Check for updates

RESEARCH ARTICLE

REVISED Relationships between variance in electroencephalography relative power and developmental status in infants with typical development and at risk for developmental disability: An observational study [version 2; referees: 2 approved]

Andrew Hooyman[1], David Kayekjian[2], Ran Xiao[3], Crystal Jiang[2],
Douglas L. Vanderbilt [ID][4], Beth A. Smith [ID][2]

[1]Motor Behavior and Neurorehabilitation Laboratory, Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA, 90089, USA
[2]Infant Neuromotor Control Laboratory, Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA, 90089, USA
[3]Department of Physiological Nursing, School of Nursing, University of California San Francisco, San Francisco, CA, 94143, USA
[4]Department of Pediatrics, Division of General Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA, 90089, USA

Abstract

Background: Electroencephalography (EEG) is a non-invasive tool that has the potential to identify and quantify atypical brain development. We introduce a new measure here, variance of relative power of resting-state EEG. We sought to assess whether variance of relative power of resting-state EEG could predict i) classification of infants as typical development (TD) or at risk (AR) for developmental disability, and ii) Bayley developmental scores at the same visit or future visits.

Methods: A total of 22 infants with TD participated, aged between 38 and 203 days. In addition, 11 infants broadly at risk participated (6 high-risk pre-term, 4 low-risk pre-term, 1 high-risk full-term), aged between 40 and 225 days of age (adjusted for prematurity). We used EEG to measure resting-state brain function across months. We calculated variance of relative power as the standard deviation of the relative power across each of the 32 EEG electrodes. The Bayley Scales of Infant Development (3rd edition) was used to measure developmental level. Infants were measured 1-6 times each, with 1 month between measurements.

Results: Our main findings were: i) variance of relative power of resting state EEG can predict classification of infants as TD or AR, and ii) variance of relative power of resting state EEG can predict Bayley developmental scores at the same visit (Bayley raw fine motor, Bayley raw cognitive, Bayley total raw score, Bayley motor composite score) and at a future visit (Bayley raw fine motor).

Conclusions: This was a preliminary, exploratory, small study. Our results support variance of relative power of resting state EEG as an area of interest for future study as a biomarker of neurodevelopmental status and as a potential outcome measure for early intervention.

Open Peer Review

Referee Status: ✔ ✔

	Invited Referees	
	1	**2**
REVISED **version 2** published 15 Nov 2018	✔ report	
version 1 published 11 Sep 2018	? report	✔ report

1 **Wanze Xie**, Harvard Medical School, USA

2 **Caterina Piazza**, IRCCS Eugenio Medea, Italy

Discuss this article

Comments (0)

Keywords

infant development, electroencephalography, movement system, motor skills

Corresponding author: Beth A. Smith (beth.smith@usc.edu)

REVISED **Amendments from Version 1**

In response to the Referees' comments, we made changes to the manuscript to increase its clarity. All changes are described in detail in the responses to the Referees. To summarize here:

- We have added information to clarify procedures in the 'Data Analyses' section.
- We edited text in the 'Data Analyses' section to clarify statistical methods and moved some text from the Data Analyses section to the Results section.
- We added some additional limitations in the 'Limitations and Future Directions' section and made sure to clearly state: "It is also important to note that EEG is not a direct measure of cortical activity, so our proposal that higher variance may represent less organized cortical activity may or may not be valid. Future work that directly measures cortical activity is needed."
- We also added some additional suggestions for future research based on the Referee comments.

See referee reports

Introduction

Early detection of atypical neurological development increases the potential for successful intervention, as a body of basic science laboratory data supports that a wide variety of interventions, from environmental enrichment to hypothermia or implantation of stem cells, can enhance cerebral plasticity during development[1]. Emerging data also support that clinical interventions can increase the developmental potential of children, rather than presuming a predetermined potential[1]. Accordingly, early therapy intervention should have the greatest benefit on neural development and functional outcomes. However, there is a crucial roadblock here. In order to help guide and monitor interventions seeking to promote healthy brain development in the early years, we need suitable measures of fetal and infant brain function and development[2] prior to functional impairments emerging.

Electroencephalography (EEG) offers one non-invasive tool with the potential to identify and quantify atypical brain development. While EEG has been used since the early 1900s to diagnose conditions such as sleep and chronic seizure disorders, it has more recently been investigated as a screening tool in the neonatal intensive care unit for high-risk infant populations[3]. The rapidly growing field of infant EEG seeks to uncover specific abnormalities in activity patterns or key features, and whether these are predictive of short-term and long-term risks or outcomes[3].

Previous research has determined that EEG measures have some capacity in infancy to predict later functional outcomes. El-Dib and colleagues[4] demonstrated the ability of an EEG measure of continuity, minimum amplitude, bandwidth, and cycling within the first week of life to predict poor outcome (death or severe delay on Bayley Scales of Infant Development, version 2) at 4 months corrected age in 55 infants born pre-term (26–29 weeks gestational age) or with very low birth weight (less than 1500 g). For poor outcomes, EEG had a sensitivity of ~30%, specificity of ~90%, positive predictive value of ~60% and negative predictive value of ~80%[4]. They did not use cross validation to confirm accuracy of model. Hayashi-Kurabashi et al.[5] demonstrated relationships between background activity of EEG within the 36 days of life and a diagnosis of developmental delay or cerebral palsy at 12–18 months corrected age in 333 infants born pre-term (less than 36 weeks gestational age). For prediction of a later diagnosis, EEG had a sensitivity of 50–61%, specificity of 74–86%, positive predictive value of 27–38% and negative predictive value of 91–93%[5]. They did not use cross validation to confirm accuracy of the model. An additional study by Périvier and colleagues[6] also related clinical EEG data to infant outcomes. They found that out of 1744 preterm infants (less than 32 weeks gestational age), 422 had non-optimal outcomes at 2 years. A clinical rating scale that considered multiple aspects of abnormality of the EEGs performed in early infancy (up to 33 weeks post-menstrual age) had good specificity (0.95) but low sensitivity (0.16) for predicting non-optimal outcomes. Non-optimal outcomes were non-optimal neuromotor function or abnormal psychomotor development across any of a number of clinical measures[6]. Although EEG measures show some promise, to date they have only provided a piece of the puzzle. In a number of studies where outcomes were predicted using EEG it has been recommended that EEG assessment be combined with other clinical measures[4,6,7]. More effort is needed to determine the salient factors of EEG to be included for an optimally accurate and efficient prediction of neurodevelopmental outcomes, which led us to explore a new measure here.

We introduce a new measure here, variance of relative power of resting state EEG. We calculated variance of relative power as the standard deviation of the relative power across each of the 32 EEG electrodes. We postulate that higher variance may represent less organized cortical activity and be an intuitive and useful metric for identifying and quantifying atypical brain development within the first months of life. As such, higher variance may represent a salient factor of EEG to include for an optimally accurate and efficient prediction of neurodevelopmental outcomes.

Methods
Recruitment

This was a preliminary study to explore potential relationships of interest between EEG and developmental status, and we used a sample of convenience. Data were collected between 17 February 2015 and 18 June 2016. A total of 22 infants with typical development (TD) participated, between 38 and 203 days of age (Table 1). There were 2 infants with TD measured once, with the other 20 infants measured once per month for 3 to 6 visits. A total of 11 infants broadly at risk (AR) for developmental disability participated (6 high-risk pre-term, 4 low-risk pre-term, 1 high-risk full-term), aged between 40 and 225 days of age (adjusted for prematurity; Table 1). Infants AR were assessed once per month for 3 to 5 visits. Assessments started as close to 1 month of age as possible, and continued until the infant successfully reached and grasped a toy with high skill. Inclusion criteria (TD): infants were from singleton, full-term births (over 38 weeks). Exclusion criteria (TD): infants experiencing complications during birth, or with any known visual, orthopedic or neurologic impairment at the time of assessment, or with a score at or below the 5th percentile for their age on the

Table 1. Infant characteristics.

Visit	Group	Descriptive statistic	Age (in days, adjusted for prematurity as applicable)	Weight (kg)	Body length (cm)	Head circumference (cm)
1	At risk, full term	N	1	1	1	1
		Mean	168	6.3	64.5	39.7
		Std. Deviation				
		Minimum	168	6.3	64.5	39.7
		Maximum	168	6.3	64.5	39.7
	Preterm, high risk	N	6	6	6	6
		Mean	91	5.5	59.6	38.9
		Std. Deviation	35	0.9	4.6	1.7
		Minimum	68	4.5	53.5	37.5
		Maximum	162	7.1	67.5	42.0
	Preterm, low risk	N	4	4	4	4
		Mean	73	5.8	60.5	37.8
		Std. Deviation	38	0.8	4.9	2.1
		Minimum	40	4.9	56.3	35.0
		Maximum	127	6.8	67.5	40.0
	Typical development	N	22	22	22	22
		Mean	105	6.7	62.6	40.1
		Std. Deviation	44	1.0	3.5	1.9
		Minimum	38	5.0	57.5	37.5
		Maximum	203	8.9	69.5	43.6
3	At risk, full term	N	1	1	1	1
		Mean	230	7.6	68.0	40.2
		Std. Deviation				
		Minimum	230	7.6	68.0	40.2
		Maximum	230	7.6	68.0	40.2
	Preterm, high risk	N	6	6	6	6
		Mean	156	6.7	63.6	41.2
		Std. Deviation	35	0.8	3.8	1.5
		Minimum	132	6.0	59.0	40.0
		Maximum	225	8.1	70.0	44.0
	Preterm, low risk	N	4	4	4	4
		Mean	137	7.2	63.1	40.6
		Std. Deviation	37	0.7	1.9	1.1
		Minimum	103	6.3	61.0	39.0
		Maximum	190	7.8	65.0	41.2
	Typical development	N	20	20	20	20
		Mean	158	7.7	66.9	41.5
		Std. Deviation	34	0.9	3.8	1.6
		Minimum	97	6.2	56.0	38.0
		Maximum	203	9.1	71.5	45.0

Bayley Scales of Infant Development, 3rd edition[8] would have been excluded at the time of testing. Inclusion criteria (AR): AR infants were born before 36 weeks of gestation (low risk) or defined as at high risk for developmental delay per the definition of the state of California[9]. Infants AR were a broad group and not homogenous to one risk profile or level for this preliminary study. Exclusion criteria (AR): infants with unstable medical conditions would have been excluded. Infants were recruited by a member of the research team in-person at the Eisner Health Clinic (Los Angeles, CA, USA) and Children's Hospital of Los Angeles (Los Angeles, CA, USA). Infants were also recruited by referral from Ventura County Medical Center (Ventura, CA, USA), through fliers distributed or posted at the University of Southern California (USC), and by word of mouth. This study was approved by the Institutional Review Board of the USC (HS-14-00690). A parent or legal guardian signed an informed consent form prior to their infants' participation.

Assessment

Infants were measured primarily in the family's home, in the morning. Per the family's preference, three families came to the laboratory at the USC Health Science Campus for some of their visits. Each visit lasted for around 1 hour. At each visit, the infant's weight, body and limb lengths, and head and limb circumferences were measured. Motor, cognitive and language development were assessed using the Bayley Scales of Infant Development, 3rd edition[8]. A small wearable sensor was placed on each arm, and 5 minutes of video of the infant's spontaneous movement in supine was recorded. Wearable sensors remained on for the rest of the day. Wearable sensor data are reported in previous publications[10,11] and are not discussed further here. The parent or guardians' highest level of education completed was recorded. Families were compensated for each visit. Data were stored on a password-protected server or in a REDCap electronic database (version 6.14.2) hosted by USC.

Electroencephalography assessment

During each visit, EEG data were acquired using a Biosemi system with 32-electrode infant headcaps (standard 10/20 system) at sampling rate of 512 Hz. Infants sat on the lap of a caregiver. First, 2 trials of 20-second resting-state EEG data were recorded. During resting state recording, a lighted, spinning globe toy was presented out of participants' reach to attract their visual attention and minimize head and body movement. This is standard in infant EEG data collection[6,7]. Next, arm reaching skill was assessed using 20-second blocks where a toy was presented at midline within reaching distance of the infant alternating with 20-second blocks without a toy to reach for. This was repeated five times. Finally, another session of resting-state EEG data were collected, similar to the first session.

Data analyses

EEG analyses. EEG analysis methods are described in detail in a previous publication[12]. Only resting state EEG data were analyzed here, ranging from 14–82 seconds. Resting-state EEG variables explored here are individual power, relative power, and variance of relative power. Briefly, EEG data from all electrodes were re-referenced to the average of T7 and T8. Next, a bandpass infinite impulse response filter (0.3–30 Hz)

was applied to the re-referenced data. Resting EEG segments were epoched and noisy segments were rejected. After rejection, remaining EEG data from 11 infants AR and 22 infants with TD were: AR visit 1 = 11, AR visit 3 = 9, TD visit 1 = 21, TD visit 3 = 13. Power spectral density (PSD) was estimated on these preprocessed EEG data using the "pwelch" function in MATLAB (ver. 2016A, MathWorks Inc., Natick, MA, USA). PSDs were transformed into relative powers so that spectral activities from all individual sessions were directly comparable. The relative powers were calculated between 0 and 30 Hz. For each frequency bin within this range and each electrode, relative power was computed by dividing PSD by the sum PSD from all bins. Variance of relative power was calculated as the standard deviation of the 32 relative power measurements for each infant, calculated by taking the standard deviation of peak power across each channel.

Bayley scales of infant development. Bayley scales of infant development version 3 raw scores for gross motor, fine motor, expressive language, receptive language, and cognition were transformed into composite scores and percentile ranks by age corrected for gestational age less than 38 weeks for motor, cognitive and language domains. Bayley composite scores are determined in 2-week, age-normalized windows and created to have a range of 40–160, mean of 100 and SD of 15. Composite score classification are: 130 and above, very superior; 120–129, superior; 110–119, high average; 90–109, average; 80–89, low average; 70–79, borderline; 69 and below, extremely low[8]. An infant developing at a steady rate would be expected to have composite scores that remained steady over time.

Statistical analyses. Logistic regression was conducted to predict at-risk status of infants in the cohort using resting state EEG data recorded at visit 1. Leave-one-out cross-validation was performed as a method to confirm accuracy of logistic regression model. Multivariate linear regression was conducted to predict current (visit 1) and future (visit 3) Bayley scores using resting-state EEG data. Statistical analyses were performed using R, version 3.5.1. Bayley score models were compared using analysis of variance. It is important to note that the EEG analysis (RX) and the statistical analysis (AH) were performed independently from one another.

Prediction of AR status

The resting state data for each infant was derived into individual power and relative power readings from each electrode, 32 electrodes in all. Raw data are available on figshare[13]. Initially, all 32 power and relative measurements from visit 1 were input into various machine learning algorithms (including K-nearest Neighbor, Support Vector Machine, and Logistic Regression with L1 regularization) to predict the infant's at-risk status. Leave-one-out cross-validation was performed on each model. Then, the variance of relative powers across 32 electrodes were computed as input features for logistic regression to test their predictive efficacy for the classification task.

Prediction of same visit (1st visit) Bayley scores

Multivariate linear regression was conducted to predict current and future Bayley scores to identify if variance of relative power

made a significant contribution to prediction. We designed 12 different linear regression models with each one specific to a different category/composite of Bayley score (Table 2). First, we implemented models that only used age in days to predict each Bayley category. These models did not use variance of relative power as a predictor and thus served as the baseline models to be compared against the baseline models plus variance of relative power.

Each model was examined for assumptions of linear regression (i.e. heteroschedasticity and multicollinearity). Visual inspection of residuals and analysis of correlation between predictors revealed that each model maintained their regression assumptions. A baseline statistical model (a model that only included age in days and at-risk status) was compared to a nested model of the baseline model features plus variance of relative power to determine significant predictive effects of variance of relative power beyond baseline prediction. We used analysis of variance to determine significant predictive effects of variance of relative power across Bayley scores.

Prediction of future visit (3rd Visit) Bayley scores

A multivariate linear regression was conducted with age, at-risk status, and variance of relative power at visit 1 to predict Bayley scores at visit 3. On average, visit 3 took place 60 days after visit 1. The 3-regressor model using age, at-risk status, and variance of relative power was compared against a 2-regressor model using age and at-risk status only.

Results

Prediction of AR status

Leave-one-out cross-validation was performed on each machine learning model to predict at-risk status among 32 infants (11 at-risk) with a mean age of 90 days. Only modest accuracy was identified with typically a high false negative rate for features from conventional metrics (i.e., power and

relative power). On the other hand, variance of relative power was calculated as standard deviation of the 32 relative power measurements for each infant and was used as the only predictor within the model. A test of the full model (at-risk status ~ variance of relative power) compared to a baseline model (at-risk status ~ intercept only) was statistically significant, indicating that variance of relative power accurately classified at-risk status (chi square = 7.64, $p < 0.01$, $df = 2$, odds ratio = 1.18). Conversely, at-risk status significantly predicted variance of relative power ($p < 0.01$, $F = 8.33$, $R^2 = 0.217$, $df = 2$). A designation of at-risk was associated with higher variance of relative power. Interestingly, as shown in Figure 1, age in days did not predict variance of relative power ($p > 0.05$).

Leave-one-out cross-validation was performed using the identified logistic model to create a confusion matrix. Results demonstrated an overall accuracy of 75%, with a true negative rate of 86% (18/21) and a true positive rate of 55% (6/11). Results of the analysis demonstrated that an infant with higher variance of relative power across all EEG electrodes had a higher probability of being classified as AR (Figure 2).

Prediction of same visit (1st visit) Bayley scores

Results demonstrated that variance of relative power provided a significant contribution to 1st visit scores of Bayley raw fine motor, Bayley raw cognitive, Bayley total raw score, and motor composite score ($p < 0.05$, see Table 2).

Prediction of future visit (3rd Visit) Bayley scores

The 2-regressor model was significantly different from a baseline model ($p < 0.001$, $F = 15.61$, adjR$^2 = 0.58$, $df = 2$). Analysis of variance was used to compare the 2-regressor model to the 3-regressor model at alpha = 0.05. This result demonstrated that the addition of variance of relative power from visit 1 contributed to prediction of Bayley raw fine motor score at visit 3 ($p < 0.001$, $F = 14.13$, adjR$^2 = 0.65$, $df = 3$). Overall,

Table 2. Between-model statistical results.

Clinical test	Adj R^2 null	Adj R^2 full	BIC null	BIC full	ANOVA *p*-value
Bayley raw fine motor	0.59	0.64	159.89	157.89	0.03
Bayley raw gross motor	0.61	0.63	169.19	169.99	0.12
Bayley raw receptive language	0.11	0.13	111.22	112.99	0.22
Bayley raw expressive language	0.28	0.27	107.43	110.20	0.43
Bayley raw cognitive	0.65	0.71	164.44	161.42	0.02
Bayley total raw score	0.69	0.74	232.00	228.62	0.01
Cognitive composite	0.17	0.22	264.50	264.83	0.1
Cognitive percentile rank	0.06	0.07	277.45	279.45	0.25
Language composite	0.11	0.11	255.08	257.46	0.32
Language percentile composite	0.00	0.03	202.36	203.60	0.18
Motor composite	0.19	0.30	267.77	265.09	0.02
Motor percentile rank	0.09	0.17	308.41	307.94	0.06

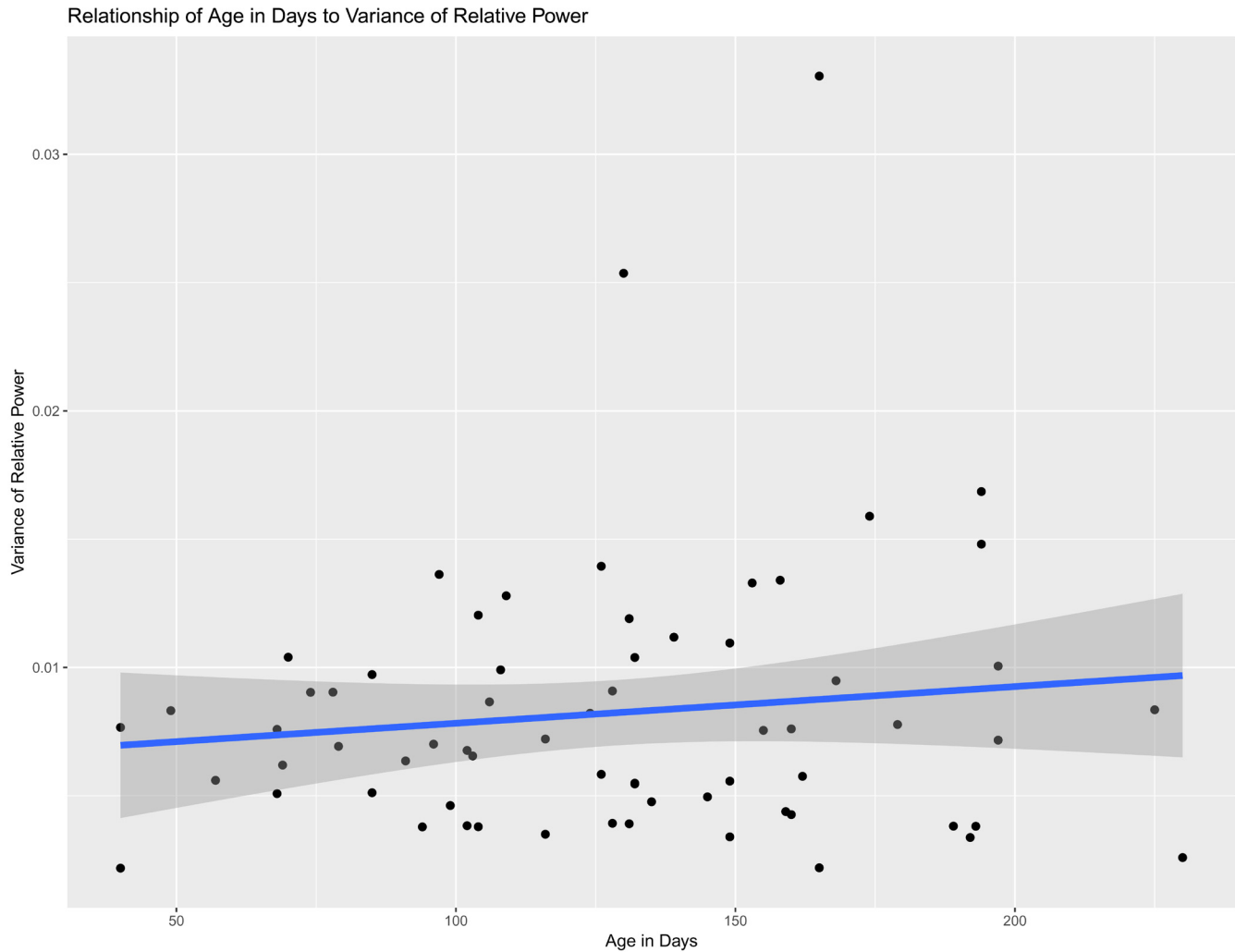Relationship of Age in Days to Variance of Relative Power

Figure 1. Age in days does not predict overall variance of relative power. Blue line represents linear regression fit, shaded area is standard error of fit.

variance of relative power was able to contribute an extra 7% of variance explained compared to a 2-regressor model using measures of age and at-risk status.

Discussion

Our main findings were: i) variance of relative power of resting state EEG can predict classification of infants as TD or AR, and ii) variance of relative power of resting state EEG can predict Bayley developmental scores at the same visit (Bayley raw fine motor, Bayley raw cognitive, Bayley total raw score, Bayley motor composite score) and at a future visit (Bayley raw fine motor).

Prediction of AR status

Higher variance of relative power predicted AR status, while age in days did not. We propose that higher variance may represent less organized cortical activity associated with an atypical trajectory of brain development. This is consistent with the use of 'EEG complexity' as a measure to distinguish infants with

TD from infants at high risk for autism spectrum disorders[14]. While age must certainly be considered—as a bias toward synaptic formation leads to a peak in synaptic density between 6–18 months of age, followed by a shift to synaptic pruning[15]— these studies imply that trajectories between populations of infants are diverging along the course of development. It is important to note that these studies both include infants who are at risk, without considering their ultimate outcomes (diagnoses). Further, we included both low- and high-risk infants in this study. We did not expect the AR infants to be a homogenous group with regards to their brain development and EEG data, rather we expected the AR infants to be different than the TD group, potentially in different ways across infants. Predicting or classifying risk status is not interchangeable with predicting future developmental outcomes/diagnoses.

Prediction of Bayley scores

Our results showed that variance of relative power provides a significant contribution to 1st visit (same visit) score prediction

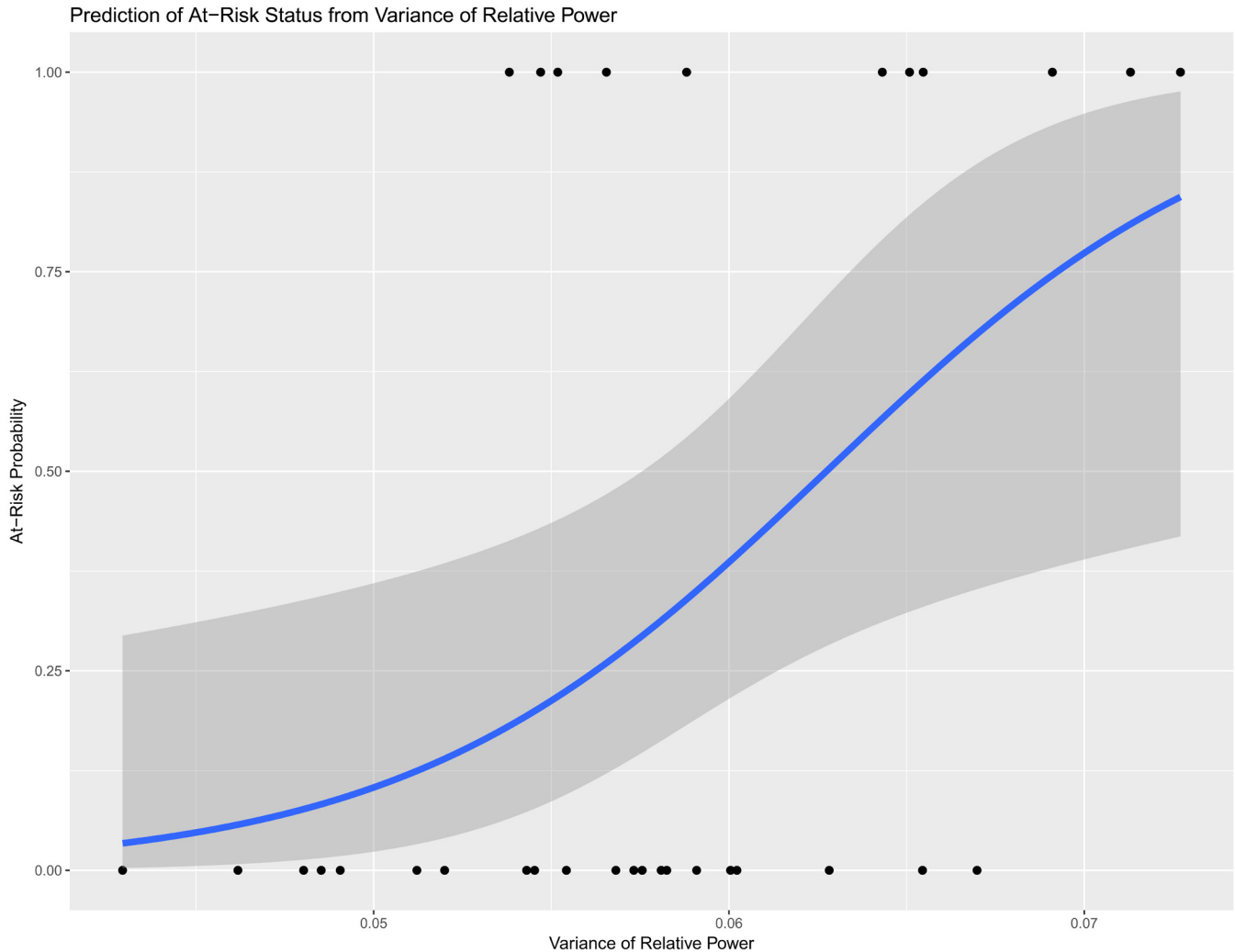Prediction of At−Risk Status from Variance of Relative Power

Figure 2. Infants with higher variance of relative power across all 32 electroencephalography electrodes had a higher probability of being classified as at-risk. Blue line represents logistic regression fit, shaded area is standard error of fit. Superior data points are variance of relative power measures for infants at risk, inferior data points are variance of relative power measures for infants with typical development.

of Bayley raw fine motor, Bayley raw cognitive, Bayley total raw score, Bayley motor composite score. Further, we found that variance of relative power from visit 1 contributes to prediction of Bayley raw fine motor score at visit 3. This is consistent with our previous work, where we found a relationship between a different measure, EEG coherence, and Bayley raw fine motor and gross motor scale scores in infants with TD (the same sample of infants with TD as included here)[12].

Previous research in infants with TD has also found relationships between EEG measures of power and coherence and motor and cognitive skill performance in infants. One study found a relationship between power in the alpha band and crawling onset in 5- to 7-month-old infants with TD[16]. Another study demonstrated differences in the power and coherence of EEG signals of 7- to 12-month-old infants with TD in relation to success with a

cognitive skill, the A-not-B task (object permanence). Infants who were successful displayed changes in frontal EEG power and increased anterior-posterior brain region coherence compared to infants who were not successful. The changes in EEG were attributed to increased organization and excitability in the frontal region[16]. The researchers also demonstrated differences in the power and coherence of EEG signals of 8-month-old typically developing infants with various amounts of crawling experience[6] and, recently, in 12-month-old typically developing infants with various amounts of walking experience[13].

Taken together, these studies link brain function, as measured by EEG, to motor and cognitive skill performance across various EEG measures and skills. Our study is unique as the infants here are younger than previous studies, and we have included infants AR in addition to infants with TD.

Limitations and future directions

This was a preliminary study in a small sample of infants. Our goal was to highlight potential relationships of interest to be pursued in future, larger, adequately powered studies. In an effort to avoid biased findings based on observations from a small data set, we conducted the EEG analysis and statistical analyses independently. In addition to the small sample size, our study is limited both by factors related to EEG as a tool and by factors related to studying infant development.

EEG as a tool has known limitations. EEG power is sensitive to non-neural factors like thickness and shape of tissues between electrodes and the cortex, as well as recording noise due to differences in hair thickness, the fit of the cap, or differing amounts of eye movement between participants. One way we addressed this was by using the relative power instead of the absolute power, another way was by showing that there were no systematic changes with age in overall variance (Figure 1). It is also important to note that EEG is not a direct measure of cortical activity, so our proposal that higher variance may represent less organized cortical activity may or may not be valid. Future work that directly measures cortical activity is needed.

There are many potential factors that likely influence developmental rate and outcomes in infants with TD and AR, and the same factor may or may not have similar effect strength in each group. Potential contributing factors to examine include: amount and type of movement experience, quality of caregiver–infant interaction, parenting style, cultural expectations, birth order, socioeconomic status, physical growth rate, nutritional status, amount and quality of sleep, personality/motivation, and genetics. Additionally, individual EEG predictors show limited power in predicting outcomes. There is the potential to aggregate these together as features to feed into machine learning algorithms for classification and prediction. We hope to pursue larger, more complex predictive models in future work with a larger sample. Adding EEG measures such as coherence and synchronization of oscillations might increase predictive power, so might including structural brain imaging data or clinic variables. Understanding the relative contribution of each factor to predicting outcomes, as well as their responsiveness to intervention, will be key to providing early intervention to reach optimal developmental potential in infants AR.

This was a preliminary, exploratory, small study of the potential importance of variance of relative power, as measured by resting state EEG data. Our results support variance of relative power as an area of interest for future study as a biomarker of neurodevelopmental status and as an outcome measure for intervention in infants AR. Higher variance may represent less organized cortical activity and be an intuitive and useful metric for identifying and quantifying atypical brain development within the first months of life. We see the potential to combine variance of relative power with other EEG and clinical measures identified in previous studies and to leverage these multiple features using machine learning techniques to improve predictive reliability.

Conclusions

Infant development is a variable and complex process. As a field, we are starting to determine how and when we can intervene in infants AR to have a positive impact on developmental outcomes. Our findings here, of the ability of variance of relative power of EEG to predict classification of infants as TD or AR and Bayley developmental scores, supports the potential of using variance of relative power of EEG to trace out and classify the developmental trajectories of the nervous system.

Data availability

A spreadsheet with resting state relative power EEG data and Bayley Scales of Infant Development Scales (version 3) scores for each participant at each assessment is available at figshare: https://doi.org/10.6084/m9.figshare.6994946[13].

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

References

1. Holt RL, Mikati MA: **Care for child development: basic science rationale and effects of interventions.** *Pediatr Neurol.* 2011; **44**(4): 239–53.
   PubMed Abstract | **Publisher Full Text**

2. Explorations GC: **Explore New Ways to Measure Brain Development and Gestational Age (Round 14).** Last accessed 08/22/18.
   Reference Source

3. Cilio MR: **EEG and the newborn.** *J Pediatr Neurol.* 2015; **07**(01): 025–43.
   Publisher Full Text

4. El-Dib M, Massaro AN, Glass P, *et al.*: **Early amplitude integrated electroencephalography and outcome of very low birth weight infants.** *Pediatr Int.* 2011; **53**(3): 315–21.
   PubMed Abstract | **Publisher Full Text**

5. Hayashi-Kurahashi N, Kidokoro H, Kubota T, *et al.*: **EEG for predicting early neurodevelopment in preterm infants: an observational cohort study.** *Pediatrics.* 2012; **130**(4): e891–7.
   PubMed Abstract | **Publisher Full Text**

6. Périvier M, Rozé JC, Gascoin G, *et al.*: **Neonatal EEG and neurodevelopmental outcome in preterm infants born before 32 weeks.** *Arch Dis Child Fetal Neonatal Ed.* 2016; **101**(3): F253–9.
   PubMed Abstract | **Publisher Full Text**

7. Shalak LF, Laptook AR, Velaphi SC, *et al.*: **Amplitude-integrated electroencephalography coupled with an early neurologic examination enhances prediction of term infants at risk for persistent encephalopathy.** *Pediatrics.* 2003; **111**(2): 351–7.
   PubMed Abstract | **Publisher Full Text**

8. Bayley N: **Bayley Scales of Infant Development: Third Edition.** San Antonio, TX: Pearson; 2005.
   Reference Source

9. California Code of Regulations: **Eligibility Criteria: Established Risk and High Risk for Developmental Delay.** Last accessed 08/21/18.
   Reference Source

10. Trujillo-Priego IA, Lane CJ, Vanderbilt DL, *et al.*: **Development of a Wearable Sensor Algorithm to Detect the Quantity and Kinematic Characteristics of Infant Arm Movement Bouts Produced across a Full Day in the Natural Environment.** *Technologies (Basel).* 2017; **5**(3): pii: 39.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Shida-Tokeshi J, Lane CJ, Trujillo-Priego IA, *et al.*: **Relationships between full-day arm movement characteristics and developmental status in infants with typical development as they learn to reach: An observational study [version 1; referees: 3 approved].** *Gates Open Res.* 2018; **2**: 17.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Xiao R, Shida-Tokeshi J, Vanderbilt DL, *et al.*: **Electroencephalography power and coherence changes with age and motor skill development across the first half year of life.** *PLoS One.* 2018; **13**(1): e0190276.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Hooyman A, Kayekjian D, Xiao R, *et al.*: **Relative Power EEG and Bayley Scales Infant data.** *figshare.* Dataset. 2018.
    http://www.doi.org/10.6084/m9.figshare.6994946.v1

14. Bosl W, Tierney A, Tager-Flusberg H, *et al.*: **EEG complexity as a biomarker for autism spectrum disorder risk.** *BMC Med.* 2011; **9**(1): 18.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Huttenlocher PR, Dabholkar AS: **Regional differences in synaptogenesis in human cerebral cortex.** *J Comp Neurol.* 1997; **387**(2): 167–78.
    PubMed Abstract | **Publisher Full Text**

16. Xiao R, Qi X, Patino A, *et al.*: **Characterization of infant mu rhythm immediately before crawling: A high-resolution EEG study.** *NeuroImage.* 2017; **146**: 47–57.
    PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Referee Status: ✔ ✔

Version 2

Referee Report 22 November 2018

https://doi.org/10.21956/gatesopenres.13984.r26765

✔ **Wanze Xie**

Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

The authors have addressed my concerns in their revision.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 01 October 2018

https://doi.org/10.21956/gatesopenres.13952.r26655

✔ **Caterina Piazza**

Bioengineering Lab, Scientific Institute, IRCCS Eugenio Medea, Lecco, Italy

The manuscript describes an interesting preliminary study proposing a new EEG measure (i.e. variance of relative power of resting state EEG) as an early biomarker for the prediction of the developmental status. Results seem to suggest the potential of using this measures to evaluate developmental trajectories, since it was able to classify infants with typical development (TD) and at risk(AR) infants and it predicted some Bayley scores both at the same time of the EEG recording and 3 months later.

The study is clearly and accurately presented. However, some specific comments that should be addressed by the authors are listed below.

1. The variance of relative power was calculated by taking the standard deviation of peak power in the frequency band 0-30Hz. The position of this peak should be reported, I expected that in most cases it was found in the low frequencies. Why did you decide to use this broad band? Did you try to calculate the same measure in the different EEG frequency bands (i.e. theta, alpha beta, etc.).

2. It is not clear if the resting-state EEG measures you took into account (i.e. individual power, relative power and variance of relative power) were computed on all the resting state trails you recorded (i.e. the 2 trails at the beginning of your experiment and the trial at the end of the procedure).

The variance of relative power seem to be a promising biomarker. However I think that your hypothesis that higher variance may represent less organized cortical activity should be better investigated. EEG scalp-based data measure surface potential changes caused by a combination of underlying signals from various sources within the brain, as well as extra-brain sources. In my opinion there are a lot of factors that may potentially contribute to higher variance at the scalp level.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Neuroscience, EEG

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 07 Nov 2018
Beth A. Smith,

Thank you for your insightful review. In response:

1. We have looked at power variance in the individual frequency bins and peak power is typically within theta. One analysis looking at the alpha frequency was also predictive of 3rd visit scores using 1st visit resting state EEG. We did try prediction within each frequency brand specifically and we found looking across the spectrum to be the most robust.

2. The resting state data came from both periods, to include as much clean data as possible. We have added a sentence in the data analyses section, Only resting state EEG data were analyzed here, ranging from 14-82 seconds". The first part of the sentence addresses a concern from another Reviewer.

Final comment: We agree! We have been careful to say that higher variance in EEG power may represent less organized cortical activity, as opposed to state that it does represent less organized

cortical activity. We have added the possibility of noisier data due to extensive eye movements for some infant participants in the limitations section. We also added a sentence in the limitations section, "It is also important to note that EEG is not a direct measure of cortical activity, so our proposal that higher variance may represent less organized cortical activity may or may not be valid. Future work that directly measures cortical activity is needed."

Competing Interests: No competing interests were disclosed.

Referee Report 24 September 2018

? **Wanze Xie**

Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

The manuscript describes an EEG-study investigating the prediction of developmental status in infants (TD or AR for developmental disability) and their cognitive outcomes (Bayley scores) using variance in EEG power across electrodes. There were 22 typical developing infants and 11 infants broadly at risk for developmental disability included in the EEG analysis. The variance of EEG relative power across 32 electrodes was calculated using resting-sate EEG data. A logistic regression model was used to predict infants' AR status, and multivariate linear regressions were conducted to predict the effects of variance in EEG power on infants' concurrent and future Bayley scores. The authors found that variance of EEG power can classify infants' developmental status and predict the outcomes/scores of a few Bayley subscales.

There is accumulating evidence supporting the capacity of EEG measures to predict functional outcomes in childhood, and thus it is important to seek for efficient and suitable EEG metrics to optimize the prediction. While this work has the potential to advance the field further, there are several issues requiring revision or clarification.They are listed below in the order of appearance in the manuscript.

1. Why does higher variance in EEG power represent less organized cortical activity? This assumption has been made in the introduction and discussion without solid evidence supporting it. Please either cite studies for this assumption or analyzing the network organization to examine the association between variance in EEG power and cortical organization. This is important because other factors can also contribute to higher variance of EEG power across electrodes, such as noisier data in the frontal electrodes due to extensive eye movements for some infant participants.

2. What is the rationale for using variance of relative EEG power to predict developmental status in infants, as well as their cognitive outcomes? Is it because this measure has been tested and validated in the adult literature, or it is just one of the few measures tested by the authors that worked the best?

3. The presentation of the EEG paradigm could be clearer. My understanding is that 2 trials of 20s resting-state EEG data were recorded, followed by a couple of "arm reaching" trials, and then an additional 2 resting-state trials were presented. Is this correct? If so, did the authors use all the data (2 RS + the Arm R trials + 2 RS?) or just the 4 resting-state trials?

4. What is the length of the epochs?

5. It is stated that the variance of EEG relative power was calculated by taking the standard deviation of the peak power of the entire frequency band (0 – 30 Hz) across the channels. I wonder if it would make sense to use the variance of EEG power for a certain frequency band (e.g., theta, alpha, beta) to control for the potential effect of artifacts (e.g., movements) on the high frequency bins. Given the PSD distribution for infants at this age, is it likely that the peak power will always reside in the low frequency bins in the theta band if there is no artificial effect on the higher frequency bins?

6. What is the attrition rate of the current EEG analysis?

7. Under the section of "prediction of AR Status", machine learning results were reported. They should belong to the results section.

8. It seems to me that the authors tested a few machine learning algorithms using the EEG raw and relative power as predictors. Selection among these machine learning algorithms with the full dataset can render circular an otherwise appropriate analysis and "the best" results, which is called "circular analysis" or "double dipping" – the use of the same data for selection and data analysis. There are a few issues associated with circular analysis (Kriegeskirte et al., 2009) and neuroscientists tend to avoid spurious effects related to double dipping by using separate datasets for model selection and testing. In your situation, if using EEG variance of power and machine learning approaches to predict cognitive outcomes is one of the ultimate goals, I would recommend you select a machine learning algorithm (e.g., SVM or K-nearest Neighbor) based on a portion of the dataset (e.g., a few subjects) and then only apply this algorithm to the rest (or all) of the data in your future research.

9. Why did the authors use a logistic regression for variance of power but machine learning approaches for raw and relative power?

10. There is typo at the beginning of the section "Prediction of same visit (1st visit) Bayley scores. "Multivariable linear regression". Should it be "multivariate linear regression" or "multiple linear regression"?

11. Page 5, "Corresponding null models to null models plus variance of relative power … Bayley scores." This sentence is confusing to me. Please clarify.

12. In table 2, do those ANOVA p-values represent the significance of the whole multivariate regression model, or the t-test p-values for one independent factor, i.e., variance of relative power? My understanding is the latter, but please clarify it.

13. Many analyses were done with these data (N >= 24?). Is the .05 significance level appropriate? Should it be adjusted?

14. Does "at-risk status" also predict concurrent or future Bayley scores? It would be great to see the results and discussion for this factor.

15. Why is variance of EEG power predictive of some concurrent subscales of Bayley (e.g., raw fine motor, raw cognitive score, and total raw score) but not the others? Why is it only predictive of future Bayley raw fine motor score? This would be an interesting point for further discussion.

16. This is a perplexing situation. On Page 8, I agree with the authors' statement that individual EEG predictors may have limited power in predicting cognitive outcomes for infants and children. Then why not use a combination of EEG measures (e.g., power, variance, coherence, synchronization of oscillations, etc.) to predict the outcomes and infants' developmental status, as some of these variables are already available to the authors? Will aggregate these measures in one model result in the highest prediction accuracy? Does any of the measures outperform the other measures? I feel these are important questions relevant to the current study and should be tested or discussed.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 07 Nov 2018
Beth A. Smith,

Thank you for your insightful review. In response:

1. We appreciate this point and have been careful to say that higher variance in EEG power may represent less organized cortical activity, as opposed to state that it does represent less organized cortical activity. We have added the possibility of noisier data due to extensive eye movements for some infant participants in the limitations section. With such factors in mind, we carefully designed the experiment, e.g., introducing a spinning globe, to attract subjects' attention, and rejected eye movement artifacts during preprocessing steps using independent component analysis to minimize the effect from EEG artifacts. We also added a sentence in the limitations section, "It is also important to note that EEG is not a direct measure of cortical activity, so our proposal that higher variance may represent less organized cortical activity may or may not be valid. Future work that directly measures cortical activity is needed."

2. This is an exploratory study to shed light on new potential metrics in EEG that could serve as features to capture early risk stratification and outcome prediction for infants at risk of developmental disability. Out of the available data, that being power and relative power of each electrode, variance of relative power was the measure that worked best.

3. The reviewer is correct regarding the study design. Only resting state data were analyzed here. We have added a sentence in the data analyses section, Only resting state EEG data were analyzed here, ranging from 14-82 seconds".

4. We have added a sentence in the data analyses section, Only resting state EEG data were analyzed here, ranging from 14-82 seconds".

5. We have looked at power variance in the individual frequency bins and peak power is typically within theta. One analysis looking at the alpha frequency was also predictive of 3rd visit scores using 1st visit resting state EEG. We did try prediction within each frequency brand specifically and we found looking across the spectrum to be the most robust.

6. A total of 22 infants with TD and 11 infants AR participated. After rejection, remaining EEG data were: AR visit 1 = 11, AR visit 3 = 9, TD visit 1 = 21, TD visit 3 = 13. We have added information to a sentence in the "Data analyses" section so that it now reads, "After rejection, remaining EEG data from 11 infants AR and 22 infants with TD were: AR visit 1 = 11, AR visit 3 = 9, TD visit 1 = 21, TD visit 3 = 13."

7. We deleted the text "and only modest accuracy was identified with typically a high false negative rate" from the methods section and added the sentence "Leave-one-out cross-validation was performed on each machine learning model to predict at-risk status among 32 infants (11 at-risk) with a mean age of 90 days. Only modest accuracy was identified with typically a high false negative rate for features from conventional metrics (i.e., power and relative power)" to the results section.

8. We understand your concern, but we do not feel the initial use of machine learning on the entire data set would be considered double dipping. This is an exploratory study and the data set here was generated for the purposes of selecting a machine learning/statistical model. Future data collections of this work would use the same data analysis to further confirm and add to the existing data set as an independent set. We are really only advocating that variance of power is a variable of interest to be pursued.

9. We tested various machine learning algorithms including logistic regression, one of the simplest linear models, using conventional EEG metrics (power and relative power) as input features for the AR classification task. We have updated the description in our methods and results to clarify this. All models yielded poor performance with high false negative rate for these EEG metrics. On the other hand, with the proposed variance of relative power, the simple logistic regression already demonstrated improved accuracy over the other EEG metrics, for the classification task.

10. Thank you for catching this, we have corrected it as "multivariate linear regression".

11. We have clarified this statement as follows: "A baseline statistical model (a model that only included age in days and at-risk status) was compared to a nested model of the baseline model features plus variance of relative power to determine significant predictive effects of variance of

relative power beyond baseline prediction. We used analysis of variance to determine significant predictive effects of variance of relative power across Bayley scores."

12. The ANOVA p-values represent the probability that the more complex multivariable linear regression model, baseline plus variance of relative power, is significantly better than the baseline model alone. Specifically, the p-value is from the computation of an F-test between the two models to determine if the sum of squares for the more complex model is significantly different than the simpler model. This is only appropriate when the complex model is a nested version of the simple model, which is what we have done here.

13. No. The many analyses here were each is done independently of one another. One test for each outcome variable, one dependent variable to each independent variable. A case for alpha value adjustment, such as a Bonferroni correction, to ensure an adequate false positive rate, would be in the case of multiple comparison where it is multiple independent variables for one dependent variable. For example, multiple t-tests to determine if several independent variables are significantly different than the dependent variable.

14. At-risk status does not predict concurrent Bayley scores but it does predict future (3rd visit) Bayley score of raw receptive language, $p < .05$, $R^2 = .18$.

15. We agree with the Reviewer that this is a very interesting consideration. Perhaps movement requires a more *stable* brain than other measures? Or the movement measures are more sensitive of brain behavior than the other tests? We are really not comfortable speculating on this in the manuscript, though, as this is such an exploratory study.

16. We hope to explore this in future work. We have added in the Limitations and Future Directions section, "We hope to pursue larger, more complex predictive models in future work with a larger sample. Adding EEG measures such as coherence and synchronization of oscillations might increase predictive power, so might including structural brain imaging data or clinic variables."

Competing Interests: No competing interests were disclosed.