Developmental Science

WILEY

Developmental Science

SHORT REPORT

Quantifying Repetitive Hand Flapping Kinematics in Autistic and Non-Autistic Toddlers Using Video-Based Pose Estimation

Jan Stenum[1] | Elizabeth Eiler[2] | Ryan T. Roemmich[1,3] | Rebecca Landa[2,4] | Rachel Reetzke[2,4]

[1]Department of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA | [2]Center for Autism Services, Science and Innovation, Kennedy Krieger Institute, Baltimore, Maryland, USA | [3]Center for Movement Studies, Kennedy Krieger Institute, Baltimore, Maryland, USA | [4]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Correspondence: Rachel Reetzke (Reetzke@kennedykrieger.org)

Received: 3 April 2025 | **Revised:** 25 November 2025 | **Accepted:** 2 March 2026

Keywords: autism | early identification | kinematics | pose estimation | repetitive behaviors

ABSTRACT

While repetitive behaviors such as hand flapping are associated with autism, non-autistic toddlers also exhibit this behavior (e.g., when excited), making clinically meaningful differences difficult to discern prior to age 2. Computer vision, a subfield of artificial intelligence (AI), has the potential to capture subtle movement differences that may otherwise be overlooked. We examined whether movement amplitude and frequency of repetitive hand flapping differed between autistic and non-autistic toddlers. Based on prior literature, we hypothesized that autistic toddlers would exhibit more intense or vigorous flapping, as reflected by increased amplitude and frequency of repetitive hand flapping. We analyzed 81 manually annotated video segments from 28 toddlers (autism spectrum disorder (ASD): $n = 14$; no ASD: $n = 14$), ages 13–16 months ($M = 14.78$, $SD = 0.85$), participating in a 3-min bubble play activity (1–10 events per child). Using AlphaPose, we automatically identified arm key points and quantified amplitude and frequency. Event-level analyses revealed that, compared to non-autistic toddlers, autistic toddlers showed significantly greater amplitude ($p = 0.028$) but did not significantly differ in the frequency of repetitive hand flapping ($p = 0.464$). In contrast, event-averaged results showed no group differences in amplitude ($p = 0.997$) or frequency ($p = 0.383$). These findings highlight the variability of repetitive hand flapping and suggest that averaging may mask potentially important nuances of repetitive hand flapping. Clinically, the dynamic nature of these early, subtle behaviors may require more fine-grained observation to identify meaningful differences during this highly dynamic period of early development.

1 | Introduction

Repetitive hand flapping is characterized by repeated hand/arm movements over time (Calkins and Fox 2002; Leekam et al. 2011). Although repetitive hand flapping is associated with autism (American Psychiatric Association 2013), non-autistic toddlers also exhibit this behavior, especially during moments of excitement (Larkin et al. 2017; Thelen 1979). Consequently, it can be challenging to discern whether hand flapping before age 2 is a natural part of typical development versus an early behavioral marker of atypical development (O'Loghlen et al. 2025).

1.1 | Identifying Early Autism-Related Behaviors Through Video-based Analysis

One method for studying the early identification and emergence of autism-related behaviors is video-based analysis, which

Summary

- Video-based pose estimation was used to examine kinematic differences in repetitive hand flapping during a 3-min bubble play activity in autistic and non-autistic toddlers.
- Event-level analyses of all repetitive hand flapping instances revealed a significantly elevated amplitude of hand flapping in autistic toddlers relative to non-autistic toddlers.
- Event-averaged analyses appeared to mask group amplitude differences, suggesting that an event-level analysis of repetitive hand flapping may better capture clinically meaningful kinematic differences.
- Findings underscore the possible added value of examining the *quality* of repetitive hand flapping to identify clinically meaningful differences during a highly dynamic developmental period.

involves observational coding systems whereby multiple research staff are trained to reliably manually annotate video recordings of the type, number, and/or duration of events of operationally defined behaviors (Yoder and Symons 2010). Researchers have applied this methodology to *retrospective* home videos collected from families of children on the autism spectrum, as well as to *prospective*, longitudinal studies of infants at elevated likelihood (EL) for autism. Prospective studies allow researchers to track the developmental trajectories of infants with an older sibling with autism, of whom approximately 20% receive a diagnosis of autism themselves (Ozonoff et al. 2011, 2024), and to identify critical windows of developmental divergence. Both retrospective and prospective video-based analyses have demonstrated that infants/toddlers later diagnosed with autism show reduced social engagement as reflected by fewer manually annotated gazes to faces and directed vocalizations (Osterling and Dawson 1994; Ozonoff et al. 2010; Werner et al. 2000; Werner and Dawson 2005), a finding replicated over decades and reflected in the autism diagnostic criteria (American Psychiatric Association 2013).

Repetitive behaviors, including hand flapping, have been less reliable in differentiating autistic from non-autistic children during early toddlerhood. One of the earliest studies using parent report via the Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994) found that parent report of "hand/finger mannerisms," including hand flapping, did not distinguish 20-month-olds with autism from those without. Another study analyzed retrospective home videos of children at 12 and 24 months and found group differences in social engagement but not in repetitive movements among early-onset ASD (n = 21), regressive ASD (n = 15), and typical development (TD; n = 20) (Werner and Dawson 2005).

In contrast, evidence from *prospective* designs has revealed significant differences in repetitive hand flapping. One prospective investigation manually annotated repetitive and symbolic behaviors from 20-min Communication and Symbolic Behavior Scales Developmental Profile (CSBS) behavior samples of toddlers (Wetherby and Prizant 2002; 18–24 months; ASD = 50, developmental delay [DD] = 25, TD = 50). This study found more frequent and a longer duration of body-focused repetitive movements (including hand flapping) in the autism group, relative to the DD and TD groups (Watt et al. 2008). Another study observed 12-month-old infants in 10–15 min lab-based sessions and found that those with a 24-month autism diagnosis showed higher rates of repetitive behaviors compared to infants at elevated and low likelihood without a later autism diagnosis (Elison et al. 2014). Finally, a study using the Autism Observation Scale for Infants (AOSI) at 12 and 18 months found significantly more frequent hand flapping in infant siblings who later received an autism diagnosis compared to non-autistic siblings or controls (Loh et al. 2007). The discrepancy in findings across studies (e.g., absent or minimal group differences in some early toddler participants versus apparent differences in older cohorts) could reflect variations in measurement approaches, including parent report, retrospective home-video analysis, and structured lab observations. Differences may also reflect heterogeneity in participant characteristics (e.g., developmental level or family history) or general developmental maturation, which makes repetitive hand flapping appear more clearly atypical during the second and third years of life (Cox et al. 1999; Morgan et al. 2008). Indeed, neurotypical 26-month-olds with elevated repetitive behaviors have been found to exhibit lower language and social cognitive abilities (Larkin et al. 2017).

1.2 | Automated Detection and Quantification of Repetitive Hand Flapping Using Computer Vision: Advances and Gaps

Assessing the intensity and frequency of repetitive hand flapping, in addition to its presence or absence, has been proposed as a potential way to support the distinction between typical and atypical repetitive movements (Richler et al. 2007). However, reliably quantifying these movement parameters can be challenging. Recent developments in computer vision, a subfield of artificial intelligence, have the potential to assist in quantifying such behaviors, as pose estimation algorithms can automatically identify and track anatomical landmarks in digital videos of human movement (often called "pose data"; Cao et al. 2017; Kidziński et al. 2020). The last decade has seen a rapid increase in the application of pose estimation algorithms in autism research. To date, this work has predominantly focused on using the pose data to automatically detect clinically relevant repetitive behaviors, rather than quantifying specific movement parameters (e.g., amplitude, and frequency), and has been conducted exclusively in autistic children without including a non-autistic comparison group (Barami et al. 2024; Gonçalves et al. 2012; Lakkapragada et al. 2022; Lemler et al. 2025). These automated detection algorithms have achieved a relatively high accuracy (approximately 84–96%) in identifying repetitive behaviors, including hand flapping, from video recordings across autistic children aged 1.4 to ~9 years (Barami et al. 2024; Gonçalves et al. 2012; Lakkapragada et al. 2022; Lemler et al. 2025). One recent large-scale investigation (Barami et al. 2024) analyzed 319 clinical assessments (~40 min each) from 241 autistic children (age range: 1.4–8.0 years). After applying pose estimation to identify anatomical landmarks, an action recognition algorithm detected repetitive movements (including hand flapping) in 7352 manually annotated segments, achieving 92.5% recall and 66.8% precision.

Despite these significant computational advances, several gaps remain. First, apart from the work by Barami et al. (2024), most studies have focused on autistic preschoolers or school-aged children, raising questions about how well pose estimation algorithms can automatically identify anatomical keypoints in toddlers. This is important because toddlerhood is a critical developmental period for early screening, diagnosis, and intervention (Guthrie et al. 2013; Landa et al. 2020, 2022; Ozonoff et al. 2015). Second, although the extant behavioral literature suggests that measuring the quality of repetitive hand flapping (e.g., amplitude, and frequency) could aid in distinguishing autistic from non-autistic behaviors, few studies have pursued this approach in early toddlerhood. Third, long recordings from multiple cameras may not be feasible in community-based or low-resource settings. Yet, it is unclear whether shorter videos (less than 10 min) would capture meaningful differences in repetitive hand flapping between autistic and non-autistic toddlers.

1.3 | The Current Study

The current study addresses these gaps. Using video-based markerless pose estimation, we automatically detect body keypoint (e.g., wrists, and elbows) pose data in brief (3-min), single-camera recordings of the semi-structured bubble play activity from the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999, 2012). Then, using the pose data, we quantify movement kinematics (amplitude and frequency) to examine whether these measures significantly differ between young autistic and non-autistic toddlers. Considering the extant evidence, we hypothesized that autistic toddlers would exhibit significantly more intense or vigorous arm movements during repetitive hand flapping, reflected by greater amplitude and frequency of hand oscillations.

2 | Methods

This study was approved by the Johns Hopkins Institutional Review Board. Informed consent was obtained from caregivers of all participating children prior to their participation. The study protocol was approved by the Johns Hopkins School of Medicine Institutional Review Board (IRB protocol: NA_00038069).

2.1 | Participants

Video recordings of 131 participants who completed the ADOS were randomly drawn from a larger prospective, longitudinal study of infant siblings with and without a family history of autism (Landa and Garrett-Mayer 2006). Infants and their families were recruited through an urban outpatient autism specialty clinic and research center in Baltimore, Maryland, United States. Eligibility criteria for the larger prospective study included: (1) child birth weight <1500 grams, (2) no history of severe birth trauma, severe congenital disabilities, head injury, or prenatal illicit drug or alcohol exposure. Details about ascertainment and proband autism diagnosis have been previously described for these participants (Landa and Garrett-Mayer 2006). Children were included in the current analysis if they completed (1) a 14-month assessment and (2) a confirmatory diagnostic outcome classification between 30–36 months of age by an expert clinical researcher.

Of the 131 children who met these initial eligibility criteria, 50 (38%) were identified as having at least one repetitive movement. We then applied the following exclusion criteria to each repetitive movement event to ensure high-quality pose data for the included repetitive hand flapping behaviors: (1) less than 1 s of continuous pose data (we chose 1 s to ensure a sufficient number of repetitive hand flapping cycles), (2) no repetitive hand flapping (e.g., jumping or rocking), and (3) an unequal number of frames between the original and pose estimation videos. This is because we observed scenarios in which the pose estimation algorithm failed to apply to the entire video, leading to missing frames and asynchronicity between the start and stop times of the manually annotated repetitive movements and the pose estimation output. Based on these criteria, an additional 22 (44%) were excluded to ensure robust quality of pose data for kinematic analysis. This resulted in a final analytic cohort of 28 children (ASD: $n = 14$; no ASD: $n = 14$).

2.2 | Measures

2.2.1 | Measures Used for Participant Characterization

Hollingshead Four-Factor Index of Socioeconomic Status (Hollingshead 2011**)**. A weighted combination of marital status, retired/employed status, educational attainment, and occupational prestige was used to calculate a family SES score. Scores range from 8 to 66. Higher scores reflect more socioeconomic advantage (e.g., advanced educational attainment and higher-prestige occupations), while lower scores indicate relatively fewer socioeconomic resources.

The **Mullen Scales of Early Learning (MSEL; Mullen, 1995)** is a standardized, norm-referenced developmental test for children from birth to 68 months of age. Four subscales were administered to assess children's development: Fine Motor, Receptive Language, Expressive Language, and Visual Reception. Verbal and Nonverbal Developmental Quotients (DQ)s were calculated by dividing age-equivalent subscale scores by the child's chronological age and multiplying by 100 (Messinger et al. 2013). Verbal (i.e., the average of the Receptive and Expressive Language DQs) and nonverbal (i.e., the average of the Fine Motor and Visual Reception DQs) DQs were derived.

The **ADOS** is a standardized, semi-structured interaction and observation tool developed to assess the presence or absence of autism-related social interaction, communication, play, and restricted, repetitive behaviors (Lord et al. 2012). It was administered to all children in this study as part of a clinical research assessment by research-reliable examiners unaware of participant group membership. All videos of the ADOS were recorded at 30 frames per second. Pixel resolution varied across videos. Of the 28 included participant videos, one was 240 (height) x 350 (width), six were 480 × 350, 19 were 480 × 720, one was 720 × 1290, and one was 1080 × 1920. The bubble play activity was segmented from the ADOS videos for subsequent manual annotation, pose estimation, and kinematic analysis.

Clinical best estimate and group classification. Clinical research autism classification (ASD vs. no ASD) was made by an expert clinical researcher conducting the child's assessment based on (a) ADOS classification, (b) the child's developmental history and behavioral data (including parent report forms), and (c) the child's behavior during the evaluation session.

2.3 | Manual Annotation of Hand Flapping

Behavioral Observation Research Interactive Software (BORIS) software (Friard and Gamba 2016) was used to manually annotate bubble play activity video segments from the ADOS. Videos were manually annotated frame-by-frame. A manual with operational definitions of behaviors was developed to identify each behavior's start and stop times (see Table S1). Using this manual, two research assistants with bachelor's degrees in psychology were trained to reliably annotate participants' repetitive behaviors. They were trained and supervised by R.L. and R.R., both of whom are recognized early autism experts. The research assistants' training consisted of learning the operational definitions of the repetitive behaviors of interest (as described in Table S1), reviewing frame-level start/stop rules, and manually annotating a practice/calibration subset of participants. Start times of hand flapping were coded as the exact moment (frame) the body part performing the repetitive movement began. Stop times were coded as the exact moment (frame) that the body part returned to its original position. The manual annotators were unaware of each participant's diagnostic group membership to prevent bias in manual annotations. All annotation discrepancies were resolved through consensus meetings between the manual annotators; when consensus was not reached, the manual annotators consulted with R.L. and R.R. to reach consensus.

All 131 bubble-play videos (~3 min each) were double-coded to ensure reliability in manual annotation of repetitive behaviors. The annotators coded all 131 videos to identify whether repetitive behaviors occurred and to mark their onset and offset, including hand flapping, before conducting pose estimation analyses. Kinematic features (amplitude, and frequency) were computed within these consensus-coded windows.

2.4 | Pose Estimation

We used AlphaPose trained on the COCO wholebody dataset (Jin et al. 2020) to track 133 keypoints (i.e., pixel coordinates) of the body, face, and hands in the bubble play videos (Figure 1(A), (B)). AlphaPose tracked keypoints of all persons (child, examiner, and caregiver) visible in each frame of the video; however, AlphaPose does not assign a consistent person identifier throughout the video. To address the issue of person identification, we wrote custom code to manually identify the pose estimation data associated with the child. The output of the pose estimation analysis was: (1) *x-y* pixel coordinates of keypoints of the child for all frames of the video, and (2) a video in which stick figures representing keypoint coordinates were overlaid on the original video.

2.5 | Kinematic Analysis

We analyzed repetitive hand flapping kinematics within the manually annotated start and stop times. To quantify repetitive hand flapping, we used the averaged vertical position of hand key points (wrist, interphalangeal joints, and fingertips) of the relevant hand (either both hands or only one hand). Next, we removed drift and offset based on a linear regression fit and normalized to torso height (using shoulder and hip keypoints). We removed drift to account for scenarios in which the child changed their position during repetitive hand flapping, which may affect the vertical keypoint positions; we removed offset to center the averaged hand keypoint position around zero; we normalized to torso height to account for differences in body size and variation in the appearance of the size of the child at various distances from the camera. We used the vertical position because in our dataset hand flapping was predominantly expressed in the up-down direction and less in the side-to-side direction (Figure 1(C)): we found that up-down amplitude was greater by a normalized magnitude of 0.42 (amplitude magnitude normalized by torso height) compared to side-to-side amplitude and that this difference was statistically different using a paired *t*-test ($p < 0.001$; 95% C.I. of 0.33–0.50). We calculated two kinematic parameters: amplitude and frequency of hand flapping (Figure 1(D)). Amplitude was calculated as the range (maximum normalized vertical position subtracted from minimum normalized vertical position) of the hand trajectory. Because amplitude was normalized to torso height, an amplitude value of one suggests that the up-down movement of hand flapping spanned a distance equivalent to the length of the torso; values less than one suggest a shorter distance, and vice versa for values greater than one. Frequency was calculated from the dominant frequency using a Fast Fourier Transformation. Frequency values reflect the number of hand flaps per second.

2.6 | Statistical Analysis

All statistical analyses were performed in RStudio (Version 2024.04.1 + 748; R Version 4.4.0; R Core Team 2024). We summarized hand-flapping event counts per participant by group using the mean and standard deviation and compared groups with Welch's *t*-test to allow for potential variance differences. To examine group differences, we first conducted an event-level, repeated-measures ANOVA for each dependent variable (normalized amplitude and frequency of hand flapping), which included each hand flapping event. In these analyses, we treated each instance of hand flapping as an individual observation to capture moment-to-moment variability between the ASD and no ASD outcome groups. In these models, the between-subject factor was group (ASD vs. no ASD), and the error term was specified as *Error (participant)* to account for repeated observations within each child. Because hand flapping bouts also varied in laterality by participant (i.e., some hand flapping instances were produced with one arm vs. both), we also fit an event-level model that included arm (left/right) as a within-participant factor and tested whether there was a significant group by arm interaction effect on frequency and amplitude (see Table S2).

Recognizing that hand flapping event-level analyses may influence within-participant variability, we also derived

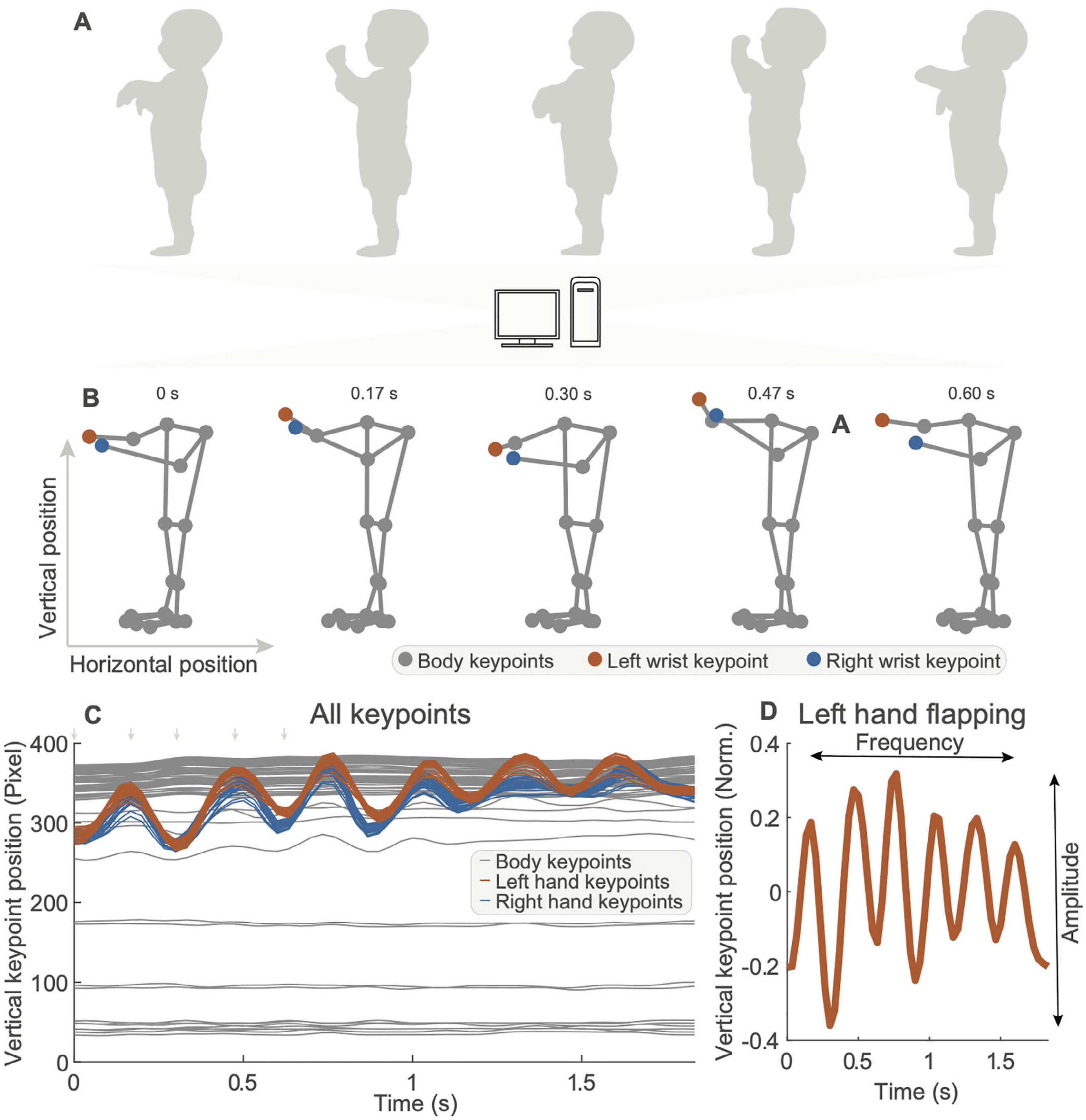

FIGURE 1 | Pose estimation of repetitive hand flapping and processing example. Digital videos of 14-month-old children during bubble play activity segments recorded as part of the ADOS (A) were tracked by a pose estimation algorithm (AlphaPose) that manually identifies key points (i.e., anatomical landmarks) of the body (B). Vertical pixel position of hand keypoints represent the up-down motion of the hand during repetitive hand flapping (C). Note that the arrows on the top of panel C correspond to the five pose images shown in panel B. We calculated amplitude and frequency from the vertical keypoint position (normalized to torso height) to quantify the kinematics of repetitive hand flapping (D).

event-averaged measures by averaging amplitude and frequency across all hand flapping instances by arm by participant (i.e., each participant then had only one amplitude and frequency metric). We then used a one-way ANOVA to examine group-level differences at the participant level. All statistical analyses were two-tailed, with significance defined as $p < 0.05$.

3 | Results

3.1 | Participant Characteristics

Participant demographic and clinical characteristics are presented in Table 1. Participants ranged in age from 13.1 to 16.7

TABLE 1 | Demographic and Clinical Characteristics by Group ($N = 28$).

Variable	ASD ($n = 14$)	No ASD ($n = 14$)	Total ($N = 28$)	p
Age, *M (SD)*	14.80 (1.03)	14.84 (0.69)	14.82 (0.86)	0.92
ASD sibling				0.002
No	0 (0.0%)	7 (50.0%)	7 (25.0%)	
Yes	14 (100.0%)	7 (50.0%)	21 (75.0%)	
Sex				0.45
Female	5 (35.7%)	7 (50.0%)	12 (42.9%)	
Male	9 (64.3%)	7 (50.0%)	16 (57.1%)	
Race				0.22
Asian	2 (14.3%)	0 (0.0%)	2 (7.1%)	
Black or African American	0 (0.0%)	2 (14.3%)	2 (7.1%)	
White	10 (71.4%)	11 (78.6%)	21 (75.0%)	
Multiracial	2 (14.3%)	1 (7.1%)	3 (10.7%)	
Hispanic				0.14
No	12 (85.7%)	14 (100.0%)	26 (92.9%)	
Yes	2 (14.3%)	0 (0.0%)	2 (7.1%)	
Maternal education				0.40
Bachelor's degree	7 (50.0%)	5 (35.7%)	12 (42.9%)	
Graduate degree	4 (28.6%)	6 (42.9%)	10 (35.7%)	
High school graduate	0 (0.0%)	1 (7.1%)	1 (3.6%)	
Some college	1 (7.1%)	2 (14.3%)	3 (10.7%)	
Some graduate school	2 (14.3%)	0 (0.0%)	2 (7.1%)	
Hollingshead, *M (SD)*	58.43 (6.66)	54.36 (12.78)	56.39 (10.21)	0.30
ADOS module				0.23
2 – Mod T	8 (57.1%)	11 (78.6%)	19 (67.9%)	
G—Mod 1	6 (42.9%)	3 (21.4%)	9 (32.1%)	
ADOS CSS, *M (SD)*	6.07 (2.73)	2.79 (1.42)	4.43 (2.71)	< 0.001
MSEL verbal DQ, *M (SD)*	68.29 (17.79)	83.62 (12.21)	75.67 (16.97)	0.02
MSEL nonverbal DQ, *M (SD)*	93.79 (14.91)	106.38 (13.02)	99.85 (15.18)	0.03

Note: Values are expressed as *M* (SD) for continuous variables and *n* (%) for categorical variables. *ADOS CSS* = Autism Diagnostic Observation Schedule comparison scores;*MSEL* = Mullen Scales of Early Learning; *DQ* = Developmental Quotient

months (M = 14.8, SD = 0.9). Compared to children in the no ASD group, those with ASD had significantly higher ADOS Calibrated Severity Scores ($p < 0.001$) and lower MSEL Verbal ($p = 0.02$) and Nonverbal Composite scores ($p = 0.03$), as well as a higher proportion of children with an older autistic sibling. Groups did not differ significantly in sex ($p = 0.45$), race/ethnicity ($p > 0.14$), maternal education ($p = 0.40$), Hollingshead socioeconomic status ($p = 0.30$), or ADOS Module ($p = 0.23$).

3.2 | Manual Annotation Results

There were 81 repetitive hand flapping events of either hand or both hands (64 with both hands, two with only the left hand, and 15 with only the right hand), resulting in 145 hand flapping trajectories. The number of repetitive hand flapping events per individual child ranged from one to ten (15 children had one event, two had two events, five had four events, one had five events, one had six events, three had seven events, and one had ten events). Of the 81 repetitive hand flapping events, 43 events were from children with no ASD, and 38 events were from children with ASD. To confirm that event-level analyses were not biased by unequal sampling, we compared the number of events per participant across groups and found no statistically significant difference between groups (no ASD: $M = 6.14$, $SD = 4.01$; ASD: $M = 5.43$, $SD = 3.67$; $t(22.79) = 0.36$, $p = 0.72$). The duration of repetitive hand flapping ranged from 1 to 19.8 s with a median of 2.0 s. There was no statistically significant difference in the duration of repetitive hand flapping between the no ASD and ASD groups (no ASD: $M = 2.48$ s, $SD = 1.95$ s; ASD: $M = 3.35$ s, $SD = 3.62$ s; $t(79) = -1.36$, $p = 0.18$).

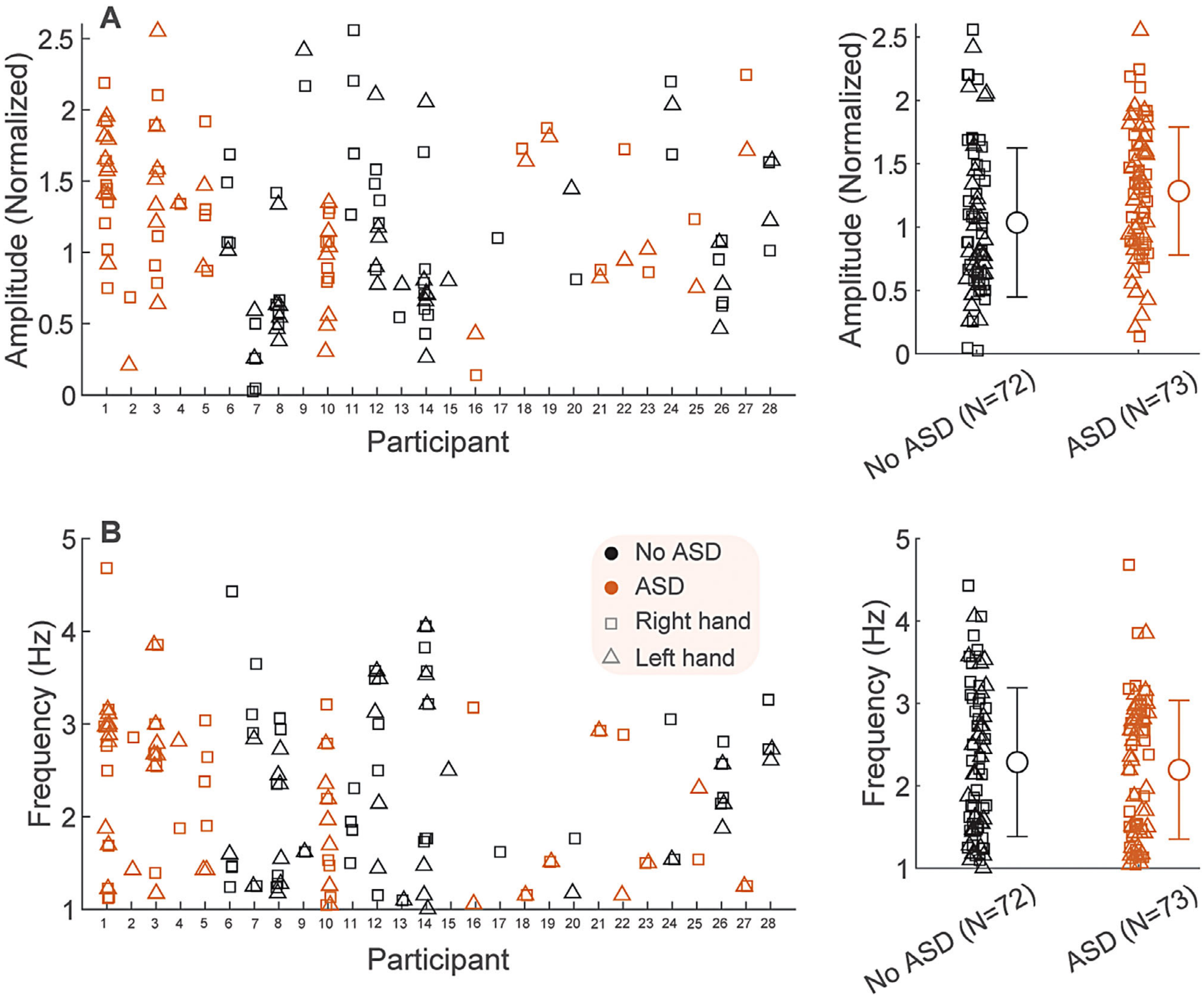

FIGURE 2 | Amplitude and frequency of repetitive hand flapping. Amplitude (A) and frequency (B) are shown for all repetitive hand flapping events for each individual (1–28). The color and shape of data points denote whether the participants had ASD (red) or no ASD (black) and which hand performed the repetitive flapping (square, right; triangle, and left). Large data points with error bars in panels on the right show mean values with standard deviation. Amplitude of hand flapping was normalized to torso height.

3.3 | Kinematic Analysis Results

Amplitude values ranged from 0.02 to 2.29 (Figure 2(A); no ASD: $M = 0.96$, $SD = 0.60$; ASD: $M = 1.18$, $SD = 0.49$), and frequency values ranged from 1.00 to 4.68 Hz (Figure 2(B); no ASD: $M = 2.14$ Hz, $SD = 0.84$ Hz; ASD: $M = 2.27$ Hz, $SD = 0.87$ Hz). Most participants exhibited a substantial range of amplitudes and frequencies across flapping events (e.g., one child had amplitudes ranging from 0.53 to 2.18 and frequencies ranging from 1.12 to 4.68 Hz).

3.4 | Event-Level Results

In a repeated-measures ANOVA with group (ASD vs. no ASD) as the between-subject factor, there was a significant main effect of group on amplitude, $F(1,142) = 4.92$, $p = 0.028$, $\eta^2_p = 0.035$. Toddlers with a later ASD diagnosis demonstrated a significantly elevated amplitude of flapping compared to those without a later ASD diagnosis. Behaviorally, this suggests that toddlers with an ASD outcome moved their hands through a larger vertical distance, corresponding to more intense or broader up-and-down flapping motions. We did not observe a significant group effect on frequency ($p = 0.464$).

3.4.1 | Event-Averaged Results

When normalized amplitude and frequency were aggregated within participant (i.e., averaged across all events of hand flapping across both arms by participant resulting in one value per child), there were no significant group differences [Amplitude: $F(1,26) = 0.00$, $p = 0.997$; Frequency, $F(1,26) = 0.79$, $p = 0.383$].

4 | Discussion

The present study examined whether toddlers with a later autism diagnosis exhibited quantifiable differences in the amplitude and frequency of repetitive hand flapping compared to toddlers who did not receive an autism diagnosis. Our event-level analysis revealed that autistic toddlers exhibited significantly greater hand flapping amplitude in a brief (∼3-min) bubble play activity. In contrast, no group differences emerged in the frequency of repetitive hand flapping. We found no significant group differences when amplitude and frequency were averaged across each participant's hand flapping events.

Our findings suggest that measuring the amplitude of repetitive hand flapping, rather than simply noting its presence or absence, may help differentiate typical from atypical repetitive movements (Richler et al. 2007). However, the significant group differences in amplitude we observed did not persist when amplitude was averaged across arms within individual participants. One possible explanation for this discrepancy is the inherently dynamic and context-dependent nature of toddler behavior; a child may exhibit one or more bouts of highly vigorous hand flapping (leading to an elevated amplitude during specific events) yet show more subdued movements in other bouts, ultimately "averaging out" these differences when all events are combined. Indeed, physiological factors (e.g., arousal level) or environmental factors (e.g., excitement around bubbles) may contribute to variability in moment-to-moment hand flapping movements (Joosten et al. 2009; O'Loghlen et al. 2025). Future research should systematically investigate how these factors affect amplitude (and possibly frequency) in larger, prospective cohorts across different environmental contexts. This approach may further clarify how moment-to-moment variability in hand flapping during brief play-based activities could support the early identification of autism-related behaviors in toddlers.

Another possible explanation for the subtle group effect observed in the current study (i.e., a significant group effect on event-level amplitude in hand flapping with no other significant group effects observed) may be due, in part, to the nature of our sample. That is, half of the no ASD group in the current study were younger siblings of an autistic proband. Prospective studies have consistently demonstrated that approximately 20% of siblings at elevated likelihood for autism will go on to receive an autism diagnosis themselves (Ozonoff et al. 2011, 2024). Although a subset of younger siblings do not meet diagnostic criteria for autism, they may display early-emerging autistic-like characteristics and later elevated autism-related behaviors, relative to children who are not at elevated familial likelihood for autism (Georgiades et al. 2013; Landa et al. 2020; Messinger et al. 2013; Ozonoff et al. 2014). For example, in a multisite cohort study, approximately 19% of infants at elevated likelihood for autism, who did not receive an autism diagnosis by age 3 years, were identified at 12 months as exhibiting a relatively high level of autism-related behaviors. These same infants later showed decreased social communication abilities at age 3 years (Georgiades et al. 2013). Consistent with our findings, Georgiades et al. reported only a small, nonsignificant difference in the parent-report ADI-R repetitive behaviors domain at age 3 between groups ($ES = 0.28$, $p = 0.11$). This observed pattern of early subclinical autism-related behaviors has also been observed through data-driven identified profiles of risk and resiliency in infants at elevated likelihood. Indeed, Landa et al. identified a profile of 14-month-olds with a greater number of restricted and repetitive behaviors as well as lower language, social, and play skills, even though only a fraction of children in that profile ultimately received an ASD diagnosis (25%; Landa et al. 2020). Considering this evidence in the context of the current study, possible subclinical but elevated autism-related behaviors among participants in our no ASD group may have decreased the likelihood of observing significant differences between groups (i.e., in the frequency of hand flapping or in the event-averaged analyses).

Finally, our pose estimation analysis presents several challenges and opportunities. The digital videos used in the present study were collected more than 20 years ago as part of a larger prospective longitudinal study of infant siblings with and without a family history of autism (Landa and Garrett-Mayer 2006; Landa et al. 2020, 2022). Digital video technology is widely available today and a core feature of smartphones and tablets with higher pixel resolutions (typically 1080 × 1920) than the videos included in this study. It is, therefore, fair to expect that pose estimation models may more accurately track anatomical key points in newly collected videos. In addition, pose estimation models are also likely to improve in the future and may be able to track key points in 3D from a single camera (Shin et al. 2024) as opposed to 2D in the present study. There are also opportunities to improve the standardization of recording conditions (e.g., camera placement and viewpoint angle, lighting conditions), which may enhance the ability to quantify hand flapping amplitude and frequency. For example, hand flapping amplitude may appear smaller if the movement is captured from a higher camera angle as opposed to a centered angle. However, such constraints may be challenging to implement with toddlers, especially those on the autism spectrum, who are highly active (Miller et al. 2020; Reetzke et al. 2021). Finally, *longer* video recordings of different play-based activities may provide more opportunities to observe various repetitive behaviors, including repetitive hand flapping. For example, Barami et al. (2024) observed that 52% of children exhibited repetitive hand flapping across a ∼36-min assessment, which is relatively higher than the 21% of children who exhibited repetitive hand flapping in the current study's 3-min bubble play activity. However, much longer video recordings require additional computational resources for pose estimation analyses and greater child compliance (e.g., remaining in view of the camera while attending to activities).

Our study has many strengths, including leveraging a prospective, phenotypically well-characterized cohort, integrating traditional manual annotation of behaviors with video-based markerless kinematic analysis using pose estimation, focusing on a relatively naturalistic yet standardized bubble play activity (from the gold-standard ADOS), and being the first to apply this specific method of kinematic analysis to young toddlers with later autism diagnoses. Despite these strengths, our approach was constrained by a relatively small sample size and a focus on only one behavior (hand flapping) in a single, brief activity (bubble play). Small clinical research cohorts are common in prospective, longitudinal studies of infants at elevated likelihood for autism (Zwaigenbaum et al. 2007), making it challenging to gather the large, heterogeneous datasets needed for robust machine-learning applications (Crippa et al. 2015). Multisite collaboration, data-sharing initiatives, and open-source repositories of annotated videos could help address this hurdle by expanding the quantity and heterogeneity of datasets. As a next step, integrating action detection methods (e.g., as in Barami et al. 2024) to automate the identification of a broader range of repetitive behaviors would reduce the reliance on manual coding and enable more comprehensive, large-scale analyses of multiple autism-related behaviors across various clinical assessment sessions.

5 | Conclusions

In summary, our findings provide preliminary evidence for the application of video-based pose estimation to conduct kinematic analyses, showing that this approach has the potential to reveal subtle yet meaningful differences in the amplitude of repetitive hand flapping between autistic and non-autistic toddlers. Larger, prospective cohorts are needed to test whether the event-level amplitude finding holds and to clarify whether this group difference in amplitude of repetitive hand flapping is stable versus transitory. It is also important to examine how environmental context and varying observation lengths influence the feasibility of our computational approach. While these findings highlight the promise of video-based pose estimation kinematic analyses in capturing nuanced aspects of toddler behavior, such techniques are not a substitute for best-practice diagnostic approaches, including direct observation, autism-specific assessments, thorough developmental and medical history, and cognitive testing. Instead, our findings underscore the potential added value of examining the *quality* of repetitive movements, such as hand flapping, to identify clinically meaningful differences during a highly dynamic developmental period.

Acknowledgments

We wish to acknowledge Brianna Hicks and Jessica Morrel for their assistance with data curation and manual annotation. We thank the children and families who participated in the original study. This work was funded by a National Institute of Mental Health R01MH59630 grant (PI: Landa), a Kennedy Krieger Institute Goldstein Innovation Award 94217 (MPIs: Reetzke, Landa), and a Simons Foundation Autism Research Initiative Human Cognitive and Behavioral Science Award 970534 (MPIs: Reetzke, Landa). We additionally acknowledge resources provided by the Kennedy Krieger Institute Intellectual and Developmental Disabilities Research Center (P50 HD103538).

Conflicts of Interest

The authors have no conflicts of interest to report.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request. A data use agreement will be required.

References

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition (DSM-5)*, 17 American Psychiatric Association.

Barami, T., L. Manelis-Baram, H. Kaiser, et al. 2024. "Automated Analysis of Stereotypical Movements in Videos of Children With Autism Spectrum Disorder." *JAMA Network Open* 7, no. 9: e2432851–e2432851.

Calkins, S. D., and N. A. Fox. 2002. "Self-Regulatory Processes in Early Personality Development: A Multilevel Approach to the Study of Childhood Social Withdrawal and Aggression." *Development and Psychopathology* 14, no. 3: 477–498.

Cao, Z., T. Simon, S.-E. Wei, and Y. Sheikh. 2017. "Realtime Multi-Person 2d Pose Estimation Using Part Affinity Fields." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7291–7299.

Cox, A., K. Klein, T. Charman, et al. 1999. "Autism Spectrum Disorders at 20 and 42 Months of Age: Stability of Clinical and ADI-R Diagnosis." *Journal of Child Psychology and Psychiatry* 40, no. 5: 719–732.

Crippa, A., C. Salvatore, P. Perego, et al. 2015. "Use of Machine Learning to Identify Children With Autism and Their Motor Abnormalities." *Journal of Autism and Developmental Disorders* 45, no. 7: 2146–2156.

Elison, J. T., J. J. Wolff, J. S. Reznick, et al. 2014. "Repetitive Behavior in 12-Month-Olds Later Classified With Autism Spectrum Disorder." *Journal of the American Academy of Child & Adolescent Psychiatry* 53, no. 11: 1216–1224.

Friard, O., and M. Gamba. 2016. "BORIS: A Free, Versatile Open-Source Event-Logging Software for Video/Audio Coding and Live Observations." *Methods in Ecology and Evolution* 7, no. 11: 1325–1330.

Georgiades, S., P. Szatmari, L. Zwaigenbaum, et al. 2013. "A Prospective Study of Autistic-Like Traits in Unaffected Siblings of Probands With Autism Spectrum Disorder." *JAMA Psychiatry* 70, no. 1: 42–48.

Gonçalves, N., J. L. Rodrigues, S. Costa, and F. Soares. 2012. "Automatic Detection of Stereotyped Hand Flapping Movements: Two Different Approaches." In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 392–397. IEEE.

Guthrie, W., L. B. Swineford, C. Nottke, and A. M. Wetherby. 2013. "Early Diagnosis of Autism Spectrum Disorder: Stability and Change in Clinical Diagnosis and Symptom Presentation." *Journal of Child Psychology and Psychiatry* 54, no. 5: 582–590.

Hollingshead, A. B. 2011. "Four Factor Index of Social Status." *Yale Journal of Sociology* 8, no. 11: 21–51.

Jin, S., L. Xu, J. Xu, et al. 2020. "Whole-Body Human Pose Estimation in the Wild." In *European Conference on Computer Vision*, 196–214. Springer International Publishing.

Joosten, A. V., A. C. Bundy, and S. L. Einfeld. 2009. "Intrinsic and Extrinsic Motivation for Stereotypic and Repetitive Behavior." *Journal of Autism and Developmental Disorders* 39: 521–531.

Kidziński, Ł., B. Yang, J. L. Hicks, A. Rajagopal, S. L. Delp, and M. H. Schwartz. 2020. "Deep Neural Networks Enable Quantitative Movement Analysis Using Single-Camera Videos." *Nature Communications* 11, no. 1: 4054.

Lakkapragada, A., A. Kline, O. C. Mutlu, et al. 2022. "The Classification of Abnormal Hand Movement to Aid in Autism Detection: Machine Learning Study." *JMIR Biomedical Engineering* 7, no. 1: e33771.

Landa, R., and E. Garrett-Mayer. 2006. "Development in Infants With Autism Spectrum Disorders: A Prospective Study." *Journal of Child Psychology and Psychiatry* 47, no. 6: 629–638.

Landa, R. J., R. Reetzke, C. B. Holingue, D. Herman, and C. R. Hess. 2022. "Diagnostic Stability and Phenotypic Differences Among School-Age Children Diagnosed With ASD Before Age 2." *Frontiers in Psychiatry* 13: 805686.

Landa, R. J., R. Reetzke, M. Tahseen, and C. R. Hess. 2020. "Early Behavioral Profiles Elucidating Vulnerability and Resiliency to Later ASD Outcomes." *Development and Psychopathology* 32, no. 4: 1217–1229. https://doi.org/10.1017/S0954579420000814.

Larkin, F., E. Meins, L. C. Centifanti, C. Fernyhough, and S. R. Leekam. 2017. "How Does Restricted and Repetitive Behavior Relate to Language and Cognition in Typical Development?" *Development and Psychopathology* 29, no. 3: 863–874.

Leekam, S. R., M. R. Prior, and M. Uljarevic. 2011. "Restricted and Repetitive Behaviors in Autism Spectrum Disorders: A Review of Research in the Last Decade." *Psychological Bulletin* 137, no. 4: 562.

Lemler, C., S. K. Kleber, L. Polzer, et al. 2025. "Semi-Automated Multi-Label Classification of Autistic Mannerisms by Machine

Learning on Post Hoc Skeletal Tracking." *Autism Research* 18, no. 4: 833–844.

Loh, A., T. Soman, J. Brian, et al. 2007. "Stereotyped Motor Behaviors Associated With Autism in High-Risk Infants: A Pilot Videotape Analysis of a Sibling Sample." *Journal of Autism and Developmental Disorders* 37: 25–36.

Lord, C., M. Rutter, P. DiLavore, and S. Risi. 1999. *Autism Diagnostic Observation Schedule-Generic (ADOS-G)*. WPS.

Lord, C., M. Rutter, P. DiLavore, S. Risi, K. Gotham, and S. Bishop. 2012. *Autism Diagnostic Observation Schedule–2nd Edition (ADOS-2)*. Western Psychological Corporation.

Lord, C., M. Rutter, and A. Le Couteur. 1994. "Autism Diagnostic Interview-Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals With Possible Pervasive Developmental Disorders." *Journal of Autism and Developmental Disorders* 24, no. 5: 659–685.

Messinger, D., G. S. Young, S. Ozonoff, et al. 2013. "Beyond Autism: A Baby Siblings Research Consortium Study of High-Risk Children at Three Years of Age." *Journal of the American Academy of Child & Adolescent Psychiatry* 52, no. 3: 300–308.e1. https://doi.org/10.1016/j.jaac.2012.12.011.

Miller, M., S. Austin, A.-M. Iosif, et al. 2020. "Shared and Distinct Developmental Pathways to ASD and ADHD Phenotypes Among Infants at Familial Risk." *Development and Psychopathology* 32, no. 4: 1323–1334.

Morgan, L., A. M. Wetherby, and A. Barber. 2008. "Repetitive and Stereotyped Movements in Children With Autism Spectrum Disorders Late in the Second Year of Life." *Journal of Child Psychology and Psychiatry* 49, no. 8: 826–837.

O'Loghlen, J., M. McKenzie, C. Lang, and J. Paynter. 2025. "Repetitive Behaviors in Autism and Obsessive-Compulsive Disorder: A Systematic Review." *Journal of Autism and Developmental Disorders* 55, no. 7: 2307–2321.

Osterling, J., and G. Dawson. 1994. "Early Recognition of Children With Autism: A Study of First Birthday Home Videotapes." *Journal of Autism and Developmental Disorders* 24, no. 3: 247–257.

Ozonoff, S., A.-M. Iosif, F. Baguio, et al. 2010. "A Prospective Study of the Emergence of Early Behavioral Signs of Autism." *Journal of the American Academy of Child & Adolescent Psychiatry* 49, no. 3: 256–266.

Ozonoff, S., G. S. Young, A. Belding, et al. 2014. "The Broader Autism Phenotype in Infancy: When Does It Emerge?" *Journal of the American Academy of Child & Adolescent Psychiatry* 53, no. 4: 398–407.

Ozonoff, S., G. S. Young, J. Bradshaw, et al. 2024. "Recurrence Risk for Autism Spectrum Disorder: Updated Estimates From the Baby Siblings Research Consortium." *Pediatrics*.

Ozonoff, S., G. S. Young, A. Carter, et al. 2011. "Recurrence Risk for Autism Spectrum Disorders: A Baby Siblings Research Consortium Study." *Pediatrics* 128, no. 3: e488–e495. https://doi.org/10.1542/peds.2010-2825.

Ozonoff, S., G. S. Young, R. J. Landa, et al. 2015. "Diagnostic Stability in Young Children at Risk for Autism Spectrum Disorder: A Baby Siblings Research Consortium Study." *Journal of Child Psychology and Psychiatry* 56, no. 9: 988–998.

R Core Team. 2024. R: A Language and Environment for Statistical Computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Reetzke, R., A. Iosif, B. Hatch, et al. 2021. "Patterns of Objectively Measured Motor Activity Among Infants Developing ASD and Concerns for ADHD." *Journal of Child Psychology and Psychiatry*.

Richler, J., S. L. Bishop, J. R. Kleinke, and C. Lord. 2007. "Restricted and Repetitive Behaviors in Young Children With Autism Spectrum Disorders." *Journal of Autism and Developmental Disorders* 37: 73–85.

Shin, S., J. Kim, E. Halilaj, and M. J. Black. 2024. "Wham: Reconstructing World-Grounded Humans With Accurate 3d Motion." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2070–2080.

Thelen, E. 1979. "Rhythmical Stereotypies in Normal Human Infants." *Animal Behaviour* 27: 699–715.

Watt, N., A. M. Wetherby, A. Barber, and L. Morgan. 2008. "Repetitive and Stereotyped Behaviors in Children With Autism Spectrum Disorders in the Second Year of Life." *Journal of Autism and Developmental Disorders* 38, no. 8: 1518–1533.

Werner, E., and G. Dawson. 2005. "Validation of the Phenomenon of Autistic Regression Using Home Videotapes." *Archives of General Psychiatry* 62, no. 8: 889–895.

Werner, E., G. Dawson, J. Osterling, and N. Dinno. 2000. "Brief Report: Recognition of Autism Spectrum Disorder Before One Year of Age: A Retrospective Study Based on Home Videotapes." *Journal of Autism and Developmental Disorders* 30, no. 2: 157.

Wetherby, A. M., and B. M. Prizant. 2002. *Communication and Symbolic Behavior Scales: Developmental Profile*. Paul H Brookes Publishing.

Yoder, P., and F. Symons. 2010. *Observational Measurement of Behavior*. Springer Publishing Company.

Zwaigenbaum, L., A. Thurm, W. Stone, et al. 2007. "Studying the Emergence of Autism Spectrum Disorders in High-Risk Infants: Methodological and Practical Issues." *Journal of Autism and Developmental Disorders* 37: 466–480.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supplementary File 1: Table S1. Guidelines for the Manual Annotation of Repetitive Behaviors; Table S2. Event-level ANOVA including "arm" (left/right) as a within-participant factor