Quantifying respiratory sinus arrhythmia: Effects of misspecifying breathing frequencies across development

TIFFANY M. SHADER, LISA M. GATZKE-KOPP, SHEILA E. CROWELL, M. JAMILA REID, JULIAN F. THAYER, MICHAEL W. VASEY, CAROLYN WEBSTER-STRATTON, ZIV BELL, AND THEODORE P. BEAUCHAINE

The Ohio State University; Pennsylvania State University; University of Utah; Incredible Years, Seattle; and University of Washington

Abstract

Low resting respiratory sinus arrhythmia (RSA), and to a lesser extent excessive RSA reactivity to emotion evocation, are observed in many psychiatric disorders characterized by emotion dysregulation, including syndromes spanning the internalizing and externalizing spectra, and other conditions such as nonsuicidal self-injury. Nevertheless, some inconsistencies exist. For example, null outcomes in studies of RSA–emotion dysregulation relations are sometimes observed among younger participants. Such findings may derive from use of age inappropriate frequency bands in calculating RSA. We combine data from five published samples (N = 559) spanning ages 4 to 17 years, and reanalyze RSA data using age-appropriate respiratory frequencies. Misspecifying respiratory frequencies results in overestimates of resting RSA and underestimates of RSA reactivity, particularly among young children. Underestimates of developmental shifts in RSA and RSA reactivity from preschool to adolescence were also observed. Although correlational analyses revealed weak negative associations between resting RSA and aggression, those with clinical levels of externalizing exhibited lower resting RSA than their peers. No associations between RSA reactivity and externalizing were observed. Results confirm that age-corrected frequency bands should be used when estimating RSA, and that literature-wide overestimates of resting RSA, underestimates of RSA reactivity, and underestimates of developmental shifts in RSA and RSA reactivity may exist.

High-frequency heart rate variability, defined by cyclic oscillations in heart rate across successive respiratory cycles (see, e.g., Malik & Camm, 1995; Zisner & Beauchaine, 2016), is of interest to developmental psychopathologists given well replicated associations with numerous aspects of psychological well-being and dysfunction (see, e.g., Beauchaine, 2001; Beauchaine & Thayer, 2015; Berntson et al., 1997; Obradovic & Boyce, 2012). As outlined below, high-frequency heart rate variability is quantified by measuring respiratory sinus arrhythmia (RSA), the component of HRV associated with breathing. Over the past two decades, consistent findings have emerged demonstrating positive associations between RSA and numerous psychological adjustment outcomes among children, adolescents, and adults. These outcomes include, effective behavior regulation during social challenges (e.g., Hastings, Nuselovici, et al., 2008; Hastings, Sullivan, et al., 2008), empathic responding to others who are in distress (Fabes, Eisenberg, & Eisenbud, 1993; Liew et al., 2011), positive social interactions with partners (Diamond, Fagundes, & Cribbet, 2012), proficient executive function (Thayer, Hansen, Saus-Rose, & Johnsen, 2009), social competence (Eisenberg et al., 2008), sustained attention capabilities (Suess, Porges, & Plude, 1994), temperamental complaisance (Huffman et al., 1998), and protection against psychopathology in adverse rearing conditions (e.g., El-Sheikh, Harger, & Whitson, 2001; Shannon, Beauchaine, Brenner, Neuhau, & Gatzke-Kopp, 2007). Thus, RSA is associated positively with a broad array of behavioral competencies.

RSA and Psychopathology

In contrast to positive associations between RSA and both psychological adjustment and resilience, abnormally low resting RSA, and to a lesser extent excessive RSA reactivity (i.e., withdrawal) to behavioral challenges, especially emotion evocation tasks, are observed in many psychiatric disorders that are characterized by emotion dysregulation, and correlate with several psychopathological traits (see Beauchaine,
2015a, 2015b; Beauchaine & Thayer, 2015; Chambers & Allen, 2007). These include anxiety disorders (e.g., Hastings, Nuselovici, et al., 2008; Hastings, Sullivan, et al., 2008; Kemp et al., 2014; Thayer, Friedman, & Borkovec, 1996), attention problems (see Rash & Aguirre-Camacho, 2012), autism (Neuhaus, Bernier, & Beauchaine, 2014; Patriquin, Scarpa, Friedman, & Porges, 2015), bipolar disorder (Moon, Lee, Kim, & Hwang, 2013; Quintana et al., 2015), borderline personality disorder (Austin, Riniolo, & Porges, 2007), callousness (de Wied, van Bokxel, Mathys, & Meews, 2012), conduct disorder (e.g., Beauchaine, Gatzke-Kopp, & Mead, 2007; Beauchaine, Katkin, Strassberg, & Snarr, 2001), depression (e.g., Koenig, Kemp, Beauchaine, Thayer, & Kaess, 2016; Rottenberg, 2007; Rottenberg, Salomon, Gross, & Gotlib, 2005; Rottenberg, Wilhelm, Gross, & Gotlib, 2002), nonsuicidal self-injury (e.g., Crowell et al., 2005), panic disorder (e.g., Asmundson & Stein, 1994), phobia (e.g., Ahs, Sollers, Furmark, Fredrikson, & Thayer, 2009), psychopathy (Hansen, Johnsen, Thornton, Waage, & Thayer, 2007), schizophrenia (see Clamor, Lincoln, Thayer, & Koenig, 2015; Quintana et al., 2015), and trait hostility (Sloan et al., 1994), among others. Moreover, heterotopic comorbidity is associated with greater reductions in RSA during emotion evocation than either internalizing or externalizing symptoms alone (Calkins, Graziano, & Keane, 2007; Pang & Beauchaine, 2013).

This lengthy, transdiagnostic list indicates that low resting RSA and excessive RSA reactivity to emotion evocation mark one or more fundamental self-regulatory functions that are disrupted across various manifestations of psychopathology (see Beauchaine, 2001, 2015a, 2015b; Beauchaine & Thayer, 2015). Reduced tonic RSA is nearly always found among adolescent and adult samples with clinical levels of psychopathology, and excessive RSA reactivity is often observed among clinical samples during emotion evocation tasks. Several important questions emerge from these findings. First, what specific trait does RSA mark that captures vulnerability to such a broad range of psychopathological conditions, and second, what are the neural substrates of this trait? Addressing these questions is important given that improved understanding of etiology often leads to more effective interventions (see, e.g., Beauchaine & Cicchetti, 2016; Beauchaine, Neuhaus, Brenner, & Gatzke-Kopp, 2008). In addition, how does this trait develop, and do its neural substrates exhibit plasticity, which may also have implications for treatment? These questions have been at the forefront of our research for almost two decades, but especially in recent years (see, e.g., Beauchaine, 2001, 2015a, 2015b; Beauchaine, Gartner, & Hagen, 2000; Beauchaine & McNulty, 2013; Beauchaine & Thayer, 2015; Beauchaine et al., 2001, 2007, 2013; Conradt et al., 2016; Crowell et al., 2005, 2014; Neuhaus, Bernier, & Beauchaine, 2016).

**RSA, Emotion Dysregulation, and Prefrontal Cortex Function**

It now well recognized that, like deficiencies in RSA, emotion dysregulation is observed across a wide range of psychiatric conditions (e.g., Beauchaine, 2001, 2015b; Beauchaine & Thayer, 2015; Cole, Hall, & Hajal, in press). An ever-expanding literature indicates that difficulties with emotion regulation (ER) predispose to, and are observed in, most forms of psychopathology (e.g., Beauchaine, 2015a; Cole, Dennis, Hall, & Martin, 2008; Kaufman et al., 2016). Emotion regulation is often defined as the set of processes through which individuals shape their emotional experience and expression in support of adaptive behavior (see, e.g., Aldao, 2013; Thompson, 1990). Shaping of emotions is effected through various interrelated mechanisms, some automated, some effortful, and some social, including attention allocation and other forms of cognitive control, dyadic interactions in core social relationships, operant reinforcement of physiological responding, and self-directed behaviors such as avoidance, reappraisal, and problem solving (e.g., Aldao, Nolen-Hoeksema, & Schweizer, 2010; Beauchaine, & Zalewski, 2016; Butler & Randall, 2013; Crowell et al., 2016).

In contrast to ER, emotion dysregulation is a pattern of emotional experience and/or expression that interferes with/obstructs effective goal-directed behavior (e.g., Beauchaine, 2015a; Beauchaine & Gatzke-Kopp, 2012). Across nearly all forms of psychopathology, at least one negative emotion (e.g., anger, anxiety, panic, rage, sadness) is experienced the sympathetic nervous system to operate unopposed, resulting in large increases in cardiac output to cope with environmental demands. Inferences about emotion regulation/dysregulation should therefore be gleaned from emotion evocation tasks, not attention allocation tasks (see Zisner & Beauchaine, 2016). Second, normative individual differences in what we think of as symptoms of psychopathology often mark nonsympathetic, temperament constructs such as behavioral inhibition, shyness, exuberance, and appetitive engagement with the environment, not maladaptation (see, e.g., Beauchaine, 2001; Degnan et al., 2011).
Quantifying RSA across development

3

either so intensely or so persistently that it interferes with adaptive function (see Beauchaine, 2015a). As we have reviewed elsewhere, and as outlined briefly above, RSA is a reliable autonomous nervous system marker of emotion dysregulation (see, e.g., Austin et al., 2007; Beauchaine, 2001, 2015a, 2015b; Porges, 1995, 2007).

Most first-generation theories of emotion regulation and dysregulation considered the parasympathetic nervous system (PNS) to be a self-regulatory system in its own right (see, e.g., Fox, 1994). According to these theories, both RSA and self-regulatory capacity emanate from the PNS, with central nervous system source nuclei in the brainstem (see Porges, 1995, 1996). Although these brainstem structures are clearly implicated in cardiorespiratory coupling, as reflected in RSA (see below; e.g., Garcia, Koschnitzky, Dashevskiy, & Ramirez, 2013), more recent models propose that associations between ER and RSA are affected through top-down, prefrontal control over subcortical neural networks (see Beauchaine, 2015a; Beauchaine & Thayer, 2015; Thayer et al., 2009). According to these newer models, brainstem nuclei specified in earlier theories of ER serve as final common pathways from prefrontal regions to the PNS, but are not responsible for ER itself.

Although full articulation of the cortical substrates of ER and RSA is beyond the scope of this article, several findings, many of which derive from extensive and accumulating neuroimaging research that was unavailable when first-generation models were formulated, provide strong support for prefrontal influences over both ER and RSA. For example, an increasing literature base indicates that neural regulation of fear and anxiety occurs through lateral prefrontal inhibition of amygdalar activity and reactivity, and that neural regulation of impulsivity occurs through orbitofrontal and dorsolateral prefrontal inhibition of striatal activity and reactivity (see, e.g., Davidson, 2002; Heatherton, 2011; Heatherton & Wagner, 2011; Macdonald, Goines, Novacek, & Walker, 2016). Furthermore, RSA has been linked to prefrontal cortex function through several research paradigms and methods, including cerebral blood flow as assessed using both positron emission tomography and fMRI (see Thayer, Ahs, Fredrikson, Sollers, & Wager, 2012), lesion studies (Buchanan et al., 2007; Katona & Jih, 1975; McCabe, Yongue, Ackles, & Porges, 1985; Santucci et al., 2008; Thayer, Hansen, & Johnsen, 2010).

Time versus frequency domain assessment

The simplest approach to computing HRV is to use time domain measures. These include the standard deviation of R-R intervals, and the mean successive difference in R-R intervals, among other metrics. Time domain measures calculate HRV directly from the interbeat interval time series. These measures are rarely used in the child and adolescent literature. For this reason, and because they are not on the same metric as frequency domain measures (described below) and therefore cannot be compared directly, we do not discuss them further.

Other commonly used measures of HRV are computed in the frequency domain. Spectral analysis, including fast Fourier transform (FFT) and autoregressive (AR) techniques,
is the most commonly used method (Poliakova et al., 2014). Frequency domain methods quantify the amount, or “power,” of HRV within specified frequency bands, which are sometimes referred to as “windows” (see, e.g., Grossman, 1992; Parati, Saul, Di Rienzo, & Mancia, 1995). Fast-Fourier analysis converts time-related phenomena into a spectral density function across all observed frequencies. As depicted in Figure 1, FFT can isolate and quantify spectral power within specific frequency bands, which may have important implications for identifying peripheral and central nervous system sources of health and disease (see above). By convention, adult spectral density functions are typically divided into low-frequency HRV (<0.04 Hz), mid-frequency HRV (0.04–0.15 Hz), and high-frequency HRV (>0.15 Hz). At normal adult breathing rates, high-frequency HRV yields an estimate of RSA. Provided that the RSA band is defined to match respiratory rates of participants, high-frequency spectral power quantifies vagal efference to the heart (see Grossman, 1992; Grossman, Karemaker, & Wieling, 1991; Ritz, 2009). However, breathing rates vary substantially both within and between individuals and samples. This is especially problematic when ages of participants are dispersed widely, since children breathe at much higher frequencies than adults (see below). To avoid misspecifying respiratory frequencies, some researchers who work with adults have turned to AR spectral analysis to bypass population-based selection of appropriate respiratory windows (see Thayer, 2002; Thayer et al., 2010; Thayer, Peasley, & Muth, 1996). AR methods apply true components analysis to the interbeat interval time series to empirically derive the appropriate respiratory frequency for each participant (see Kay & Marple, 1981). Autoregressive spectral analysis should therefore be more accurate than spectral analysis conducted without correction for respiration.

**Developmental Shifts in RSA and Links to Psychopathology**

To date, AR spectral analysis has not been used in child or adolescent samples. Furthermore, a large proportion of studies in the child development and developmental psychopathology literatures have used RSA frequency bands that do not match breathing rates of participants. Although our intent is not to criticize individual researchers or labs (our own work suffers from the same limitation), a brief literature search yields dozens of examples of respiratory frequencies being misspecified in research with both normal and psychopathological samples of children and adolescents. These misspecifications include (a) use of adolescent and adult frequency bands to assess RSA among infant and child samples (the most common misspecification), (b) use of child frequency bands to evaluate RSA among adolescents and adults, (c) use of a single frequency band to evaluate RSA among cross sectional samples with broad age ranges, and (d) use of a single frequency band to evaluate development of RSA in longitudinal samples. Such misspecifications may yield artifactual RSA–behavior relations (Type I errors), and/or obscure real
RSA behavior relations (Type II errors). Given that 4-year-olds breathe at over twice the respiratory frequency of adults, such effects could be substantial.

With these caveats in mind, evidence collected to date suggests that among individuals without psychopathology, resting RSA increases linearly with age from preschool to adolescence (see Beauchaine et al., 2001; Crowell et al., 2006; Zisner & Beauchaine, 2016). After young adulthood, RSA decreases gradually across the life span, although such decreases are offset somewhat by physical fitness (e.g., Byrne, Fleg, Vaitkevicius, Wright, & Porges, 1996). As discussed above, however, accurate assessment of RSA depends on appropriate specification of respiration rates, which decrease substantially from childhood to adulthood (Fleming et al., 2011; Wallis, Healy, Undy, & Maconochie, 2005). Table 1 summarizes observed respiration rates (at the 2.5 percentile, which captures 97.5% of individuals) for children and adolescents, ages 4–16 years. Most evidence for increases in RSA across development derives from assessments that have not used age-corrected respiratory frequencies. Thus, the magnitude (and existence) of developmental shifts in RSA may be uncertain. As shown in Figure 2, using an adult respiratory band with children is likely to overestimate RSA, since doing so adds spectral power that is not respiratory related. Accordingly, one objective of this article is to evaluate age effects on RSA by reanalyzing existing data from our labs that were not analyzed originally with proper respiratory frequencies.

Evidence also suggests that age moderates links between RSA both internalizing and externalizing psychopathology. Group differences between externalizing samples and controls in resting RSA are often not observed until middle school (e.g., Beauchaine et al., 2001, 2007; Crowell et al., 2006). Similar findings have recently been reported for depression (Koenig et al., 2016). According to developmental theories of emotion dysregulation, links between RSA and psychopathology emerge in part through protracted operant conditioning of emotional lability within families. From this perspective, deficiencies in RSA develop over time through repeated aversive interaction patterns that slowly canalize abnormal patterns of physiological activity and reactivity (see Beauchaine, 2015a, Beauchaine & McNulty, 2013; Beauchaine & Zalewski, 2016; Crowell et al., 2014). Given the above discussion, it is important to eliminate the possibility that apparent age effects on RSA–psychopathology relations are artifactual. Thus, we evaluate whether RSA–externalizing relations differ as a function of age when age-appropriate frequency bands and AR spectral analysis are used to compute RSA. Addressing this question may help to determine whether existing findings are valid despite widespread misspecification of respiratory frequencies.

Finally, developmental changes in RSA reactivity (i.e., withdrawal) to emotion evocation are also observed, becoming less pronounced among healthy controls from preschool to adolescence (see Beauchaine et al., 2001; Crowell et al., 2006; Pang & Beauchaine, 2013). This suggests that healthy control children become better able to regulate their physiological reactions to emotion evocation as they mature. In contrast, children with externalizing behavior problems become more reactive as they mature (see Pang & Beauchaine, 2013). Accordingly, we also evaluate age effects on RSA reactivity using developmentally appropriate respiratory frequencies and AR spectral analysis.

### Method

#### Data sets

All analyses were conducted using data from a sample that combined participants from five separate studies (total \( N = 559 \)). The age range of participants was 4 to 17 years. All data sets included assessments of baseline RSA, RSA reactivity, and externalizing psychopathology. Although each study included additional measures, these are not the focus of this article, so they are not discussed. Interested readers are referred to original sources. Methods of recruitment and specific tasks used in each study, which are presented in order of increasing age, are described below. As is customary in psychophysiological research, RSA reactivity was computed as task minus baseline, and RSA data were log transformed to normalize their distributions (see Zisner & Beauchaine, 2016).

In Crowell et al.’s (2006) study, participants between ages 4 and 6 years were recruited to investigate autonomic correlates of attention-deficit/hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) in early childhood. Patterns of RSA and RSA reactivity were compared between 18 participants with ADHD and ODD versus 20 controls.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Breaths/Min</th>
<th>Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>20</td>
<td>0.33</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>0.32</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>0.30</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>0.28</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>0.28</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>0.27</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>0.25</td>
</tr>
<tr>
<td>11</td>
<td>14</td>
<td>0.23</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>0.23</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>0.22</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>0.20</td>
</tr>
<tr>
<td>15</td>
<td>12</td>
<td>0.20</td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>0.18</td>
</tr>
<tr>
<td>Adult</td>
<td>9</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 1. Respiration rates (2.5 percentile; breaths/minute), and corresponding cycles per second (Hz) by age

Note: Respiration rates were evaluated among 1,109 children, adolescents, and adults, as reported by Wallis et al. (2005). The 2.5 percentile captures 97.5% of individuals.
The current analysis includes 35 of these participants (17 ADHD/ODD: 8 female, 9 male; 18 control: 9 female, 9 male). Three participants were dropped (1 ADHD/ODD, 2 control) due to file corruption. The sample was 65.7% Caucasian.

The ADHD/ODD group was ascertained via interview using the Child Symptom Inventory (CSI; Gadow & Sprafkin, 1997a), which yields diagnostic cutoffs for most DSM-IV (American Psychiatric Association, 1994) syndromes. Parent reports on the ADHD and ODD scales of the CSI were used. Parents also completed the aggression, attention problems, and anxious/depressed subscales of the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1991). Children were included in the externalizing group if they met DSM-IV criteria for both ADHD (hyperactive/impulsive or combined subtype) and ODD on the CSI, and scored at or above the 98th percentile on the attention problems and aggression subscales of the CBCL. Control group children were required to score below the 60th percentile on the CBCL scales, and to remain medication free throughout the intervention. Among the 99 treated participants, 48 also met criteria for ODD, which was not an exclusion criterion. The current analysis included 121 participants (86 ADHD: 21 female, 65 male; 35 control: 11 female, 24 male). Nineteen participants were dropped (13 ADHD, 6 control) due to problems with either data acquisition or data corruption that resulted during transfer of files to a new server. The sample was 76.0% Caucasian.

As in the Crowell et al. (2006) study, all children underwent a stimulus-free, 5-min resting baseline in a sound-attenuated room. Baseline RSA was extracted from the last minute of this rest period. Children also participated in a challenging 7-min block building task that is quite difficult for 4- to 6-year-olds to complete, and pulls for both parental and child frustration, as assessed by dyadic conflict and expressions of negative affect (Beauchaine, Strassberg, Kees, & Drabick, 2002). RSA was defined as spectral power above 0.15 Hz.

In the Fortunato, Gatzke-Kopp, and Ram (2013) study, participants were drawn from a clinical trial of a preventive intervention for high-risk children (see Gatzke-Kopp, Greenberg, & Bierman, 2015, for details). Briefly, all children who were enrolled in kindergarten among nine schools in a single school district were screened using teacher reports of aggressive/oppositional behavior from the Teacher Observation of Child Adaptation—Revised (Werther-Larsson, Kellam, & Wheeler, 1991). Children who scored in the upper quartile of externalizing behavior within each classroom (n = 207) were recruited as a part of a high-risk externalizing group (diagnostic procedures were not used), and children who scored in the bottom quartile in matched classrooms (n = 132) were recruited into a low-risk comparison group. Among high-risk children, 100 were assigned randomly to a social skills training program that took place during the second half of
kindergarten and the first half of first grade. The remaining 107 were assigned to a no-treatment control group. Unlike the Fortunato et al. manuscript, data included in the present analyses were taken from a 2-year follow-up when children were in second grade. Given evidence that interventions for externalizing behavior can affect autonomic function (e.g., Beauchaine et al., 2015), children in the intervention condition were excluded from analyses.

At the time of physiological assessments, second-grade teachers completed the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) to assess current levels of externalizing and internalizing symptoms. Of the original sample, 180 children remained in the school district and had teacher-supplied symptom ratings. Physiological data were missing for 3 participants, resulting in a total of \( n = 177 \) for the present analyses (100 high risk: 37 female, 63 male; 77 low risk: 24 female, 53 male). The sample was 30.5% Caucasian.

Resting RSA was computed while participants were asked to sit still for 2-min while they “travelled through space” to play a game. A moving star field video was presented on a computer screen during the recording. Participants first completed a cognitive game (these data are not evaluated here), before “traveling through space” again for another 2 min, after which they viewed four emotion-induction film clips extracted from the movie The Lion King (see Fortunato et al., 2013; Gatzke-Kopp, Jetha, & Segalowitz, 2014, for details). In the present analyses, baseline RSA was extracted from the final minute of the first 2-min baseline (before the child participated in any tasks), and RSA reactivity was assessed only during the sad film clip, to maintain consistency with the other studies.

The Pang and Beauchaine (2013) study included 207 children, ages 8–12 years, who participated in a longitudinal evaluation of depression and conduct problems across middle childhood (also see Beauchaine, Hong, & Marsh, 2008; Kopp & Beauchaine, 2007; Shannon et al., 2007; Vasilyev, Crowell, Beauchaine, Mead, & Gatzke-Kopp, 2009). Based on results of interviews with parents using the CSI, children were placed into one of the following four groups: control (\( n = 69 \)) could not meet CSI criteria for ODD, conduct disorder (CD), major depressive disorders (MDD), or dysthymia (DYS); conduct problems (\( n = 30 \)) met CSI criteria for CD and/or ODD, did not meet CSI criteria for DYS or MDD; depression/dysthymia (\( n = 28 \)) met CSI criteria for DYS or DYS and failed to meet CSI criteria for CD and ODD; comorbid (\( n = 80 \)) met inclusion criteria for both the depressed and conduct problems groups. Parents also completed the aggression, attention problems, and anxious/depressed subscales of the CBCL. Scores at or above the 95th percentile on CBCL scales that were relevant to each group (e.g., aggression for the conduct problems group, anxious/depressed for the depression/dysthymia group) were also required for admission. The current analysis included 162 participants (83 conduct problems/comorbid, 23 depression/dysthymia: 31 female, 75 male; 56 control: 30 female, 26 male). Of the original 207 participants, 45 participants were dropped due to file corruption when data were transferred to a new server (27 conduct problems/comorbid, 5 depression/dysthymia, 13 control). The sample was 61.7% Caucasian.

All participants underwent a stimulus-free, 5-min resting baseline in a sound-attenuated room. Baseline RSA was extracted from the last minute, 2-min window following the 3-min baseline window. Next, participants watched a film scene from The Champ, a film that depicts a boy with his dying father. This film is effective in eliciting strong emotional reactions (see Marsh, Beauchaine, & Williams, 2008). RSA was defined as spectral power above 0.15 Hz.

The Beauchaine et al. (2001) study compared patterns of psychophysiological responding among boys, ages 12–17 years, with either (a) ADHD (\( n = 17 \)), (b) ADHD and CD (\( n = 20 \)), or (c) no psychiatric disorder (\( n = 22 \)). Diagnoses were ascertained via parent interviews using the Adolescent Symptom Inventory (Gadow & Sprafkin, 1997b). Adolescents were included if they met DSM-IV diagnostic criteria for their respective group. Adolescents in the ADHD group were also required to score at or above the 95th percentile on the CBCL attention problems scale, below the 70th percentile on the CBCL delinquency scale, and endorse no CD criteria. Adolescents in the ADHD/CD group were required to score at or above the 95th percentile on the delinquency and aggressive behavior subscales of the CBCL. Control group participants could not meet DSM-IV criteria for any disorder, or score above the 60th percentile on any CBCL scale. The current analysis includes 64 participants (18 ADHD, 22 ADHD/CD: 0 female, 40 male; 24 control: 0 female, 24 male). Although the original analysis included 59 participants, data from 5 additional participants were collected following publication. The sample was 79.7% Caucasian.

Resting RSA was measured during the last minute of a 5-min, stimulus free baseline in a sound-attenuated room. RSA reactivity was measured while participants watched a 2-min video of an escalating conflict between age- and race-matched peers. RSA was defined as spectral power above 0.15 Hz.

**Original analyses**

As noted above, RSA was computed in a similar fashion across data sets. Although the second-grade data from the Fortunato et al. (2013) paper have not been reported elsewhere, all other studies used FFT, and defined RSA as spectral power above 0.15 Hz. These windows are all appropriate for older adolescents and adults, not for young children or early adolescents. As described by Zisner and Beauchaine (2016), windowing should differ depending on participant age, given age changes in respiration (see above; Table 1).

**Current analysis**

Kubios HRV software (version 2.2) was used to analyze all data. Kubios HRV includes options for analysis in which (a) FFT frequency bands are specified manually, and (b) AR spectral analysis is used (Tarvainen, Niskanen, Lipponen, Ranta-aho, & Karjalainen, 2014). Prior to analysis, a medium artifact correction was applied, and smoothing priors were
used both to reduce distortion of data end points and to remove the lowest frequency trend components (λ = 300; Tarvainen, Ranta-aho, & Karjalainen, 2002). All data were analyzed using both age-corrected FFT (FFT AC), based on respiratory frequencies specified by Wallis et al. (2005), and AR. For the FFT AC analyses, the upper end for all frequency windows was set at 0.50 Hz, regardless of age. For example, 6-year-old children’s frequency windows were set at 0.30–0.50 Hz, whereas 16-year-old children’s frequency windows were set at 0.18–0.50 Hz. As noted in Table 1, these age corrections capture respiratory frequencies for 97.5% of individuals. All spectral density functions were inspected visually to ensure that respiratory peaks and their surrounding distributional dispersion were captured using an upper bound of 0.50 Hz. In all but two cases, the 0.50-Hz upper boundary was effective. For these two exceptions, the upper bound was moved manually. For AR spectral analysis, respiratory spectral power distributions were identified for each participant by Kubios. As outlined above, all baseline estimates of RSA were computed during the last minute of each individual study task. Task estimates of RSA were computed across the entirety of each individual study task.

In some cases, we had baseline but not task data for a participant, and in other cases we had task data but no baseline data. In these circumstances, we imputed missing data following recommendations outlined by Graham (2009). Baseline data were imputed for 8 participants (0.01% of total data), and task data were imputed for 19 participants (0.03% of total data).

Our index of externalizing behavior was restricted by available measures. The only measure of externalizing that was available for at least four of the five studies was the CBCL aggression subscale (Beauchaine et al., 2001, 2013, Crowell et al., 2006; Pang & Beauchaine, 2013). The conceptually closest available measure for the Fortunato et al. (2013) study was SDQ conduct problems. Accordingly, SDQ conduct problem scores were converted to T scores to place them on the same metric as CBCL aggression scores (see Youth in Mind, 2013).

In order to test relations between RSA and externalizing scores, correlations were first calculated between RSA and both CBCL and SDQ scales. Analysis of variance (ANOVA) was used to evaluate potential differences between methods for both baseline and reactivity scores. Next, we compared those with clinical levels of externalizing behavior (T ≥ 70; 98th percentile) with the remainder of the sample on both baseline RSA and RSA reactivity in one-way ANOVAs. This ensured that any effects that were restricted to participants with clinical levels of psychopathology would be captured rather than swamped by correlations computed across the entire sample (see Footnote 3).

**Results**

Before computing inferential statistics, distributions of age and externalizing were assessed for normality. Skew and kurtosis values, respectively, were 0.49 and 0.01 for age, and 0.57 and –0.45 for externalizing. All of these values fall within acceptable ranges. Assessing normality was important given that each data set included separate clinical/high risk and control groups, and therefore did not represent the entire range of externalizing behavior in the population. Sex is not included in any of the analyses presented below, since it did not alter directions of association, change significance levels of correlations, or affect results in the group-based analyses.

**Comparisons of analysis method across data sets**

**Correlational analyses.** The total sample included 559 children and adolescents (171 female, 388 male) between ages 4 and 17 years (M = 8.19, SD = 2.94). There were significant positive correlations of large effect size among RSA analysis methods (FFT, FFT AC, and AR) for both baseline scores (all rs ≥ .72; see Table 2), and reactivity scores (all rs ≥ .58; see Table 3). At baseline, there were also significant positive correlations between age and RSA for both FFT AC and AR (both rs ≥ .13; see Table 2). For reactivity scores, age and RSA were correlated positively across all scoring methods (all rs ≥ .20; see Table 3). Thus, children became less reactive with age. At baseline, the correlation between age and RSA using FFT AC was significantly larger than the correlations between both age and RSA using AR (z = 2.54, p = .01) and FFT (z = 4.20, p < .01). Correlations between age and RSA using AR, and between age and RSA using FFT, did not differ (z = 1.66, p = .09). For reactivity, the correlation between age and RSA using AR was larger than the correlations between age and RSA using both FFT AC (z = 1.90, p = .06) and FFT (z = 2.06, p < .05). Correlations between age and RSA using FFT AC and FFT did not differ (z = 0.16, p = .87).

**Table 2. Correlations among baseline RSA values and age**

<table>
<thead>
<tr>
<th>Measure</th>
<th>FFT</th>
<th>FFT AC</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT AC</td>
<td>.723**</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>.780**</td>
<td>.842**</td>
<td>1.000</td>
</tr>
<tr>
<td>Age</td>
<td>.027</td>
<td>.272**</td>
<td>.126**</td>
</tr>
</tbody>
</table>

Note: N = 559. RSA, respiratory sinus arrhythmia; FFT, fast Fourier transform; FFT AC, FFT with age correction; AR, autoregressive. **p ≤ .01.

**Table 3. Correlations among and RSA reactivity values and age**

<table>
<thead>
<tr>
<th>Measure</th>
<th>FFT</th>
<th>FFT AC</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT AC</td>
<td>.575**</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>.595**</td>
<td>.738**</td>
<td>1.000</td>
</tr>
<tr>
<td>Age</td>
<td>.203**</td>
<td>.212**</td>
<td>.318**</td>
</tr>
</tbody>
</table>

Note: N = 559. RSA, respiratory sinus arrhythmia; FFT, fast Fourier transform; FFT AC, FFT with age correction; AR, autoregressive. **p ≤ .01.
Resting RSA comparisons. Next, original FFT scores, which did not use age-corrected respiratory frequencies, were compared with FFTAC and AR scores at rest. There were significant main effects for all three contrasts, including the original analysis versus FFTAC comparison, \( F(1, 554) = 306.11, p < .001, \eta^2_p = 0.36 \), the original analysis versus AR comparison, \( F(1, 554) = 69.92, p < .001, \eta^2_p = 0.11 \), and the FFTAC versus AR comparison, \( F(1, 554) = 161.20, p < .001, \eta^2_p = 0.23 \). As shown in Figure 3, these main effects indicate that baseline RSA scores were highest for the original FFT analyses and lowest for FFTAC.

These main effects must be interpreted in the context of significant Analysis Method \( \times \) Data Set interactions for both the original analysis versus FFTAC comparison, \( F(4, 554) = 14.35, p < .001, \eta^2_p = 0.09 \), and the FFTAC versus AR comparison, \( F(4, 554) = 12.14, p < .01, \eta^2_p = 0.08 \). In addition, the Analysis Method \( \times \) Data Set interaction for the original analysis versus AR comparison approached significance, \( F(4, 554) = 2.15, p = .07, \eta^2_p = 0.02 \). These interactions indicate that (a) original FFT overestimated RSA compared with both FFTAC and AR among younger participants, and (b) differences between FFTAC and AR were larger among younger participants (see Figure 3).

RSA reactivity comparisons. For the reactivity analyses, there were significant main effects for the original FFT versus AR comparison, \( F(1, 554) = 66.39, p < .001, \eta^2_p = 0.11 \), and the FFTAC versus AR comparison, \( F(1, 554) = 65.82, p < .01, \eta^2_p = 0.12 \), but not for the original FFT versus FFTAC comparison, \( F(1, 554) = 0.94, p = .33, \eta^2_p < 0.01 \). As shown in Figure 4, estimates of RSA withdrawal were smaller for both original FFT and FFTAC than for AR.

Significant Analyses Method \( \times \) Data Set interactions were also observed for all three contrasts, including the original analysis versus FFTAC comparison, \( F(4, 554) = 8.78, p < .001, \eta^2_p = 0.06 \), the original analysis versus AR comparison, \( F(4, 554) = 17.47, p < .001, \eta^2_p = 0.11 \), and the FFTAC versus AR comparison, \( F(4, 554) = 5.50, p < .01, \eta^2_p = 0.04 \). These interactions indicate that (a) original FFT underestimated RSA reactivity among younger participants, and (b) differences between FFTAC and AR were larger among younger participants (see Figure 4).

Relations among RSA, RSA reactivity, and externalizing behavior

RSA externalizing correlations. As shown in Table 4, in the combined data set there were significant negative correlations at baseline between RSA and externalizing scores for both the original analysis method (\( r = -.13 \)) and AR (\( r = -.12 \)). Thus, lower baseline RSA was associated with higher externalizing scores using these methods. These associations were carried primarily by the oldest sample, for whom correlations between RSA and externalizing scores ranged from \( r = -.17 \) (FFT) to \( r = -.23 \) (FFTAC). In the other data sets, no correlations between baseline RSA and externalizing scores were significant at any age (see Table 4). No significant correlations were observed between RSA reactivity and externalizing scores at any age, regardless of measure.

Group differences. As noted in Footnote 3, low RSA and/or excessive RSA reactivity among participants with clinical levels of psychopathology could be swamped in correlational analyses across the larger sample, thereby going undetected.

Figure 3. Relations between data set and spectral power, by analysis method. Data sets are plotted along the x-axis by increasing age. Significant main effects and interactions were observed for all comparisons. Thus, original analyses, which were conducted without age-corrected respiratory frequencies, overestimated baseline respiratory sinus arrhythmia compared with both age-corrected fast Fourier transform and autoregressive techniques, especially among younger samples. See text for additional details.
Thus, we also evaluated group differences between those whose externalizing $T$ scores in the clinical range ($T \geq 70$; 98th percentile) based on national norms ($n = 178$) versus the rest of the sample ($n = 381$). For baseline RSA, significant group differences were observed for both FFT, $F(1, 557) = 6.53, p = .01$, and AR, $F(1, 557) = 7.08, p = .008$. For FFT$_{AC}$, the group effect was not significant, $F(1, 557) = 2.20, p = .14$. No group differences were found for RSA reactivity, all $Fs(1, 557)$ $\leq 0.91$, all $ps \geq .34$. Nearly identical findings were observed when the clinical cutoff was placed at the 95th as opposed to the 98th percentile.

**Discussion**

Our objectives in conducting this study were twofold. First, we sought to characterize effects of misspecifying respiratory frequencies on estimates of RSA across development. Second, we sought to evaluate whether RSA–externalizing relations are observed when FFT$_{AC}$ and/or AR spectral analysis are used, in order to determine whether existing evidence for such relations is credible, or a possible artifact of inappropriate scoring methods. Below we discuss these objectives, and consider how our findings apply to the existing
literature on the development of RSA, and its links to externalizing behavior and emotion dysregulation. We conclude with recommendations for future research.

Effects of misspecifying respiratory frequencies on estimates of RSA across development

Findings from this study indicate clearly that FFTAC or AR spectral analysis should be used when computing RSA among developmental samples. Even though original, uncorrected FFT baseline RSA values correlated highly with both FFTAC and AR estimates, uncorrected FFT values did not correlate with age, and therefore obscured developmental increases in resting RSA. No such effect was observed for RSA reactivity, for which all computational methods correlated positively with age, indicating less PNS withdrawal to eliciting stimuli across development. Smaller reductions in RSA to emotion evocation across development are consistent with the notion that emotion regulation improves as children mature.

There were also significant interactions between analysis method and data set for both baseline and reactivity measures of RSA. Without age-corrected respiratory frequencies, FFT overestimated baseline RSA, especially among the younger samples. In contrast, FFT without age-corrected respiratory frequencies underestimated RSA reactivity (withdrawal) to emotion evocation among younger samples. Finally, FFT without age-corrected respiratory frequencies underestimated developmental changes in both RSA and RSA reactivity from early childhood to adolescence. These findings suggest that literature-wide overestimates of resting RSA, underestimates of RSA reactivity, and underestimates of developmental shifts in RSA and RSA reactivity may exist. Accordingly, either age-corrected respiratory frequencies should be used when estimating RSA from FFT analysis or AR spectral analysis should be used. As reviewed above, FFTAC involves selecting appropriate respiratory frequencies for each participant based on tabulated values from large samples (Wallis et al., 2005), whereas AR spectral analysis identifies observed respiratory frequencies for each participant empirically (see Kay & Marple, 1981).

There were also significant differences between FFTAC and AR. Specifically, FFTAC produced lower estimates of resting RSA, and higher estimates of RSA reactivity, especially among younger participants. Since AR identifies respiratory frequencies for each individual empirically rather than extracting an age-appropriate respiration rate from tabulated values, it may be more accurate than FFTAC. If this is the case, AR spectral analysis might be considered the best option in future research.

A potential concern is that the respiratory frequencies we used for FFT and FFTAC truncated spectral power distributions for some individuals, resulting in underestimates of baseline RSA and RSA reactivity, especially among younger participants. Even though we used respiration frequencies that capture 97.5% of the population (see Table 1), these values define respiratory peaks. Spectral power is disbursed around those peaks. To ensure that we captured this dispersion, we inspected all spectral density plots visually. These inspections revealed only two cases for which the high-frequency cutoff had to be shifted to capture the entire high-frequency spectral power distribution. Thus, truncating spectral power distributions was not a significant issue in this sample. The homogeneity of variance assumption was not violated in any of the ANOVAs, all Fs (4, 554) ≤ 1.52, all ps ≥ .19. However, visual inspection is quite time-consuming, which provides another argument in favor of AR.

Links between RSA and externalizing behavior

In the combined sample, there were significant negative correlations, albeit of small effect size, between baseline RSA and externalizing scores for both original FFT and AR spectral analysis, but not for FFTAC. Consistent with existing literature (e.g., Beauchaine et al., 2001; Mezzacappa et al., 1997; Pang & Beauchaine, 2013), lower resting RSA was associated with greater externalizing behavior. These findings were carried by RSA–externalizing correlations observed in the oldest sample, which were generally larger than those observed in the younger samples, particularly for FFTAC and AR.

As noted elsewhere, sympathetic nervous system deficiencies, which reliably characterize externalizing samples, are evident as early as preschool (see Beauchaine, 2015a; Crowell et al., 2006). In contrast, PNS deficiencies tend to emerge later in development (Beauchaine & Gatzke-Kopp, 2012; Beauchaine et al., 2007; Fortunato et al., 2013). Similar findings have since been reported for depressed youth, whose deficiencies in RSA emerge in adolescence, not childhood (Koenig et al., 2016). According to operant reinforcement accounts of emotion dysregulation, altered PNS activity and reactivity develop over time in families with both externalizing and internalizing children through social reinforcement mechanisms. In many of these families, escalation of dyadic conflict/invalidation and associated physiological arousal are negatively reinforced because conflict and invalidation terminate aversive interactions (see Beauchaine & Zalewski, 2016; Crowell et al., 2013, 2014). Through these negative reinforcement mechanisms, emotion dysregulation and associated deficiencies in RSA emerge over time among affected children (Beauchaine, 2015a). Our findings indicate that age-related changes in associations between externalizing behavior and deficiencies in resting RSA are not an artifact of using age-inappropriate respiratory frequencies when calculating RSA.

Despite very small effects in the correlational analyses, baseline RSA among those with externalizing T scores in the clinical range was lower than in the rest of the sample. This finding applied to both FFT and AR, but not to FFTAC. Although we do not know why FFTAC was less sensitive to group differences than the other methods, these findings are potentially important, as they suggest a nonlinearity
in RSA–externalizing relations. As outlined in Footnote 3, normative variation on externalizing measures may be best construed as temperamentally driven rather than a marker of psychopathology. Thus, RSA may mark emotion deregulatory processes primarily among those with significant psychopathology. Similar nonlinearities apply to certain neural substrates of ADHD (Plichta & Scheres, 2014).

There were no significant associations between RSA reactivity and externalizing scores, which is inconsistent with several studies of RSA reactivity among clinical samples (see above). At first glance, these findings also appear to be inconsistent with original results from three of the five constituent studies (Beauchaine et al., 2001, 2013; Pang & Beauchaine, 2013). However, the original analyses evaluated more inclusive measures of externalizing behavior, not solely aggression. In contrast, we were forced to use a measure that was available for four of the five studies.

Implications for emotion dysregulation research

Effects of misspecifying respiratory frequencies in estimating RSA may have important implications for the emotion dysregulation literature. Perhaps most important, developmental changes in RSA and RSA reactivity have likely been underestimated in previous research. Developmental increases in baseline RSA and developmental decreases in RSA reactivity to emotion evocation both suggest improved emotion regulation as children mature. According to our analyses, these developmental changes may be steeper than previously thought (cf. Zisner & Beauchaine, 2016). Without proper age corrections for respiratory frequencies, baseline RSA is overestimated, and RSA reactivity is underestimated in younger samples. Our findings indicate that variance in RSA reactivity may have been truncated in many young samples. Since RSA reactivity to emotion evocation may mark emotion regulatory capacity (see Beauchaine 2015a, 2015b; Beauchaine & Thayer, 2015), these findings suggest that effect sizes in studies of RSA reactivity–emotion regulation associations may have been underestimated, and should therefore be revisited.

Limitations and future directions

As alluded to above, seeming discrepancies between these findings and those from previous analyses could be related to differences in externalizing measures. Unfortunately, we could not use DSM diagnoses, continuous DSM symptom counts, or broadband CBCL externalizing scores, since they were not consistently available across studies. Thus, CBCL aggression scores were used. Broadband CBCL externalizing scores include both aggression and delinquency items, and therefore reflect a broader array of problem behaviors, as do DSM conduct disorder scales. Thus, smaller effect sizes than those reported in the original analyses, those reported in the literature more broadly, may reflect use of a more circumscribed measure of externalizing conduct.

In addition to limitations already discussed, we were missing data for 70 participants, largely as a result of file corruption that occurred when transferring data to a new server. This reduced sample size and affected statistic power, and rendered data slightly different from those in the original analyses. Furthermore, different baseline and task measures were used across samples. Fortunato et al. (2013) used a “vanilla” baseline (watching a neutral video), whereas all other studies used true baselines. As noted above, vanilla baselines can result in underestimates of RSA reactivity. In addition, Crowell et al. (2006) and Beauchaine et al. (2013) used frustration tasks, whereas Fortunato et al. (2013) and Pang and Beauchaine (2013) used sad emotion inductions, and Beauchaine et al. (2001) used an escalating peer conflict. Frustration tasks are more demanding metabolically, which can affect RSA reactivity. However, this cannot explain null outcomes. As we have described elsewhere (e.g., Beauchaine, 2012; Zisner & Beauchaine, 2016), and as alluded to in Footnote 3, stimulus conditions should be considered carefully when interpreting psychophysiology–behavior relations, and different tasks cannot be considered equivalent. In developmental research, this presents a difficult problem, since the same stimuli cannot be used to evoke similar psychological states and reactions at all ages. Thus, even though all tasks used in the constituent studies evoked emotion, differences in tasks are unavoidably confounded with the ages of participants, and may have contributed to apparent developmental shifts in RSA reactivity from preschool to adolescence.

Finally, heterogeneity of clinical attributes across samples likely limits generalizability of our findings, particularly with respect to externalizing outcomes. Nevertheless, the finding that misspecifying respiratory frequencies among younger participants has especially large effects on accuracy of both RSA assessment and estimates of developmental shifts in RSA and RSA reactivity is likely valid.

Caveats aside, age-appropriate respiratory frequencies should be used when estimating RSA, whether applied manually via FFTAC or automatically using AR spectral analysis. Doing so will eliminate overestimates of resting RSA and underestimates of RSA reactivity among young children. Our findings apply only to situations in which adult or adolescent respiratory frequencies are used with children. Our literature search also yielded studies in which toddler (and in some cases infant) respiratory frequencies were used for child and adolescent samples. Although one might surmise that such studies would yield an opposite pattern of results, this is an empirical question that should be examined in future research. Misspecifying respiratory frequencies across development may also affect estimates of inter- and intraindividual stability in longitudinal studies. Reliabilities in developmental samples may be underestimated if incorrect respiratory frequencies are used. It is our hope that this article stimulates additional research on these and other questions.
References


keeping the panes clean, and viewing the neural topography. Biological Psychology, 34, 131–161.


Quantifying RSA across development


