yes, the hosts are the children with CHD. We believe this terminology is broadly used and, as such, have not amended the manuscript.

No changes made to manuscript.

It is not quite clear why the chapters “Measures” and “Follow-up Assessments” are presented separately because the instruments presented in the chapter “Follow-up Assessments” are measures, so why aren’t they presented in the respective chapter? Consider merging these two chapters and grouping the respective information on the annual questionnaires and the face-to-face assessment.

Many thanks for this suggestion for improvement. Follow-up Assessments refers to the procedure of follow-up. We have measures section, commencing page 14, as per your suggestion.

Table 1:

Column “comments”: the number of age-appropriate questionnaires is mentioned (e.g., 21 for the ASQ). Does “questionnaires” refer to items? Or scales? I don’t think “questionnaires is the correct term?

Many thanks for this question. Many of these additional comments or instructions are taken from the corresponding test manual. Eg. ASQ uses the wording questionnaire.

No changes made to the manuscript.

I may have missed it but I could not find the “^” and the “§” in the Table but only in the footnote

Many thanks for pointing this out. These symbols have now been added.

Table 2

Typo : Fatigue: “.5-point Likert scale”

Apologies, this has been corrected.

“^” missing in the Table?

Apologies, this has been added next to #.

P. 36 : what do “procedures for escalating efforts to reach participants” entails?

Many thanks for raising this question. Our SOP for contacting parents, is based on Needham’s work/website https://www.improvelto.com/cohort-retention-tools/. The following information has been included on page 40:

, including varying contact modes and reminders.

P. 39: What is meant by “exploring social determinant interactions”?

We apologise for the confusion. We have amended this sentence on page 42:

Furthermore, this cohort allows for exploring which sociodemographic variables predict with neurodevelopment in a large binational cohort.

VERSION 2 – REVIEW

REVIEWER	Flavia Wehrle University Children’s Hospital Zurich, University Children's Hospital Zurich
REVIEW RETURNED	07-Aug-2023

GENERAL COMMENTS	Thank you for the careful and detailed revision of the manuscript. The newly submitted version resolves the concerns I had raised previously. The only remaining comment I have: Please add appropriate references to the following statements: "Executive functions begin to emerge during infancy and are core skills critical for the life-course, including success in school and in life." "These impairments in children with CHD are important indicators of school readiness, with increasing awareness of the need to obtain an adequate developmental assessment before school entry so that education, family and child supports can be put into place to optimise outcomes." All the best with the study, I'm looking forward to interesting reports on the results in the future!