Received: 26 June 2024 | Revised: 12 November 2024 | Accepted: 8 November 2024

DOI: 10.1111/psyp.14733

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

# Heterogeneity in pediatric resting EEG data processing and analysis: A state of the field

**Sonya V. Troller-Renfree**[1] | **Santiago Morales**[2] | **George A. Buzzell**[3,4] | **Aislinn Sandre**[5]

[1]Department of Human Development, Teachers College, Columbia University, New York, New York, USA

[2]Department of Psychology, University of Southern California, Los Angeles, California, USA

[3]Department of Psychology, Florida International University, Miami, Florida, USA

[4]Center for Children and Families, Florida International University, Miami, Florida, USA

[5]Department of Biobehavioral Sciences, Teachers College, Columbia University, New York, New York, USA

**Correspondence**
Sonya V. Troller-Renfree, Department of Human Development, Teachers College, Columbia University, 525 W. 120th St., New York, NY, 10027, USA.
Email: svt2110@tc.columbia.edu

**Funding information**
National Institute of Child Health and Human Development, Grant/Award Number: R00HD104923

**Abstract**
Developmental, resting electroencephalography (EEG) is gaining rapid popularity with implementation in large-scale studies as well as a recent WHO report naming resting EEG as a gold standard measure of brain health. With an increased interest in resting EEG as a potential biomarker for neurocognition, it is paramount that resting EEG findings are reliable and reproducible. One of the major threats to replicability and reproducibility stems from variations in preprocessing and analysis. One of the primary challenges facing the field of developmental EEG is that it can be challenging to acquire data from infants and children, which commonly makes data cleaning and analysis difficult and unstandardized. The goal of the present manuscript is to take a state of the field of the methods experts in resting EEG report they would use to clean and analyze a hypothetical data set. Here we report on the responses of 66 self-identified experts in developmental psychophysiology, none of which submitted identical preregistrations. As expected, there were areas of more and less consensus, but ultimately, we believe our findings highlight opportunities for core methodological work and field-level efforts to establish consensus.

**KEYWORDS**
baseline, child development, EEG, preregistration, resting

## 1 | INTRODUCTION

Developmental resting electroencephalography (EEG) is a method of measuring a child's brain activity while "at rest," which is thought to convey information about a child's neurodevelopment. Broadly, resting EEG is generally recorded while an infant or child sits with a caregiver or in a chair and watches a neutral stimulus. As children become older, it is also common to assess resting EEG under eyes-open and eyes-closed conditions. Following these recordings, brain activity is often quantified in terms of the magnitude of brain oscillations (power) in different frequency bands (denoted by Greek letters: delta, theta, alpha, beta, and gamma)— which are thought to be functionally significant for neurocognition (Begus & Bonawitz, 2020; Bell & Cuevas, 2012; Saby & Marshall, 2012; Tan, Troller-Renfree, et al., 2023; Troller-Renfree et al., 2023). Resting EEG measures in children are commonly cited as valuable early markers of learning and cognition (e.g., language, executive function, IQ); with many studies showing prospective associations with neurocognition (Benasich et al., 2008; Brito et al., 2016; Gou et al., 2011; Tan, Tang, et al., 2023; Wilkinson et al., 2019) and mental health (Brito et al., 2019; McLaughlin et al., 2010; Troller-Renfree et al., 2017). Furthermore, early resting EEG measures have been shown to be sensitive to early experiences known to have an outsized impact on neurodevelopment, such as maternal stress (Pierce et al., 2019; Troller-Renfree et al., 2020, 2023), caregiving disruption (Debnath, Tang, et al., 2020; Marshall et al., 2004; Vanderwert et al., 2016), and socioeconomic status (Brito et al., 2016; Otero et al., 2003; Sandre et al., 2024; Tomalski et al., 2013). Even more promising, early resting brain activity appears to be sensitive to early interventions, perhaps providing a possible marker for intervention success (Debnath, Tang, et al., 2020; Rockers et al., 2023; Troller-Renfree et al., 2022). With the immense promise and possible utility of early resting EEG measures, resting EEG recordings are being incorporated into some of the largest investigations of child development, including the Environmental Influences on Child Health Outcomes (Blaisdell et al., 2022) and Healthy Brains and Child Development (Norton et al., 2021; Fox et al., 2024), as well as within mobile applications (Troller-Renfree et al., 2021) and low-resource contexts (Jensen et al., 2021; Rockers et al., 2023)

One of the primary challenges facing the field of developmental EEG is the difficulty of acquiring data from infants and children, which often leaves developmental researchers with difficult and messy data to clean and process. As a result, the field of developmental EEG commonly needs to modify tried-and-true neuroscientific standards used in adult EEG processing to obtain usable data from pediatric populations. These concessions occur at every level, including stimuli creation, task programming and adaptation, recording environments, and recording durations, but nowhere is this more evident than in EEG data cleaning and analysis. Infant and child EEG data are commonly plagued by unsystematic artifacts, which are difficult, if not impossible, to completely remove during data cleaning, which can lead to substantial data loss (van der Velde & Junge, 2020). Adding complexity, data patterns and properties of real EEG signals tend to change as the brain develops (e.g., alpha peak movement, changes in absolute power as fontanelle closes). Developmental EEG researchers must strike a delicate balance between removing artifacts and retaining enough data for reliable analyses. The tension between data retention and quality has led to various data cleaning and analysis techniques tailored or unique to developmental EEG. While the genesis of many such practices (e.g., hand editing or private, customized processing pipelines) is in pursuit of asking good scientific questions with high-quality data, they also threaten the reproducibility and replicability of developmental EEG work (Lopez et al., 2023). Towards the long-term goal of increasing convergence in developmental EEG methods, the goal of the present manuscript is to assess the current extent of variation in the methods preferred by developmental EEG researchers. To this end, we sought to characterize the current state of the field by reporting the results of a survey assessing the preferred processing and analytic methods of current developmental EEG experts.

The field of developmental EEG has historically seen the publication of relatively less methodological work (compared with work with adults), with less consensus on analytic approaches. In adult research, there have been a number of articles aimed at establishing field-level consensus in EEG processing (e.g., Babiloni et al., 2020; Keil et al., 2022), which has been less true in the developmental literature. However, many efforts are underway to improve analytic techniques, reliability, and replicability in developmental EEG (Buzzell et al., 2023). The last decade of work has seen an uptick in the publication of developmentally focused, standardized preprocessing pipelines for larger and smaller electrode montages, such as MADE (Debnath, Buzzell, et al., 2020; Troller-Renfree et al., 2021), HAPPE (Gabard-Durnam et al., 2018; Lopez et al., 2022), and APICE (Fló et al., 2022), as well as some methods-based work in areas such as test–retest reliability (Lopez et al., 2023), ICA-based artifact removal (Leach et al., 2020), time–frequency work (Morales & Bowers, 2022), and mobile EEG (Troller-Renfree et al., 2021). However, despite these advances, there still seems to be little consensus in cleaning, analytic, and statistical approaches across studies. This lack of consensus spans a variety of settings, including cleaning settings (e.g., epoch lengths, artifact thresholds), but also includes ways of summarizing the data (e.g., frequency band boundaries), data transformations, statistical test selection, different ways of estimating measures (e.g., types of EEG power), and corrections for multiple comparisons.

Reproducibility and replicability in neuroscience studies have historically been low (Botvinik-Nezer et al., 2020; Poldrack et al., 2017). Neuroimaging (fMRI) research suffers from high levels of false-positive results (Poldrack et al., 2017) partly because neuroscientific analyses often have a high degree of flexibility in analysis methods (Botvinik-Nezer et al., 2020). Such issues are also applicable to EEG, given the similarly high degree of flexibility in analysis methods (e.g., filter parameters, epoch overlap, band definitions, and power transformations). There have been many proposed solutions to these problems, including preregistration, registered reports, and establishing field-level consensus (Garrett-Ruffin et al., 2021; Poldrack et al., 2017). However, while there is an increased availability of preregistered reports in journals, many preregistrations do not include their data cleaning and analysis plans.

The present study aims to assess the current degree of consensus/disagreement in preprocessing and data analysis approaches for the field of developmental resting EEG. We surveyed self-identified experts in developmental EEG and asked them to preregister a processing and analysis stream for the same hypothetical data set. By identifying areas where consensus currently does and does not exist, we hope this endeavor will spark collaborative efforts to establish consensus in areas where researchers currently disagree, ultimately, increasing the reproducibility and reliability of the field. Moreover, we hope these efforts will inspire much-needed methodological work elucidating the impact of data processing and analytic practices on developmental EEG data.

## 2 | METHODS

### 2.1 | Participants and participant inclusion

The present study aimed to survey researchers who self-identify as experts in developmental EEG processing. In total, 102 researchers who self-identified as meeting our eligibility criteria consented to participate in the EEG survey. Participants were recruited via social media posts, personal emails, advertisements at developmental conferences, word of mouth, and asking survey participants for collaborators. To ensure adequate expertise, we required participants to have a career level of at least doctoral trainee.

Of those who consented, 24 participants dropped out of the study before completing any EEG-related questions, leaving 78 participants who completed at least one EEG-related question. Of these 78 participants, we removed responses from participants who did not provide valid responses to at least 25% of the EEG-related questions. This resulted in the exclusion of an additional 12 participants. In sum, we were left with 66 participants who were included for analysis. While participants could complete the survey anonymously or enter a false identity, most did not. For participants who volunteered their personal information, our team was able to confirm that 89% (59 out of 66) of respondents were individuals currently engaged in developmental EEG research based on publicly available information.

Of the 66 participants included in our final analytic sample, 32% of participants identified as male, 65% identified as female, and 3% identified as other. For ethnicity, 8% of participants identified as Hispanic or Latino/a, and 83% identified as not Hispanic or Latino/a. Additionally, 6% of participants identified with ethnicities other than Hispanic or Latino/a, and 3% chose not to report their ethnicity. A total of 12% of participants identified as Asian, and 85% identified as White. The remaining experts identified with other racial groups or preferred to keep their racial background private. In terms of career stage, the distribution was as follows: 30% were doctoral students, 24% were postdocs, 8% were research or staff scientists, 2% were adjunct professors, 17% were assistant professors, 14% were associate professors, and 6% held the position of full professors. The range of published resting pediatric EEG articles was from 0 ($N=8$) to more than 30 ($N=3$), with an average of more than five reported publications per expert.

All research activities were approved by the Institutional Review Board at Teachers College, Columbia University.

### 2.2 | EEG survey design

The EEG portion of the survey consisted of two parts (See Appendix A for a full survey). First, participants were asked to preregister their preprocessing and analysis parameters for a low-density data set. Next, they were asked what they would change for a high-density data set (see supplemental materials for exact text, and parenthetical text below for high-level details). The prompts were as follows:

> Great News! The EEG Fairy has dropped a data set on your doorstep! Your job is to analyze these data to the best of your ability. It turns out these data are from a Randomized Control Trial (RCT) with two groups—a Treatment group and a Control group. You will be examining group-level differences as well as individual differences using these data in the Theta, Alpha, Beta, and Gamma bands. However, to analyze these data, you have to preregister your entire processing and analysis stream. Here is what you know about the data set:

1. Participants: 600 infants split between two groups.
2. Age of participants: 12 Months.
3. Sampling Rate: 500 Hz.
4. Recording Length: 5 min.
5. Number of Electrodes: 18, plus 2 mastoid references (High density: 124 [128-electrode cap with no eye/face electrodes])

### 2.3 | Data cleaning

Responses were largely left untouched. However, in a few instances, experts indicated ranges (e.g., 1–1.5 Hz low-pass filter) or added some additional information (e.g., "...it depends, but I usually use a low-pass of .1 Hz"). When this was the case, we would take the midpoints of provided ranges or the most concretely provided value. Furthermore, when this procedure did not make sense, intent was confirmed via consensus interpretation among authors.

For the high-density data set, we asked participants whether they would change their responses from the low-density data set. For each question, we provided the option of "I don't remember what I said for the low-density dataset" and then requested the settings they would use. During data cleaning, if participants indicated they would not change their response from the low-density data set, we populated their high-density responses with their responses to the low-density data set.

## 3 | RESULTS

In the following section, we go over each processing step and what participants indicated they would preregister. Results are presented in two different categories: Preprocessing and Analyses. All reported percentages were rounded to the nearest whole number, which sometimes results in cumulative responses summing to 101% due to rounding differences. All percentages were calculated as the total number of responses in a category divided by the number of valid responses (items left blank by respondents were removed from the denominator). Generally, we indicate what participants preregistered for the low-density data set. Responses to the high-density data set were largely consistent except in a few areas; given space constraints, results of the high-density data set are only noted briefly in the main text (below) and are more thoroughly reported in Supplement S1. Furthermore, to conserve space, some areas of the low-density data are discussed in more detail in Supplement S2. Throughout, we report consensus among expert. For items that were open response or had more than two categories consensus was as follows: low consensus (<35%), moderate consensus (35%–69%), and high consensus (70%–100%) of consensus. For items that were dichotomous (besides don't know and prefer not to answer), consensus was as follows: low (<60%), moderate (60%–75%), high (75%–100%). Finally, Figures 1–4 provide visual companions to many of the findings discussed below, but they also provide precise counts and percentages of responses.

### 3.1 | Preprocessing parameters

#### 3.1.1 | Filtering

For filtering, we had high consensus with 92% of our experts indicating they would filter their data as part of preprocessing. When asked about high-pass filter settings, reported ranges were between 0.01 Hz and 1–1.5 Hz. The mode response was .1 Hz (35% of respondents with moderate consensus; Figure 1a). Low-pass filter responses showed more variability, ranging from 25 Hz to 250 Hz (Nyquist frequency), and some endorsement of no low-pass filter. The mode response was 50 Hz (24% of valid numerical responses reflecting low consensus; Figure 1b). See Supplement S2.1 for more details.

No expert changed their reported filter settings for the high-density data set, suggesting a consensus that filter settings should be consistent across montages.

#### 3.1.2 | Ica

Next, we asked experts if they would use ICA to remove artifacts, with a reminder that there would be no eye/face electrodes as infants do not commonly tolerate such electrodes. For the low-density data set, there was low consensus among experts with 55% reporting they would use ICA, while 39% reported they would not use ICA, and 6% were unsure if they would use ICA. As a note, experts reported they were more likely to use ICA for higher density montages (71% report they would use ICA for higher density montages, reaching moderate consensus; See Supplemental Material 1).

#### 3.1.3 | Epoching

For epoching, there was moderate consensus between experts, with 71% reporting they would epoch their data before power decomposition. For those who endorsed they would epoch their data ($N=47$; Figure 1c), epoch lengths ranged from .5 seconds to 4 seconds. The mode epoch length was 1 second (36%, moderate consensus), with 2 seconds being the second most common response (32%; See Supplemental Material S2.2 for more details). Of those who would epoch their data, there was low consensus between respondents (47%) that they would allow their epochs to overlap, 38% would not allow epoch overlap, and 15% were unsure. Of those who would allow their epochs to overlap, there was high consensus (82% of respondents) for allowing epochs to overlap by 50%.

There was moderate consensus surrounding whether experts would baseline correct their epochs and when they would apply this transformation. And 39% reported they would baseline correct their epochs before artifact detection, 26% reported they would baseline correct after artifact detection, 27% would not baseline correct, 6% were not sure, and 2% reported other responses.

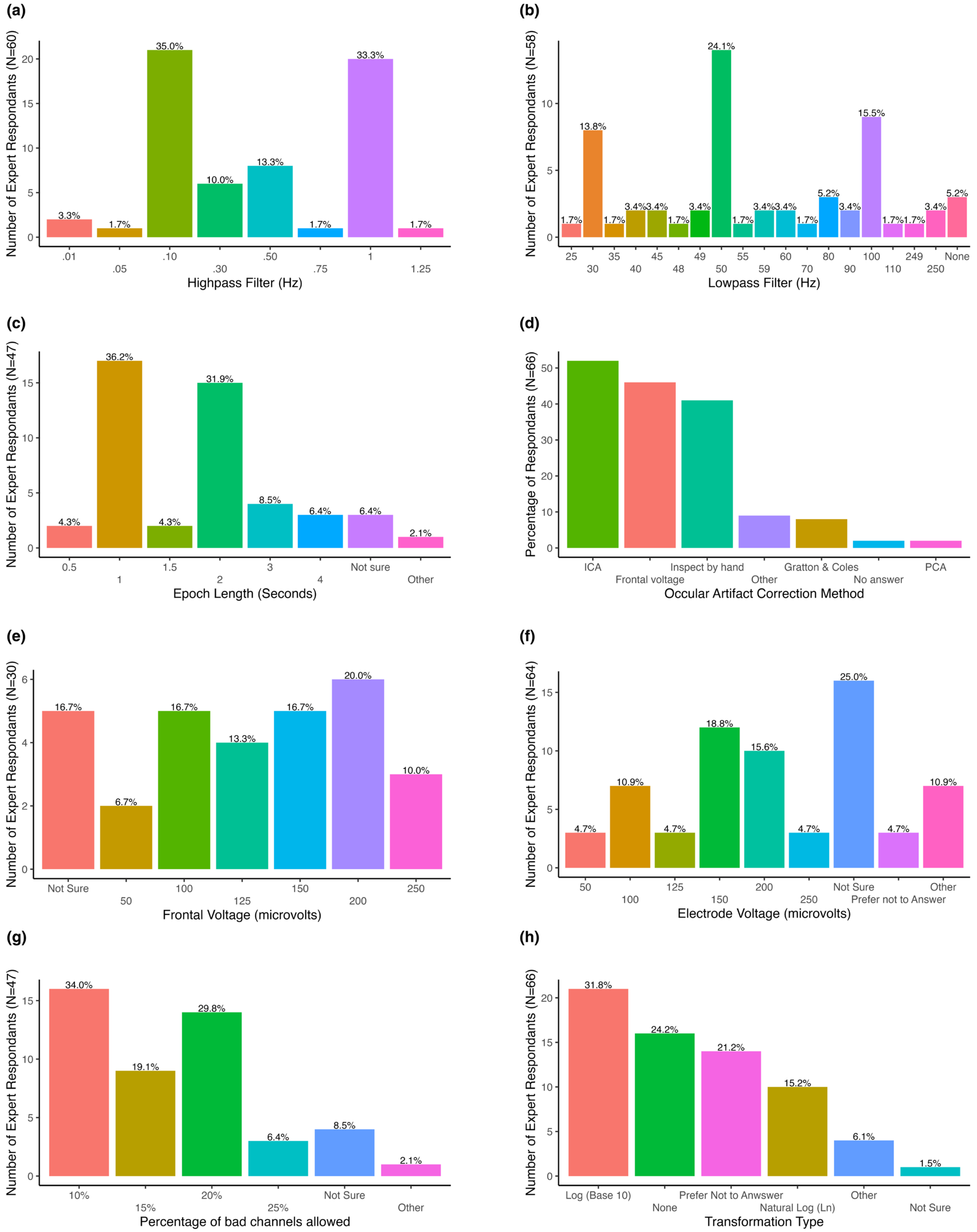


**FIGURE 1** (a–h). Histograms of provided responses. Number of valid responses is indicated on the *y*-axis. Numbers above bars indicate the percentage of valid responses where appropriate. No percentages are provided for 1D because experts could select multiple corrections.

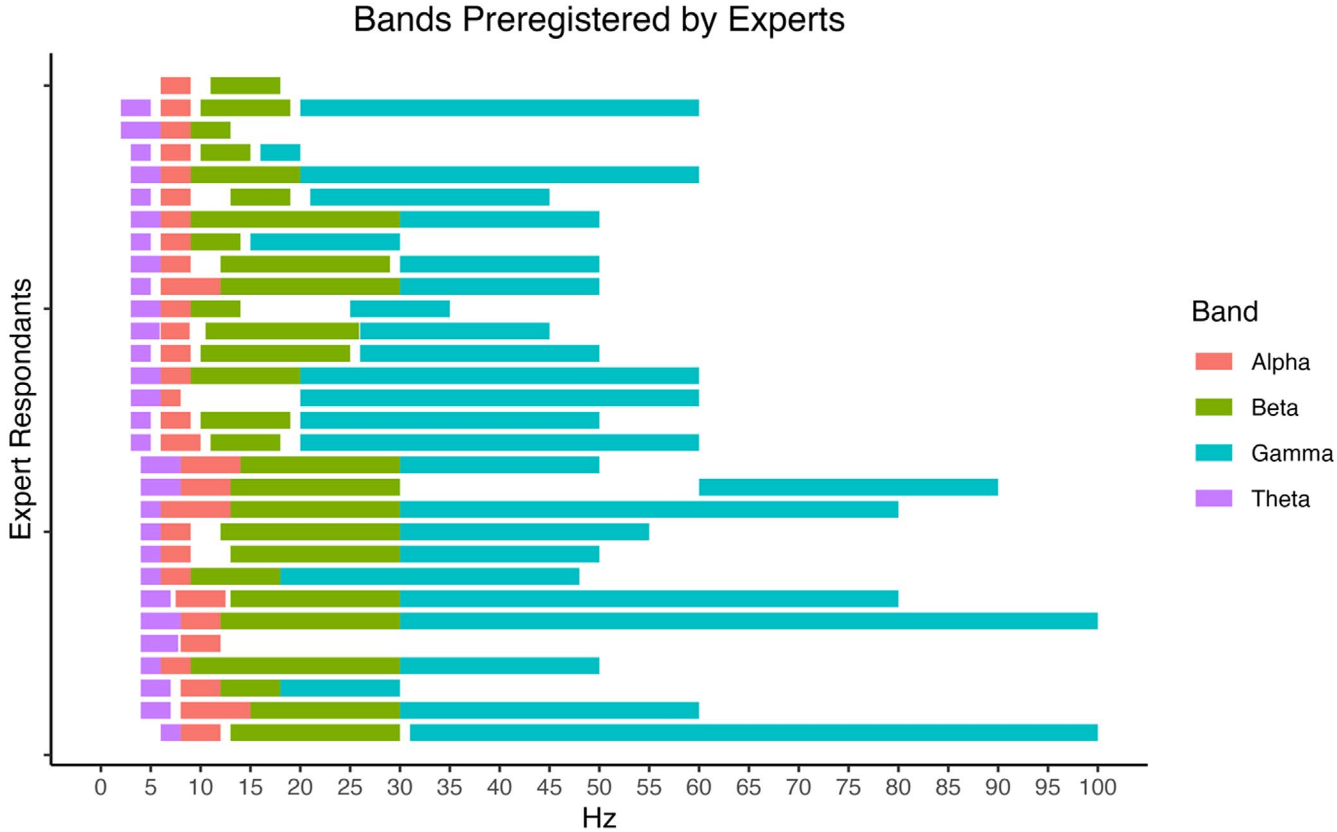


**FIGURE 2** Frequency bands preregistered by expert respondents. Some experts chose not to preregister some bands and, as such, those bars may be missing from the figure.

Finally, we asked whether respondents would apply any detrending operations to their data. Largely, participants were not sure whether they would perform this processing step (46%).

### 3.1.4 | Artifact removal

To understand how experts would remove ocular artifacts, we provided a list of common removal operations and allowed experts to select as many as they would like. About 41% reported they would inspect epochs by hand, 46% reported they would use a voltage inspection of frontal electrodes, 52% reported they would use ICA, 2% reported they would use PCA, 8% reported they would use the regression-based method of Gratton and Coles (Gratton et al., 1983), 10% reported they would use another method, and 2% preferred not to answer (Figure 1d). For those who reported they would use a voltage inspection of frontal electrodes, we asked what voltage threshold they would use, and responses ranged from 50 to 250 microvolts (Figure 1e). The most common response was 200 microvolts (20%; low consensus), followed closely by 100 and 150 microvolts (both reported by 17% of respondents).

In addition to ocular artifacts, we inquired how experts would remove other artifacts from their data (e.g., ICA or hand inspection). Most experts (82%) indicated they would use some other method to remove artifacts, while 17% would not use another correction method, and 2% were unsure. Of those who said they would use other methods, some of the most frequently reported methods were ICA (33%), inspection by hand (32%), and/or voltage inspection of certain electrodes (13%).

### 3.1.5 | Rejection of bad channels

Next, we inquired about the methods respondents would use to identify bad channels. And 85% of respondents (high consensus) reported they would check to see if channels surpass a specific voltage (e.g., +/− 100 μV), 89% reported they would check to see if channels were flat (e.g., a range of less than one μV over the segment), 83% reported they would check to see if the channel "jumps" (i.e., a certain μV change within a specific time window), and 23% reported they would use another inspection which included visual inspection for bad electrodes, checking experimental notes, using impedance values before and after the experiment, using a quality metric, using a Hurst gradient,

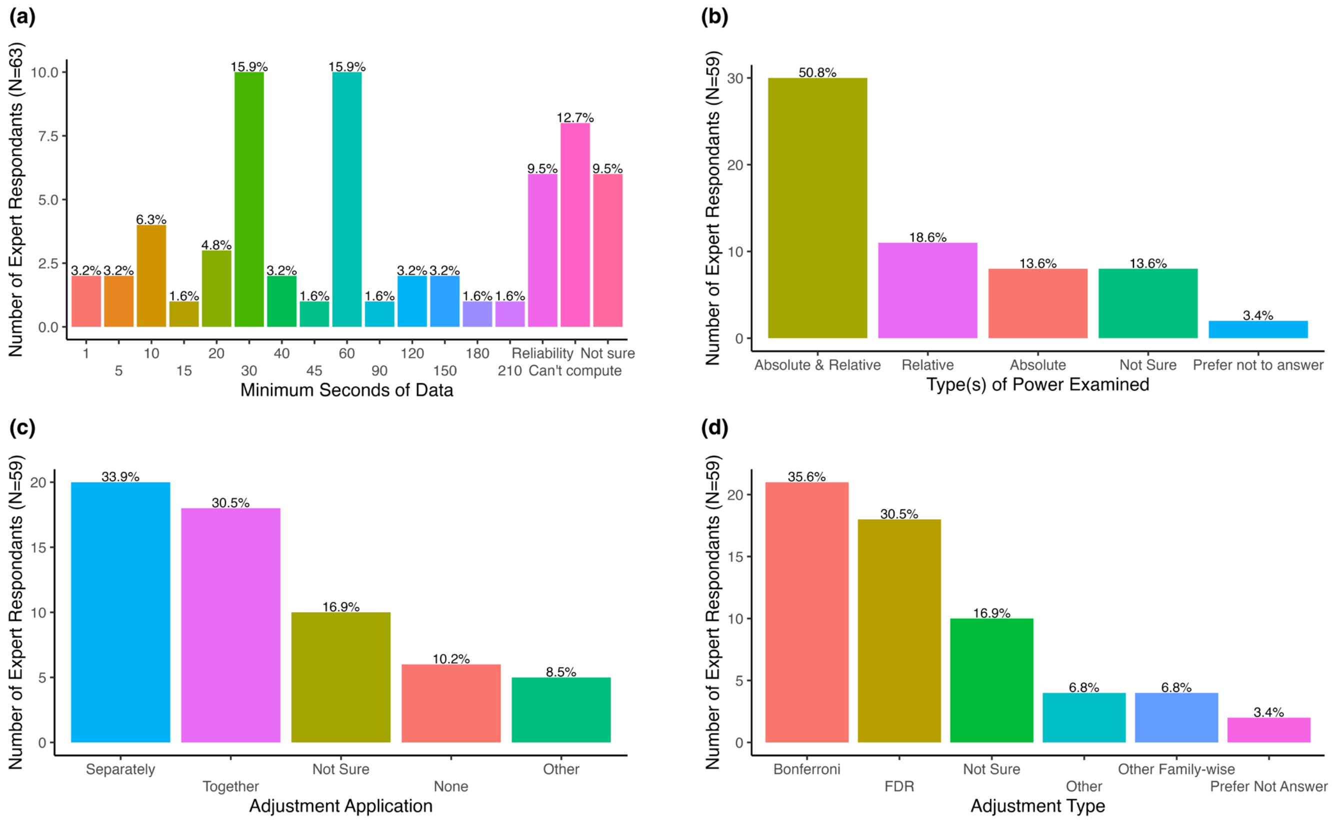


**FIGURE 3** (a–d). Histograms of provided responses. Number of valid responses is indicated on the Y-axis. Numbers above bars indicate the percentage of valid responses where appropriate. At times, responses shown reflect multiple independent questions (see S1 for survey).

checking for 60-Hz and very high-frequency noise, and identifying channels that exceed a set number of standard deviations.

To follow-up on voltage rejection, respondents were asked how they would use a voltage-based rejection to remove any remaining artifacts (following the prior processing steps already described). This question was asked separately for those who did and did not endorse using ICA to assess whether artifact thresholds were lower for those who utilized ICA. However, this was not the case (both responses fell in the same range; Figure 1f), and responses collapsed. Together, artifact threshold responses ranged from 50 to 250 microvolts with a mode artifact threshold of 150 microvolts (19% of respondents; low consensus). Other responses included 50 (5%), 100 (11%), 125 (5%), 200 (16%), and 250 (5% of respondents) microvolts. Furthermore, 25% of respondents indicated they were unsure of what threshold they would use, 5% preferred not to answer, and 11% indicated they would use other inspections, including using step functions in addition to absolute thresholds, a minimum threshold, kurtosis measures, thresholds for different frequency bands, z-score thresholds, and calculating a deviation from median absolute voltage values.

#### 3.1.6 | Epoch exclusion and channel interpolation

Following inquiries regarding bad channel rejection, we asked respondents about what they would do with these bad channels. When a channel was identified as bad, 68% of respondents reported they would interpolate any bad data, while 23% would not interpolate, and 9% were unsure.

Next, we inquired as to what respondents would do with epochs that contained bad electrodes. Most respondents (71%) reported they would establish a threshold for the number of bad electrodes allowed to be included in each epoch, 21% of respondents reported they would throw out any epoch that had a bad electrode (no bad electrodes allowed in an epoch), 3% reported they would keep all epochs (as long as one or more electrodes is/are good, the epoch is retained), 3% were unsure, and 2% preferred not to answer. For those who reported they would utilize a threshold ($N=47$, Figure 1g), 10% bad channel allowance was the most common (34%; low consensus), with most responses falling within a range of 10 to 25%.

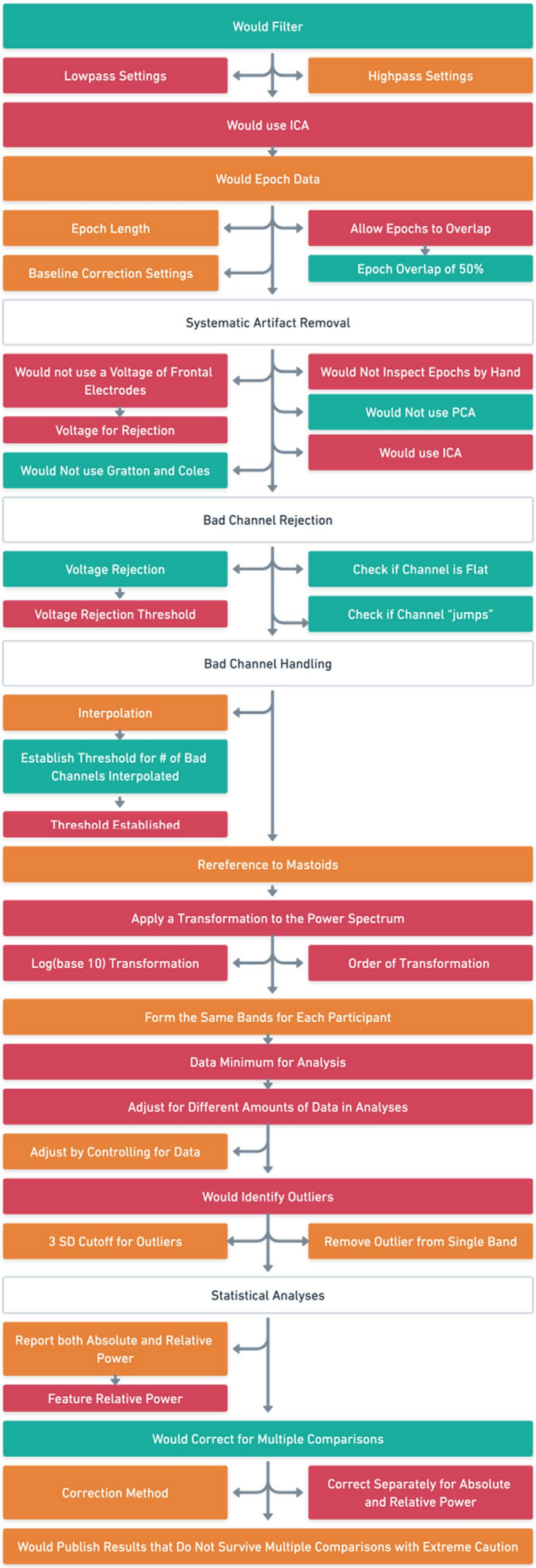


**FIGURE 4** The general consensus among experts across various processing steps. For items that were open responses or had more than two categories, consensus was as follows: Low consensus (Red: <35%), moderate consensus (Orange: 35%–69%), and high consensus (Green: 70%–100%) of consensus. For items that were dichotomous (besides don't know, other, and prefer not to answer) consensus was as follows: Low (Red: <60%), moderate (Orange: 60%–75%), high (Green: 75%–100%). The order of processing steps maps largely onto the order of the survey and not the order in which experts would perform these steps during processing. These consensus values are meant to be illustrative to show a general summary and patterns of consensus.

#### 3.1.7 | Re-referencing

Next, respondents reported on whether and where they would re-reference. The most common reference location was the mastoids (42%; moderate consensus), followed by the average reference (39%). Other less common responses included no re-referencing (3%), Cz (6%), other (5%), and not sure (5%). It is important to note that preferred re-referencing location did vary by montage size, such that more respondents reported using the average reference for the larger montage (see Supplementary Material 1 for high-density responses).

### 3.2 | Analysis parameters

#### 3.2.1 | Power transformations

Following data cleaning procedures, we asked respondents whether they would apply any transformations to their power spectrum. There was low consensus (56%) among experts that they would apply a transformation (Figure 1h; see Supplemental Material S2.3 for more information). Of those who endorsed applying a transformation (N=37), Log (base 10) (57%; low consensus) and natural log (27%) transformations were the most common, with the rest of experts endorsing other transformation, saying they were not sure, or not providing a response. We then asked experts *when* they would apply this transformation—a detail that is underreported in the literature—and identified great variability. About 30% reported they would apply their transformation after the *fast Fourier transform* (FFT) but before band power calculations (low consensus), 24% reported they would apply their transformation after band power is calculated but that the adjustment would be applied across bands, 24% reported they would apply the transformation after band power was computed and that the transformation would be applied within each band (e.g., separate transformations for each power band), 8%

reported they would apply their transformation at another time than those listed, 11% were unsure, and 3% preferred not to answer.

#### 3.2.2 | Band definition

Next, we asked how respondents would quantify their power spectrum. About 55% of respondents stated they would use the same bands for each participant based on existing literature, 30% of respondents stated they would use the same band for everyone based on data inspection or a data-driven method to identify the window of interest, 8% stated they would define different frequency bands for each individual based on inspecting their data, and 8% preferred not to answer.

For those who reported they would use bands to quantify their data (Figure 2), we asked for the band boundaries they would preregister. For the lower bound of the theta band, the mode was 3 Hz with a minimum of 2 Hz and a maximum of 6 Hz. For the upper bound of theta, the mode was 6 Hz with a minimum of 5 Hz and a maximum of 8 Hz. For the lower bound of alpha, the mode response was 6 Hz with a range of 6–8 Hz. For the upper bound of alpha, the mode was 9 Hz with a minimum of 8 Hz and a maximum of 15 Hz. For the lower bound of beta, the mode was 9 Hz with a range of 9–15 Hz. For the upper bound of beta, the mode response was 30 Hz, with a range of 13–30 Hz. For the lower bound of gamma, the mode was 30 Hz with a range of 18–60 Hz. For the upper bound of gamma, the mode response was 50 Hz with a range of 20–100 Hz. Finally, it is important to note that a few respondents were unsure of the boundaries of a few bands (treated as missing), three respondents were unsure of all bands, and one respondent noted they would report contiguous smaller bands (e.g., 2 Hz) from 1 to 100 Hz.

#### 3.2.3 | Data minimums and adjustments

Next, we requested respondents to report the minimum amount of data they require for a participant to be included in statistical analyses. Respondents could report either in epochs or seconds. Responses in epochs were converted to seconds of data. Responses ranged from no minimum (1 second) to 210 seconds, with a mean of 52 seconds and a standard deviation of 49 seconds (Figure 3a). Mode responses were 30 and 60 seconds (16% of experts; low consensus). In addition, 10% reported they would run reliability analyses to determine a threshold, 10% of respondents were unsure of their minimum, and 13% of experts would use other thresholds that could not be converted into seconds of data (e.g., epochs reported without an epoch length).

In addition, we inquired as to whether respondents would account for different amounts of data contributed by participants. 48% of respondents indicated they would correct for different amounts of data (low consensus), while 47% would not adjust, and 5% preferred not to answer. Of those who would perform an adjustment (N=30), there was moderate consensus for controlling for the amount of data contributed (e.g., controlling for epochs; 63% of respondents). Interestingly, for the experts who endorsed a weighting procedure, some reported a weighting procedure where participants with more data mattered more for analyses (13% of respondents), while others endorsed performing a weighting procedure so that all participants contributed about the same amount of data (17%), and 7% of respondents preferred not to answer.

#### 3.2.4 | Outlier identification and exclusion

We also inquired as to whether and how respondents would deal with outlier power values in their data set. There was low consensus (50%) among respondents as to whether they would identify outliers (see S2.4 for other reported methods). For a cutoff, there was moderate consensus among respondents with 55% endorsing they would use a three standard deviation cutoff. Finally, we inquired how respondents would apply outlier adjustments, and there was moderate consensus (65%), experts endorsed that if a power value in any single band exceeded their prespecified cutoff, they would remove that participant from analyses for that band only.

#### 3.2.5 | Power examinations

Moving to statistical analyses, we first asked respondents what types of power they would examine. About 51% (moderate consensus) reported they would examine both absolute and relative power, 14% reported they would examine absolute power only, 19% reported they would examine relative power only (Figure 3b).

Next, we posed a hypothetical to respondents, which was as follows, "It turns out you have to highlight one type of power when you present your RCT results in your manuscript, and the other type will be put into your supplement as an additional analysis." Following this prompt, 41% of respondents reported they would report relative findings in the main manuscript (low consensus), 31% reported they would report absolute power in the main manuscript, 20% were unsure, and 9% preferred not to answer.

#### 3.2.6 | Corrections for multiple comparisons

To understand how respondents would correct for multiple comparisons, we provided another hypothetical, which was as follows, "In your manuscript, you perform 8 tests looking for group differences in EEG power (absolute theta, absolute alpha, absolute beta, absolute gamma, relative theta, relative alpha, relative beta, and relative gamma). Would you correct for multiple comparisons across these tests?" 78% of respondents (high consensus) reported they would correct for multiple comparisons, 10% would not correct, and 12% were unsure (see Supplemental Material S2.5 for a more nuanced discussion). Next, for those who would correct (N=46), we asked how they would apply their multiple comparison correction. 44% of experts reported they would apply separate corrections for absolute and relative power (2 corrections for 4 tests; low consensus), 39% reported they would correct for all analyses together (1 correction for 8 tests; Figure 3c), 7% were unsure, and 11% endorsed some other correction technique. For the correction procedure, responses were rather split, with 36% stating, they would use Bonferroni (moderate consensus), 31% stating they would use FDR, and 7% reporting they would use another family-wise correction technique like Westfall-Young (Figure 3d).

Finally, we posed one more hypothetical question that is common among developmental EEG studies with smaller sample sizes: "It turns out you have findings that are conventionally significant but do not survive correction for multiple comparisons. Your effect size is small (exactly .2). Would you still publish these results?" With moderate consensus (58%), experts reported they would still publish, but only with extreme caution used throughout the manuscript, 17% said they would publish without extreme caution, 10% reported they would not publish these findings, and 15% were unsure. Figure 4 shows a summary of consensus across processing steps.

#### 3.2.7 | Other considerations

We also inquired what kind of code or analysis software experts would use for preprocessing and responses were as follows: 40% would use a published preprocessing pipeline (e.g., HAPPE, MADE, Automagic), 24% would use publicly available code, 14% reported they would use lab-developed code that is not publicly available, 10% would use a software program that came with their EEG system, 9% reported they would use a software program that did not come with their EEG system, and 3% preferred not to answer. Of those that would use a published pipeline, 4% reported they would use BEAPP, 18% reported they would use HAPPE (or a subsidiary), 50% reported they would use MADE, 14% reported they would use EEGLAB, and 14% reported they would use another pipeline.

## 4 | DISCUSSION

The present study was designed to evaluate the current state of the field for developmental resting EEG by asking self-identified experts to agree across various processing and analysis decisions for developmental resting EEG data. To this end, experts were asked to preregister their preprocessing and analysis steps for a hypothetical infant resting EEG data set. Despite the number of experts surveyed, many with collaborative relationships, no two individuals submitted identical preregistrations. As expected, there were areas of more and less consensus, but ultimately, we believe our findings highlight opportunities for core methodological work and field-level efforts to establish consensus. We believe that making researchers aware of areas in which more or less consensus exists will enable developmental EEG researchers to appropriately evaluate the literature, and ultimately increase the utility and interpretability of developmental EEG research.

### 4.1 | Insights about preprocessing

Within preprocessing, a few areas showed high consensus. For instance, filter settings were rather consistent across experts. There was high consensus among respondents (90%) endorsing they would utilize filters to clean their data. For high-pass filters, most experts stated they would utilize this setting, and the range was quite small (0.01–1.5 Hz). For low-pass filters, the range was much broader (25 Hz to no filter), but all responses were valid and logical (e.g., some electing to filter under line noise (50/60 Hz), while others only electing a notch filter). Overall, variations in these settings are not likely to have large effects on band-specific absolute power. As long as frequency bands are specified around the filter settings, the responses from experts are unlikely to produce large differences in band power (although differences in band boundaries will, as discussed later). That said, these settings may have larger influences on relative power computations as where the data are filtered will impact the total power in the denominator of relative power computations.

In contrast to the filter settings, the use of ICA to remove systematic artifacts was less uniform across experts. The lack of consensus is likely a reflection of the uncertainty surrounding the appropriateness of ICA for developmental recordings, especially for infants. This uncertainty likely stems from several sources. First, as reflected in our data, the size of the recording montage clearly guides the decision as more experts endorsed using ICA for larger recording montages. Second, the quality and veracity of ICA decompositions and automated labeling scripts for developmental data have received rightful scrutiny with some efforts to optimize settings for developmental data (Haresing et al., 2021; Leach et al., 2020). Furthermore, it is important to note that some publicly available pipelines include ICA (e.g., MADE) while others do not (e.g., HAPPE 2.0), which is also likely to increase variability in responses. Core methodological work examining the use of ICA for developmental data is needed as high levels of artifact are one of the primary problems facing developmental EEG data. This problem is particularly difficult to solve as it is likely that ICA may be more or less effective based on the type of artifact, electrode montage, developmental age, recording length, and signal quality. Additionally, as many infants and children do not tolerate electrodes on their faces, ICA decompositions are much harder to interpret, and especially for low-density montages, disambiguating frontal ocular artifacts from frontal brain activity can be difficult. Furthermore, the use of ICA is likely to have large downstream effects on bad channel/epoch identification and data rejection, making methodological work and field consensus even more important.

Beyond ICA, other steps for artifact removal and data cleaning showed a similar lack of consensus. Specifically, we asked about voltage-based rejections as these are commonly used to remove both stereotyped (e.g., eyeblinks, cardiac artifact, saccades) and non-stereotyped artifacts (e.g., body movements, jaw movements, muscle). The range for voltage-based data rejection was wide (50–250 microvolts), and no single threshold was endorsed by more than 20% of experts. Surprisingly, many experts were unsure of what threshold they would use. One reason for this uncertainty may be the delicate balance between data cleanliness (i.e., lower thresholds) and the amount of data retained. Also, although we asked experts to preregister for a single age, it is true that the magnitude of some artifacts (e.g., eyeblinks) change across age, and as such, experts may have been inclined to report a voltage threshold for the age range in which they usually conduct research. In these authors' experience, it is not uncommon to get the "feel" for a data set before choosing a threshold. The information researchers use to establish a threshold varies. Some manually review their data to observe the size of various artifacts while others try multiple thresholds to see the impact on data retention and/or reliability. The problem is that different researchers may use different criteria, leading to differential data and participant inclusion/exclusion, and even worse, it could lead to p-hacking. While some amount of field-level flexibility may be prudent for different research populations (e.g., different ages or special populations), field consensus for thresholds would only improve the replicability of developmental EEG studies. To establish consensus, more methodological work is needed to understand the impact of different thresholds on power values, signal reliability, and inclusion/exclusion of different participants. Furthermore, it is important to consider that consensus on these settings may not be a specified threshold (e.g., everyone uses 100 μV) – particularly as threshold values are likely to vary across systems and developmental periods (e.g., 4-year-old children are likely to provide much cleaner data than 18-month-old toddlers). Rather, consensus may be achieved by a replicable framework for making thresholding decisions, such as based on specific data quality metrics, or an acceptable percentage of children or data lost considering the sample size or recording length. Finally, of course, it is important to understand which artifacts are removed/remain in the data at different thresholds (e.g., eye blinks are removed, but small body movements are not) to have an understanding of noise remaining in data. Voltage thresholds for identifying artifacts were not substantively different across individuals who did and did not elect to use ICA.

Although there was a lack of consensus on thresholds for identifying bad electrodes, there was more consensus on what to do with identified bad electrodes. Most experts (71%) agreed that a threshold should be used for the number of bad electrodes allowed in an epoch, and that threshold should probably be somewhere between 10 and 20% of electrodes. Similarly, most experts (68%) agreed that any channels identified as "bad" should be interpolated. While these opinions are far from unanimous, they do provide good guidance for future projects and a good starting place for confirmatory methodological work.

Re-referencing is another area ready for methodological work and consensus. While over 90% of experts endorsed that they would re-reference, there was almost an even split between re-referencing to the mastoids and an average reference. As long as the reference meets all of the necessary assumptions, the reference location is unlikely to change statistical significance for resting EEG. There are a few considerations for both reference types when it comes to developmental research. Mastoid references convey a unique benefit as they are not sensitive to montage size and two studies using the same system with mastoid references should be directly comparable (Luck, 2014). However, many of the common child-friendly fast-capping systems have been hypothesized to have imperfect mastoid references that can be easily contaminated by brain activity if cap fit is not optimal. In contrast, average references, are likely sensitive to montage size (better with larger montages; Dien, 1998). This may be particularly true with small montages as a single electrode with artifact or noise might be more likely to bias the reference for smaller montages in comparison to larger montages. However, ultimately, the reference location is likely to change topographic heat maps and the magnitude of absolute power values, which makes it harder to compare findings across studies.

Moving from data cleaning to power calculations, experts had contrasting perspectives on how to calculate power. First, somewhat surprisingly, there was rather low agreement on whether and how EEG power values should be transformed. Just over 50% of experts endorsed that they would apply a transformation to their power spectra or values, but the kind of transformations experts endorsed were split across two different transformations (Log and Ln). Such variability in power transformation applied (i.e., Log vs. Ln) is commonly reported in the literature. However, what is not typically reported—but is critical for reproducibility—is that experts differed in their reported preferences for when in the processing stream (in what order) these Log/Ln transformations should be applied. Specifically, some experts differed on whether they would: (1) apply the transformation to the entire power spectrum before computing band-specific power, (2) compute band-specific power first and then apply transformations, or (3) apply transformations separately within each frequency band (e.g., separate transformations for theta and alpha). Together, the variability in the types of transformations and the order of when such transformations are applied will lead to different absolute and relative power values (see Supplement 3 for an example). Such differences are particularly problematic and a threat to reproducibility across studies, especially given that differences in the order in which power transformations are applied are not always reported in published reports. Converging on consensus in this area is likely to be complicated as data transformations are commonly driven either by theory or statistical properties, which may differ across studies. However, basic methodological work may also help inform how transformations should be applied.

## 4.2 | Insights concerning data analysis

Moving beyond data cleaning and power computation, experts varied in how they would make decisions surrounding which participants would ultimately be included in statistical analyses. For instance, there was a very wide range of data minimums (1–210 seconds of data) required for inclusion, which would undoubtedly result in different children making it into final statistical analyses. Critically, these differences are likely not random as data yield is usually a result of both trait (e.g., reactions to novelty, irritability, ease of soothing, neurodevelopmental differences, hair type and styling) and state (e.g., hunger, tiredness, mood), and individual (e.g., hairstyle, race, age) differences between children (Adams et al., 2024; van der Velde & Junge, 2020). Furthermore, expert respondents differed in what they would do when participants contributed varying amounts of data (e.g., nothing, weighting procedures, covarying), which could result in different participants driving effects of interest. We believe coming to a high-quality field-level consensus for these settings is critical for ensuring participant representation and reducing sample bias. Possible avenues for this include weighting procedures for individual differences and standard operations for establishing signal reliability (e.g., Clayson & Miller, 2017; Sandre et al., 2024; Troller-Renfree et al., 2021) for cutoff identification. Critically, these methodological innovations should include flexibility for differences in measures, recording setups, study populations, and age.

There was moderate agreement among experts as to how to identify and correct outliers. About half the experts reported they would use a specified cutoff, with most of those experts (55%) endorsing a cutoff of three standard deviations. Furthermore, most experts (65%) stated that when a participant had outlying power values in one band, they would remove them from analyses for that band only (not all analyses). On the one hand, although such pairwise deletion approaches yield different numbers of participants for each analysis, such procedures are consistent with recent recommendations to avoid biased statistical inferences that can arise from listwise deletion (Heise et al., 2022; Little & Rubin, 2019; Schafer & Graham, 2002). On the other hand, such pairwise deletion introduces questions about what to do for examinations of relative power, as relative power calculations require power values from all bands in the denominator. Furthermore, it raises questions surrounding at what level outliers should be identified (e.g., regional vs. whole-brain). Given that most developmental EEG studies to date employ modest sample sizes that can be easily influenced by outliers, it is important to converge on both how outliers are identified and corrected to ensure associations of interest are not driven by extreme values.

Perhaps, the most surprising area of variability among experts was in their preregistered statistical approach, which is largely independent of preprocessing and data cleanliness decisions. While most experts (51%) agreed they would report both absolute and relative power values, there was not much consensus on whether absolute or relative power should be highlighted more prominently. Furthermore, while most experts (>50%) agreed they would correct for multiple comparisons, there was no consensus on what type of correction (e.g., Bonferroni, FDR, etc.) and how corrections would be applied (e.g., across power type vs. within each type of power). However, experts did agree that effects that are conventionally significant but do not survive correction for multiple comparisons should still be published and interpreted with extreme caution. However, it is important to note that 10–25% of our experts reported they were not sure or would not publish findings that are marginally significant after correction for multiple comparisons, which suggests there may be a bias for publishing only statistically significant findings.

Altogether, we believe the present manuscript highlights both areas of strength and areas of improvement in terms of the current state of consensus with the field of developmental EEG. However, the present investigation is not without its limitations. First, as with all survey research, the validity of data generated is dependent on the degree to which the sample is representative of the target population. While we confirmed the expert status of most of our respondents (89%), it is possible that some respondents were not experts in developmental EEG. In addition, it is impossible to know whether the cross-section of experts surveyed generalizes to the entire field of developmental EEG, although we believe the variety of processing pipelines, software, and code reportedly used by our surveyed experts demonstrates that our surveyed sample reflects a diversity of expert respondents. Second, we had an initial dropout rate of 23.5% experts ($N=24$) prior to answering any EEG questions, perhaps reflecting a bias in who completed the survey. This rate of drop-off is near or slightly below the expected average drop-off rate for online surveys (Galesic, 2006). Third, while we have reports of what experts would do hypothetically, it is hard to know whether these responses reflect what is currently being used in the field. As such, more methodological work examining the variability of preprocessing approaches and their implications for research findings is needed, similar to work that has been done in fMRI (Botvinik-Nezer et al., 2020) and is currently being done with adult EEG (Algermissen et al., 2023). Fourth, it is important to note that we had a low percentage of "do not know" responses for most of our questions, which are hard to interpret. These responses may reflect a whole amalgam of situations (e.g., experts are unsure of setting thresholds, experts who are unfamiliar with various processing steps, and experts who feel unsafe sharing their opinions), meaning we can only draw limited interpretations from such "do not know" responses. Fifth, it is important to note that for some sub-questions, cell sizes were rather small (e.g., in the 20–30s experts), impacting generalizability.

## 5 | SUMMARY AND CONCLUSIONS

The field of developmental EEG is currently in a time of rapid growth, with many large-scale projects utilizing EEG-derived measures for documentation of developmental change as well as targets for intervention. As such, it is imperative that developmental EEG continues to become more replicable and reliable. Here, in a state-of-the-field snapshot of developmental resting EEG research, we identify some areas of consensus but also highlight areas currently lacking consensus, which would benefit from collaborative efforts to improve the rigor and reproducibility of developmental EEG research.

## AUTHOR CONTRIBUTIONS

**Sonya V. Troller-Renfree:** Conceptualization; data curation; formal analysis; funding acquisition; methodology; visualization; writing – original draft; writing – review and editing. **Santiago Morales:** Conceptualization; methodology; writing – review and editing. **George A. Buzzell:** Conceptualization; methodology; writing – review and editing. **Aislinn Sandre:** Methodology; writing – review and editing.

## ACKNOWLEDGMENTS

First and foremost, the authors would like to thank their colleagues, mentors, and friends for taking the time to complete this survey. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R00HD104923 (to STR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Thank you to Isabel Kovacs for her help in organizing and programming the survey.

## FUNDING INFORMATION

U.S. Department of Health and Human Services. National Institutes of Health. National Institute of Child Health and Human Development. R00HD104923.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Given the small nature of the field, making the entirety of the data set publicly available could easily result in the possible reidentification of expert respondents. As such, we will make individual variables and responses available upon request in a limited format to prevent reidentification.

## CODE SHARING

As the entirety of the analyses presented are descriptive, we have no substantial statistical code. However, the code for the descriptive statistics is available upon request.

## ORCID

*Sonya V. Troller-Renfree* https://orcid.org/0000-0001-9979-4696
*Santiago Morales* https://orcid.org/0000-0002-9850-042X
*George A. Buzzell* https://orcid.org/0000-0003-3324-3183

## REFERENCES

Adams, E. J., Scott, M. E., Amarante, M., Ramírez, C. A., Rowley, S. J., Noble, K. G., & Troller-Renfree, S. V. (2024). Fostering inclusion in EEG measures of pediatric brain activity. *npj Science of Learning*, *9*(1), 1–8. https://doi.org/10.1038/s41539-024-00240-y

Algermissen, J., Busch, N., Cesnaite, E., Marshall, T., Nilsonne, G., Pascarella, A., Puoliväli, T., Senoussi, M., Trübutschek, D., Vinding, M. C., Vitale, A., Yang, Y.-F., & Yeaton, J. (2023). EEGManyPipelines: Robustness of EEG results across analysis pipelines. https://doi.org/10.17605/OSF.IO/42K5H

Babiloni, C., Barry, R. J., Başar, E., Blinowska, K. J., Cichocki, A., Drinkenburg, W. H. I. M., Klimesch, W., Knight, R. T., Lopes da Silva, F., Nunez, P., Oostenveld, R., Jeong, J., Pascual-Marqui, R., Valdes-Sosa, P., & Hallett, M. (2020). International Federation of Clinical Neurophysiology (IFCN)—EEG research workgroup: Recommendations on frequency and topographic analysis of resting state EEG rhythms. Part 1: Applications in clinical research studies. *Clinical Neurophysiology*, *131*(1), 285–307. https://doi.org/10.1016/j.clinph.2019.06.234

Begus, K., & Bonawitz, E. (2020). The rhythm of learning: Theta oscillations as an index of active learning in infancy. *Developmental Cognitive Neuroscience*, *45*, 100810. https://doi.org/10.1016/J.DCN.2020.100810

Bell, M. A., & Cuevas, K. (2012). Using EEG to study cognitive development: Issues and practices. *Journal of Cognition and Development*, *13*(3), 281–294. https://doi.org/10.1080/15248372.2012.691143

Benasich, A. A., Gou, Z., Choudhury, N., & Harris, K. D. (2008). Early cognitive and language skills are linked to resting frontal gamma power across the first 3 years. *Behavioural Brain Research*, *195*(2), 215–222. https://doi.org/10.1016/j.bbr.2008.08.049

Blaisdell, C. J., Park, C., Hanspal, M., Roary, M., Arteaga, S. S., Laessig, S., Luetkemeier, E., & Gillman, M. W. (2022). The NIH ECHO program: Investigating how early environmental influences affect child health. *Pediatric Research*, *92*(5), 1215–1216. https://doi.org/10.1038/s41390-021-01574-8

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Brito, N. H., Elliott, A. J., Isler, J. R., Rodriguez, C., Friedrich, C., Shuffrey, L. C., & Fifer, W. P. (2019). Neonatal EEG linked to individual differences in socioemotional outcomes and autism risk in toddlers. *Developmental Psychobiology*, *61*(8), 1110–1119. https://doi.org/10.1002/dev.21870

Brito, N. H., Fifer, W. P., Myers, M. M., Elliott, A. J., & Noble, K. G. (2016). Associations among family socioeconomic status, EEG power at birth, and cognitive skills during infancy. *Developmental Cognitive Neuroscience*, *19*, 144–151. https://doi.org/10.1016/j.dcn.2016.03.004

Buzzell, G. A., Morales, S., Valadez, E. A., Hunnius, S., & Fox, N. A. (2023). Maximizing the potential of EEG as a developmental neuroscience tool. *Developmental Cognitive Neuroscience*, *60*, 101201. https://doi.org/10.1016/j.dcn.2023.101201

Clayson, P. E., & Miller, G. A. (2017). ERP reliability analysis (ERA) toolbox: An open-source toolbox for analyzing the reliability of event-related brain potentials. *International Journal of Psychophysiology*, *111*, 68–79. https://doi.org/10.1016/j.ijpsycho.2016.10.012

Debnath, R., Buzzell, G. A., Morales, S., Bowers, M. E., Leach, S. C., & Fox, N. A. (2020). The Maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology*, *57*(6), e13580. https://doi.org/10.1111/psyp.13580

Debnath, R., Tang, A., Zeanah, C. H., Nelson, C. A., & Fox, N. A. (2020). The long-term effects of institutional rearing, foster care intervention and disruptions in care on brain electrical activity in adolescence. *Developmental Science*, *23*(1), e12872. https://doi.org/10.1111/desc.12872

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 34–43. https://doi.org/10.3758/BF03209414

Fló, A., Gennari, G., Benjamin, L., & Dehaene-Lambertz, G. (2022). Automated pipeline for infants continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. *Developmental Cognitive Neuroscience*, *54*, 101077. https://doi.org/10.1016/j.dcn.2022.101077

Fox, N. A., Pérez-Edgar, K., Morales, S., Brito, N. H., Campbell, A. M., Cavanagh, J. F., & Yoder, L. (2024). The development and structure of the HEALthy Brain and Child Development (HBCD) Study EEG protocol. *Developmental Cognitive Neuroscience*, *69*, 101447. https://www.sciencedirect.com/science/article/pii/S1878929324001087

Gabard-Durnam, L. J., Mendez Leal, A. S., Wilkinson, C. L., & Levin, A. R. (2018). The Harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Frontiers in Neuroscience*, *12*, 97. https://doi.org/10.3389/fnins.2018.00097

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, *22*(2), 313.

Garrett-Ruffin, S., Hindash, A. C., Kaczkurkin, A. N., Mears, R. P., Morales, S., Paul, K., Pavlov, Y. G., & Keil, A. (2021). Open science in psychophysiology: An overview of challenges and emerging solutions. *International Journal of Psychophysiology*, *162*, 69–78. https://doi.org/10.1016/j.ijpsycho.2021.02.005

Gou, Z., Choudhury, N., & Benasich, A. A. (2011). Resting frontal gamma power at 16, 24 and 36 months predicts individual differences in language and cognition at 4 and 5 years. *Behavioural Brain Research*, *220*(2), 263–270. https://doi.org/10.1016/j.bbr.2011.01.048

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Haresing, M., Phillips, E., Whitehorn, M., Noreika, V., Jones, E. J. H., Leong, V., & Wass, S. V. (2021). Automatic classification of ICA components from infant EEG using MARA. *Developmental Cognitive Neuroscience*, *52*, 101024. https://doi.org/10.1016/j.dcn.2021.101024

Heise, M. J., Mon, S. K., & Bowman, L. C. (2022). Utility of linear mixed effects models for event-related potential research with infants and children. *Developmental Cognitive Neuroscience*, *54*, 101070. https://doi.org/10.1016/j.dcn.2022.101070

Jensen, S. K. G., Xie, W., Kumar, S., Haque, R., Petri, W. A., & Nelson, C. A. (2021). Associations of socioeconomic and other environmental factors with early brain development in Bangladeshi infants and children. *Developmental Cognitive Neuroscience*, *50*, 100981. https://doi.org/10.1016/j.dcn.2021.100981

Keil, A., Bernat, E. M., Cohen, M. X., Ding, M., Fabiani, M., Gratton, G., Kappenman, E. S., Maris, E., Mathewson, K. E., Ward, R. T., & Weisz, N. (2022). Recommendations and publication guidelines for studies using frequency domain and time-frequency domain analyses of neural time series. *Psychophysiology*, *59*(5), e14052. https://doi.org/10.1111/psyp.14052

Leach, S. C., Morales, S., Bowers, M. E., Buzzell, G. A., Debnath, R., Beall, D., & Fox, N. A. (2020). Adjusting ADJUST: Optimizing the ADJUST algorithm for pediatric data using geodesic nets. *Psychophysiology*, *57*(8), e13566. https://doi.org/10.1111/psyp.13566

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Lopez, K. L., Monachino, A. D., Morales, S., Leach, S. C., Bowers, M. E., & Gabard-Durnam, L. J. (2022). HAPPILEE: HAPPE in low electrode electroencephalography, a standardized pre-processing software for lower density recordings. *NeuroImage*, *260*, 119390. https://doi.org/10.1016/j.neuroimage.2022.119390

Lopez, K. L., Monachino, A. D., Vincent, K. M., Peck, F. C., & Gabard-Durnam, L. J. (2023). Stability, change, and reliable individual differences in electroencephalography measures: A lifespan perspective on progress and opportunities. *NeuroImage*, *275*, 120116. https://doi.org/10.1016/j.neuroimage.2023.120116

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT Press.

Marshall, P. J., Fox, N. A., & BEIP Core Group. (2004). A comparison of the electroencephalogram between institutionalized and community children in Romania. *Journal of Cognitive Neuroscience*, *16*(8), 1327–1338.

McLaughlin, K. A., Fox, N. A., Zeanah, C. H., Sheridan, M. a., Marshall, P., & Nelson, C. A. (2010). Delayed maturation in brain electrical activity partially explains the association between early environmental deprivation and symptoms of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *68*(4), 329–336. https://doi.org/10.1016/j.biopsych.2010.04.005

Morales, S., & Bowers, M. E. (2022). Time-frequency analysis methods and their application in developmental EEG data. *Developmental Cognitive Neuroscience*, *54*, 101067. https://doi.org/10.1016/j.dcn.2022.101067

Norton, E. S., MacNeill, L. A., Harriott, E. M., Allen, N., Krogh-Jespersen, S., Smyser, C. D., Rogers, C. E., Smyser, T. A., Luby, J., & Wakschlag, L. (2021). EEG/ERP as a pragmatic method to expand the reach of infant-toddler neuroimaging in HBCD: Promises and challenges. *Developmental Cognitive Neuroscience*, *51*, 100988. https://doi.org/10.1016/j.dcn.2021.100988

Otero, G. A., Pliego-Rivero, F. B., Fernández, T., & Ricardo, J. (2003). EEG development in children with sociocultural disadvantages: A follow-up study. *Clinical Neurophysiology*, *114*(10), 1918–1925. https://doi.org/10.1016/S1388-2457(03)00173-1

Pierce, L. J., Thompson, B. L., Gharib, A., Schlueter, L., Reilly, E., Valdes, V., Roberts, S., Conroy, K., Levitt, P., & Nelson, C. A. (2019). Association of Perceived Maternal Stress during the perinatal period with electroencephalography patterns in 2-month-old infants. *JAMA Pediatrics*, *173*(6), 561–570. https://doi.org/10.1001/jamapediatrics.2019.0492

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Rockers, P. C., Leppänen, J. M., Tarullo, A., Coetzee, L., Fink, G., Hamer, D. H., Yousafzai, A. K., & Evans, D. (2023). Evaluation of a community health worker home visit intervention to improve child development in South Africa: A cluster-randomized controlled trial. *PLoS Medicine*, *20*(4), e1004222. https://doi.org/10.1371/journal.pmed.1004222

Saby, J. N., & Marshall, P. J. (2012). The utility of EEG band power analysis in the study of infancy and early childhood. *Developmental Neuropsychology*, *37*(3), 253–273. https://doi.org/10.1080/87565641.2011.614663

Sandre, A., Troller-Renfree, S. V., Giebler, M. A., Meyer, J., & Noble, K. G. (2024). Prenatal family income, but not parental education, is associated with resting brain activity in one-month-old infants. *Scientific Reports*, *14*(1), 13638.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Tan, E., Tang, A., Debnath, R., Humphreys, K. L., Zeanah, C. H., Nelson, C. A., & Fox, N. A. (2023). Resting brain activity in early childhood predicts IQ at 18 years. *Developmental Cognitive Neuroscience*, *63*, 101287. https://doi.org/10.1016/j.dcn.2023.101287

Tan, E., Troller-Renfree, S., Morales, S., Buzzell, G., McSweeney, M., Antúnez, M., & Fox, N. (2023). *Theta activity and cognitive functioning: Integrating evidence from resting-state and task-related developmental electroencephalography (EEG) research*. OSF. https://doi.org/10.31219/osf.io/egb6w

Tomalski, P., Moore, D. G., Ribeiro, H., Axelsson, E. L., Murphy, E., Karmiloff-Smith, A., Johnson, M. H., & Kushnerenko, E. (2013). Socioeconomic status and functional brain development—Associations in early infancy. *Developmental Science*, *16*(5), 676–687. https://doi.org/10.1111/desc.12079

Troller-Renfree, S. V., Buzzell, G. A., & Fox, N. A. (2020). Changes in working memory influence the transition from reactive to proactive cognitive control during childhood. *Developmental Science*, *23*(6), e12959. https://doi.org/10.1111/desc.12959

Troller-Renfree, S. V., Costanzo, M. A., Duncan, G. J., Magnuson, K., Gennetian, L. A., Yoshikawa, H., Halpern-Meekin, S., Fox, N. A., & Noble, K. G. (2022). The impact of a poverty reduction intervention on infant brain activity. *Proceedings of the National Academy of Sciences*, *119*(5), e2115649119. https://doi.org/10.1073/pnas.2115649119

Troller-Renfree, S. V., Morales, S., Leach, S. C., Bowers, M. E., Debnath, R., Fifer, W. P., Fox, N. A., & Noble, K. G. (2021). Feasibility of assessing brain activity using Mobile, in-home collection of electroencephalography: Methods and analysis. *Developmental Psychobiology*, *63*, e22128.

Troller-Renfree, S. V., Sperber, J. F., Hart, E. R., Costanzo, M. A., Gennetian, L. A., Meyer, J. S., Fox, N. A., & Noble, K. G. (2023). Associations between maternal stress and infant resting brain activity among families residing in poverty in the U.S. *Biological Psychology*, *184*, 108683. https://doi.org/10.1016/j.biopsycho.2023.108683

Troller-Renfree, S. V., Zeanah, C. H., Nelson, C. A., & Fox, N. A. (2017). Neural and cognitive factors influencing the emergence of psychopathology: Insights from the Bucharest early intervention project. *Child Development Perspectives*, *12*, 28–33. https://doi.org/10.1111/cdep.12251

van der Velde, B., & Junge, C. (2020). Limiting data loss in infant EEG: Putting hunches to the test. *Developmental Cognitive Neuroscience*, *45*, 100809. https://doi.org/10.1016/j.dcn.2020.100809

Vanderwert, R. E., Zeanah, C. H., Fox, N. A., & Nelson, C. A., 3rd. (2016). Normalization of EEG activity among previously institutionalized children placed into foster care: A 12-year follow-up of the Bucharest early intervention project. *Developmental Cognitive Neuroscience*, *17*, 68–75. https://doi.org/10.1016/j.dcn.2015.12.004

Wilkinson, C. L., Levin, A. R., Gabard-Durnam, L. J., Tager-Flusberg, H., & Nelson, C. A. (2019). Reduced frontal gamma power at 24 months is associated with better expressive language in toddlers at risk for autism. *Autism Research*, *12*(8), 1211–1224. https://doi.org/10.1002/AUR.2131

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article. **Data S1.**

**How to cite this article:** Troller-Renfree, S. V., Morales, S., Buzzell, G. A., & Sandre, A. (2025). Heterogeneity in pediatric resting EEG data processing and analysis: A state of the field. *Psychophysiology*, *62*, e14733. https://doi.org/10.1111/psyp.14733