Feasibility of accelerometer technology with individuals with autism spectrum disorder referred for aggression, disruption, and self injury

Mindy Scheithauer [a,b,*], Shruthi Hiremath [c], Audrey Southerland [c], Agata Rozga [c], Thomas Ploetz [c], Chelsea Rock [b], Nathan Call [a,b]

[a] Emory University School of Medicine, 1920 Briarcliff Rd, Atlanta, GA 30329, USA
[b] Children's Healthcare of Atlanta, Marcus Autism Center, 1920 Briarcliff Rd, Atlanta, GA 30329, USA
[c] Georgia Institute of Technology, 756 W Peachtree St NW, Atlanta, GA 30308, USA

A B S T R A C T

Background: Most research on aggression, self-injury, and disruption in autism spectrum disorder (ASD) has relied on caregiver report or direct observation, both of which have limitations. Past studies demonstrate preliminary evidence for direct detection of these behaviors using accelerometers, but additional research is needed to determine the feasibility during actual clinical assessments and times when a therapist cannot be present for direct observation, as measurement during these periods has the most applied significance.
Aims: This study addressed these gaps by evaluating the feasibility of accelerometer use with children with ASD and severe aggression, self-injury, and disruption in clinical and home contexts.
Methods and Procedures: We evaluated the feasibility of individuals with ASD wearing accelerometers during behavioral assessments following structured habituation procedures. We also evaluated the feasibility of caregivers applying sensors to individuals with ASD in the home setting.
Outcomes and Results: Most participants passed habituation and tolerated sensors during behavioral assessments (e.g., functional analyses). Caregivers applied sensors in the home with variable fidelity with wear time duration and data-collection.
Conclusions and Implications: The feasibility of using accelerometers with this population is promising and should be explored further in future research.

What does this paper add?

This paper evaluates the feasibility of using accelerometers with individuals with ASD who engage in severe aggression, self-injury, and disruption with details on habituation outcomes. We evaluate feasibility in a novel context where participants worse sensors during actual behavioral clinical assessment procedures (i.e., functional analyses). Last, we extend feasibility analyses to situations where caregivers are responsible for applying sensors in the home setting. These are all necessary extensions to expand research on

* Corresponding author at: Emory University School of Medicine, 1920 Briarcliff Rd, Atlanta, GA 30329, USA.
  E-mail addresses: Mindy.Scheithauer@choa.org (M. Scheithauer), shruthi.hiremath@cc.gatech.edu (S. Hiremath), asoutherland3@gatech.edu (A. Southerland), agata@gatech.edu (A. Rozga), thomas.ploetz@gatech.edu (T. Ploetz), Chelsea.Rock@choa.org (C. Rock), Nathan.Call@choa.org (N. Call).

wearable technology to automatically detect behavior in clinically meaningful contexts.

1. Introduction

Individuals with Autism Spectrum Disorder (ASD) have a higher prevalence of behavior that interferes with adaptive development (e.g., aggression, self-injury, disruption) compared to individuals without ASD. Prevalence estimates vary based on the measurement system and sample, but large-scale studies suggest that 36 % of youth with ASD engage in aggressive behavior (Edelson, 2021), 42 % in self-injury (Steenfeldt-Kristensen et al., 2020), and 12 % have comorbid disruptive/impulse-control/conduct disorders (Lai et al., 2019),.

The majority of research on the assessment and treatment of aggression, self-injury, and disruption (i.e., targeted behavior[1]) relies on caregiver completed questionnaires or direct observation of the behavior. Both caregiver report and direct observation are useful tools that facilitate research, but there are limitations to both approaches. Caregiver report is an essential tool for treatment development but may be burdensome for caregivers who already have multiple demands on their time. Additionally, caregiver report often relies on retrospective recall, which has the potential for report bias based on extraneous variables such as parenting stress (Stokes et al., 2011) and issues with recalling behavior over extended periods. Direct observation accounts for some of these biases, but it is incredibly resource intensive, requiring one to two observers to score behavior for several hours for a representative sample (Sharp et al., 2015; Tiger et al., 2013). This resource intensiveness often limits the accessibility of services that rely on direct observation to guide assessment and treatment decisions.

Technology that can automatically detect behaviors with accelerometers may assist in circumventing these issues. Accelerometers are sensors, often worn on the body, that detect the magnitude and direction of acceleration to measure movement in three-dimensional (3-D) space (Arvidsson et al., 2019). One use of accelerometry data is detecting the intensity of the signal via the magnitude of the raw 3-D signal.

Using sequential data analysis techniques from the machine learning domain, researchers can then model continuous streams of raw, tri-axial acceleration data to detect occurrences of target higher level behavior. Sequential analysis models such as Hidden Markov Models (Khreich et al., 2012), or more recently Recurrent Neural Networks such as Long Short Term Memory and its variants (Yu et al., 2019), make use of probabilistic modeling to learn associations between underlying patterns in the data stream and the high-level activities of interest. In the supervised learning setting, the analysis uses human generated annotations of the targeted behavior (often called "ground truth") to develop parameters of the sequential models for the behavior of interest. Automated assessment models are trained using portions of labeled data, spanning over a few sessions, and validated using different paradigms. In most cases, if enough data is available, a leave one subject out paradigm is employed to ascertain the robustness of the developed model across different populations and demographics. Good performance on a leave-one-session out protocol is indicative of effective personalization of the developed model. The model training spans two main procedures: segmentation and classification. Change Point Detection methods are utilized to identify the boundary between when the activity of interest occurs, whereas in the classification stage, a label is assigned to the segmented-out activity instance. Automated systems developed in this manner can segment portions of data of interest from the continuous stream and classify them as one of the target behavior categories when deployed for use.

In previous work (Albinali et al., 2009) studying repetitive behavior, an implicit windowing technique is used to segment data streams into 1-second windows. As is the norm, feature engineering (Plötz et al., 2011) helps learn useful information from underlying raw data streams and makes the learning procedure robust to noise. A cross-validation approach can then be used to find an appropriate classification backend based on performance metrics such as precision, recall, and averaged F1 scores. An explicit segmentation technique can be used (Plötz et al., 2012) that makes use of both energy and orientation change information collected from the sensor data stream to identify seed points that are representative of segments of activity instances. Similar to the procedure above, the analysis learns features over these identified segments and a classification backend assigns a behavior label to each of these segments.

Accelerometry data, combined with sequential analysis methods, have been successfully used in general pediatric healthcare research, measuring variables such as sleep, physical activity, and seizures (e.g., Breitenstein et al., 2020; Milošević et al., 2017; Voss et al., 2017). This approach can feasibly and accurately measure similar variables in children with ASD (Garcia et al., 2019; Wachob & Lorenzi, 2015) and has been extended to other clinically significant behavior with individuals with ASD, such as repetitive behavior (e. g., Gilchrist et al., 2018; Goodwin et al., 2011; Rad et al., 2018).

Extending the use of accelerometers to behavior such as aggression and self-injury has several potential advantages, including the capability to collect a direct measure of behavior that can corroborate caregiver report, while reducing the resources required for direct observation. Specifically, if accelerometers can eventually automatically detect behaviors such as aggression and self-injury, therapy centers can reduce personnel required for collecting direct observation data during assessment and treatment, reducing resources required for these services. Accelerometers may also allow for a direct measurement system in settings where observation by trained observers is difficult (e.g., home/school settings or overnight).

---

[1] Aggression, self-injury, and disruption have historically been referred to as "problem behavior" or "challenging behavior". Recently, these terms have received criticism for their negative and value-laden interpretation from self-advocacy groups and others focused on the important neurodiversity movement. Topography-specific language is suggested (e.g., Bottema-Beutal et al., 2020). However, given that we are discussing a cluster of behavioral responses, listing each topography throughout the manuscript is verbose. Given that we are discussing this behavior as it is the direct target of assessment and treatment, we decided the term "targeted behavior" is most appropriate to ensure conciseness while avoiding negative connotations.

Children with ASD who engage in aggression and/or self-injury are more likely to experience sensory sensitivities (van den Boogert et al., 2021) and distress with changes to their environment (Edelson, 2021), which may place them at risk for noncompliance with wearing accelerometers. Fortunately, there is some preliminary evidence suggesting that children with ASD who engage in these behaviors will wear accelerometers (Cantin-Garside et al., 2020; Goodwin et al., 2019) and other devices worn on extremities (e.g., Ferguson et al., 2019; Northrup et al., 2022). Goodwin et al. (2019) used accelerometers with 20 individuals with intellectual and developmental disabilities in an inpatient facility and successfully predicted aggressive episodes using accelerometry data, combined with other physiological measures. Cantin-Garside et al. (2020) used accelerometers with children with ASD to measure self-injury in a laboratory setting. Ploetz et al. (2012) provides additional evidence for automatic detection of targeted behaviors by demonstrating that accelerometers can automatically detect aggression and self-injury using confederate researchers engaging in simulated behavior.

This past research is very promising regarding the potential use of accelerometers to automatically detect targeted behaviors such as aggression and self-injury. Thus, the goals of the current study are to extend this past work. First, past studies evaluating accelerometer feasibility with children with ASD and targeted behavior incorporated an inclusion criterion of tolerating the sensors, with some studies mentioning the use of a desensitization (i.e., habituation) procedures to promote tolerance Cantin-Garside et al., 2020; Goodwin et al. (2011); Rad et al. (2018)). However, those studies did not consistently report the time required for habituation or percentage of participants excluded based on issues with tolerating the sensors. This information is crucial in planning for large-scale studies to build on this past work.

Second, if accelerometer technology is to replace other measures of behavior (e.g., direct observation) in actual clinical settings, such as treatment clinics using applied behavior analysis (ABA), it is important to establish the feasibility of clients wearing the accelerometer sensors in these settings and during behavioral assessments. For example, clients will need to tolerate accelerometers during functional analyses, which are the gold standard assessment to guide interventions for behavior reduction (Saini et al., 2019). Functional analyses set up antecedents/establishing operations that are likely to evoke targeted behavior, as such, it is possible that participants' compliance with sensors will diminish in this context. For example, if a child engages in escape-maintained targeted behavior, the child is likely to exhibit these behaviors during a functional analysis condition when a therapist presents demands. It is possible that noncompliance with wearing sensors (e.g., attempting to remove sensors) will occur in the same response class as other escape-maintained behaviors. Thus, the child may attempt to remove sensors in this context when they previously tolerated them during habituation procedures with no demands. The current evaluation assesses whether participants will continue to wear sensors post-habituation procedures during functional analyses of targeted behavior.

Third, if accelerometer technology is successful in detecting targeted behavior, an impactful future application is when a clinician is not present to conduct direct observations (e.g., home settings). Thus, another important question is whether caregivers (e.g., parents) are able and willing to apply accelerometer sensors in the home when a clinician is not present. Thus, we sent accelerometers home with caregivers and asked them to apply them to their children in the home setting.

The current project aims to advance the research in this area by assessing a) duration and success rate of habituation procedures; b) whether participants continue to wear the sensors post habituation during clinical behavioral assessments; and c) caregiver integrity with applying the sensors with their child in the home.

2. General methods

The study consisted of two parts with different participants with ASD. Part 1 participants attended in-clinic sessions and, following habituation procedures, wore accelerometer sensors during behavioral assessments conducted at an intensive outpatient clinic (aims a and b above). Part 2 participants completed most of the study in their homes, where the participants' legal guardians (referred to as caregivers throughout the manuscript) applied the sensors (outcome c above).

2.1. Compliance with ethical standards

This National Institute of Health funded the study through a grant awarded to the last author (R21 HD086491). None of the authors have any conflicts of interest with the study method or outcomes. The institutional review board (IRB) of the first author's primary institution approved the study. Caregivers of all participants consented to participation in the study after a detailed informed consent conversation with an investigator. The IRB waved formal assent for all participants due to the participants' level of intellectual functioning, but researchers reviewed procedures with participants using developmentally appropriate language prior to participation. Researchers performed the study in accordance with the ethical standards outlined by the 1964 Declaration of Helsinki and later amendments.

3. Method and Material: Part 1 – clinic evaluation

3.1. Participants

Researchers recruited Part 1 participants from clients admitted to an intensive outpatient clinic (30 hrs a week) that specializes in the treatment of aggression, self-injury, disruption, and other targeted behavior using function-based ABA strategies in the Southeastern United States. This clinic admits youth with an intellectual and developmental disability whose behavior has caused serious risk of physical harm to self or others and/or has demonstrated resistance to other less intensive treatments. The clinic is equipped with padded austere session rooms with one-way observation windows and is staffed with three behavior technicians per client and a

doctorate level behavior analyst (BCBA-D®) supervising each case. We approached clients who met the following inclusion criteria: ages 4–13; English speaking caregivers; a caregiver reported diagnosis of ASD; and a primary concern of aggression, self-injury, or disruption with extremities (i.e., involving arm or leg movement). Recruitment occurred from 2017 to 2019.

The researcher approached the caregivers of 35 potentially eligible participants, of which 33 consented. For consented participants, caregiver reported demographics showed a mean age of 8.90 years ($SD = 2.29$); 78.69 % male ($n = 26$); 42.42 % White ($n = 14$), 42.42 % Black ($n = 14$), 6.06 % more than one race ($n = 2$), and 9.09 % Asian ($n = 3$).

24 participants completed Part 1 of the study in its entirety, meaning they successfully completed the habituation protocol and wore the sensors during behavioral assessments (Fig. 1). Demographics of these 24 participants showed a mean age of.82 years ($SD = 2.11$); 57.58 % male; 50 % White, 37.5 % Black, 8.33 % Asian, and 4.17/5 more than one race. Mean height was 51.67 in ($SD = 9.02$) and weight was 78.61 lbs ($SD = 37.71$). A researcher confirmed the ASD diagnosis for all participants with the Childhood Autism Rating Scale (CARS; Schopler, Wellman, & Love, 2010). All participants engaged in multiple topographies of targeted behavior with extremities including aggression (87.5 %), disruption (75.0 %), and self-injury (78.17 %). Participants also engaged in other behaviors, including pica (16.67 %), elopement or dropping (29.17%), screaming or inappropriate vocalizations (29.17 %), dangerous acts (20.83 %), and disrobing (4.17 %). Parents of 21 of the participants completed the Vineland II at the start of the study and all participants exhibited deficits in adaptive skills ($M = 53.77$, range 26–78, $SD = 12.37$).

Demographics of the nine excluded participants showed a mean age of 9.11 years ($SD = 2.85$); 77.78 % male; 55.56 % Black, 22.22 % White, 11.11 % more than one race, 11.11 % Asian.

3.2. Procedures

3.2.1. Accelerometer sensor information

The accelerometer sensors included small (6 mm × 21.5 mm×31.5 mm) sensors (Axivity AX3© data loggers) housed in a fabric pouch designed for this project to ensure the sensors fit snuggly in each pouch (Fig. 2). The researcher attached the sensors with wristbands (plastic bands that enclosed with a small plastic push-through enclosure or Tyvek® bands if the push-through posed a pica risk) that looped through the back of the fabric pouch. The researcher also encircled the pouch with masking tape to further prevent removal of the sensor. Each sensor came with an embedded micro electromechanical system (MEMS) 3-axis accelerometer and a 512 MB Flash based onboard memory that supported up to 14 days of continuous logging at 100 Hz. Battery life exceeded 14 days, thus the limiting factor for duration of continuous recording was memory. The sensor was hermetically encapsulated in a tough macromelt polymer, and was shock proof, food safe, wipe clear, and sterilizable using alcohol.

3.2.2. Habituation

We conducted habituation between assessment and treatment time during the participant's normal treatment day at the clinic, generally during a group activity or lunch. Depending on the participant's clinic protocol, researchers conducted habituation either in austere session rooms with padded floors/walls or communal areas with several toys and other individuals present.

The researcher started each habituation session seated near and interacting with the participant and gave the participant preferred
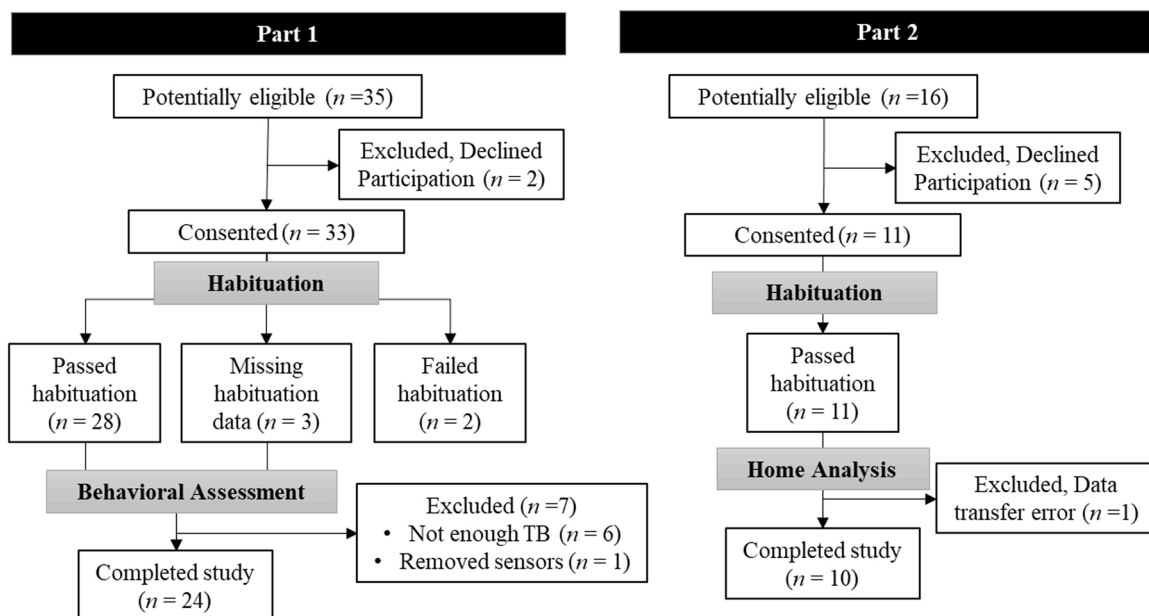

Fig. 1. CONSORT diagram for recruitment and exclusion.

Fig. 2. Accelerometer sensors, pouches, and wrist/ankle bands.

toys and activities. Researchers selected Toys, activities, and type of attention based on suggestions from the clinical team. The clinical team consisted of trained behavioral therapist who previously conducted preference assessments with clients as part of their routine clinical admission. Thus, while we did not conduct formal preference assessments as part of the study, we selected items informed by these assessments. The initial interaction (before introducing the accelerometer devices) lasted approximately 5 min, unless the participant engaged in avoidant movement or negative vocalizations (see definitions below), in which case the researcher continued in this phase (i.e., close proximity without presentation of the sensors) until these behaviors stopped for at least 2 min. After this, the therapist showed the child the sensors and wristbands, explained the sensors and how they are worn, and modelled attaching the

Table 1
Operational definitions.

Behavior	Definition
Aggression	Any instance or attempt in which the child's:
	• Hand, head, or foot contacts another person from ≥ 6"
	• Throws an object a distance of ≥ 6" that comes within 2' of a person
	• Fingers grasp another person's body and squeezes or clothing.
	• Nails contact another person's skin and drag ≥ 1" (scratching).
	• Teeth contacts another person's body and/or clothing.
	• Fingers grasp another person's hair and pulls.
	• Chin presses against someone.
	• Hands pulls or pushes someone ≥ 6" from their starting position.
Self-Injury	Any instance or attempt in which the child's:
	• Hand(s), open or closed, comes in contact with his head from ≥ 6"
	• Fingernails come in contact with his body and drags across the skin ≥ 1"
	• Head comes into contact with surface or object from ≥ 6".
	• Teeth make contact with any part of his/his body or clothing.
	• Body slams into objects or walls from ≥ 12".
Disruption	Any attempt or instance in which the child is (excludes appropriate toy play):
	• Throws an item ≥ 6", outside of aggression.
	• Rips an item.
	• Tilts furniture ≥ 45 degrees or moves it ≥ 2'
	• Foot contacts a surface or object ≥ 6"
	• Hand/arm contacts an object or surface from ≥ 6"

sensors to wrists and ankles.

Next, the formal habituation session began. The researcher applied the four sensors, one to each wrist and ankle, while the participant maintained access to preferred items and attention. When possible, the researcher applied the wristbands with the sensors below clothing (e.g., under a long sleeve shirt) to discourage removal attempts. The researcher blocked removal attempts and redirected the participant to preferred items. All researchers held certification in procedures to safely manage and redirect targeted behavior. The researcher collected frequency data in 1 min intervals on the child's targeted behavior (operational definitions developed by the clinical team; an example of standard operational definitions is in Table 1), negative vocalizations (defined as screaming, cursing, and negative verbalizations about the researcher or procedures, such as "go away" or "stop"), and avoidant movements (defined as fingers grasping or scratching at the sensor, moving at least two steps away from the researcher, or placing hands between the sensor and the wrist/ankle). Screaming was omitted for some participants if the clinical team reported screaming in most settings, even with access to preferred items and attention (suggesting automatically maintained repetitive vocalizations). If the participant's clinical protocol included blocking specific behavior (e.g., self-injury), this continued in habituation.

If the participant went five consecutive intervals (i.e., 5 min) without negative vocalizations or avoidant movements, the researcher terminated the session, removed the sensors, and the participant "passed" habituation. If the participant did not pass, the researcher removed the sensors after 30 min. Sessions also terminated early if a participant's schedule dictated he/she should return to clinical sessions. Researchers initiated a new session at a later time based on the clinical schedule for up to three attempts per day for up to three days (nine total attempts). Researchers terminated habituation prior to nine attempts if negative vocalizations and avoidant movements were not on a decreasing trend or behavior was unsafe (compared to the level of targeted behavior the child emitted on a regular basis – a decision made by the principal investigator in collaboration with the clinical team). If habituation stopped before meeting the pass criteria, the participant failed habituation and did not complete any subsequent study procedures.

3.2.3. Feasibility of wearing sensors during behavioral assessments

After habituation, a researcher met with the therapy team and participant during regularly scheduled treatment hours. Data collection occurred in the early portion of the participants' admissions when therapists conducted functional analyses (Beavers et al., 2013; Iwata et al., 1994) or baseline assessments (sessions very similar to functional analyses; Scheithauer et al., 2020). Thus, all data collection from Part 1 of the study occurred during already occurring clinical activities under contingencies likely to evoke targeted behaviors. All sessions occurred in austere rooms equipped with a one-way window and adjoining observation space. Per clinic procedures, at least two clinical staff members accompanied participants, one who conducted the assessment sessions while the other collected data on targeted behavior. Material in the room varied for the specifics of the assessment but often included a table, chair(s), preferred items, and demand materials. During a break between therapy sessions, the researcher applied the sensors to the participants' wrists and ankles and then remained in the observation space for the rest of the data collection period to record any removal attempts or other issues with the study procedures.

We recorded all behavioral assessments using cameras built into the session room. Most data collection was done by the participant's therapy team using Behavior Data Analysis and Collection System (BDACS), which allows therapists to collect real time data on behavior using keystrokes creating time stamped data. During a few times when therapists could not collect live data (e.g., the data collector entered the room to assist with managing behavior or technical issues), the researcher viewed the video after completion of the observation and scored it using a single viewing with BDACS at real time speed (to replicate live coding). Data-collectors scored the frequency of targeted behavior with continuous data-collection.

Data collection continued until the child engaged in 100 instances of targeted behavior with their extremities, including aggression, self-injury, and disruption (based on live coding). The length of each data collection period varied based on the frequency of the participant's behavior and the availability of research staff. We excluded participants if the clinical team discontinued assessment or moved to a common area that did not have the appropriate video technology before capturing 100 instances of targeted behavior. The researcher would have also discontinued participation if the therapy team indicated study procedures interfered with the delivery of best practice clinical care.

3.3. Data analysis

To measure the feasibility of recruitment, we calculated the percentage of presumably eligible participants that consented to the study (consented participants over all participants approached). For feasibility of habituation, we calculated the percentage of participants who consented that passed habituation. Additionally, we scored the duration of habituation and frequency of negative vocalizations, avoidant movements, and targeted behavior in habituation.

Last, we determined the percentage of participants who passed habituation but we subsequently excluded for not tolerating wearing the sensors during the behavioral assessment. or excluded because the clinical team indicated sensors interfered with clinical work. We also tracked participants excluded because we did not observe enough targeted behavior (defined as 100 instances).

We determined an a priori power analysis on the amount of data needed for machine learning analyses was not feasible as we did not have sufficient information on anticipated effect sizes. Past research has collected a mean of 20–30 instances of targeted behavior per participant (Cantin-Garside et al., 2020; Goodwin et al., 2019). However, these studies evaluated fewer total topographies (self-injury or aggression alone) compared to the current evaluation. Thus, we decided to take a conservative estimate of 100 instances of targeted behavior per participant to account for the potential of having several model parameters evaluated separately for each behavior class and to account for potential variability in model parameters across participants. Additionally, from a feasibility perspective, we wanted participants to wear the sensors for a sufficient sample of behavior to draw a reasonable conclusion that the

participant would continue to tolerate the sensors throughout behavioral assessment and treatment evaluations. For clinical utility, participants will need to wear the sensors throughout assessment/treatment for the technology to replace live data collection.

4. Results – Part 1

4.1. Recruitment feasibility

The researcher discussed the study with the caregivers of 35 potentially eligible participants. Of these, two chose not to participate: 94.29 % of potentially eligible participants consented (Fig. 1).

4.2. Habituation data

Of the 33 participants who enrolled, two did not pass habituation, both due to persistent attempts to remove the sensors that was not a decreasing trend (habituation pass rate of 93.94%). We did not conduct habituation prior to applying the sensors during the behavioral assessment for one participant due to a protocol error; however, this participant tolerated the sensors during the behavioral assessment despite this error. There was missing habituation data for two participants (also due to a protocol error), but researchers marked each as passing habituation.

For the remaining 28 participants, habituation took a mean of 11.61 min (range 5–45), which occurred in an average of 1.64 sessions (range 1–6) held across 1.39 days (range 1–3). The 28 participants engaged in a mean of 2.57 instances (0.15 RPM) of targeted behavior and 12.59 instances (0.37 RPM) of avoidant movements and negative vocalizations during habituation with significant variability (range 0–30 instances for targeted behavior and 0–103 instances for avoidant movement and negative vocalizations).

4.3. Feasibility of wearing sensors during behavioral assessments

Of the 31 participants who wore sensors during the behavioral assessment, seven met exclusion criteria. Only one participant removed the sensors during the behavioral assessment. Of note, this participant also engaged in the most avoidant movements and negative vocalizations (103) and took longest to pass habituation (45 min).

The remaining six participants met exclusion because they engaged in fewer than 100 instances of targeted behavior prior to the treatment team electing to stop assessment in a session room where video was available, either because the clinical team moved assessment to a more naturalistic environment or started treatment. No participants met exclusion due to interference with clinical care. Thus, 24 participants successfully completed the study (68.57 % of individuals approached; 72.73 % of participants consented; and 77.52 % of participants passing habituation). These 24 participants worse sensors during behavioral assessments for a mean of 6.74 h with significant variability across participants ($SD = 5.18$ h, range 32 min to 18.30 h).

5. Method: Part 2 - home data-collection

5.1. Participants

Researchers recruited Part 2 participants from clients waiting for services in an outpatient treatment program (different program but the same clinic as Part 1 participants). This program offered weekly caregiver training services on behavioral assessment and treatment strategies for targeted behavior and referrals generally represent youth with mild-to-moderate severity targeted behavior. The researcher reviewed the medical records of children on the waitlist for this program and contacted families with English-speaking caregivers based on the presence of targeted behavior with extremities, a caregiver reported diagnosis of ASD, and ages 4–13 years. Eleven participants enrolled in Part 2, with one participant subsequently excluded due to issues with data-transfer from the sensors to a hard drive after participation ended.

Demographics of the ten participants showed a mean age of 7.55 years ($SD = 1.66$); 90.00 % male; 70.00 % Black and 30.00 % White. A researcher confirmed the ASD diagnosis with a CARS for seven participants. Two participants recently completed diagnostic evaluations with an ADOS (n = 1) or SCQ (n = 1) that confirmed ASD and one participant had missing data with no diagnostic confirmation. Caregivers of nine participants completed the Vineland 2 and all reported delays in adaptive skill development ($M = 66.33$, range 53–79, $SD = 10.48$). The caregiver in the study was the individual who cared for the child the majority of the time when he/she was in the home. This was a biological parent for nine participants and a grandparent for one.

5.2. Procedures

5.2.1. In-clinic appointment

Participants attended one appointment in the clinic. During this appointment, the researcher implemented the habituation protocol, with a few modifications to replicate how habituation might be executed if primarily conducted by a caregiver in a future study. Specifically, researchers implemented the same contingencies as described in Part 1, but the researcher gradually shifted the role of blocking sensor removal to the caregiver. Instead of determining passing of habituation using a predetermined criterion, each participant wore the sensors until the caregiver confirmed that he/she was confident that the participant would keep them on at home.

After initial habituation, the researcher removed the sensors and used behavioral skills training. to teach the caregiver how to apply

and remove the sensors and use the home data sheet. The researcher first explained how to apply the sensors, then modelled application by putting the sensors first on themselves or another researcher and subsequently the participant. After this, the researcher asked the caregiver to apply, and subsequently remove, the sensors on the participant's wrists and ankles and provided vocal feedback (and additional modelling if needed) for any errors. Steps included the caregiver applying the four sensors to the corresponding applicator pouch and extremity and correct placement on child (i.e., tightness, direction of sensor). Re-application occurred until the caregiver demonstrated independence with all steps.

The home datasheet asked the caregiver to record the time the sensors were applied and removed and to note any issues or removal attempts. When the participant wore the sensors, the researcher asked the caregiver to score 30 min partial interval data on targeted behavior (i.e., mark yes/no as to whether the behavior occurred at all in each 30 min interval). The researcher told the caregiver to apply the sensors for 4 hrs a day for 5 days when the participant was awake and in the caregiver's presence (thorough instructions were also on the data sheets). The sensors had sufficient charge and memory for the data-collection period, thus the researcher did not provide instructions for charging devices or uploading data.

5.2.2. Home visit

Approximately one week ($M = 7.4$ days) after the in-clinic visit, the researcher conducted a home visit. The researcher repeated the steps from the in-clinic appointment to ensure continued participant compliance with the sensors and maintenance of the caregiver's skills of application and removal of the sensors and data-collection. The researcher reminded the caregiver to apply the sensors at least 4 hrs a day for 5 days and to collect data on targeted behavior. The sensors were fully charged and had sufficient battery and storage for the duration of home data collection. The researcher returned in approximately one week ($M = 8.4$ days) to collect the sensors and data.

5.3. Data analysis

Similar to Part 1, for the purpose of feasibility, we measured presumably eligible participants who consented to the study. Additionally, we measured study drop out, which included the caregiver choosing to discontinue the study at any point prior to the last home visit.

To measure caregiver integrity with applying the sensors, we evaluated the caregiver reported wear time (i.e., time indicated that the participant was wearing the sensors) based on the caregiver completed data sheet as well as the wear time based on the accelerometry data extracted from the sensors. We measured several outcomes including: a) the time the caregiver reported the child wore the sensor; b) the time the caregiver reported the child wore the sensors and the accelerometry data indicated the child wore the sensors (both measures indicated wear time); c) the time the caregiver reported the child wore the sensors but the accelerometry data indicated the sensors were not worn (disagreement between caregiver report and accelerometry data); and d) the time the caregiver reported the child was not wearing the sensors but the accelerometry data indicated the sensors were worn (disagreement between caregiver report and accelerometry data). If the caregiver applied the sensors for more than 5 days, we selected the 5 days with the longest application based on caregiver report for analysis.

We analyzed data across each day and averaged across the 5 days. We also compared outcomes to the targeted time of 4 hrs a day for 5 days. Last, we evaluated the time that parents recorded the 30 min partial interval data on targeted behavior, as caregiver collected data will be important for later validation studies using accelerometer sensors in home settings.

To determine wear time based on the accelerometry data, we used the biobank Accelerometer Analysis tool (Doherty et.al., 2017) that was built to extract meaningful information from large accelerometer datasets. The tool was built and validated to analyze
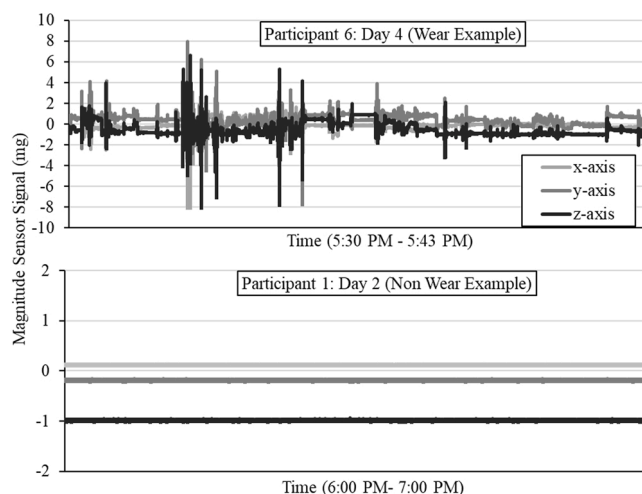

Fig. 3. Examples of accelerometer sensor data indicating wear time (top panel) and non-wear time (bottom panel) data.

physical activities in daily life of over 100,000 individuals over a period of 7 days. Data from this prior analysis (Doherty et al.) used the same sensors used for the present study. Although the past study investigated different target variables (e.g., sleep and physical activity), the tool is applicable to our data for detecting episodes of wear and non-wear times.

Data-analysis for detecting wear time included a calibration process where the acceleration signal was calibrated to local gravity. We computed a per sample Euclidean norm with the three axes, and a fourth order Butterworth low pass filter to remove machine noise (cut-off frequency of 20 Hz). To identify episodes of wear time, we separated the gravitational component from this vector magnitude. We defined a non-wear episode as consecutive stationary episodes of five second epochs, where all three axes have a standard deviation less than a certain value. Given our target population of youth with ASD was significantly different from the population that the biobank tool was built for (Doherty et al., 2017.), we investigated the first day of data collection for each participant and estimated the standard deviation that closest provided the reported wear time. The value of 20.0 mg, which is suggested by Doherty and colleagues, sufficed for all but two participants (20%). For these, we changed the value to 50.0 mg and 70.0 mg as this indicated a better fit. See Fig. 3 for examples of wear time and non-wear time data.

6. Results – Part 2

6.1. Participant recruitment and habituation

The researcher discussed Part 2 of the study with the caregivers of 16 potentially eligible participants, 11 of whom consented to the study (68.75%). All 11 caregivers mastered the skill of independently applying the sensors (100%) and all participants passed habituation and completed Part 2 of the study (100%). The data from one participant's sensors did not download correctly prior to
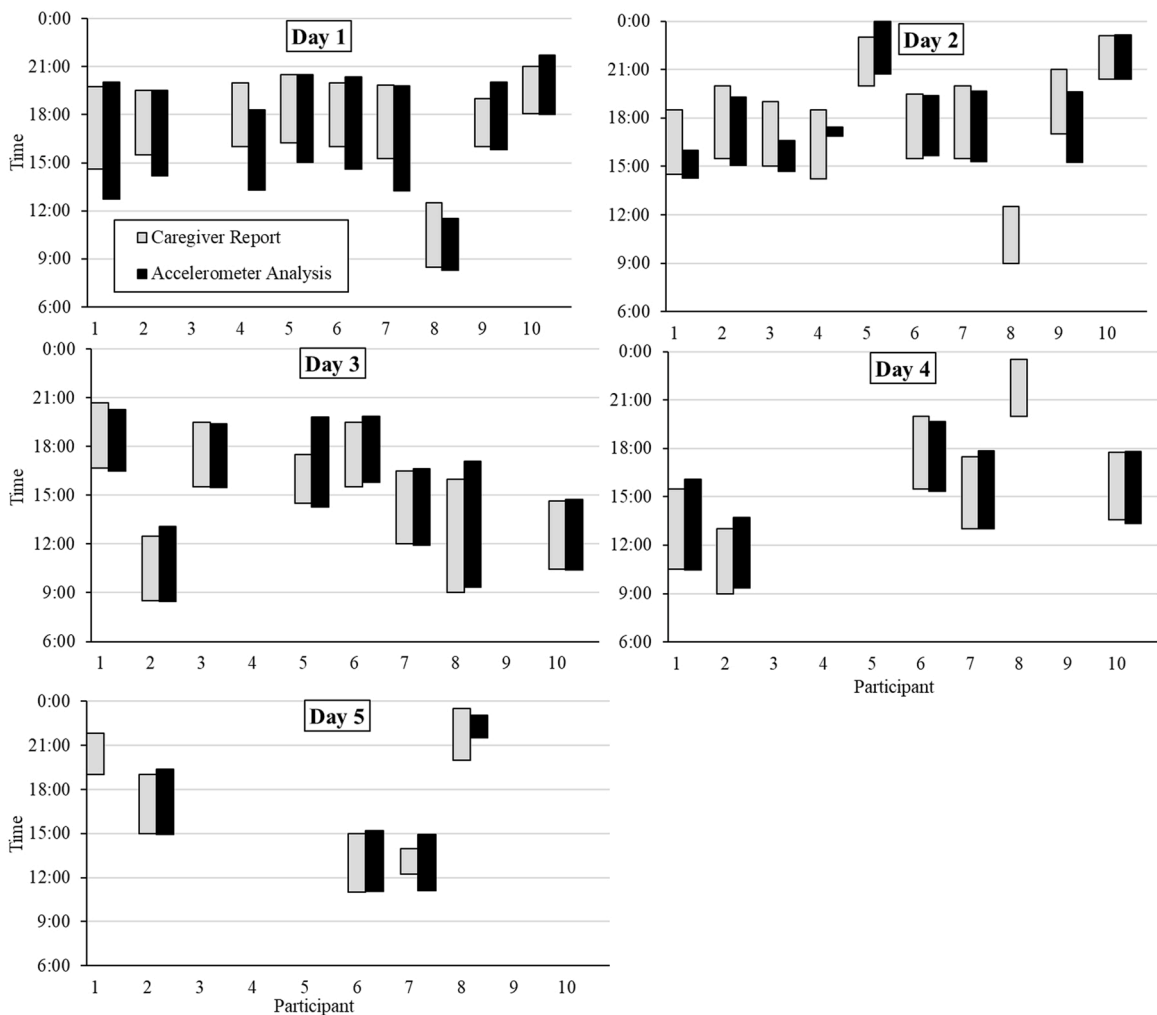

Fig. 4. Wear-time of accelerometer sensors in the home as indicated by caregiver report and the accelerometry data per participant (X-axis), across days (different panels of the graph), and based on the time of day (using a 24 hr clock, Y-axis).

researchers deleting the data from the sensors, thus we excluded this participant from future analyses, leaving 10 participants.

6.2. Home data collection

Participant 3 wore the sensors for over 12 hrs on two days and caregivers indicated that they accidently left the sensors on overnight. We removed these two outlier days from analyses.

Wear time reported by caregivers and based on the sensor data for each day per participant is in Fig. 4. Based on the on and off times recorded by the caregivers, five participants wore the sensors for five days, and three of these wore the sensors for approximately 4 hrs each of the five days (3.8 hrs or more a day). Across all ten participants, caregivers reported that participants wore the sensors for an average of 3.8 days ($SD = 1.40$) and, including days with no wear time, the wear time per day averaged 3.10 hrs ($SD = 1.12$). Thus, based on caregiver report of application time, participants wore the sensors for 77.50 % of the targeted time (caregiver reported wear time compared to the target of 4 hrs a day for five days or 20 total hours).

Caregivers collected at least some partial interval data on targeted behavior each day that they reported the sensors were applied. Compared to the targeted 4 hrs a day, participants collected a mean of 2.58 hrs of behavioral data ($SD = 1.14$) or 64.55 % of the target. Two caregivers collected at least 4 hrs of behavioral data on all 5 days.

Of the caregiver reported wear time, 81 % of the time the accelerometry data also indicated wear time ($SD = 22$ % across participants). We identified a mean of .60 hrs per day per participant ($SD = 0.66$) that caregivers reported the child was wearing the sensors but accelerometry data indicated the sensors were not worn. Interestingly, there were also times in which the accelerometry data indicated the sensors were likely worn, but caregivers did not report that the sensors were applied to the participant ($M = 0.54$ hrs per day; $SD = 0.29$).

When comparing wear time to the target of 4 hrs a day for 5 days using only time that the caregiver report and accelerometry data *both* indicated wear time, three participants wore the sensors for at least five days and two of these participants wore the sensors for approximately 4 hrs a day. Across all ten participants, considering only time that both the caregiver report and sensor data indicated wear time, participants wore the sensors for a mean of 3.50 days ($SD = 1.27$) and, including days the sensors were not worn at all, averaged 2.50 hrs a day ($SD = 1.18$). This represents 62.41% of the targeted time.

7. General discussion

Similar to past work, our overall findings suggest that it is feasible for children with ASD to wear accelerometer sensors and generalizes this finding to individuals admitted to an ABA clinic targeting severe aggression, self-injury, and/or disruption, with 93.94 % of our Part 1 participants passing habituation. Our results also provide insight into details regarding habituation for this specific population and demonstrate feasibility of wearing the sensors during actual clinical ABA assessments as well as at home when a therapist is not present.

Using a relatively conservative criterion of 5 min (consecutive) without avoidant movements or negative vocalizations to define passing habituation, participants spent an average of 11.61 min in habituation prior to passing with a mean of 12.59 instances of avoidant movements or negative vocalizations. The majority of participants exhibited at least some avoidant movements or negative vocalizations during the habituation process, demonstrating the importance of conducting systematic habituation procedures with this population prior to applying sensors for the purpose of collecting usable data. There was significant variability in the necessary duration of habituation (up to 45 min) as well as the prevalence and persistence of avoidant movements and negative vocalizations (up to 103 instances). Although the maximums in these ranges were driven by the participant that removed the sensors during the behavioral assessment, there was still considerable variability with this participant's data removed (upper range of 40 min and 49 avoidant movements or negative vocalizations). This suggests that, for a subset of individuals with ASD who engage in severe targeted behavior, it may be necessary to conduct habituation for lengthy periods of time and work through several instances of avoidant movements and negative vocalizations in habituation.

Additionally, two participants did not pass habituation. Given this small sample, we could not conduct statistical analyses to determine what characteristics might contribute to failing habituation. Descriptive analyses did not identify any consistent hypotheses as their ages (6 and 11) and parental descriptions of topographies, severity of targeted behavior, and communication abilities were not outliers compared to the rest of the sample.

In sum, out of 44 participants consented to both parts of the study, we excluded two who failed habituation, one for removing sensors in the behavioral assessment, and one due to a technological error with transferring the data. A remaining six participants were excluded as we did not observe 100 instances of targeted behavior. Future research should consider this attrition, combined with the potential duration of habituation, when planning for large-scale studies using accelerometry technology with this population.

This study also expands on past work by demonstrating the feasibility of children with ASD wearing these sensors during actual ABA assessment sessions. The long-term goal of this line of research is for accelerometry technology to automatically detect targeted behavior to provide information during clinically meaningful times. One potential use with applied significance is during ABA assessment and treatment sessions. However, these types of assessments specifically set up situations that are likely to evoke targeted behavior (e.g., removing preferred items or asking the child to complete work; Iwata et al., 1994). It is possible that a child who is compliant with wearing sensors in habituation might be less compliant when situations that are potentially aversive are presented, as is the case in ABA assessments. In other words, participants may be less compliant with instructions to keep the sensors on when the establishing operation for functional reinforcers maintaining other topographies of noncompliance and disruptive behavior are presented. Fortunately, our results found that most participants continued to wear the sensors during the behavioral assessment sessions,

with only one participant excluded for removing the sensors during this time.

Several other participants were excluded from our analysis because they did not engage in the frequency of targeted behavior deemed necessary for the sequential machine learning models (Bishop, 2006). This is important for determining recruitment numbers for future large-scale studies aimed at validating machine learning models targeting this population. However, it does not necessarily demonstrate long-term feasibility issues when considering the end goal of using the sensors clinically after these models have been sufficiently developed.

Another extension from our results is the feasibility of using accelerometer sensors in home settings when a clinician or researcher is not present. Following two rounds of behavioral skills training (one appointment in clinic and one in the child's home) to ensure caregivers mastered the skill of applying the sensors, we found that all caregivers applied the sensors to the child for at least some period. This is a promising finding that it is feasible for caregivers to get their children with ASD to wear sensors in the home setting without a researcher's or clinician's assistance.

Although this is a positive outcome, there is still work to be done in this area. We observed significant variability in the amount of time that caregivers reported applying the sensors and disagreement between caregiver reported wear time and wear time as detected directly by the accelerometry data. The researcher instructed caregivers to apply the sensors for 4 hrs a day for 5 days. Based on caregiver report alone, they were able to apply the sensors fairly close to this targeted time (77.40 % of the targeted time). However, when caregiver report and the accelerometry data both indicated wear time, that percentage drops to 62.41 % of the targeted time. This suggests there may have been fidelity errors in caregivers' report of the time the sensors were applied. In some cases, this seemed to be the result of rounding (e.g., a caregiver reporting the child wore the sensors from 3:00 PM – 7:00 PM and the accelerometry data indicating wear time was likely from 3:10 PM – 6:45 PM). In other situations, there were larger amounts of time that the caregivers indicated wear time and the accelerometry data suggested the sensors were likely not worn (Fig. 4). Thus, there were different patterns of disagreement between caregiver report and accelerometry data across participants that warrants further investigation.

It is important to note that it is not guaranteed that the wear time data as calculated from the accelerometry data are perfect. For instance, it is possible that the accelerometer sensors would not pick up on wear time if the child was sitting extremely still (e.g., if the child was asleep) or it may pick up on inaccurate movements (e.g., being carried by the caregiver). It is likely that wear time based on the accelerometry data is more accurate than caregiver report given our data analysis procedures are based on a very large-scale study with validated procedures (Doherty et al., 2017). However, it is possible that there were errors in the data analysis procedure as the analysis has not been validated with this specific population. Thus, we cannot say for certain which is ground truth (accelerometry data vs. caregiver report). Future studies would benefit from home videos to determine ground truth to validate caregiver report and accelerometry data and to identify methods of increasing caregiver fidelity.

In our study, we asked caregivers to complete 30 min partial interval data, a relatively simple data collection strategy. Despite this, integrity with data collection was worse than integrity with applying/removing the sensors (partial interval data were collected for less time than caregivers reported the participant wore the sensors). Given these integrity issues, future research evaluating the use of accelerometer sensors in home settings should include a direct observation measure of targeted behavior (e.g., video recordings) and/or focus on methods to increase fidelity with caregiver collected data (e.g., daily reminders).

There are a few other limitations that should be considered in future research. First, our study only included children ages 4–13. It is possible that younger or older individuals could also benefit from the use of accelerometer sensors to automatically detect targeted behavior but the feasibility with these ages may differ. For data-collection in the home (study 2), we recruited participants with mild-to-moderate targeted behavior. This may be a limitation as integrity may be worse for individuals with more severe behavior. Additionally, the procedures should be replicated with a larger sample to determine the generalizability of results.

Second, all researchers in the study had extensive training in working with children with severe targeted behavior and the setting of the study was built specifically for work with this unique population. The positive feasibility findings might not be replicated with participants with severe targeted behavior in different settings or with less specialized personnel.

As a third limitation, we had missing data from habituation for two participants and skipped habituation for one participant due to protocol errors. It is unclear if this data is missing at random or whether there were any systematic patterns to participants for whom we had missing habituation data. We also only included English speaking families due to study constrains with including interpreters. It is unclear if our results would generalize to non-English speaking families.

Regarding data-collection to determine reaching the criteria of 100 instances of targeted behavior, we did not collect interobserver agreement (IOA) data to ensure this target was reached. For purpose of future data-analysis, researchers will conduct ground-truth scoring of frame-by-frame video with IOA. However, given IOA was not recorded in the moment, it is possible that we over- or under-estimated 100 instances for some participants. We also selected the requirement of 100 instances as an attempt to ensure we had enough data for complex data analysis models if needed. However, this frequency is somewhat arbitrary, and may over- or under-estimates the actual need. Future research should work to identify the necessary parameters to conduct a-priori power analyses.

Another significant limitation is we have not yet evaluated the data using machine-learning techniques. We conducted a preliminary and very cursory analysis of the data to gauge whether data are appropriate for sequential data analysis techniques. A researcher viewed graphical depictions of raw data to assess for excessive noise or outliers in the data streams. Additionally, they randomly sampled instances across participants where direct observation indicated distinct instances of targeted behavior (e.g., a single hit) and analyzed the raw data-stream to assess for a matched change in movement (i.e., a peak in the data stream). This initial data-analysis was promising, suggesting the data are likely valid for additional analyses, but subsequent work is needed to ensure the quality of data. The primary goal of this study was to report on the feasibility of using accelerometer sensors with this specific population of individuals with ASD who engage in aggression, self-injury, or disruption during actual clinical assessments and in home settings. A crucial next step is to determine the machine learning models that are most appropriate to actually use the data from the

sensors to automatically detect targeted behavior. Although initial models can be explored using the data collected in the sample(s) described in this study, additional research with a larger and more heterogenous sample will be needed to validate these models. This study provides helpful feasibility information to design and implement these future studies.

Funding

CRediT authorship contribution statement

Mindy Scheithauer: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Shruthi Hiremath:** Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Audrey Southerland:** Conceptualization, Methodology, Project administration, Visualization, Writing – original draft. **Agata Rozga:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Writing – original draft. **Thomas Ploetz:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Chelsea Rock:** Data curation, Project administration, Writing – review & editing. **Nathan Call:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft.

Conflict of interest

The authors do not have any conflicts of interest to report.

References

Albinali, F., Goodwin, M.S., & Intille, S.S. (2009, September). Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In Proceedings of the 11th international conference on Ubiquitous computing (pp. 71–80).

Arvidsson, D., Fridolfsson, J., & Börjesson, M. (2019). Measurement of physical activity in clinical practice using accelerometers. *Journal of Internal Medicine, 286*(2), 137–153. https://doi.org/10.1111/joim.12908

Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 46*(1), 1–21. https://doi.org/10.1002/jaba.30

Bishop, C. M. (2006). Pattern recognition and machine learning. *7 BIS,* (681.326).

van den Boogert, F., Sizoo, B., Spaan, P., Tolstra, S., Bouman, Y. H., Hoogendijk, W. J., & Roza, S. J. (2021). Sensory processing and aggressive behavior in adults with autism spectrum disorder. *Brain Sciences, 11*(1), 95. https://doi.org/10.3390/brainsci11010095

Breitenstein, R.S., Doane, L.D., & Lemery-Chalfant, K. (2020). Children's objective sleep assessed with wrist-based accelerometers: Strong heritability of objective quantity and quality unique from parent-reported sleep. Sleep. doi: 10.1093/sleep/zsaa142.

Cantin-Garside, K. D., Kong, Z., White, S. W., Antezana, L., Kim, S., & Nussbaum, M. A. (2020). Detecting and classifying self-injurious behavior in autism spectrum disorder using machine learning techniques. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-020-04463-x

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., … Wareham, N. J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *PLOS ONE, 12*(2), Article e0169649. https://doi.org/10.1371/journal.pone.0169649

Edelson, S. M. (2021). Comparison of autistic individuals who engage in self-injurious behavior, aggression, and both behaviors. *Pediatric Reports, 13*(4), 558–565. https://doi.org/10.3390/pediatric13040066

Ferguson, B. J., Hamlin, T., Lantz, J. F., Villavicencio, T., Coles, J., & Beversdorf, D. Q. (2019). Examining the association between electrodermal activity and problem behavior in severe autism spectrum disorder: A feasibility study. *Frontiers in Psychiatry, 10,* 654.

Garcia, J. M., Leahy, N., Rivera, P., Brazendale, K., & Rice, D. J. (2019). The association among demographic factors, health behaviors and sleep quality in youth with autism spectrum disorder. *Disability and Health Journal, 13*(3), Article 100885. https://doi.org/10.1016/j.dhjo.2019.100885

Gilchrist, K. H., Hegarty-Craver, M., Christian, R. B., Grego, S., Kies, A. C., & Wheeler, A. C. (2018). Automated detection of repetitive motor behaviors as an outcome measurement in intellectual and developmental disabilities. *Journal of Autism and Developmental Disorders, 48*(5), 1458–1466. doi: /10.1007/s10803-017-3408-6.

Goodwin, M. S., Intille, S. S., Albinali, F., & Velicer, W. F. (2011). Automated detection of stereotypical motor movements. *Journal of Autism & Developmental Disorders, 1*(6), 770–782. https://doi.org/10.1007/s10803-010-1102-

Goodwin, M. S., Mazefsky, C. A., Ioannidis, S., Erdogmus, D., & Siegel, M. (2019). Predicting aggression to others in youth with autism using a wearable biosensor. *Autism Research, 12*(8), 1286–1296. https://doi.org/10.1002/aur.2151

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197–209. https://doi.org/10.1901/jaba.1994.27-197

Khreich, W., Granger, E., Miri, A., & Sabourin, R. (2012). A survey of techniques for incremental learning of HMM parameters. *Information Sciences, 197,* 105–130.

Lai, M. C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., … Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry, 6*(10), 819–829. https://doi.org/10.1016/S2215-0366(19)30289-5

Milošević, M., Van de Vel, A., Cuppens, K., Bonroy, B., Ceulemans, B., Lagae, L., … Van Huffel, S. (2017). Feature selection methods for accelerometry-based seizure detection in children. *Medical & Biological Engineering & Computing, 55*(1), 151–165. https://doi.org/10.1007/s11517-016-1506-9

Northrup, J.B., Goodwin, M.S., Peura, C.B., Chen, Q., Taylor, B.J., Siegel, M.S., & Mazefsky, C.A. (2022). Mapping the time course of overt emotion dysregulation, self-injurious behavior, and aggression in psychiatrically hospitalized autistic youth: A naturalistic study. Autism Research: Official Journal of the International Society for Autism Research, 10.1002/aur.2773. Advance online publication. doi: 10.1002/aur.2773.

Plötz, T., Hammerla, N.Y., & Olivier, P.L. (2011, June). Feature learning for activity recognition in ubiquitous computing. In Twenty-second international joint conference on artificial intelligence.

Plötz, T., Hammerla, N.Y., Rozga, A., Reavis, A., Call, N., & Abowd, G.D. (2012). Automatic assessment of problem behavior in individuals with developmental disabilities. Paper presented at the proceedings of the 2012 ACM conference on ubiquitous computing (pp. 391–400).

Rad, N. M., Kia, S. M., Zarbo, C., van Laarhoven, T., Jurman, G., Venuti, P., … Furlanello, C. (2018). Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Signal Processing, 144,* 180–191. https://doi.org/10.1016/j.sigpro.2017.10.011

Saini, V., Fisher, W. W., Retzlaff, B. J., & Keevy, M. (2019). Efficiency in functional analysis of problem behavior: A quantitative and qualitative review. *Journal of Applied Behavior Analysis, 53,* 44–46. https://doi.org/10.1002/jaba.583

Scheithauer, M., Schebell, S. M., Mevers, J. L., Martin, C. P., Noell, G., Suiter, K. C., & Call, N. A. (2020). A comparison of sources of baseline data for treatments of problem behavior following a functional analysis. *Journal of Applied Behavior Analysis, 53*(1), 102–120. https://doi.org/10.1002/jaba.549

Schopler, E., Wellman, G.J., & Love, S.R. (2010). Childhood Autism Rating Scale. 2nd edn. Torrance, CA: Western Psychological Services.

Sharp, R. A., Mudford, O. C., & Elliffe, D. (2015). Representativeness of direct observations selected using a work-sampling equation. *Journal of Applied Behavior Analysis, 48*, 153–166. https://doi.org/10.1002/jaba.193

Steenfeldt-Kristensen, C., Jones, C. A., & Richards, C. (2020). The prevalence of self-injurious behaviour in autism: A meta-analytic study. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-020-04443-1

Stokes, J., Pogge, D., Wecksell, B., & Zaccario, M. (2011). Parent–child discrepancies in report of psychopathology: The contributions of response bias and parenting stress. *Journal of Personality Assessment, 93*(5), 527–536. https://doi.org/10.1080/00223891.2011.594131

Tiger, J. H., Miller, S. J., Mevers, J. L., Mintz, J. C., Scheithauer, M. C., & Alvarez, J. (2013). On the representativeness of behavior observation samples in classrooms. *Journal of Applied Behavior Analysis, 46*(2), 424–435. https://doi.org/10.1002/jaba.39

Voss, C., Duncombe, S. L., Dean, P. H., de Souza, A. M., & Harris, K. C. (2017). Physical activity and sedentary behavior in children with congenital heart disease. *Journal of the American Heart Association, 6*(3), Article e004665. https://doi.org/10.1161/JAHA.116.004665

Wachob, D., & Lorenzi, D. G. (2015). Brief report: influence of physical activity on sleep quality in children with autism. *Journal of Autism and Developmental Disorders, 45*(8), 2641–2646. https://doi.org/10.1007/s10803-015-2424-7

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation, 31*(7), 1235–1270.