

# **An informant report behavior diary for measuring temper outbursts in an intervention setting**

Leah E. Bull<sup>1</sup>, Chris Oliver<sup>1</sup>, Penelope L. Tunncliffe<sup>1</sup> & Kate A. Woodcock<sup>2</sup>

1. Cerebra Centre for Neurodevelopmental Disorders, University of Birmingham, UK
2. School of Psychology, Queen's University Belfast, UK

Author note:

Correspondence concerning this article can be sent to Dr Kate A. Woodcock, School of Psychology, Queen's University Belfast, University Road, Belfast, BT7 1NN, UK. Email to [papers@katewoodcock.com](mailto:papers@katewoodcock.com).

This manuscript was in press with the *Journal of Developmental and Physical Disabilities*.

The final manuscript can be accessed here: Bull, L.E., Oliver, C., Tunncliffe, P. L. & Woodcock, K. A. (in press). An informant report behavior diary for measuring temper outbursts in an intervention setting. *Journal of Physical and Developmental Disabilities*. doi: 10.1007/s10882-015-9429-1

## **Abstract**

Temper outbursts are associated with several psychological disorders and comprise a high priority for intervention. However, the low frequency of outbursts presents a challenge for valid measurement. In the present study an informant report behavior diary for measuring temper outbursts was developed and its validity assessed in a case series. Caregivers of twelve individuals with the neurodevelopmental disorder Prader-Willi syndrome (PWS, in which temper outbursts are common) completed a behavior diary over four weeks, and a structured interview. Heart rate and movement data were recorded during a sample of the days subject to diary reporting. Individuals with PWS completed self-report ratings of negative emotion experience. Behavior diaries showed high concordance with the component behaviors and duration of temper outbursts reported in structured interviews; but tended to report a lower frequency. For outbursts reported in diaries during physiological recording, heart rate was consistently elevated above a resting state baseline; and was comparable to that recorded during high physical activity. Available self-report data demonstrated correspondence with the diaries but few self-report data were produced. The present results provide critical proof of principle data supporting the concurrent validity of the ecologically valid, resource efficient diaries, which can be exploited in future research.

## **Key words**

Informant report diary; behavior diary; temper outbursts; tantrums; neurodevelopmental disorder; validity

## **Introduction**

Temper outbursts (also referred to as tantrums or rages) can be defined as negative emotional episodes with concurrent expression of overt behaviors such as crying, hitting or kicking; and comprise heightened autonomic arousal (Potegal & Davidson, 2003). Outbursts show a characteristic normative trajectory in children, peaking in prevalence between 2 and 3 years, with frequency usually between once per month and once per day (Osterman & Bjorkqvist, 2010; Potegal & Davidson, 2003; Wakschlag et al., 2012).

Several neurodevelopmental and psychiatric disorders that affect children including autism spectrum disorder, Tourette's syndrome, attention deficit hyperactivity disorder, obsessive compulsive disorder, oppositional defiant disorder, depressive disorder, intellectual disability, and a number of genetically defined disorders, are associated with more frequent and severe temper outbursts, which can persist throughout childhood and into adulthood (Belden, Thomson & Luby, 2008; Carlson, Potegal, Margulies, Gutkovich & Basile, 2009; Carlson & Dyson, 2012; Chen et al., 2013; Goldin, Matson, Tureck, Cervantes & Jang, 2013; Kano, Ohta, Nagai, Spector & Budman, 2008; Konst, Matson & Turygin, 2013; Maskey, Warnell, Parr, Le Couteur & McConachie, 2013; Oliver et al., 2013; Storch et al., 2012). In addition, it appears possible to distinguish normative temper outburst behavior from that associated with psychological disorder, on the basis of intensity, duration and the nature of component behaviors, even in children at an age when normative temper outbursts are common (Belden et al., 2008; Wakschlag et al., 2012). These disordered temper outbursts are highly disruptive and are a treatment priority for many individuals and their families (Carlson et al., 2009; Eisbach, Cluxton-Keller, Harrison, Krall, Hayat & Gross, 2014; Leclerc, O'Connor, Forget & Lavoie, 2011; Woodcock, Oliver & Humphreys, 2009; Tunnicliffe, Woodcock, Bull, Oliver & Penhallow, 2014).

Few measurement instruments of temper outbursts have been developed and validated systematically and those that have focus on measurement at a single point in time. For example, the Multidimensional Assessment of Preschool Disruptive Behavior includes a “temper loss” subscale in which six-point frequency ratings of outburst component behaviors are provided by informants, but these are based on average behavior over the preceding month (Wakschlag et al., 2012); the Children’s Agitation Inventory is an observation schedule of outburst component behaviors but this has only been tested for reliability in an inpatient setting with preexisting observation arrangements in place (Carlson et al., 2009). Since temper outbursts are generally low frequency but highly intense behavior episodes with prolonged duration (e.g. one hour or more), small changes in frequency and/or changes in duration or intensity may have dramatic implications for individuals and their families (Potegal & Davidson, 2003; Tunnicliffe et al., 2014). However, based on retrospective informant report, existing measures are limited in their sensitivity to ongoing, potentially subtle changes in the nature of temper outbursts. Such measures are also potentially subject to relatively large response biases in intervention settings, where large placebo effects on temper outbursts have been demonstrated (Whalley & Hyland, 2013). In addition, the low frequency of temper outbursts relative to many other forms of disruptive behavior further limits the utility of many widely applied methods for measuring behavior in intervention contexts (Eisbach et al., 2014; Ervin, Radford, Bertsch, Piper, Ehrhardt & Poling, 2001; Tarbox, Wallace, Tarbox, Landaburu & Williams, 2004; Whitaker, Walker & McNally, 2004). The present study aimed to overcome these challenges via the development of a resource efficient and ecologically valid measurement tool for the assessment of temper outbursts in an intervention setting.

Behavior diaries, in which defined behaviors and associated contextual characteristics are recorded immediately after they happen, provide one way to overcome the factors

outlined above that limit the application of existing measures of temper outbursts to an intervention setting. Diary entries are not made retrospectively so memory biases are minimised, and behavior is described using objective criteria rather than subjective ratings so potential responder demand characteristics are minimised. Importantly, diaries provide the means to track behaviors as they occur over time, allowing even small changes in the frequency, duration or nature of the behavior to be measured. This tracking capability of behavior diaries is particularly important given the low frequency of temper outbursts. The utility of such diaries in a temper outburst context would suggest that these may be effectively developed for other low frequency clinically relevant behaviors.

Mother report behavior diaries of operationally defined infant behaviors have been widely administered and validated using audio and video recording of infant behavior, and researcher observations (Barr, Kramer, Boisjoly, McVeywhite & Pless, 1988; Barr, Kramer, Pless, Boisjoly & Leduc, 1989; Hunziker & Barr, 1986; St James-Roberts, Hurry & Bowyer, 1993; St James Roberts & Peachey, 2011; Anzman-Frasca, Liu, Gates, Paul, Rovine, Birch, 2013).

In the present study, an informant report behavior diary for measuring temper outbursts was developed. Operational definitions of temper outburst component behaviors were included in a similar way as in previously validated infant behavior diaries. Temper outbursts appear to be sensitive to the presence of unfamiliar observers (Woodcock et al., 2011), relevant populations are generally fully mobile, and temper outbursts are low frequency (but high intensity) behaviors. Thus, validity testing using the methods reliant on short periods of recording using an immobile, obtrusive, audio or video recording device, which have been applied to infant behavior diaries, would be largely inappropriate or unfeasible. However, since heightened negative emotional arousal is comprised in the temper outburst construct, convergent validity of the diaries was assessed via measurements of heart

rate and physical activity, which can be used to index emotional arousal; and participant self-reports of negative emotional experience. In addition, a structured informant report interview that included report of frequency, duration and component behaviors of outbursts, was administered as a further assessment of convergent validity.

## **Method**

### ***Participants***

A case series of 12 males with a diagnosis of Prader-Willi syndrome and their primary caregivers (*Table 1*) participated in the study. Prader-Willi syndrome is a genetic neurodevelopmental disorder affecting around one in every ten thousand individuals, which is caused by a mutation impacting on the paternally derived q11.2-q13 region of Chromosome 15 (Whittington, Holland, Webb, Butler, Clarke & Boer, 2001). The syndrome has a well characterized physical phenotype (Holm et al., 1993) and causes a mean shift in cognitive ability downwards by approximately 40 IQ points (Whittington, Holland, Webb, Butler, Clarke & Boer, 2004). Particularly relevant to the present context, Prader-Willi syndrome is associated with elevated incidence of frequent temper outbursts, which have been demonstrated in more than 80% of several samples of individuals with the syndrome (Dimitropoulos, Feurer, Butler & Thompson, 2001; Holland, Whittington, Webb, Boer & Clarke, 2003; Walz & Benson, 2002; Woodcock et al., 2009). Importantly, the temper outburst behavior shown by individuals with Prader-Willi syndrome represents one of the most functionally impairing aspects of the disorder and a key intervention target (Einfeld et al., 2014). Finally, since the present definition of temper outbursts comprises heightened emotional arousal, it is notable that temper outbursts in individuals with Prader-Willi syndrome have been similarly linked to heightened emotional arousal (Woodcock et al., 2009; Tunnicliffe et al., 2014).

Inclusion criteria comprised caregiver report of current difficulties temper outbursts at time of recruitment. In addition, some of the present participants went on to take part in further studies by the authors, thus an additional inclusion criterion stated that a proportion of participants' outbursts were triggered by changes to routines or plans. Further, no participants had a behavioral treatment plan in place over the course of the study (though several went on to engage with a specific plan following study completion). Gender was not controlled due to the restricted size of the population, but the inclusion of only males corresponds well with the male heavy gender ratios apparent in diagnoses of several psychological disorders associated with temper outbursts (Petresco et al., 2014). Participants were recruited via the Prader-Willi Syndrome Association UK and via a group of residential homes in the UK.

(Table 1)

### ***Measures***

#### ***Structured caregiver report interview***

A structured caregiver report interview administered to caregivers to comprehensively describe possible triggers and the behavioral sequence of temper outbursts has been described elsewhere<sup>1</sup> [withheld for blind review]. Relevant here, the interview also required caregivers to report the component behaviors observed in a typical outburst, the present frequency of

---

<sup>1</sup> The cited study described the range of percentages of temper outburst behaviors reported in the interviews of 8 of the present participants, which were also reported in their behavior diaries. However, convergent validity of the two measures was not assessed in detail nor is the type of agreement assessed previously reported in the present study.

temper outbursts (with reference to the period of time within which the informant would expect to see the next outburst: 15 minutes; 1 hour; 1 day; 1 week; 1 month), the typical duration of outbursts (< 1 minute; < 5 minutes; < 15 minutes; < 1 hour; > 1 hour), and the duration of the longest outburst observed in the past month. Caregivers of Participant BC (*Table 1*) did not wish to take part in the structured interview. The interviews took place in a single session and the mean duration was 50 minutes (range: 21-90 minutes). Two researchers coded 10 interviews and showed 100% agreement on the items reported here.

### ***Behavior diary***

Temper outburst component behaviors were identified for each individual based on such behaviors that have been previously reported in the literature (Oliver, Woodcock & Humphreys, 2009; Potegal & Davidson, 2003; Wakschlag et al., 2012) and a semi-structured interview with caregivers (*see supplementary materials*). Thus, individualised diary report cards were created for each participant which included lists of possible temper outburst component behaviors, antecedent events (triggers), and following events. These lists were then discussed with caregivers in operational terms and described in the report card using language familiar to the caregiver (*see supplementary materials*). “Other” items were included in all lists to ensure they were exhaustive and space was provided for additional description of these. In addition, the diary report cards included columns for entry of the date, the time, the duration of the outburst and an intensity rating (Likert-type 1-10, 10 being most intense ever observed). Primary caregivers were asked to complete the diary immediately following any temper outbursts observed over four weeks. Where participants’ time was split between two settings (e.g. home and school), this was the primary caregiver from each environment (*Table 1*).



### ***Physiological measures***

For two days over the course of the month, participants were asked to wear an Actiwatch by CamNtech around one wrist, and a heart rate monitor (Polar RS400) comprising a chest strap and recording device around the other wrist. Heart rate was recorded in average beats per minute (bpm) every second. The Actiwatch comprised an accelerometer, which produces an electric current when movement is detected. Resultant changes in voltage were measured as an average activity count every ten seconds, which indexes movement. Recording days were selected based on practical concerns for each participant, were not necessarily consecutive, and were days which caregivers anticipated would be most reflective of a normal day (e.g. not days on which one off activities were planned).

### ***Self-report emotion ratings***

Following temper outbursts, caregivers were asked to prompt participants to complete a visual rating scale of their present emotional experience (*see supplementary materials*). The scale was adapted from the Distress Thermometer that was developed for rapid screening of negative mood in cancer patients (Roth, Kornblith, Batel-Copel, Peabody, Scher, Holland, 1998). A recent meta-analysis across a total of 14,808 adult cancer patients in 42 studies showed pooled sensitivity of 81% (95% confidence intervals 79-82%) and specificity of 72% (95% confidence intervals 71-72%) of the Distress Thermometer, against reference diagnostic measures for anxiety and depression (Ma et al., 2014). It cannot be assumed that these validity data, collected from previously healthy adults, are relevant in the present population. However, to our knowledge, no similarly simple visual scales of negative emotion experience have been validated in developmentally disabled populations. The present adaptations were designed to decrease the general cognitive demands for completion

and comprised reducing the available rating points from 11 to 5; including schematic faces to represent each point; and adding a written description of the associated emotional state (OK, a little bad; quite bad; very bad; terrible).

### ***Procedure***

Ethical approval was obtained [*withheld for blind review*] Review Committee. All adult participants and caregivers provided written consent, and children also provided written assent. The semi-structured interview was conducted via telephone by the first author. Caregivers then received a further call to go over the operational definitions. Following this, caregivers were visited by a researcher. Forms were passed on to caregivers, behavior diary requirements were reviewed and instructions were provided on use of the Actiwatch and heart rate monitor. The physiological recording equipment was then sent on agreed days along with written instructions. The structured caregiver report interview was administered to the primary caregiver between 18 months before and 18 months after behavior diary completion, by the third author face to face (Participants AC & CB) or via telephone. For participants CC, DB and DC (living in group accommodation) this was the support worker who knew the participant best. The time discrepancy between structured interview and diary assessment was necessitated by resource limitations (the implications of this discrepancy are discussed further below).

### ***Analysis***

Agreement between component behaviors, frequency and duration of temper outbursts reported in the diaries compared to the structured interviews was described descriptively.

Increases in heart rate can index increased physiological arousal (Ekman, Levenson & Friesen, 1983; Rainville, Bechara, Naqvi & Damasio, 2006; Fernandez, Pascual, Soler, Elices,

Portella & Fernandez-Abascal, 2012). However, heart rate is heavily influenced by physical activity (Iellamo, 2001). Thus, over the course of a day, any periods of increased physical activity would be associated with increased heart rate. In addition, increased heart rate would result from periods of increased emotional arousal not linked to temper outbursts (e.g. excitement about a pleasant stimulus or event; Bernat, Cadwallader, Seo, Vizueta & Patrick, 2011). To provide support for the convergent validity of the behavior diaries, measured heart rate should therefore be sensitive to reported temper outbursts (should be increased). However, we would not expect heart rate to act as a specific indicator of temper outbursts. Thus, convergent validity of the behavior diaries with heart rate data does not require elevated heart rate to occur during reported outburst periods *only*.

To obtain an estimate of changes in heart rate over the course of the recording day, which was appropriate for visual comparison, mean heart rate per minute was calculated based on the 60 corresponding data points recorded. Similarly, mean activity count per minute was calculated based on the six corresponding data points. Beats per minute for heart rate represents a scale with high interpretability. We therefore divided the standard deviation of the distribution of mean activity counts for each participant by that of the corresponding distribution of mean heart rates, and then divided each mean activity count value by the resultant scaling factor. Following this, we added a constant to the mean activity count values, which depended on the minimum recorded mean heart rate value for the corresponding participant. This procedure resulted in a transformed distribution of mean activity counts, which showed the same standard deviation as the corresponding distribution of mean heart rates, and was anchored at an arbitrary positive value which allowed easy visualisation on the same axis as the mean heart rates.

The resting state baseline was assumed to reflect the person's heart rate during periods of relaxed inactivity. These baselines were calculated with reference to each participant's

activity data collected from the Actiwatch on the day of the reported outburst. Four periods were selected which: showed an activity count consistently close to zero (in the transformed activity data this equated to a count close to the relevant constant); did not include the ascending component of peaks in activity following included periods of relative inactivity; lasted at least 10 minutes; and showed no overlap with the reported outburst. The resting state baseline was then computed as the mean heart rate over the four selected time periods. Similarly, high activity baseline heart rates were calculated based on four periods of elevated activity, at least 10 minutes in duration, which did not include apparent outlying activity count values or the reported temper outburst period. The high activity baseline was assumed to reflect heart rate during periods of physical exertion. Standard deviations of heart rate measurements comprised in the computation of each baseline were calculated as a measure of variability.

A total of 23 self-report ratings were made following outbursts reported by caregivers in the diaries. However, 19 of these reports were obtained from a single participant (the other four from different participants). The correspondence between this individual's self-reports and the behavior diary entries made by their caregiver was therefore assessed descriptively.

## **Results**

### ***Structured interviews***

For five out of eight participants for which at least one outburst was recorded in the diary and a structured interview was completed, 66% or more of the outburst component behaviors reported across all corresponding diary entries, were also reported in the interview (*Table 2*). For Participant BB however, a single outburst was recorded in the diary, which comprised shouting and him breaking his glasses. Destruction of property was not reported in the corresponding interview (hence the 50% agreement) but it is possible that the

destruction observed was part of an act of physical aggression, which was described in the interview as manifesting in varied ways. The primary discrepancy for Participant AA resulted from a distinction being made in the diary between shouting and screaming (both reported), but not in the interview (only shouting reported). In addition, one outburst (of 10) involved refusing to move; and another involved the participant scratching their face, not reported in the interview. Similarly, for Participant AB the diary made a distinction between shouting and screaming (both reported), which was not reported in the interview (only shouting). In addition, one outburst (of 12) recorded in the diary involved destruction of property; and another involved stamping, which were not reported in the interview.

(Table 2)

The boundaries of numbers of outbursts that correspond to the available response categories for the interview question on frequency are reported in parentheses in the sixth column of *Table 2*. Only five of the eleven participants showed recorded diary frequencies that fell within the corresponding interview specified boundaries; and reported frequencies tended to be higher in the interviews relative to the diaries. However, only the data for Participant DA showed a discrepancy of more than one interview defined category.

In the eight participants for whom duration information was available from diaries and interviews, there was agreement for all eight between the modal diary-recorded frequency and typical frequency in the corresponding interview. The duration of the longest outburst reported across diaries and interviews corresponded for six participants. However, the longest outbursts recorded in the diaries for Participants AB and CA, were discrepant from the corresponding interview data by a single category only.

### ***Physiological data***

Two participants whose caregivers completed the behavior diary expressed their desire not to wear the physiological recording equipment on either recording day. An additional two participants were only willing to wear the recording equipment on one day. During the resultant total of 16 days when physiological recordings were made, three temper outbursts were recorded in the diaries (three different participants). For Participant AA the recorded outburst lasted five minutes (intensity 8/10), was triggered by a change in expectation, and comprised crying and face-stroking. The outburst recorded for Participant AB lasted two minutes (intensity 2/10), was triggered by teasing, and comprised crying. Finally, the outburst recorded for Participant AC lasted 50 minutes (intensity 8/10), was triggered by a disagreement, and comprised crying, shouting and verbal aggression.

Fulfilling the minimum criterion necessary to provide support for the convergent validity of the behavior diaries, heart rate during the course of all three reported temper outbursts was consistently above the resting state baseline (*Figure 1*). For Participant AC, the reported outburst period comprised a period of four minutes when heart rate dropped slightly below the resting state baseline, but this occurred towards the end of the period and corresponded to only 8% of the total reported duration. Participants AB and AC also demonstrated a heart rate above the high activity baseline for at least half of the reported outburst period, and consistently within one standard deviation of the high activity baseline. However, the reported temper outburst period for Participant AA only comprised the initial minute of heart rate within one standard deviation of the high activity baseline. The time necessary for heart rate to drop to resting state following the outburst period was 15 minutes for Participant AA; 1 hour 51 minutes for Participant AB; and one hour for Participant AC.

*(Figure 1)*

### ***Self-report ratings***

Participant BA made the 19 self-report ratings available for analysis (*Table 3*). Caregivers prompted BA to make self-reports after outbursts but BA also made spontaneous ratings at other times (not appropriate for analysis here). None of the diary reported temper outburst periods were associated with the lowest available negative emotion self-rating (1= “OK”). Further, the difference between outbursts associated with a self-report rating of “a little bad” (2) and those associated with a more negative self-report rating (3, 4 or 5) were distinguishable based on corresponding diary reports of median duration, the longest reported duration, median intensity and the highest reported intensity. However, there was no difference between outbursts associated with the three most negative self-report ratings (quite bad, very bad, terrible) that could be seen consistently across these duration and intensity diary report indices.

(*Table 3*)

### **Discussion**

The convergent validity of informant report behavior diaries of temper outbursts occurring during a four week period was assessed using informant report structured interviews; recordings of heart rate and physical activity; and self-report rating scales of negative emotion experience. There was high agreement between diaries and interviews on temper outburst component behaviors and durations. However, report of outburst frequency tended to be higher in the interviews relative to the diaries. In line with the negative emotional arousal assumed to be comprised in the temper outburst construct, all temper outbursts reported in the diaries during physiological recording were associated with heart rate consistently elevated above the resting state baseline. In addition, for two of the three reported outbursts, heart rate was elevated above a high activity baseline reflecting periods of

high physical exertion. Only one participant completed enough self-report ratings for analysis but for this person there was high correspondence between three levels of self-reported negative emotion experience and the duration and intensity of corresponding outbursts described in the behavior diary.

Thus, the data presented provide convergent support at three levels for an association between a diary-recorded outburst period, and the substantive occurrence of an outburst (correspondence with component behaviors and durations reported at interview; increased heart rate; and self-report of increased negative emotion). The diaries were quick to administer (anecdotally caregivers reported entry completion time as 2-5 minutes) and flexible enough to be administered across varied natural environments (including family homes, group homes and schools), and therefore satisfied the low resource, high ecological validity objectives for the measure.

However, there are fewer data available on how far the diary entries described all of the outbursts that actually occurred, and the convergence between frequency estimates across the diary and the interview was not strong. Of note, diaries generally under reported the frequency of outbursts relative to that described in the interview. On face value, this frequency discrepancy comprises two potential threats to the validity of the diaries. Firstly, caregivers may not record all outbursts they observe. Secondly, caregivers may not observe all outbursts or may not recognize as outbursts all of the same events as another person (including the participant themselves). Nevertheless, for measurement of temper outbursts in an intervention setting, it is worth noting that it is the validity of any reported changes in frequency (rather than that of the reported absolute frequency) that is most relevant, which may be more robust to these limitations. Circumscribed trajectories of outbursts can be identified from vocal recordings (Green, Whitney & Potegal, 2011), thus future validity testing could be achieved using continuous audio recording with a discreet device worn by



participants. In addition, providing diary forms on a portable pocket size device may further reduce the time and memory demands on informants (increasing resource efficiency), and make it possible to ascertain reports from multiple informants on the same behaviors to assess inter-rater reliability.

However, in discussing the apparent discrepancies in frequency between structured interview and diary measures, the relatively long period of time (up to 18 months) between the conduct of these two assessments, is important to highlight. This time discrepancy represents a limitation of the present study in that temper outburst behavior may be expected to change over time. The data on such changes are limited in individuals with Prader-Willi syndrome though in recently reported 2 year longitudinal data (Bull, 2013), more than 75% of a sample of more than 50 individuals showed no change in the incidence of aggression over the period (given the nature of the behavior profile associated with the syndrome, aggression could be expected to occur primarily in the context of temper outbursts, e.g. Holland et al. 2003). It is notable that concurrence between the structured interviews and diaries in behavioral composition and duration of the reported outbursts was quite high despite the time discrepancy between the conduct of the two assessments. It is possible that whilst the potential for day to day events to trigger temper outbursts in individuals with Prader-Willi syndrome and the incidence of such events in individuals' daily lives may vary over time, once an outburst has been triggered, the profile of behaviors demonstrated follows a similar course. This suggestion is supported by work indicating that temper outbursts in individuals with Prader-Willi syndrome may occur downstream from a fundamental deficit in emotion regulation (Tunnicliffe et al., 2014; Woodcock et al., 2009).

Two of the present outbursts that occurred while heart rate was recorded were followed by a period of at least one hour before heart rate returned to resting state, and for all three reported outbursts longer durations were followed by longer recovery time. This is

interesting as extended periods of repeated apologies have been reported to follow temper outbursts (Oliver et al., 2009; Tunnicliffe et al., 2014). One possibility that warrants further study is whether negative emotional experience for an individual may continue long after their overt outburst behaviors have subsided. This is particularly relevant for intervention programs for temper outbursts that may specify actions to follow an outburst as part of the program.

When considering the physiological data however, it is important to highlight that the present measure of heart rate was confounded by physical activity (Iellamo, 2001). We addressed this limitation by calculating the baselines with reference to concurrently recorded activity. However, for future research, portable devices are now available that allow concurrent measurement of heart rate and activity in the single device, which enables indices to be calculated from the heart rate data that are more sensitive to arousal than to movement (e.g. Koehler, de Marees, Braun & Schaenzer, 2011).

It is interesting to note that the self-report scale of negative emotional experience appeared to comprise three (of five) rating points that could be distinguished using the corresponding informant diary reports. One of the few fields that has systematically developed self-report scales for individuals with intellectual disabilities is the study of anger (a specific negative emotion), and these have generally comprised three or four rating points (Hagiliassis, Gulbenkoglou, Di Marco, Young & Hudson, 2005; Novaco & Taylor, 2008; Rose et al., 2013; Rose, 2013). Of interest for future research is whether an optimal number of rating points for self-report scales can be ascertained in relation to measured characteristics of a group of participants.

Additional limitations most relevant here include the number of days during which physiological recording could take place, which resulted from practical constraints due to the number of devices available. However, one advantage of the method we employed was that

the participants who demonstrated outbursts during physiological recording were not necessarily those with caregivers most adept in keeping the diary. If caregivers were requested to allow physiological recording over a longer period, it is likely that particularly highly motivated caregivers would be more highly represented in such a sample. In addition, due to constraints on participants' availability and researchers' time, the period between administration of the diary and the structured interview was substantial for many participants. In future research a second measure of outbursts could be incorporated into the diaries as regular retrospective questions on the temper outbursts observed over the previous specified period (e.g. three days). This would much more closely approximate an assessment of test-retest reliability. Finally, the threat to generalizability most important to consider here is that caregivers were recruited from a support organisation for families of individuals with PWS, and thus may be particularly highly motivated to complete the diaries to the best of their ability. An important next step will be to assess the validity of these behavior diaries in groups of caregivers selected for diverse reasons.

### **Acknowledgements**

[withheld for blind review]

### **Research involving human participants**

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## References

- Anzman-Frasca, S., Liu, S. W., Gates, K. M., Paul, I. M., Rovine, M. J., & Birch, L. L. (2013). Infants' Transitions out of a Fussing/Crying State Are Modifiable and Are Related to Weight Status. *Infancy*, 18(5), 662-686. doi: 10.1111/infa.12002
- Barr, R. G., Kramer, M. S., Boisjoly, C., McVeywhite, L., & Pless, I. B. (1988). Parental diary of infant cry and fuss behavior. *Archives of Disease in Childhood*, 63(4), 380-387.
- Barr, R. G., Kramer, M. S., Pless, I. B., Boisjoly, C., & Leduc, D. (1989). Feeding and temperament as determinants of early infant crying fussing behavior. *Pediatrics*, 84(3), 514-521.
- Belden, A. C., Thomson, N. R., & Luby, J. L. (2008). Temper tantrums in healthy versus depressed and disruptive preschoolers: Defining tantrum behaviors associated with clinical problems. *Journal of Pediatrics*, 152(1), 117-122. doi: 10.1016/j.jpeds.2007.06.030
- Bernat, E. M., Cadwallader, M., Seo, D., Vizueta, N., & Patrick, C. J. (2011). Effects of Instructed Emotion Regulation on Valence, Arousal, and Attentional Measures of Affective Processing. *Developmental Neuropsychology*, 36(4), 493-518. doi: 10.1080/87565641.2010.549881
- Bull (2013). *Longitudinal behaviour changes with age in Prader-Willi syndrome*, Understanding and changing behaviour in Prader-Willi syndrome, Chapter 3, pp 60-111, Unpublished PhD thesis, University of Birmingham, UK.
- Carlson, G. A., & Dyson, M. (2012). Diagnostic Implications of Informant Disagreement About Rage Outbursts: Bipolar Disorder or Another Condition? *Israel Journal of Psychiatry and Related Sciences*, 49(1), 44-51.
- Carlson, G. A., Potegal, M., Margulies, D., Gutkovich, Z., & Basile, J. (2009). Rages-What Are They and Who Has Them? *Journal of Child and Adolescent Psychopharmacology*, 19(3), 281-288. doi: 10.1089/cap.2008.0108
- Chen, K., Budman, C. L., Herrera, L. D., Witkin, J. E., Weiss, N. T., Lowe, T. L., . . . Mathews, C. A. (2013). Prevalence and clinical correlates of explosive outbursts in Tourette Syndrome. *Psychiatry Research*, 205(3), 269-275. doi: 10.1016/j.psychres.2012.09.029
- Dimitropoulos, A., Feurer, I. D., Butler, M. G., & Thompson, T. (2001). Emergence of compulsive behavior and tantrums in children with Prader-Willi syndrome. *American Journal on Mental Retardation*, 106(1), 39-51. doi: 10.1352/0895-8017(2001)106<0039:eocbat>2.0.co;2
- Einfeld, S. L., Smith, E., McGregor, I. S., Steinbeck, K., Taffe, J., Rice, L. J., . . . Guastella, A. J. (2014). A Double-Blind Randomized Controlled Trial of Oxytocin Nasal Spray

- in Prader Willi Syndrome. *American Journal of Medical Genetics Part A*, 164A(9), 2232-2239. doi: 10.1002/ajmg.a.36653
- Eisbach, S. S., Cluxton-Keller, F., Harrison, J., Krall, J. R., Hayat, M., & Gross, D. (2014). Characteristics of temper tantrums in Preschoolers with Disruptive Behavior in a Clinical Setting. *Journal of Psychosocial Nursing and Mental Health Services*, 52(5), 32-40. doi: 10.3928/02793695-20140110-02
- Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous-system activity distinguishes among emotions. *Science*, 221(4616), 1208-1210. doi: 10.1126/science.6612338
- Ervin, R. A., Radford, P. M., Bertsch, K., Piper, A. L., Ehrhardt, K. E., & Poling, A. (2001). A descriptive analysis and critique of the empirical literature on school-based functional assessment. *School Psychology Review*, 30(2), 193-210.
- Fernandez, C., Pascual, J. C., Soler, J., Elices, M., Portella, M. J., & Fernandez-Abascal, E. (2012). Physiological Responses Induced by Emotion-Eliciting Films. *Applied Psychophysiology and Biofeedback*, 37(2), 73-79. doi: 10.1007/s10484-012-9180-7
- Goldin, R. L., Matson, J. L., Tureck, K., Cervantes, P. E., & Jang, J. (2013). A comparison of tantrum behavior profiles in children with ASD, ADHD and comorbid ASD and ADHD. *Research in Developmental Disabilities*, 34(9), 2669-2675. doi: 10.1016/j.ridd.2013.04.022
- Green, J. A., Whitney, P. G., & Potegal, M. (2011). Screaming, Yelling, Whining, and Crying: Categorical and Intensity Differences in Vocal Expressions of Anger and Sadness in Children's Tantrums. *Emotion*, 11(5), 1124-1133. doi: 10.1037/a0024173
- Hagiliassis, N., Gulbenkoglul, H., Di Marco, M., Young, S., & Hudson, A. (2005). The Anger Management Project: A group intervention for anger in people with physical and multiple disabilities. *Journal of Intellectual & Developmental Disability*, 30(2), 86-96. doi: 10.1080/13668250500124950
- Holland, A. J., Whittington, J. E., Webb, B. T., Boer, H., & Clarke, D. (2003). Behavioral phenotypes associated with specific genetic disorders: evidence from a population-based study of people with Prader-Willi syndrome. *Psychological Medicine*, 33(1), 141-153. doi: 10.1017/s0033291702006736
- Holm, V. A., Cassidy, S. B., Butler, M. G., Hanchett, J. M., Greenswag, L. R., Whitman, B. Y., & Greenberg, F. (1993). Prader-willli syndrome - consensus diagnostic-criteria. *Pediatrics*, 91(2), 398-402.
- Hunziker, U. A., & Barr, R. G. (1986). Increased carrying reduces infant crying - a randomized controlled trial. *Pediatrics*, 77(5), 641-648.
- Iellamo, F. (2001). Neural mechanisms of cardiovascular regulation during exercise. *Autonomic Neuroscience-Basic & Clinical*, 90(1-2), 66-75. doi: 10.1016/s1566-0702(01)00269-7

- Kano, Y., Ohta, M., Nagai, Y., Spector, I., & Budman, C. (2008). Rage attacks and aggressive symptoms in Japanese adolescents with Tourette syndrome. *Cns Spectrums*, 13(4), 325-332.
- Koehler, K., de Marees, M., Braun, H. & Schaenzer, W. (2011) Evaluation of two portable sensors for energy expenditure assessment during high-intensity running. *European Journal of Sport Science*, 13(1), 31-41. doi: 10.1080/17461391.2011.586439
- Konst, M. J., Matson, J. L., & Turygin, N. (2013). Comparing the rates of tantrum behavior in children with ASD and ADHD as well as children with comorbid ASD and ADHD diagnoses. *Research in Autism Spectrum Disorders*, 7(11), 1339-1345. doi: 10.1016/j.rasd.2013.07.023
- Leclerc, J., O'Connor, K. P., Forget, J., & Lavoie, M. E. (2011). Behavioral Program for Managing Explosive Outbursts in Children with Tourette Syndrome. *Journal of Developmental and Physical Disabilities*, 23(1), 33-47. doi: 10.1007/s10882-010-9213-1
- Ma, X. L., Zhang, J., Zhong, W. N., Shu, C., Wang, F. T., Wen, J. N., . . . Liu, L. (2014). The diagnostic role of a short screening tool-the distress thermometer: a meta-analysis. *Supportive Care in Cancer*, 22(7), 1741-1755. doi: 10.1007/s00520-014-2143-1
- Maskey, M., Warnell, F., Parr, J. R., Le Couteur, A., & McConachie, H. (2013). Emotional and Behavioral Problems in Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 43(4), 851-859. doi: 10.1007/s10803-012-1622-9
- Novaco, R. W., & Taylor, J. L. (2008). Anger and assaultiveness of male forensic patients with developmental disabilities: Links to volatile parents. *Aggressive Behavior*, 34(4), 380-393. doi: 10.1002/ab.20254
- Oliver, C., Adams, D., Allen, D., Bull, L., Heald, M., Moss, J., Wilde, L. & Woodcock, K. (2013) Causal models of clinically significant behaviors in Angelman, Cornelia de Lange, Prader-Willi and Smith-Magenis syndromes. *International Review of Research in Developmental Disabilities*, Volume 44: Challenging Behavior, Elsevier. doi: 10.1016/B978-0-12-401662-0.00006-3
- Osterman, K., & Bjorkqvist, K. (2010). A cross-sectional study of onset, cessation, frequency, and duration of children's temper tantrums in a nonclinical sample. *Psychological Reports*, 106(2), 448-454. doi: 10.2466/pr0.106.2.448-454
- Petresco, S., Anselmi, L., Santos, I. S., Barros, A. J. D., Fleitlich-Bilyk, B., Barros, F. C., & Matijasevich, A. (2014). Prevalence and comorbidity of psychiatric disorders among 6-year-old children: 2004 Pelotas Birth Cohort. *Social Psychiatry and Psychiatric Epidemiology*, 49(6), 975-983. doi: 10.1007/s00127-014-0826-z
- Potegal, M., & Davidson, R. J. (2003). Temper tantrums in young children: 1. Behavioral composition. *Journal of Developmental and Behavioral Pediatrics*, 24(3), 140-147.

- Rainville, P., Bechara, A., Naqvi, N., & Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1), 5-18. doi: 10.1016/j.ijpsycho.2005.10.024
- Rose, J. (2013). A Preliminary Investigation into the Influence of Therapist Experience on the Outcome of Individual Anger Interventions for People with Intellectual Disabilities. *Behavioral and Cognitive Psychotherapy*, 41(4), 470-478. doi: 10.1017/s1352465812000823
- Rose, J., Willner, P., Shead, J., Jahoda, A., Gillespie, D., Townson, J., . . . Hood, K. (2013). Different Factors Influence Self-Reports and Third-Party Reports of Anger by Adults with Intellectual Disabilities. *Journal of Applied Research in Intellectual Disabilities*, 26(5), 410-419. doi: 10.1111/jar.12037
- Roth, A. J., Kornblith, A. B., Batel-Copel, L., Peabody, E., Scher, H. I., & Holland, J. C. (1998). Rapid screening for psychologic distress in men with prostate carcinoma - A pilot study. *Cancer*, 82(10), 1904-1908. doi: 10.1002/(sici)1097-0142(19980515)82:10<1904::aid-cnrc13>3.0.co;2-x
- St. James-Roberts, I., Hurry, J., & Bowyer, J. (1993). Objective confirmation of crying durations in infants referred for excessive crying. *Archives of Disease in Childhood*, 68(1), 82-84.
- St James-Roberts, I., & Peachey, E. (2011). Distinguishing infant prolonged crying from sleep-waking problems. *Archives of Disease in Childhood*, 96(4), 340-U381. doi: 10.1136/adsc.2010.200204
- Storch, E. A., Jones, A. M., Lack, C. W., Ale, C. M., Sulkowski, M. L., Lewin, A. B., . . . Murphy, T. K. (2012). Rage Attacks in Pediatric Obsessive-Compulsive Disorder: Phenomenology and Clinical Correlates. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(6), 582-592. doi: 10.1016/j.jaac.2012.02.016
- Tarbox, J., Wallace, M. D., Tarbox, R. S. F., Landaburu, H. J., & Williams, W. L. (2004). Functional analysis and treatment of low rate problem behavior in individuals with developmental disabilities. *Behavioral Interventions*, 19(3), 187-204. doi: 10.1002/bin.158
- Tunnicliffe, P., Woodcock, K., Bull, L., Oliver, C., & Penhallow, J. (2014). Temper outbursts in Prader-Willi syndrome: causes, behavioral and emotional sequence and responses by carers. *Journal of Intellectual Disability Research*, 58(2), 134-150. doi: 10.1111/jir.12010
- Wakschlag, L. S., Choi, S. W., Carter, A. S., Hullsiek, H., Burns, J., McCarthy, K., . . . Briggs-Gowan, M. J. (2012). Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 53(11), 1099-1108. doi: 10.1111/j.1469-7610.2012.02595.x

- Walz, N. C., & Benson, B. A. (2002). Behavioral phenotypes in children with Down syndrome, Prader-Willi syndrome, or Angelman syndrome. *Journal of Developmental and Physical Disabilities, 14*(4), 307-321. doi: 10.1023/a:1020326701399
- Whalley, B., & Hyland, M. E. (2013). Placebo by proxy: the effect of parents' beliefs on therapy for children's temper tantrums. *Journal of Behavioral Medicine, 36*(4), 341-346. doi: 10.1007/s10865-012-9429-x
- Whitaker, S., Walker, T., & McNally, C. (2004). The use of time base lag sequential analysis to look at the relationship between environmental events and challenging behavior in people with learning disabilities. *Behavioral and Cognitive Psychotherapy, 32*(1), 67-76. doi: 10.1017/s1352465804001079
- Whittington, J., Holland, A., Webb, T., Butler, J., Clarke, D., & Boer, H. (2004). Cognitive abilities and genotype in a population-based sample of people with Prader-Willi syndrome. *Journal of Intellectual Disability Research, 48*, 172-187. doi: 10.1111/j.1365-2788.2004.00556.x
- Whittington, J. E., Holland, A. J., Webb, T., Butler, J., Clarke, D., & Boer, H. (2001). Population prevalence and estimated birth incidence and mortality rate for people with Prader-Willi syndrome in one UK Health Region. *Journal of Medical Genetics, 38*(11), 792-798. doi: 10.1136/jmg.38.11.792
- Woodcock, K., Oliver, C., & Humphreys, G. (2009). Associations between repetitive questioning, resistance to change, temper outbursts and anxiety in Prader-Willi and Fragile-X syndromes. *Journal of Intellectual Disability Research, 53*, 265-278. doi: 10.1111/j.1365-2788.2008.01122.x
- Woodcock, K. A., Oliver, C., & Humphreys, G. W. (2011). The relationship between specific cognitive impairment and behavior in Prader-Willi syndrome. *Journal of Intellectual Disability Research, 55*, 152-171. doi: 10.1111/j.1365-2788.2010.01368.x



*Table 1: Participant demographics*

PP	Age (years: months)	Place of residence	Settings for diary administration	Genetic status <sup>2</sup>	VABS: daily living skills age equivalence score (years: months)
AA	10:3	with family	home, school	mUPD	10:1
AB	12:6	with family	home, school	mUPD	6:6
AC	43:1	with family	home, day centre <sup>3</sup>	Del + Transloc	6:5
BA	9:5	with family	home, school	unknown	5:2
BB	33:4	group home	group home, family home	mUPD	5:6
BC	11:8	with family	home, school	mUPD	3:7
CA	9:7	with family	home	mUPD	5:1
CB	10:9	with family	home	mUPD	3:9
CC	47:10	group home	group home	unknown	14:1
DA	16:4	with family	home, school	unknown	5:3
DB	45:8	group home	group home	UPD	12:9
DC	39:7	group home	group home	unknown	8:7

<sup>2</sup> Caregiver reported previous genetic diagnosis of maternal uniparental disomy of chromosome 15 (mUPD), a deletion (del) or a translocation (transloc) in the C15q11-q13 region, or genetic test result unknown.

<sup>3</sup> No diaries were returned by the day centre

Table 2: Convergent validity assessment of behavior diaries relative to structured interviews conducted between 18 months before and 18 months later

PP	Months between diary and interview <sup>4</sup>	Diary: # of component behaviors reported	Diary behaviors also in interview	Diary: Total # outbursts <sup>5</sup>	Interview: expect to next see within ¼ hour; 1 hour; 1 day; 1 week; 1 month	Diary modal duration	Interview duration: typical	Diary: longest outburst	Interview: longest outburst in last month
AA	7	6	3 (50%)	10	1 week (4-2)	≤5	<5	10	<15
AB	-7	6	3 (50%)	12	1 day (31-5)	≤5	<5	90	<60
AC	-18	7	6 (86%)	17	1 day (31-5)	≥60	>60	all day	>60
BA	-8	6	4 (66%)	3	1 day (31-5)	≤15	<15	15	<15
BB	-4	2	1 (50%)	1	too unpredictable	15	<15	15	<15
BC	****	1	****	5	****	≤5	****	5	****
CA <sup>6</sup>	12	9	8 (88%)	20	1 hour (252-22)	≤15	<15	45	> 60
CB	-17	5	4 (80%)	10	1 day (31-5)	≤15	<15	20	<60
CC	0	None	N/A	None	1 week (4-2)	N/A	<15	N/A	<15
DA	17	7	5 (71%)	5	1 hour (372-32)	≤15	<15	>60	>60
DB	-5	None	N/A	None	> 1 month (< 1)	N/A	<15	N/A	>60
DC	-3	None	N/A	None	1 week (4-2)	N/A	<15	N/A	<15

<sup>4</sup> Positive numbers indicate that the diary data were collected before the interview data

<sup>5</sup> For participant AA and AC, data were recorded for 4 weeks and 2 days with 2 days being separate from a 4 week block, and comprising the physiological recording days. Thus, data for this comparison are taken from the 4 week block.

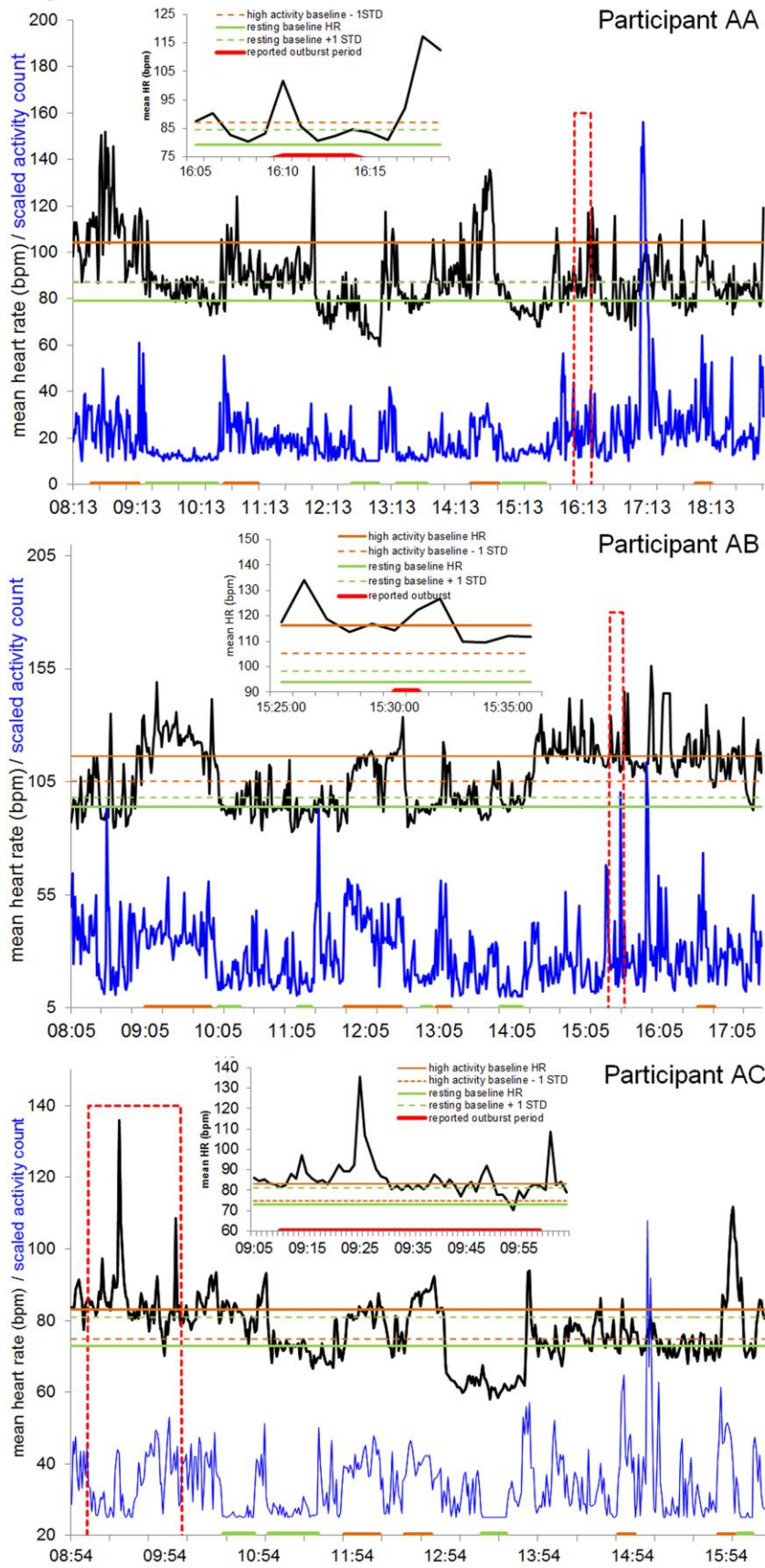
<sup>6</sup> Diary data were only available for 3 weeks from this informant due to other commitments. Boundaries for the categories for frequency report in the interviews (column 6) are therefore adjusted accordingly.

*Table 3:* Informant report diary entries corresponding to each category of temper outbursts as classified using Participant BA's self-report ratings of negative emotion experience.

Informant report diary entries on corresponding outbursts	Participant BA: self-report rating of emotional experience				
	OK (1)	A little bad (2)	Quite bad (3)	Very bad (4)	Terrible (5)
Number of outbursts	0	2	6	2	9
Median duration (minutes)	N/A	3	6	7.5	5
Range of durations	N/A	3 - 3	3 - 10	5 - 10	3 - 15
Median intensity (1-10 Likert)	N/A	5.5	9	7.5	9
Range of intensities	N/A	4 - 7	4 - 10	5 - 10	5 - 10

*Figure 1:* mean heart rate (black) and scaled mean activity counts (blue) for the three participants reported to show a temper outburst during physiological recording, across the entire recording day when the outburst was reported. Horizontal green lines represent resting state baselines and 1 standard deviation above the baseline (dashed); and horizontal orange lines represent high activity baselines and 1 standard deviation below the baseline (dashed). Green and orange marks along the X-axes show the periods used to calculate the corresponding baselines.

Figure 1



## Supplementary methods

### *Semi-structured Interview Schedule for recruitment and creating the behaviour diary*

#### *(Telephone)*

#### Temper outburst/tantrum –

*Highly emotional response. Period of crying, screaming, angry ranting, shouting, stamping feet, or kicking. Can last for a prolonged period of time.*

1. Does \_\_\_\_\_ ever display temper outburst behaviour?
2. If so, what behaviours does \_\_\_\_\_ show during a typical episode?
3. How often do the temper outbursts occur?
4. Think about the last time a temper outburst occurred, what seemed to trigger the behaviour in this example?
5. In the example you thought about, how did you respond to the temper outburst?
6. Roughly, how many times does the trigger you mentioned actually result in a temper outburst?
7. Are there times when this particular trigger does not actually trigger a temper outburst?
8. Roughly, how many times after a temper outburst would you respond in the way that you mentioned in the example?
9. What behaviours does \_\_\_\_\_ typically show after an outburst?
10. Think of other examples in which a temper outburst occurred, what seemed to trigger this and how did you respond?
11. How long roughly do the temper outbursts last for?

#### Review

12. Finally go over behaviours, antecedents and consequences listed.  
Any more behaviours, antecedents, consequences?

Can these antecedents and consequences be categorised? Do this with parent/carer.

*Example of a behaviour diary report card*

Date	Time	Behaviour Observed	Duration	Intensity 1-10	Before	After
		<input type="checkbox"/> Shouted <input type="checkbox"/> Kicked <input type="checkbox"/> Spat at someone <input type="checkbox"/> Red Face <input type="checkbox"/> Stamped feet <input type="checkbox"/> Other_____			<input type="checkbox"/> Change to routine <input type="checkbox"/> Change to expectation <input type="checkbox"/> Somebody told him off <input type="checkbox"/> Somebody told him off about food <input type="checkbox"/> Did not get some food he wanted <input type="checkbox"/> Other_____	<input type="checkbox"/> Cried <input type="checkbox"/> Said sorry repeatedly <input type="checkbox"/> Blamed somebody else <input type="checkbox"/> Wanted a cuddle <input type="checkbox"/> Tired <input type="checkbox"/> Other _____

*Self-report negative emotional experience sheet*

Date:

Time:



1. How bad did you feel?

Colour up to a point on the thermometer or mark a point

		<b>Terrible</b> Extremely angry, worried, annoyed or upset	
		<b>Very bad</b> Very angry, worried, annoyed or upset	
		<b>Quite bad</b> Angry, worried, annoyed or upset	
		<b>A little bad</b> A little angry, worried, annoyed or upset	
		<b>OK</b> I didn't feel bad at all	