PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cohort profile: Understanding Pregnancy Signals and Infant Development (UPSIDE), a pregnancy cohort study on prenatal exposure mechanisms for child health
AUTHORS	o'Connor, Thomas; Best, Meghan; Brunner, Jessica; Ciesla, Allison; Cunning, Allison; Kapula, Ntemena; Kautz, Amber; Khoury, Leena; Macomber, Allison; Meng, Ying; Miller, Richard; Murphy, Hannah; Salafia, Carolyn; Vallejo Sefair, Ana; Serrano, Jishyra; Barrett, Emily

VERSION 1 – REVIEW

REVIEWER	Antje Horsch University of Lausanne, Switzerland
REVIEW RETURNED	17-Oct-2020

GENERAL COMMENTS	This study protocol describes an impressive multi-centre study aimed at better understanding the mechanisms involved in the prospective relationship between prenatal stress exposure and child outcomes. Its strengths are repeated assessments across trimesters and up to 4 years following childbirth, analysis of multiple biological pathways in question, and detailed analysis of placental structure and function as mediators of child health outcomes. There is no doubt that the results of this study will make an important contribution to the literature and provide many opportunities for external collaborations on the enormous body of data generated by this study. I suggest that the authors add some more details and clarifications to their study protocol before publication. Introduction It would be helpful to have a clear definition of “stress” – sometimes you also use the term “distress”. “Central to this hypothesis is the concept that early exposures have a privileged – or different – effect on biological systems than those occurring later in development.” Could you expand on this and provide references? Methods If you’re particularly interested in the impact of prenatal stress exposure on child outcomes, what was the rationale for not selecting a high-risk sample or comparing a high-risk with a low-risk sample? This is my main concern regarding the methodology of the study.

	Could you provide a rationale for the sample size (power calculation)? Could you provide a rationale for the choice of time points for the study assessments? “Some of these assessments (e.g. Bayley Scales of Infant Development-III) have been widely used in hundreds of paediatric studies and provide global measures of development, whereas others are more targeted, selected based on previous work linking them to prenatal stress, inflammation, or sex steroid exposure.” Could you provide a more detailed rationale for the choice of child outcomes and include some references of previous studies?

REVIEWER	Larissa Rossen University of British Columbia, Canada
REVIEW RETURNED	21-Nov-2020

GENERAL COMMENTS	Thankyou for the opportunity to review the cohort profile paper of the UPSIDE study (Understanding Pregnancy Signals and Infant Development), a pregnancy cohort study on prenatal exposure mechanisms for child health. There are several key strengths of the study, which are clearly articulated in the paper. The study collects intensive, serial bio-specimen and questionnaire collection from the first trimester of pregnancy through age 4 that will allow the authors to test hypotheses regarding the pathways by which psychosocial stress impacts children’s development during critical and sensitive periods. Another strength of the study is the intensive longitudinal follow-up of mother-child dyads with a focus on repeated measures to assess intra-individual changes over time. I really commend the authors on a very comprehensive and valuable contribution to the literature and I am anticipating some interesting findings coming from this cohort. Here are some detailed comments to be addressed for the study: Page 8, Line 18: There needs to be a space between “period” and “were” The BSID-III is administered at 3 time points postpartum. Wondering how learning from children completing repeated measures of the task will be accounted for? Page 17, line 29-30: It states that a third measurement is obtained when the first two differ by more than a pre-specified amount. Could you please detail what the pre-specified amount is for the present study? The cord blood collection rate (88%) is quite a bit less than the placenta collection rates (96%). Wondering if there are any consistent reasons for this difference that you have found? References 17 and 19 require some formatting. Partners are noted in the demographics table, but are not discussed at detail throughout the study. Wondering why partners aren’t a focus of this study? Perhaps a clear rationale for only including mothers could be provided.

	My biggest question or concern for this study is that, although the introduction specified maternal prenatal stress and the placenta as a key focus for this study, there is an inordinate amount of data being collected in addition to these variables with no real theoretical basis provided. Of course, one can assume there are a lot of hypothetical mechanisms or pathways for these variables to be included, but the rationale for inclusion is not clearly provided in any great detail throughout the paper. As a result, the study design comes across as a large “fishing expedition” for variables that could be of interest to developmental associations in childhood and beyond. The authors would need to address this clearly and succinctly for credibility and rigour. Although the research question is clear at the start of the paper, the inclusion of so many other variables makes this research question unclear.

REVIEWER	Gonzalez Casanova, Ines Indiana University Bloomington, Applied Health Science
REVIEW RETURNED	03-Dec-2020

GENERAL COMMENTS	This manuscript describes an exciting new cohort designed to look at associations between maternal stress and child growth and development. The cohort is very relevant, the design seems strong and the follow-up so far has been high. However, the reporting of the work needs to be better organized and strengthened to better convey the many strengths of the cohort. More specific suggestions include: - Include in the introduction the gap that this cohort is addressing. Why is it important to have another observational study of these associations? What is the primary and secondary questions that this cohort will answer (or contribute to answer)? There are some suggestions of where this study is going in the text but it is never explicitly described. -Also in the introduction describe what are the specific mechanisms that this cohort will help elucidate. Maybe a diagram that includes the developmental programming model, the mechanistic outcomes (like the placenta) and then the functional outcomes like growth and development. Methods: - Separate the different methods by timepoint. Sometimes it is hard to tell what was measured when. -Categorize measurements into outcomes (final and intermediate), exposures and confounders Results -It was not clear to me which activities have already been conducted and which ones are planned but not finalized. It would be helpful to have that information at the beginning of the results. In the methods some are written in past tense but the studies have not been conducted. -A statistical analysis plan to address the primary study questions with power calculations should be included. Future directions - the ECHO consortium should be described earlier, perhaps as part of the introduction and rationale for this cohort. In the future directions also main analyses could be discussed.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Antje Horsch, University of Lausanne, Lausanne University Hospital

Comments to the Author:
This study protocol describes an impressive multi-centre study aimed at better understanding the mechanisms involved in the prospective relationship between prenatal stress exposure and child outcomes. Its strengths are repeated assessments across trimesters and up to 4 years following childbirth, analysis of multiple biological pathways in question, and detailed analysis of placental structure and function as mediators of child health outcomes. There is no doubt that the results of this study will make an important contribution to the literature and provide many opportunities for external collaborations on the enormous body of data generated by this study.

Thank you for your positive comments about the manuscript and the project more generally. Please note that the current manuscript describes the UPSIDE cohort located in Rochester, NY, USA and not the multi-cohort component of the national ECHO study.

Introduction

It would be helpful to have a clear definition of "stress" – sometimes you also use the term "distress".

We appreciate the reviewer's concern and agree with the implicit concern about terminology – which, like the measures used in studies in this area, varies across papers. Our aim in using "distress" was to provide a single term that would avoid awkward phrasing such as "symptoms of anxiety or stress" or other more descriptive and inclusive terms that are a bit more cumbersome. We have revised the paper so that we now explicitly note in the introduction the variation across studies in terms and measures, and our measurement strategy prioritizes the core constructs. We continue to prefer a more general term for use in the paper, but hopefully the additional text helps clarify how we have operationalized the construct.

"Central to this hypothesis is the concept that early exposures have a privileged – or different – effect on biological systems than those occurring later in development." Could you expand on this and provide references?

We have revised this section and provided additional references. (Please note the references from the prior sentence also illustrate this point.) The key point here is that the developmental origins hypothesis proposes that in utero exposures may have lasting effects, and that is the central focus of the study.

p. 5, lines 91-97: "In other words, when exposures occur very early in development, physiology may change (either adaptively or pathologically) resulting in long-lasting or permanent impacts on health and well-being. A classic example of this is the "thrifty phenotype" whereby nutrient deprivation during prenatal development may lead to reduced fetal growth and metabolic changes to conserve energy. In the presence of subsequent nutrient surplus (characteristic of the modern Western diet), this metabolic conservation may lead to obesity and metabolic disease."

Methods

If you're particularly interested in the impact of prenatal stress exposure on child outcomes, what was the rationale for not selecting a high-risk sample or comparing a high-risk with a low-risk sample? This is my main concern regarding the methodology of the study.

The revised paper more clearly describes that we oversampled from a high psychosocial risk (but medically normal risk) population as identified by sociodemographic factors of the constituent patient populations at the participating clinics. The result is that we have obtained a sample with considerable variation in psychosocial risk. We had considered a more clinical focus (e.g., comparing clinically anxious and non-anxious samples) but we think our approach provides stronger generalization to the full spectrum of maternal psychosocial status and logistically, it provided a more streamlined way to integrate our study into the workflow of busy obstetric clinics. We have added the text below to address this point:

p. 9, lines 199-201: "No screening for distress was conducted prior to consent; instead we recruited from clinics who serve women at high psychosocial risk."

Could you provide a rationale for the sample size (power calculation)?

We have added text on power calculations and selection of the sample size as follows.

p. 20, lines 470-486: "Power calculations were informed by results from our prior cohort studies and indicated that the study would be appropriately powered with a sample size of approximately 290 mother-child dyads. These original power calculations were designed to address hypotheses related to sex steroid pathways, as those were the first set of funded Aims at the time of cohort establishment. For example, with an anticipated correlation of 0.19, we would have 90% power to detect a significant slope in the regression of maternal anxiety (PSWQ scores) on concentrations of estriol, an estrogen of primarily placental origin. For our hypothesis on PSWQ scores in relation to anogenital distance (a marker of prenatal sex steroid activity), with an estimated correlation of 0.25 between maternal PSWQ scores and girls' AGD, we would have 81% power to detect a slope≠0 and 86% power to detect a sex-anxiety interaction (with boys' slope = -0.13). Retention of 226 children at age 12 months would provide 89% power to detect an association between maternal PSWQ scores and play behavior in girls, with weaker or no associations expected in boys. These power calculations are provided as illustrative analyses with the recognition that there will be variation in power based on the particular question under consideration. Additionally, for some highly novel analyses (e.g. maternal serial inflammatory markers in relation to child MRI data), unfortunately there is a lack of effect size data on which to power the study."

Could you provide a rationale for the choice of time points for the study assessments?

Thank you for this important point. We have added the following text (pp. 10, lines 216-220): "Child outcome timepoints were chosen based on consideration of several key criteria: 1) developmental milestones and critical windows; 2) coincidence with routine well-child appointments; 3) spacing of visits to allow for repeated measures within domains over time, while minimizing participant burden and loss to follow-up; and 4) constraints of funding timelines."

"Some of these assessments (e.g. Bayley Scales of Infant Development-III) have been widely used in hundreds of paediatric studies and provide global measures of development, whereas others are more targeted, selected based on previous work linking them to prenatal stress, inflammation, or sex steroid exposure." Could you provide a more detailed rationale for the choice of child outcomes and include some references of previous studies?

This sentence has been deleted in the revision, however we note that there are references provided for all neurodevelopmental outcome measures and when the measures are specific to one biological mechanism (e.g. sex steroids and sex-dependent development) we have noted that in the text. The revised manuscript emphasizes that our assessment strategy was to include leading measures of core constructs for neurodevelopment, such as general cognition, language, and executive function; the reviewer is correct in noting that only some of these measures were used in prior studies of sex differences in infants and young children.

Reviewer: 2
Dr. Larissa Rossen, UBC

Comments to the Author:
Thank you for the opportunity to review the cohort profile paper of the UPSIDE study (Understanding Pregnancy Signals and Infant Development), a pregnancy cohort study on prenatal exposure mechanisms for child health. There are several key strengths of the study, which are clearly articulated in the paper. The study collects intensive, serial bio-specimen and questionnaire collection from the first trimester of pregnancy through age 4 that will allow the authors to test hypotheses regarding the pathways by which psychosocial stress impacts children's development during critical and sensitive periods. Another strength of the study is the intensive longitudinal follow-up of mother-child dyads with a focus on repeated measures to assess intra-individual changes over time. I really commend the authors on a very comprehensive and valuable contribution to the literature and I am anticipating some

interesting findings coming from this cohort. Here are some detailed comments to be addressed for the study.

We appreciate your positive feedback- thank you!

Page 8, Line 18: There needs to be a space between "period" and "were"

We have made this correction.

The BSID-III is administered at 3 time points postpartum. Wondering how learning from children completing repeated measures of the task will be accounted for?

The rationale for repeated assessments is to accommodate the in-built error and challenge in assessing young children. The evidence is that correlations between cognitive assessments in very young children are modest, probably both because of the nature of development and the challenges of assessing very young children. The more critical point is that, given the spacing between assessments (6m, 12m, 24m), there would not be substantive concerns about practice effects. Those kinds of concerns are notable, but really only with much more closely spaced assessments (e.g., up to a few weeks apart), and all of the concerns that we have seen on this issue are with older children. Also, we know that the infants will not know if they got an item correct (formal feedback is not allowed in the assessment) and so, technically, there would not be a feedback for learning. We have added the text below to make that point.

P. 15, lines 350-353: "Importantly, although practice effects sometimes occur in older children when measures are closely spaced (i.e. several weeks apart), this is unlikely to occur in children this young with visits spaced many months or years apart."

Page 17, line 29-30: It states that a third measurement is obtained when the first two differ by more than a pre-specified amount. Could you please detail what the pre-specified amount is for the present study?

The threshold for triggering a third measurement varied by measure and age. For instance, for skinfold thickness measures which are quite small at birth (reference range ~2.0-6.0 mm), a third measurement was taken if the first two differed by more than 1.0 mm. By contrast, head circumference measures are larger (in the range of 30-38 cm at birth) and a third measurement was taken when the first two differed by 1.0 cm. These guidelines were developed to help the study staff quickly determine when to take a third measurement without having to calculate percent difference. We have used this approach in prior work and updated the text as specified below.

P. 18, lines 419-421: "In general, measurements are collected in duplicate at each time point, with a third measurement obtained when the first two differ by more than a pre-specified amount (which varied by specific measure and age)."

The cord blood collection rate (88%) is quite a bit less than the placenta collection rates (96%). Wondering if there are any consistent reasons for this difference that you have found?

Thank you for this question. We have two approaches to collecting cord blood. First we ask that the nursing staff collect cord blood directly from the cord at delivery. However this cannot always be obtained due to complications during the delivery or the nursing staff is unable to collect due to competing clinical duties. Second, we collect cord blood from the surface vasculature of the placenta. In that case, if there are any delays in collection and processing, the surface vasculature can drain (if the cord is not clamped) or clot. These difficulties in cord blood collection account for the lower success rate compared to overall placenta collection. We have added the text below to the discussion.

P. 22, lines 522-525: "The discrepancy between success in cord blood collection and placenta collection results from the more intensive immediate processing required for the former as even short delays can result in draining or clotting, making collection impossible."

References 17 and 19 require some formatting.

We have fixed the formatting.

Partners are noted in the demographics table, but are not discussed at detail throughout the study. Wondering why partners aren't a focus of this study? Perhaps a clear rationale for only including mothers could be provided.

This is an excellent question and indeed, we would have liked to study the partners more thoroughly. Based on our prior longitudinal studies of women recruited through the participating clinics, we knew that a moderate-high proportion would not have stable partners and that partners attended prenatal visits (at which all prenatal study activities including consent occurred) infrequently. Thus although we were interested in the role of partners, we elected to prioritize data and biospecimen collection from the women themselves as those would provide the most insight into the biological mechanisms of interest. The limited data on biological fathers was obtained from report by the mothers. We have added this as a limitation.

P. 24, lines 567-571: "Finally, although the biological and psychosocial contributions of partners is of great interest and relevance to children's development, our prior work in this population suggested that partner attendance at prenatal visits was likely to be low, making consent and data collection quite difficult. Thus like many pregnancy cohorts, our data on partners is limited to information provided by the participating women."

My biggest question or concern for this study is that, although the introduction specified maternal prenatal stress and the placenta as a key focus for this study, there is an inordinate amount of data being collected in addition to these variables with no real theoretical basis provided. Of course, one can assume there are a lot of hypothetical mechanisms or pathways for these variables to be included, but the rationale for inclusion is not clearly provided in any great detail throughout the paper. As a result, the study design comes across as a large "fishing expedition" for variables that could be of interest to developmental associations in childhood and beyond. The authors would need to address this clearly and succinctly for credibility and rigour. Although the research question is clear at the start of the paper, the inclusion of so many other variables makes this research question unclear.

We appreciate the reviewer's perception that the previous version did not sufficiently emphasize the core measures and hypotheses. We have done extensive revisions to clarify that there are focused aims that drove the study design (see pp. 8-9, lines 166-189). This section ("Primary Aims") further describes that UPSIDE was developed to meet the needs of several complementary research projects, thus further explaining the need for broad data collection. Each individual project has additional specific Aims customized to that project (not shown), however in the interests of brevity, we offer a high-level summary of the Aims guiding the main projects. In addition, our involvement with the NIH ECHO program (now described much earlier in the text), required collection of standard measures (beyond our specific aims) across all involved cohorts. Because this is the manuscript that will "introduce" our cohort to the scientific community, we wish to characterize data collection as comprehensively as possible even when it goes beyond the specific aims. Nevertheless, to reduce the appearance of an unfocused study we have: (1) more explicitly noted primary study aims; and (2) put greater emphasis on primary exposures and outcomes, while limiting discussion of other data points (including covariates). We hope this satisfies the reviewer's concern about the broad scope of data collection.

Reviewer: 3
Dr. Ines Gonzalez Casanova, Indiana University Bloomington, Emory University

Comments to the Author:
This manuscript describes an exciting new cohort designed to look at associations between maternal stress and child growth and development. The cohort is very relevant, the design seems strong and the follow-up so far has been high. However, the reporting of the work needs to be better organized and strengthened to better convey the many strengths of the cohort.

Thank you for your enthusiasm and constructive comments which we have addressed below.

More specific suggestions include:
- Include in the introduction the gap that this cohort is addressing. Why is it important to have another observational study of these associations?

We have added text in the introduction to more explicitly state the research gap that this cohort fills. As we noted in the original paper and as we have strengthened in this revision, key innovations of the paper include the detailed assessment of placenta structure and function and the collection of complementary mechanisms linking maternal distress to child health and development, i.e., stress physiology, sex steroids, inflammation. The statements of the "gaps" in the current literature are noted below:

P. 6, lines 128-133: "Although many current and past pregnancy cohort studies have examined the relationship between maternal psychosocial measures and child outcomes, few have gone beyond the HPA axis to examine additional biological pathways. Accordingly, a first major methodological and conceptual strength of the UPSIDE study is the assessment of biomarkers relevant to alternative pathways (e.g. cytokine profiles, steroidogenic activity) across pregnancy and in multiple biological sample types (e.g. maternal blood, cord blood, placenta)."

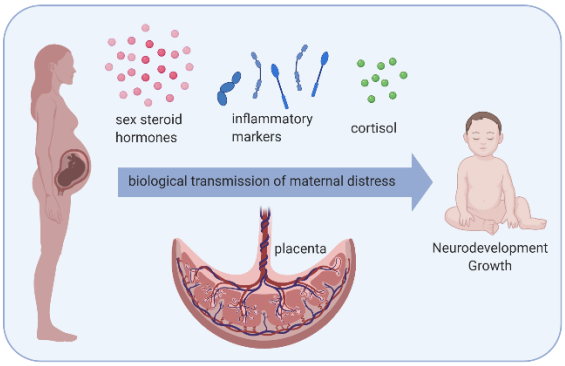
p. 7, lines 136-138: "Despite the placenta's critical role in transmitting maternal signals to the developing fetus, direct measurement of the placenta has been notably absent from the vast majority of studies on prenatal distress and child development."

p.7, lines 151-153: "What has been missing from this field are prospective pregnancy cohort studies that track mother-child dyads from early gestation through early childhood, while also collecting detailed placenta data."

What is the primary and secondary questions that this cohort will answer (or contribute to answer)? There are some suggestions of where this study is going in the text but it is never explicitly described. -Also in the introduction describe what are the specific mechanisms that this cohort will help elucidate. Maybe a diagram that includes the developmental programming model, the mechanistic outcomes (like the placenta) and then the functional outcomes like growth and development.

We thank the reviewer for pointing out the lack of clarity from the prior version. We have now better explained that there are multiple research grants that informed the study design and aims (one focused on sex steroids, the other on inflammation). We now explain this history to provide greater context and describe the main aims guiding the cohort (pp. 8-9, lines 166-189).

As suggested, we developed a Figure (below) to provide a high level overview of the study premise. We do not feel that it adds a lot to the text, but if the reviewers/editors feel it is important to include, we can add it.



Methods:

- Separate the different methods by timepoint. Sometimes it is hard to tell what was measured when.

We appreciate this comment and considered this organizational structure when we originally drafted the manuscript. However our emphasis on repeated measures meant that this structure would be extremely repetitive. Instead, we chose to illustrate the "schedule" of assessments in Table 3 so that the reader could visualize the activities by visit over time. We have now added a reference to that table in the introductory paragraph on child study activities (P. 15, lines 342-343) saying: "Activities conducted are displayed by visit timepoint in Table 3."

-Categorize measurements into outcomes (final and intermediate), exposures and confounders.

Given the quite large number of planned analyses to address the many Aims of the projects funded through this cohort, the same variable may plan multiple roles (outcome, exposure, confounder) depending on the analysis of interest. For instance, depending on the particular study question, maternal inflammatory markers may be the primary exposure or outcome, or even a confounder or mediator. Nevertheless, we appreciate the reviewer's point that more clarity is needed throughout the methods section of the manuscript, we have more clearly identified primary exposures of interest as well as the main child outcome measures (neurodevelopment and growth/physical development). We have removed extended discussion of the secondary measures and instead just briefly mention those constructs (with an appropriate reference) to avoid confusion.

Results -It was not clear to me which activities have already been conducted and which ones are planned but not finalized. It would be helpful to have that information at the beginning of the results. In the methods some are written in past tense but the studies have not been conducted.

Thank you for this comment. We appreciate the lack of clarity on this given that we are very much in the midst of conducting study visits. This is explained on p. 15 lines 339-341, which is now updated to read, "At present, all birth and 1 month visits have been completed, whereas 6, 12, 24, and 36 month visits are ongoing, and 48 month visits will start in early 2021." As such, we have referred to prenatal and birth visits in the past tense and child visits in the present tense. We have also indicated in Table 3 (Summary of Child Assessments), which visits are ongoing.

-A statistical analysis plan to address the primary study questions with power calculations should be included.

We have added a "Statistical analysis and power calculations" section which gives an overview of our proposed approach. As discussed in our responses to Reviewers 1 & 2, the UPSIDE cohort was developed to address the Aims of several funded projects and to include analysis plans that are comprehensive of even the primary aims is beyond the scope of this paper. We have provided exemplar power calculations to address this point and future data-driven papers will include greater detail on specific analytic strategies. (see p. 20; lines 462-486).

Future directions - the ECHO consortium should be described earlier, perhaps as part of the introduction and rationale for this cohort.

We now discuss participation in the ECHO consortium earlier as suggested and believe that it provides greater context for our broad data collection spanning multiple exposures and outcomes.

p. 8, lines 171-178: "Soon after, additional study activities were funded through the NIH's ECHO program the largest American study of early childhood health and development ever undertaken, with up to 50,000 other participating mother-child dyads from cohort studies around the U.S. (UG3/UH3OD023349). The ECHO funding allowed us to expand the contributions of the cohort to consider inflammatory mechanisms, extend child follow-up to age 4, and add a more intensive battery of outcome measures. Additionally, data and biospecimens from our study are harmonized with those of the other participating cohorts in order to address ECHO-wide scientific priorities."

In the future directions also main analyses could be discussed.

Thank you for this suggestion as this is indeed our most urgent future direction. We have added a sentence to the beginning of future directions (p. 24, lines 584-585) saying: "As data are cleaned and final outcome data become available, our highest priority is to address the primary study aims for the multiple projects that support this cohort.

VERSION 2 – REVIEW

REVIEWER	Prof. Antje Horsch University of Lausanne and Lausanne University Hospital, Switzerland
REVIEW RETURNED	10-Feb-2021

GENERAL COMMENTS	Many thanks for addressing all of my comments in a satisfactory way. I wish you good luck with this study and look forward to seeing the results published.