

Convergence of behavioral and cardiac indicators of distress in toddlerhood: A systematic review and narrative synthesis

Jordana A. Waxman,¹ Miranda G. DiLorenzo,¹ and
Rebecca R. Pillai Riddell^{1,2,3} 

Abstract

The objective of the current study was to systematically review the available literature on the relation between behavioral and cardiac indicators used to measure distress in toddlerhood. After ascertaining the eligibility of 2,424 articles through a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and PROSPERO (International Prospective Register of Systematic Reviews) guided search process, 22 articles ($N = 2,504$) that investigated associations between behavioral and cardiac indicators of distress in toddlerhood were identified. The narrative synthesis described the overall relation (direction [positive, negative], strength [Cohen's D]) between behavioral and cardiac indicators and was organized by cardiac indicator (i.e., heart rate [HR], heart period, respiratory sinus arrhythmia, pre-ejection period) and type of behavior measured (i.e., coding for expressed emotion behaviors vs. emotion regulation behaviors). Methodological characteristics (i.e., timing of measurement for behavioral and cardiac indicators [concurrent, predictive], length of measurement epochs, inclusion of covariates and moderators) were also described. HR was consistently positively ($D = .05$ to $.54$) related to expressed emotion behaviors. No other cardiac and behavioral indicators were consistently related. Methodological differences related to behavioral and cardiac indicators utilized, timing of measurement, and length of measurement epochs may be responsible for heterogeneity in findings. The findings suggest that researchers might get divergent results depending on whether distress is measured with cardiac or behavioral indicators of distress in toddlerhood. Suggestions for future psychophysiological research with young children are offered.

Keywords

Behavior, heart rate, heart rate variability, psychophysiology, systematic review

Introduction

Measurement of responses during distress involves examining one's behavioral and biological initial reaction to changes in the environment (i.e., reactivity) or the behaviors and biologic processes (e.g., attention, behavioral approach/avoidance, self-soothing) used to recover from emotional challenges and return to homeostasis (i.e., regulation) (Kopp, 1982; Rothbart & Derrberry, 1981). While distress reactivity tends to be highly dependent on antecedent circumstances (e.g., genetics, sensory thresholds, caregiver factors), the process of distress regulation is more complex and dynamic, as it involves the conscious and unconscious sequencing of emotion, behavioral regulatory strategies, and biology (Calkins & Leerkes, 2004; Ekas et al., 2018). These are considered distinct yet interrelated facets of distress (Barrett, 2013; Cole et al., 2004; Thompson et al., 2008).

It is believed that reactivity and regulatory responses are especially malleable period early in life (Campos et al., 1989; Cole et al., 2004), with behavioral, physiological, attentional, emotional, cognitive, and interpersonal or social processes having unique yet interrelated developmental trajectories that emerge across infancy, toddlerhood, and early childhood (Calkins & Fox, 2002). Calkins (2010) suggested that research investigating distress reactivity and regulation utilize a biopsychosocial theoretical framework, which considers how the child's biology, behavior, and social context

fundamentally change one another continuously over the course of time. Indeed, systematic research using a behavioral indicator of distress suggest that there is significant variability in how infants regulate from distress across the first year of life (Pillai Riddell et al., 2013). As such, it is important for research on the development of distress regulation to not collapse across developmental stages (i.e., infancy, toddlerhood, early childhood).

The focus of the current systematic review will be on distress reactivity and regulation in toddlerhood (i.e., second and third years of life). Toddlerhood is a critical developmental phase to examine distress reactivity and regulation, as it coincides with developments in motor, language, and representation skills, and encompasses the transition from passive caregiver-directed regulation to more active and purposeful self-regulation (Kopp, 1982; Rothbart et al., 1992).

¹ York University, Canada

² The Hospital for Sick Children, Canada

³ University of Toronto, Canada

Corresponding author:

Rebecca R. Pillai Riddell, Department of Psychology, York University, 4700 Keele Street, OUCH Laboratory, 2004/6 Sherman Health Sciences Building, Toronto, ON M3J 1P3, Canada.

Email: rpr@yorku.ca

Infants initially depend on their primary caregiver to understand and scaffold adaptive regulation from distressing events (Ekas, 2018; Pillai Riddell & Racine, 2009). However, by the end of the first year of postnatal life, the child is able to deploy strategies aimed at modifying several aspects of the emotion, including the intensity and duration of the emotions, and also impact caregivers' behaviors when they need help regulating their emotions (Ekas et al., 2018). The purpose of the current review is to provide a more nuanced understanding of the patterning of cardiac and behavioral responses in toddlerhood during negatively valenced events. Foundational to a review of this kind requires an understanding of the development of behavioral and biological reactivity and regulatory responses to distress and its associations with developmental outcomes.

Development of Behavioral and Biological Regulatory Responses

Behavioral regulation skills are developing from birth, with a shift from utilizing primitive mechanisms of self-soothing and signaling (e.g., sucking, moving away, crying) to more voluntary, simple actions to modify arousal levels (e.g., self-initiated distraction) (Kopp, 1982; Rothbart et al., 1992). Maturation of different biological support systems is implicated in the increasingly sophisticated emotions and behavioral regulation strategies observed across infancy and early childhood, as physiological changes are hypothesized to prepare the individual for actions associated with emotional events (Calkins & Hill, 2007). The majority of empirical work that focuses on underlying physiological components highlights the maturation of the autonomic nervous system as playing a fundamental role in emotion reactivity and regulation (Santucci et al., 2008). Both the parasympathetic nervous system (PNS) and sympathetic nervous system (SNS) have been implicated in the development of distress responses (both reactivity and regulatory). Indeed, the Polyvagal theory (Porges, 1995) specified these two branches as serving different evolutionary stress responses in mammals. The vegetative vagus (i.e., SNS activity) represents the more primitive branch, which mediates the coping strategies of reptiles. Higher levels of SNS activity tend to be associated with physiological arousal, metabolic output, and fight or flight responding during distress (Porges, 2007). In contrast, the smart vagus (i.e., PNS activity) is more evolved and mediates cardiac activity when demands require more sophisticated coping. Specifically, decreasing levels of PNS activity tend to be associated with orienting, self-soothing behaviors, responding to challenges, and increased metabolic output (Porges, 2007).

Associations Between Behavioral and Cardiac Reactivity/Regulation and Developmental Outcomes

The limited research available suggests that there is moderate stability of behavioral and cardiac reactivity and regulatory responses across the life span (Gunnar & Vasquez, 2006; Matthews & Phillips, 2010; Seifer et al., 1996; Stifter & Fox, 1990) and that response patterns developed in toddlerhood can be reliably linked to future developmental outcomes (Beauchaine, 2001; Gendron & Feldman, 2009; Qu & Leerkes, 2018). Indeed, certain reactivity and regulation patterns after a distressing event (e.g., partial vagal withdrawal leading to SNS activation, more effective behavioral strategies employed) early in life have been linked to traits that promote

optimal learning and development, such as high inhibitory control, executive function, and delay of gratification (e.g., Obradović & Finch, 2017; Santucci et al., 2008). Alternatively, certain reactivity and regulation patterns after a distressing event (e.g., excessive vagal withdrawal, ineffective regulatory strategies employed) are associated with negative emotional traits (e.g., depression, anxiety, and aggression) and states (e.g., Beauchaine, 2001; Enebrink et al., 2005; Frick et al., 2003; Shirtcliff et al., 2009). As such, an understanding of how these distress-related behavioral and cardiac responses are associated has important implications for child development.

Challenges With Assessment of Distress Regulation in Toddlerhood

Developmental scientists have utilized a variety of paradigms to observe the intensity and temporal features (e.g., speed of onset or recovery from distress) of negative emotions. Although there is agreement on the multifaceted nature of emotions, considerable controversy exists regarding the degree to which different aspects of emotion (i.e., emotion expression, emotion regulatory behaviors, biology) coordinate across a distressing event (Cole et al., 2004). This debate is complicated further in infancy and toddlerhood, as unlike behavioral responses to distress, individuals mount a variety of physiological responses that are not directly observable. Indeed, responses during distress responding is a dynamic process that involves the coordination of cognitive, cortical, cardiac, and endocrine systems, in addition to emotion and behavioral strategies (Ekas et al., 2018). Understanding how these different systems coordinate is a critical area of study. However, to better understand how these systems coordinate, challenges with behavioral observation systems and cardiac indicators and methodologies need to be addressed.

Phases of distress responding. As mentioned previously, seminal research by Rothbart and Derryberry (1981) provided an important distinction between distress reactivity and regulation in relation to infant temperament, which could fruitfully be applied to behavioral and cardiac indicators of distress. While reactivity tends to be highly dependent on innate, biological factors, the development of distress regulation seems to emerge during the second year of life and coincides with the development of certain cognitive capacities (Rothbart & Derryberry, 1981). Thus, behavioral and physiological responses underlying distress reactivity and regulation may not be identically associated across time. Despite the importance of considering phases of distress responding when attempting to capture toddlers' distress, it is unclear how the available literature has examined phases of distress responding. As such, the length (in seconds) and timing (concurrent [behavior and cardiac data acquired simultaneously within the same distress paradigm] or predictive [behavior and cardiac data acquired at different times within the same appointment]) of measurement epochs¹ will be important considerations in the review.

Behavioral distress responding. Various methodologies have been used to measure behavioral responses to distress across development. Researchers tend to observe the latency, intensity, frequency, and duration of emotional behaviors, the attentional and behavioral strategies used to manage such reactions, or a combination of these emotional reactions and regulatory strategies. Complicating the

picture further, researchers have measured behavioral outcomes differently, using either continuous composite scores or extremes of a dimension. Cole and colleagues (2004) have described this debate regarding how to best distinguish emotion expression behaviors from emotion regulation behaviors themselves and suggest that independent measurement of the activated emotion (e.g., observed facial expressions and behaviors) and the resultant regulatory strategy (e.g., soothing, visual engagement) is necessary. It is posited that global measures that encapsulate both expressed emotions and emotion regulatory strategies fail to capture the dynamic and temporal nature of expressed emotion and emotion regulation. To continue pushing developmental literature forward, it is important for behavioral measures to be separated based on whether they measured expressed emotion behaviors (i.e., activated emotion) or emotion regulation behaviors (i.e. behaviors that serve to bring the organism back to homeostasis). This distinction will be addressed in the following review.

Cardiac distress responding. The development of behavioral distress responses has been reliably linked to continuous changes across biological systems, with the majority of empirical work highlighting the integral role of the autonomic nervous system in the regulation of emotions (Santucci et al., 2008). Specifically, brain regions associated with emotions and cognitions can influence brainstem cardiorespiratory control centers and thus affect cardiac indicators (Berntson et al., 1994). Perry and Calkins (2018) posited that physiological indicators help identify underlying mechanisms that are imperative in understanding expressed emotion responses and emotion regulation responses that cannot be communicated through behavioral observation alone. As mentioned above, both branches (i.e., PNS/SNS) have been implicated in the development of distress responses; however, the convergence (i.e., strength) and direction of the relationship between behavioral and cardiac indicators during distress has not been reliably established.

Common cardiac indicators used in the literature include heart rate (HR), heart period (HP), and heart rate variability (HRV). HR measures the number of contractions of the heart per minute and reflects both sympathetic and parasympathetic activity. Specifically, increases in sympathetic activity lead to increased HR and the focusing of attention, which is essential for generating emotional responses that are effective in modulating arousal (Bornstein & Seuss, 2000). HP refers to the interval between heartbeats (in milliseconds) and is inversely related to HR (i.e., low HP is equivalent to high HR) (Burgess et al., 2003). Regarding HRV, one of the most common indexes is respiratory sinus arrhythmia (RSA). RSA is thought to reflect the variability in the time between heartbeats that occurs at the frequency of respiration, with higher values of RSA reflecting greater levels of parasympathetic influence (Berntson et al., 2007). During times of distress, vagal influence is withdrawn which results in increased sympathetic activation (i.e., fight or flight response) (Gatzke-Kopp & Ram, 2018). Seminal research has found that decreasing levels of PNS activity during distress, as indexed by decreasing RSA, tend to be associated with orienting, responding to challenge, and increased metabolic output (Porges, 2007). Two other related HRV indicators that are used in conjunction with behavior are RSA change score (decreases in RSA from baseline to distress) and baseline RSA. RSA change score is posited to be related to physiological regulation of emotions or coping efforts, and reflective of vagal withdrawal. On the other hand, baseline or tonic RSA is thought to reflect an individual's

temperamental reactivity and emotionality (Beauchaine, 2001; Porges, 1995). Pre-ejection period (PEP) is validated as an index of SNS functioning (Sherwood et al., 1990) and reflects the time interval (in milliseconds) between the onset of ventricular depolarization and the onset of left ventricular ejection (i.e., ejection of blood into the aorta) (Berntson et al., 2007). Shorter PEP suggests greater sympathetic influence, which is functionally related to an individual's mobilization of resources to meet environmental demands associated with stress (Beauchaine, 2001). Sherwood and colleagues' (1990) guideline for impedance cardiography spurred much research utilizing PEP in children, adolescents, and adults (e.g., Beauchaine, 2001; Beauchaine et al., 2007; Brenner et al., 2005; Crowell et al., 2006). However, scientists have less frequently investigated the increase in sympathetic influence during distress in toddlerhood (e.g., Buss et al., 2005; Kahle et al., 2018). Thus, it is unclear in early childhood how sympathetic (versus parasympathetic) measures compare to traditional quantifications of reactivity and regulation responses.

Current Review

Developmental research on toddler distress-related reactivity and regulatory responses has largely focused on direct behavioral observations and parent reports of distress rather than multidimensional approaches incorporating both behavioral and physiological indicators. A review by Barrett (2015) highlighted that there may be inter- and intra-individual variability in behavioral and physiological responses to emotions within and across time. Thus, there may not be a single, consistent pattern of relations in the PNS, facial movements, or other behaviors for each emotion category. Without proper systematic reviews that examine how multiple dimensions of distress-related responses react together (such as behavior and cardiac) that takes into account the types of behavior (expressed emotion versus emotion regulation behaviors) and cardiac responses (e.g., HR, HP, RSA, PEP) across developmental stages, it is difficult to assess the current state of the field.

The main goal of the present study was to systematically review the findings on the concurrent (i.e., within-session) relation between behavioral and cardiac indicators measured during distress in toddlerhood. Based on the methodological and conceptual issues outlined above, our original aim was to conduct a meta-analysis, to quantitatively estimate the overall relation between behavioral and cardiac indexes of distress in toddlerhood and investigate how the effect varied according to certain study characteristics (e.g., timing of measurement for behavioral and cardiac indicators, length of measurement epochs, inclusion of covariates and moderators). However, based on the guidelines outlined by Higgins and Green (2011), it was concluded that a meta-analysis would not be appropriate. Specifically, many studies utilized multiple, divergent cardiac indicators (i.e., baseline, responding and/or change scores for HR, HP, RSA, and PEP) and behavioral coding schemes. Given that the studies are clinically diverse, a meta-analysis may have obscured genuine differences in effects (Higgins & Green, 2011). Therefore, we undertook a narrative synthesis, which is an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarize and explain the findings of the synthesis (Popay et al., 2006). A narrative synthesis was deemed a more appropriate analytic approach, as it involves the interpretation of statistical data, with the ability to couch the synthesis based on differences in

methodology, methodological rigor, and overall study quality (Popay et al., 2006).

The aim of our narrative synthesis was to examine the direction and magnitude of the relation between behavioral and cardiac indicators measured during distress in toddlerhood, taking into account the methodological limitations based on study characteristics that may influence the relation between indicators. In accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009), the quality of the research articles included in the synthesis was assessed. As well, the methodological rigor of physiological data collection was examined, as this has been shown to be important factor in recent meta-analyses and empirical work involving cardiac indicators (e.g., Beauchaine et al., 2019; Shader et al., 2018). Previous meta-analyses in infants, children, and adults revealed a significant association ($r = 0.09\text{--}0.15$) between cardiac indicators and self-regulation (Holzman & Bridgett, 2017; Zahn et al., 2016) that was impacted by publication bias. As such, we hypothesized a similar degree of association between behavioral and cardiac indexes measured during distress in toddlerhood. However, we expected the direction of association to vary according to cardiac indicator (i.e., positive associations with SNS indicators [i.e., HR, HP, PEP], negative associations with PNS indicators [i.e., RSA]).

There are a number of variables that are hypothesized to impact the relation between behavioral and cardiac indicators of distress in toddlerhood. A plethora of research exists that highlights the impact of sex, age, and attachment on the development of emotion regulation (see Perry & Calkins, 2018, for a review of the development of emotion regulation processes in childhood). Studies examining sex differences in emotional responding have reported mixed results. These differences in results are posited to be due to differences in methodology (e.g., level of emotional responding under study [e.g., behavior, psychophysiology], stimuli utilized [positive vs. negative emotion]) (Domes et al., 2009). With regards to age, across early development there is a dramatic growth in the acquisition and display of emotion regulation skills and abilities (Calkins, 2010). These changes in emotion regulation skills across time have been linked to differences in caregiver–infant attachment, as caregivers assist in the regulation of arousal in infancy and early childhood (Perry & Calkins, 2018). Infants with secure attachment relationships with their primary caregivers will seek and receive caregiver support during times of distress. However, infants with insecure attachment will develop ineffective strategies to heighten their call for support or suppress their emotions (Zimmer-Gembeck et al., 2017). Additionally, baseline responses (i.e., cardiac and behavior prior to task) are known to impact the size of a psychophysiological response to distress (Berntson et al., 1994; Oberlander & Saul, 2002; Waxman et al., 2016). No a priori hypotheses were made regarding the impact of behavioral coding scheme, study characteristics, or methodological considerations on the association between indexes of distress in toddlerhood, given the dearth of research in this area.

Method

Search Strategy

A systematic search was conducted using Medline, Embase, PsycINFO, and CINAHL in July 2017 for English-language references. An updated search was conducted in Embase in March 2019 to include any relevant articles published between July 2017 and March

2019. The original search was not limited by year published to encompass both historical and contemporary articles and reviews. Search terms related to distress-inducing contexts, behavioral and cardiac measures, and infancy or toddlerhood were systematically paired (see Online Supplementary Material 1). Hand-searching reference lists of relevant studies and systematic reviews on the convergence between behavioral and cardiac measures of distress were also conducted. Our review followed an a priori protocol according to the PRISMA guidelines (Moher et al., 2009). The review protocol was registered on the PROSPERO website (Booth, 2013) before data extraction (registration no. CRD42017077288).

Inclusion/Exclusion Criteria and Study Selection

English-language, prospective-observational or descriptive studies on humans who were 12 months to less than 4 years of age, undergoing an acutely distressing task, and monitored using video recording and electrocardiography were included. Our definition of observational studies included cohort studies in which participants were prospectively identified and followed up during distress tasks using behavioral and cardiac indicators as well as cross-sectional studies that observed participants during a distress task using both behavioral and cardiac measures.

Studies were excluded if they described nonhuman animal models of distress, did not include a distressing event (e.g., pain, fear, frustration), did not include both a cardiac and behavioral measure of distress in toddlerhood (i.e., 12 months to less than 4 years of age), did not measure behavioral and cardiac outcomes during the same time period (i.e., within the same age and appointment), or did not report quantitative or qualitative information regarding the relation between the behavioral and cardiac measures. Review articles, case studies, and conference abstracts were also excluded.

Two authors designed the abstract selection criteria with an initial selection of 500 abstracts. Covidence software (www.covidence.org) was used to independently read and select among all the retrieved references and abstracts. All abstracts were double-coded for reliability purposes. Percentage agreement between the reviewers before consensus coding was 97.9%. Disagreements between the two reviewers were flagged and resolved through discussion. Full texts of potentially eligible studies were retrieved (see Figure 1 for PRISMA flowchart).

Data Extraction

A database was created recording a description of the distress paradigm (e.g., fear, frustration), behavioral coding system, cardiac indicator (i.e., HR, RSA [specifying if computed through time-domain, frequency-domain, and non-linear analysis], PEP), results, and any additional variables that were included when analyzing the relation between behavioral and cardiac measures during distress (i.e., covariates, moderators). It was important to investigate covariates and moderators included in the studies because there are many variables that have been posited to affect the cardiac system or behavioral response to distress (e.g., baseline physiology, sex, age, attachment). Although previous meta-analyses (Holzman & Bridgett, 2017; Zahn et al., 2016) have found significant associations between behavioral self-regulation and HRV measured through time-domain analyses, there was not enough variability in the studies to discuss these differences in methodology. The strength of the relation between behavioral and cardiac indicators

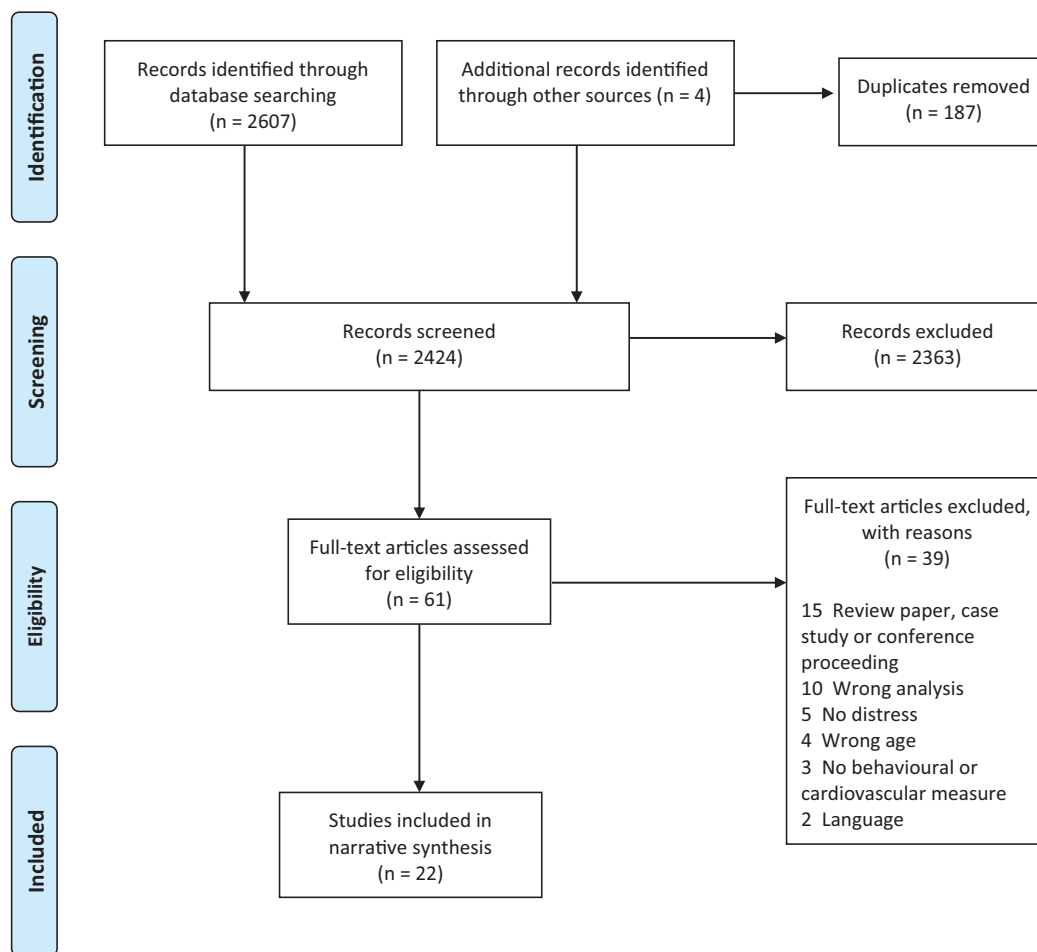


Figure 1. PRISMA Flowchart.

PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

was extracted from the available analysis (e.g., r , t -score, r^2) and converted to a standardized mean-difference effect size (i.e., Cohen's D ; Cohen, 1988). When data were not provided in the articles, the authors were contacted via e-mail on two separate occasions. All studies were extracted by two reviewers. Discrepancies were minimal and resolved through consensus.

Quality Assessment and Methodological Rigor

To evaluate the overall quality of evidence in our systematic review, a modification of the checklists designed by the National Heart, Blood, and Lungs Institute (2014), Downs and Black (1998), and Crombie and McQuay (1998) was used (see Online Supplementary Material 2). The National Heart, Blood, and Lungs Institute has provided a checklist for assessing the quality of observational cohort and cross-sectional studies, and the Downs and Black (1998) and Crombie (1996) measures were chosen based on a multidisciplinary collaborative review discussing quality in case-control, cohort, and cross-sectional studies (Sanderson et al., 2007). Checklist items were scored as *yes* (1), *no* (0), or *unable to determine*. All extractions were consensus-coded for quality scores to ensure reliability. Disagreements were minimal (reliability between authors was 90%) and resolved through discussion. The National Heart, Lung, and Blood institute stated that their checklist

is not intended to create a cutoff that allows one to arrive at a categorical judgment of quality. Instead, it suggests thinking about the items in the checklist and how each communicates something about the potential for bias in a study. These items were chosen through a consensus with the authors of the current article, who agreed that those selected were most reflective of study quality (i.e., sample size justification, predictor and outcome variables being clearly defined, reliable, valid, and implemented consistently, blinding, and key potential confounding variables measured and adjusted for their impact). A proportion score of the criteria met was also calculated. Higher proportions reflect studies that have taken greater precaution against bias and can be considered relatively higher in study quality. The lead author coded every article and one of two other authors double-coded each article with overall reliability being 94% agreement. All discrepancies were discussed and consensus was used for the final value.

In addition to assessing the overall quality of the studies, we approximated the rigor of the physiological data collection. To investigate methodological rigor, we coded whether authors indicated electrocardiography sampling rate, high frequency band used to define RSA (only for studies that used frequency-domain quantification), percentage of missing data, whether 1996 Task Force guidelines (as an index that authors adhered to psychophysiological standards) were referenced, and whether a "true" baseline condition

was utilized (i.e., stimulus free) in cardiac data collection. Checklist items were scored as *yes* (1), *no* (0), or *unable to determine*.

Narrative Synthesis Structure

The narrative synthesis included an examination of the general direction (positive, negative) and magnitude (Cohen's *D*; Cohen, 1988) of the findings. Note that the magnitude of the effect sizes is not categorized (i.e., small, medium, large), given that this categorization is meaningless in the absence of a frame of reference, which is the goal of the current study (see Funder & Ozer, 2019, for an overview of the problematic nature of effect size categorizations). Based on the articles found, studies were first organized according to the cardiac indicator (HR [baseline, task, change score], HP [baseline, standard deviation], RSA [baseline, task, change score], PEP [task]). In Beauchaine's (2001) seminal work describing an integrated model of ANS functioning and psychopathology, the author asserts that a source of confusion in the literature describing behavioral and cardiac indicators of emotion regulation is the definition of different cardiac indicators. As such, for the purposes of the current article, cardiac measures labeled "task" are reflective of when cardiac indicators are collected across the distressing procedure, while "change score" is utilized when cardiac indicators were measured during a baseline period and distress paradigm, with one score being subtracted from the other score (e.g., baseline RSA minus RSA during the distress paradigm). A large proportion of the studies included fell under the "task" category, yet did not account for participants' baseline responses. The *Law of Initial Value* asserts that the size of a psychophysiological response depends on the initial baseline level of the measure (Bernston et al., 1994; Oberlander & Saul, 2002; Waxman et al., 2016). As such, it is difficult to distinguish whether associations between behavioral and cardiac indicators are related to individual physiological differences in baseline or distress reactivity. Given the limited number of studies, we chose to include "task" measures to draw conclusions from the current state of the field. Related to RSA change scores, this term was originally conceptualized in the psychophysiology literature as task (e.g., distress paradigm) minus baseline, with negative scores indicating reduced RSA following distress (Beauchaine et al., 2019). However, a recent meta-analysis by Beauchaine and colleagues (2019) found that authors were computing RSA in differing ways (i.e., baseline minus task, regressing baseline RSA onto task RSA), which impacted the overall relations with other indicators/outcome measures. As such, we have noted how the "change score" was calculated for each study to account for this in our narrative synthesis (see Online Supplementary Material 3).

The type of behavioral coding scheme (expressed emotion behaviors, emotion regulatory behaviors) utilized in each study was the next level of organization. Given the wide variety of behavioral coding schemes used, the behavioral coding systems were categorized based on Cole and colleagues (2004) assertion to separate behaviors reflecting expressed emotions versus behaviors enacted to regulate emotions. Two authors investigated the behavioral coding schemes used in each study and made a forced-choice decision to categorize the coding scheme as representing either expressed emotion behaviors (e.g., facial fear, vocal distress, bodily fear) or emotion regulatory behaviors (e.g., self-comforting, mother-orientation, distraction, aggression/venting, constructive coping). If coding systems were mixed (i.e., included expressed emotion

behaviors and emotion regulatory behaviors), the study was categorized based on what the majority of items represented in the coding scheme. Disagreements between coders were minimal and resolved through consensus coding.

After grouping studies in conceptual categories (i.e., Cardiac Response Type [HR: baseline, responding, change score; HP: baseline, standard deviation; RSA: baseline, responding, change score; PEP: responding] by Behavioral Response Type [emotion expression behavior, emotion regulation behavior]), findings were then further contextualized by methodological differences (*timing of measurement* as either concurrent or predictive, *epoch length* [length of both behavioral and cardiac measurement epochs in seconds], and *covariates/moderators* [were additional measures controlled for or investigated in analysis]). We required at least two studies in a given category (e.g., distress HR by expressed emotion) to perform a narrative synthesis. During our initial extraction, effect sizes were not found to qualitatively differ across distress paradigms (i.e., fear, frustration). As such, results were collapsed across distress paradigms to maximize the number of studies in each category of the narrative synthesis. There were no other types of distress paradigms.

Results

Studies Included

Through the electronic searches, 2,424 unique articles were identified. These articles were then reviewed by title and abstract and were included or excluded based on the a priori selection criteria. A total of 61 articles were then examined in a full-text review and of these, 22 (involving 2504 participants) fulfilled the inclusion criteria. These studies were included in the final review.

Study Characteristics

Supplementary Online Material 3 provides an overview of the studies included, including country of origin, sample size, study design, distress paradigm, behavioral coding scheme, definition of the baseline measurement paradigm, cardiac indicator, methodological details (i.e., timing of measurement, epoch length), and quality score. Fear and frustration were the only types of distress paradigms found in the literature search. Please see Supplementary Online Material 3 for details.

Quality Assessment and Methodological Rigor

Quality assessment. The lowest quality proportion score was .43 (Provost & Gouin-Decarie, 1979), the median quality score for the papers was .55, and the highest score was .81 (Hay et al., 2017). Only one study was given an overall quality judgment of *higher* (see Figure 2 for breakdown of items on quality checklist). We decided that a more extensive description study quality was necessary, as the quality of these studies would contextualize the overall findings from the narrative synthesis. Only six of 22 studies provided a sample size justification, and approximately half of the studies relied on small sample sizes ($N < 100$) for their statistical analyses. Additionally, several studies did not include clearly defined, reliable, valid, and consistently implemented predictor (6/22) or outcome (9/22) variables. Specifically, validation studies for behavioral indicators of distress were not commonly cited and inter-rater reliability estimates were not consistently reported. As

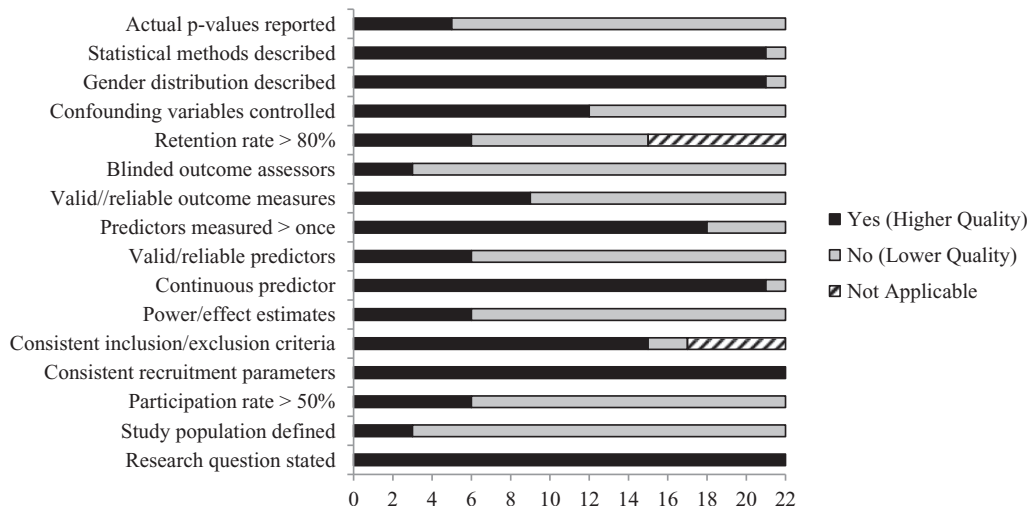


Figure 2. Bar Graph Illustrating the Number of Studies (out of 22) That Fulfilled Each Item on the Quality Assessment.

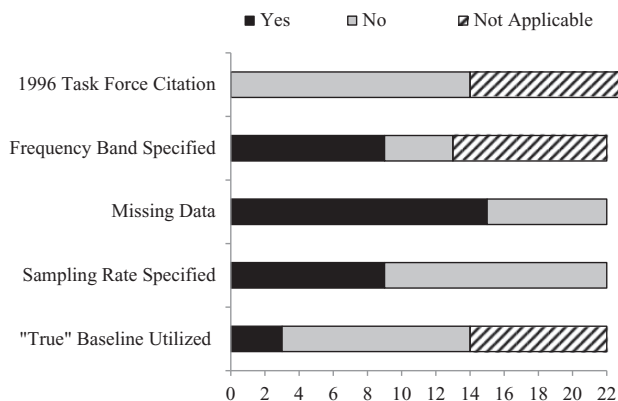


Figure 3. Bar Graph Illustrating the Number of Studies (out of 22) That Fulfilled Each Item on the Methodological Rigor Assessment.

well, for cardiac indicators, the number of coders and reliability estimates was only included for one study. Most studies did not describe blinding their coders to the study hypotheses (19/22). Finally, only 12 studies statistically accounted for at least one key potential confounding variable that is known to impact physiological variables (e.g., age, sex, baseline characteristics, time of day, activity level, respiration).

Methodological rigor. Large inconsistencies in reporting were noted when investigating the methodological rigor for physiological data collection (See Supplementary Online Material 4 and Figure 3 for breakdown of items coded for methodological rigor).

Relation Between Behavioral and Cardiac Measures of Distress

Supplementary Online Material 5 provides a descriptive catalogue of all behavioral distress measures investigated. The coding systems have been organized by whether they were coding emotion expression or emotion regulation behaviors. Tables 1, 3, and 5 summarize effect sizes for each study in the narrative synthesis.

Adjoining tables have been created (i.e., Tables 2, 4, and 6) to provide the patterns of results from each of the cells of the summary tables (Cardiac Indicator by Behavior Type). As noted above, narrative syntheses were only completed if there were at least two studies in a cell (e.g., expressed emotion behavior by task HR cell). However, summaries of findings for single studies are still indicated in Tables 1, 3, and 5.

Heart rate

HR and expressed emotion behaviors. *Baseline HR* (see Tables 1 and 2: Column 2, Row 2): The calculated Cohen's D was .14 for the relation between baseline HR and expressed emotion behaviors (Paret et al., 2015; $D = .14$). One study (Calkins & Johnson, 1998) did not provide data on the relation between indicators. These studies acquired baseline HR data prior to the distressing task (predictive measurement), utilized validated distress paradigms (i.e., Interesting but Scary paradigm, Toy Removal), and had varying epoch lengths for analyzing behavioral (i.e., 120 s) and cardiac (i.e., >120 and 300 s) indicators. Calkins and Johnson (1998) controlled for sex in their analysis and did not find differences in the strength of relation between distress indicators.

Task HR (see Table 1 and 2: Column 3, Row 2): Across the four studies included (Baker et al., 2012; Buss et al., 2005; Hay et al., 2017; Provost & Gouin-Decarie, 1979), there was a positive relation between task HR and expressed emotion behaviors. Effect sizes ranged from a $D = .05$ to $D = .54$. Regarding methodological considerations, all studies acquired behavioral and cardiac data concurrently during the distress paradigm, had varying behavioral and cardiac (i.e., 180 to 600 s) measurement epochs, the majority of studies utilized validated distress paradigms (i.e., Strange Situation procedure, Toy Removal) (Baker et al., 2012; Buss et al., 2005; Provost & Gouin-Decarie, 1979), and included covariates or moderators in the analyses (Baker et al., 2012; Buss et al., 2005; Hay et al., 2017). Sex was a covariate in three studies (Baker et al., 2012; Buss et al., 2005; Hay et al., 2017). Only one study found a significant effect of sex on the relation between behavioral and cardiac indicators of distress (Hay et al., 2017), with male sex increasing the strength of the relationship between distress indicators. Hay and colleagues (2017) also controlled for baseline physiology (i.e., HR

Table 1. Summary of Study Findings (Effect Sizes) for HR and HP With Behaviors (Emotion Expression and Emotion Regulation).

Behavioral coding scheme	Baseline HR	Task HR	HR change score	Baseline HP	Baseline HP SD
Emotion expression behaviors	Calkins & Johnson, 1998 (18 months: null [data not provided]; P, N) Paret et al., 2015 (24 months: null [.14]; P, N)	Baker et al., 2012 (12 months: positive [.54]; 24 months: positive [.46]; 36 months: positive [.46]; C, N) Buss et al., 2005 (24 months: positive [.31 to -.46]; C, N) Hay et al., 2017 (12 months: positive [.05]; C, N) Provost & Gouin-Decarie, 1979 (12 months: positive [.45 to -.54]; C, Y)	Bohlin & Hagekull, 1993 (12 months: positive [.37 to -.64]; C, Y)	Calkins & Johnson, 1998 (18 months: null [data not provided]; P, N) Garcia Coll et al., 1984 (21 months: null [data not provided]; P, Y) Stifter & Jain, 1996 (18 months: positive [.30]; P, N)	Garcia Coll et al., 1984 (21 months: null [data not provided]; P, Y)
Emotion regulation behaviors		Kagan et al., 1989 (14 and 20 months: null [data not provided]; C, N)	Spangler & Grossman, 1993 (12 months: null [data not provided]; C, Y)	Burgess et al., 2003 (24 months: null [.12]; P, N) Calkins & Fox (24 months: null [data not provided]; P, N)	

Note. Cohen's *D* was calculated for the results available, where no data are available, gray shading is used. Results displayed as (Participant Age, Direction of association (positive, negative) [effect size]); Time of measurement: Concurrent (C) = heart rate and behavior measured concurrently; Predictive (P) = heart rate or behavior collected before the other measure; Epochs: Yes (Y) = behavior or cardiovascular measure analyzed in reactivity and/or regulation phases; No (N) = behavior and cardiovascular not analyzed in reactivity and/or regulation phases (collapsed across baseline or reactivity and regulation phases); HR = heart rate; HP = heart period.

Table 2. Summary of Study Findings (Patterns of Results) for HR and HP With Behaviors (Emotion Expression and Emotion Regulation).

Behavioral coding scheme	Baseline HR	Task HR	HR change score	Baseline HP	Baseline HP SD
Emotion expression behaviors	Weak (2 studies; <i>N</i> = 106; 18–24 months)	Positive (6 studies; <i>n</i> = 620; 12, 18, 24–36 months)	Positive (1 study; <i>n</i> = 31; 12 months)	Unclear direction and effect size (3 studies; <i>N</i> = 290; 18–21 months)	Weak (1 study; <i>n</i> = 117; 21 months)
Emotion regulation behaviors		Weak (1 study; <i>n</i> = 100; 14 and 20 months)	Weak (1 study; <i>n</i> = 41; 12 months)	Weak (2 studies; <i>N</i> = 224; 24 months)	

Note. Results displayed as Overall Pattern Found: Positive/Negative, Weak = no studies with significant findings; Unclear direction and effect size = studies had differing directionality and magnitude of relationships (Number of studies, *N/n* = number of participants in studies/study, respectively, Participant Age). Where no data are available, gray shading is used. HR = heart rate; HP = heart period.

and activity) and childhood adversity in their analyses and found that holding these variables constant increased the strength of the relationship between distress indicators.

Heart period

HP and expressed emotion behaviors. Baseline HP (see Table 1 and 2: Column 5, Row 2): Results were mixed regarding the relation between baseline HP and expressed emotion behaviors, with two studies not reporting the relation between expressed emotion and baseline HP (Calkins & Johnson 1998; Garcia Coll et al., 1984), and one study (Stifter & Jain, 1996) finding a positive relation between distress indicators (i.e., $D = .3$). Studies included in the synthesis collected baseline cardiac data (Calkins & Johnson, 1998; Stifter & Jain, 1996) or cardiac data in response to auditory and visual stimuli (Garcia Coll et al., 1984), reflecting predictive measurements. Two studies included validated distress paradigms (i.e., Behavioral Inhibition task, Toy Removal); however, Stifter & Jain (1996) incorporated multiple distress paradigms (e.g., Toy Removal and Still Face Paradigm). Additionally, most studies utilized inconsistent measurement epochs for behavior (i.e., 120 to 180 s) and cardiac (i.e., 300 s or analyzed by group) data collection and did not include covariates or moderators in their analyses (Calkins & Johnson, 1998; Stifter & Jain, 1996). Garcia Coll and colleagues (1984) investigated sex as a moderator in their analyses

and found that sex did not impact the relation between distress indicators.

HP and emotion regulation behaviors. Baseline HP (see Table 1 and 2: Column 5, Row 3): A Cohen's D of .12 was found (i.e., Burgess et al., 2003) between baseline HP and emotion regulation behaviors (Burgess et al., 2003). Data were not reported for one study (Calkins & Fox, 1992). These studies collected baseline cardiac data (Calkins & Fox, 1992) or cardiac data in response to auditory and visual stimuli (Burgess et al., 2003) (predictive measurement), and both studies utilized a validated distress paradigm (i.e., Behavioral Inhibition task). Both studies used varying measurement lengths for behavioral (i.e., 420 s or not reported) and cardiac (i.e., 180 to 300 s) measurement epochs, and there were no covariates or moderators included in the analyses.

Respiratory sinus arrhythmia

RSA and expressed emotion behaviors. Baseline RSA (see Table 3 and 4: Column 2, Row 2): Most studies found near-zero relationships (i.e., $D = -.00$ to $D = .01$) between baseline RSA and expressed emotion behaviors (Bocknek et al., 2018; Calkins & Johnson, 1998; Paret et al., 2015; Voegtline, 2010). However, two studies (Calkins & Johnson, 1998; Stifter & Jain, 1996) found a positive relation among distress indicators (i.e., $D = .29$ to $.34$). Most studies utilized validated distress paradigms and collected

Table 3. Summary of Study Findings (Effect Sizes) for RSA With Behaviors (Emotion Expression and Emotion Regulation).

Behavioral coding scheme	Baseline RSA	Task RSA	RSA change score
Emotion expression behaviors	Bocknek et al., 2018 (24–31 months: null [–0.0]; N, P) Calkins & Johnson, 1998 (18 months: <i>distress task</i> : positive [.29]; <i>composite distress measure</i> : null [data not provided]; N, P) Paret et al., 2015 (36 months: null [.013]; N, P) Stifter & Jain, 1996 (18 months: positive [.34]; N, P) Voegtline, 2010 (24 months: null [–.14]; N, P)	Brooker & Buss, 2010 (24 months: null [data not provided]; N, C) Buss et al., 2005 (24 months: null [–.15]; N, C) Kahle, 2018 (42 months: <i>reactivity</i> : null [–.00 to –.02], <i>recovery</i> : null [–.08 to –.00]; Y, C)	Brooker & Buss, 2010 (24 months: null [.2]; N, C) Voegtline, 2010 (24 months: <i>fear</i> : null [–.05]; <i>frustration</i> : negative [–.15]; N, P)
Emotion regulation behaviors	Burgess et al., 2003 (24 months: null [.15]; N, P) Calkins and Fox, 1992 (24 months: null [data not provided]; N, P) Liew et al., 2011 (18 months: null [.002], 30 months: null [.002]; Y, P)	Fox, 1989 (14 months: negative [–1.61 to –.23]; N, C) Rubin et al., 1997 (24 months: negative [–.07]; N, C) Kahle et al., 2018 (42 months: null [–.00 to –.01]; Y, C)	Augustine et al., 2018 (24 months: negative [–.04]; N, C) Calkins & Dedmon, 2000 (30 months: null [data not provided]; N, C) Liew et al., 2011 (18 months: null [.04]; 30 months: negative [–.19]; Y, P)

Note. Cohen's *D* was calculated for the results available, where no data are available, gray shading is used. Results displayed as (*Participant Age*, *Direction of association* (positive, negative), *Magnitude of association* (null, small, medium, large) [effect size]); *Time of measurement*: Concurrent (C) = heart rate and behavior measured concurrently; Predictive (P) = heart rate or behavior collected before the other measure; *Epochs*: Yes (Y) = behavior or cardiovascular measure analyzed in reactivity and/or regulation phases; No (N) = behavior and cardiovascular not analyzed in reactivity and/or regulation phases (collapsed across baseline or reactivity and regulation phases); RSA = respiratory sinus arrhythmia.

Table 4. Summary of Study Findings (Patterns of Results) for RSA With Behaviors (Emotion Expression, Emotion Regulation).

Behavioral coding scheme	Baseline RSA	Task RSA	RSA Change Score
Emotion expression behaviors	Weak (4 studies; <i>N</i> = 566; 18–36 months) Positive (2 studies, <i>n</i> = 173; 18 months)	Weak (3 studies; <i>N</i> = 239; 24–42 months)	Weak (2 studies; <i>N</i> = 418; 24 months) Negative (1 study; <i>n</i> = 330; 24 months)
Emotion regulation behaviors	Weak (3 studies; <i>N</i> = 471; 18–24 months)	Weak (1 study, <i>N</i> = 83; 24–42 months) Negative (2 studies, <i>N</i> = 243; 14–42 months)	Weak (2 studies; <i>N</i> = 346; 18–30 months) Negative (1 study; <i>n</i> = 138; 18–30 months)

Note. Results displayed as *Overall Pattern Found*: Positive/Negative/Weak (near-zero associations); Unclear direction and effect size = studies had differing directionality and magnitude of relationships (*Number of studies*, *N/n* = number of participants in studies/study, respectively, *Participant Age*). RSA = respiratory sinus arrhythmia.

Table 5. Summary of Study Findings (Effect Sizes) for PEP With Behaviors (Emotion Expression and Emotion Regulation).

Behavioral coding scheme	Task PEP
Emotion expression behaviors	Buss et al., 2005 (24 months: null [.04]; N, C) Kahle et al., 2018 (42 months: <i>reactivity</i> : null [–.00], <i>recovery</i> : null [–.005 to –.001]; Y, C)
Emotion regulatory behaviors	Kahle et al., 2018 (42 months: <i>reactivity</i> : null/negative [–.00 to –.005], <i>recovery</i> : null [–.00 to –.001]; Y, C)

Note. Cohen's *D* was calculated for the results available, where no data are available, gray shading is used. Results displayed as (*Participant Age*, *Direction of association* (positive, negative), *Magnitude of association* (null, small, medium, large) [effect size]); *Time of measurement*: Concurrent (C) = heart rate and behavior measured concurrently; Predictive (P) = heart rate or behavior collected before the other measure; *Epochs*: Yes (Y) = behavior or cardiovascular measure analyzed in reactivity and/or regulation phases; No (N) = behavior and cardiovascular not analyzed in reactivity and/or regulation phases (collapsed across baseline or reactivity and regulation phases); PEP = pre-ejection period

baseline cardiac data prior to the distress paradigm (predictive measurement). Behavior (i.e., 120 to 270 s) and cardiac (i.e., 120 to 900 s) measurement epochs were variable, and only two studies included analyses with covariates or moderators (Calkins & Johnson, 1998; Paret et al., 2015). Although sex did not moderate the relationship between distress indicators (Calkins & Johnson, 1998),

Table 6. Summary of Study Findings (Patterns of Results) for PEP With Behaviors (Emotion Expression and Emotion Regulation).

Behavioral coding scheme	Task PEP
Emotion expression behaviors	Weak (2 studies; <i>N</i> = 151; 24–42 months)
Emotion regulatory behaviors	Unclear direction and effect size (1 study; <i>n</i> = 83; 42 months)

Note. Results displayed as *Overall Pattern Found*: Positive/Negative, Small/Medium/Large, No effect = no studies with significant findings; Unclear direction and effect size = studies had differing directionality and magnitude of relationships (*Number of studies*, *N/n* = number of participants in studies/study, respectively, *Participant Age*). Where no data are available, gray shading is used; PEP = pre-ejection period.

attachment status did moderate the relationship between baseline RSA and expressed emotion (Paret et al., 2015).

Task RSA (see Table 3 and 4: Column 3, Row 2): Across the three studies included (Brooker & Buss, 2010; Buss et al., 2005; Kahle et al., 2018), calculated Cohen's *D* ranged from –.15 to .00 for the relation between task RSA and expressed emotion behaviors. All studies utilized validated distress induction tasks and measured emotion expression behaviors and cardiac indicators concurrently during the distress task. The studies varied greatly in their behavioral and cardiac (i.e., 12 to 600 s) measurement epochs. No studies investigated relevant covariates.

RSA change score (see Table 3 and 4: Column 4, Row 2): Findings were mixed regarding the relationship between RSA change score and expressed emotion behaviors ($D = -.15$ to $.20$; Brooker & Buss, 2010; Voegtline, 2010). However, differences in how RSA change score was calculated (Brooker & Buss, 2010: Task-Baseline; Voegtline, 2010: Baseline-Task) were noted. Both studies utilized well-validated distress paradigms (i.e., Stranger Approach, Toy Removal). One study (Brooker & Buss, 2010) collected behavioral and cardiac data concurrently, while one study investigated RSA suppression in response to a different task (Voegtline, 2010). The studies included utilized varying measurement epochs for behavior (i.e., 120 to 150 s) and cardiac (i.e., 30 to 900 s) indicators, and only one study (Brooker & Buss, 2010) utilized covariates in their analysis. Specifically, movement was controlled for in the analysis and decreased the relationship between distress indicators.

RSA and emotion regulation behaviors. Baseline RSA (see Table 3 and 4: Column 2, Row 3): Calculated Cohen's D was between $.002$ and $.15$ for studies investigating the association between baseline RSA and emotion regulation behaviors (Burgess et al., 2003; Calkins & Fox, 1992; Liew et al., 2011). All studies utilized validated distress paradigms but did not collect baseline cardiac data consistently. Specifically, cardiac indicators were collected at baseline (Calkins & Fox, 1992) while receiving visual and auditory stimuli (Burgess et al., 2003) or while watching an empathy inducing video (Liew et al., 2011). Cardiac (i.e., 42 to 300 s) and behavioral (i.e., 120 to 420 s) epochs varied in length. Liew and colleagues (2011) were the only study to utilize covariates, and they found that although sex did not affect the relationship between distress indicators, older children had a stronger relationship between baseline RSA and emotion regulation behaviors.

Task RSA (see Table 3 and 4: Column 3, Row 3): The results were inconsistent across the three studies that investigated the relationship between task RSA and emotion regulation behaviors (Fox, 1989; Kahle et al., 2018; Rubin et al., 1997). While one study (Kahle et al., 2018) found near-zero associations between indicators (Cohen's $D = -.00$ to $.01$), two studies found a negative relation ($D = -.07$, Rubin et al., 1997; $D = -1.61$ to $-.23$, Fox, 1989). Distress paradigms were validated and all studies used concurrent measurement of the behavioral and cardiac indicators. Only one study analyzed the behavioral (i.e., 60 to 180 s) and cardiac (i.e., 12 to 25 s) data using distinct measurement epochs (Kahle et al., 2018). The other two studies (Fox, 1989; Rubin et al., 1997) used variable measurement epochs and did not describe the range in length of behavioral or cardiac measurement epochs. One study (Rubin et al., 1997) investigated maternal over-solicitousness and separation/reunion distress as moderators and found that the relationship between indicators did not differ depending on these variables. One study (Fox, 1989) investigated sex as a moderator and found that the relationship between indicators did not depend on sex.

RSA change score (see Table 3 and 4: Column 4, Row 3): The results were inconsistent from the four analyses (Augustine et al., 2018; Calkins & Dedmon, 2000; Liew et al., 2011 had two age groups representing two different analyses) investigating the relationships between RSA change score and emotion regulation behaviors. For two studies, data on the relation between indicators were not reported or findings were near-zero (Calkins & Dedmon, 2000, null findings but data not provided; Liew et al., 2011, $D = .04$). Two studies found negative, albeit sometimes near-zero, relationships between RSA suppression and emotion regulation behaviors

(compliance [$D = -.04$], comfort seeking [$D = -.19$], proximity to mother [$D = -.93$]). Two of the studies conceptualized RSA change score similarly (Baseline-Task; Augustine et al., 2018; Calkins & Dedmon, 2000), yet one study (Liew et al., 2011) utilized a unique method to calculate RSA change score (i.e., regression of baseline RSA onto task RSA). Only one study had a distress paradigm that was not validated (Augustine et al., 2018) and two of the three studies used concurrent measurement of the behavioral and cardiac indicators (Augustine et al., 2018; Calkins & Dedmon, 2000). Studies included ranged in length of behavioral (i.e., 120 to 300 s) and cardiac (i.e., 42 to 300 s) measurement epochs. One study (Liew et al., 2011) investigated sex as a covariate and found that the relationship between indicators did not depend on sex.

Pre-ejection period

PEP and expressed emotion behaviors. Task PEP (see Table 5 and 6: Column 2, Row 2): Both studies (Buss et al., 2005; Kahle et al., 2018) found near-zero relations (Buss et al., 2005, $D = .04$; Kahle et al., 2018, $D = -.005$ to $-.001$) between task PEP and expressed emotion behaviors. Distress paradigms were validated and behavioral and cardiac measures were collected concurrently during the distress paradigm. Behavioral (i.e., 60 to 120 s vs. 60 to 600 s) and cardiac (i.e., 12 to 25 s vs. 60 to 600 s) measurement epochs varied between the two studies included. Only one study had covariates in their analysis (Buss et al., 2005), which was described in a previous section.

Discussion

To our knowledge, the current study is the first to systematically review the available literature on the relation between behavioral and cardiac indicators measured during distress in toddlerhood, taking into account which type of behavioral coding scheme (i.e., emotion expression behaviors, emotion regulation behaviors) was used, which type of cardiac measure was used (i.e., HR, HP, RSA, PEP), and various methodological factors (i.e., timing of measurement [concurrent versus predictive], length of behavioral and cardiac measurement epochs, and covariates/moderators). The narrative synthesis suggested that much of the literature is mixed regarding the association between cardiac and behavioral indicators measured during distress in toddlerhood. The only consistent relationship between cardiac and behavioral indicators was that task HR and expressed emotion behaviors are positively associated with standardized effect sizes in the range of Cohen's $D = .05$ to $D = .54$. Relationships were variable between baseline HR and expressed emotion behaviors (Cohen's D not reported to $.14$), baseline HP and emotion regulation behaviors (Cohen's D not reported to $.12$), task RSA and expressed emotion behaviors (Cohen's $D = -.15$ to $.00$), and baseline RSA and emotion regulation behaviors (Cohen's $D = .002$ to $.15$). Overall effect sizes across cardiac indicators were similar to those reported in previous meta-analyses on the association between RSA and behavior across the life span (Holzman & Bridgett, 2017; Zahn et al., 2016). As such, it is possible that the magnitude of the association between behavioral and most cardiac indicators of distress may in fact be smaller than previously expected (i.e., Cohen's $D < 0.2$).

Importantly, the conclusions need to be couched within the lower quality ratings that the majority (21/22) of the studies received. Indeed, studies did not use consistent, valid, and reliable behavioral coding schemes, report reliability coefficients for HR

indicators, or use analytic techniques that accounted for variables known to impact toddler physiology. Additionally, there was substantial variability in methods of physiological data collection. Of note, studies that fell under the “task” HR, RSA, and PEP did not account for baseline physiology, and as such we cannot conclude if the lack of or consistent associations between indicators in this category is related to differences in baseline or task physiology. Additionally, many studies were not using “true” baseline conditions, which may have increased the literature-wide noise in both resting RSA and RSA change scores. Additionally, many of the sampling rates were below that which is preferred (i.e., 1024 Hz; Beauchaine et al., 2019), and high frequency bands were often not specified when RSA was measured. It is possible that studies not reporting frequency bands may be using bands appropriate for older children and adults, which will create large over- and underestimates of resting RSA and RSA change scores, respectively (Shader et al., 2018). Missing data were not reported in almost one third of the studies, which may have impacted the overall relation between indicators. Finally, the 1996 Task Force guidelines were not cited in any studies published after 1996, which may suggest that noise could have been introduced into the data by not following rigorous methodological standards. Next, we review our narrative synthesis findings and contextualize them based on past literature and methodological challenges within the studies included. We then discuss limitations of our research and implications for future research and clinical practice.

Associations Between Behavioral and Cardiac Indicators During Distress

Heart rate. The narrative synthesis suggests that task HR is consistently associated with behavioral indicators of expressed emotion behavior. Specifically, the indicators are positively associated with an overall effect size with standardized mean differences in the $D = .05$ to $.54$ range. This result suggests that physiological systems might be implicated in affective behavior (Davidson et al., 2000) and suggests that cardiac-behavior associations are strongest when the eliciting events are the same for the cardiac and behavioral measures (i.e., concurrent instead of predictive associations). Indeed, baseline HR was not related to indicators of expressed emotion across studies (i.e., $D = .14$, other study did not provide effect size data). No research has investigated the relation between HR indicators and emotion regulation behaviors, which may be linked given that emotion regulation may affect the intensive and temporal features of emotion (Thompson, 1994).

Heart period. Regarding HP, research has only investigated the relation between baseline HP and expressed emotions and emotion regulation behaviors. Although findings were mixed regarding the relation between baseline HP and expressed emotion behaviors, the relations between baseline HP and emotion regulation behaviors were consistently near zero (i.e., $D = .15$ or effect size data not provided). The results suggest that although HR and HP are often considered inversely related, there may be distinct, albeit weaker, associations between HP and behavioral indicators. It is noteworthy that we found no studies from the last 15 years that utilized HP, likely bolstering the proposition that it has not been a promising indicator in terms of relationships with behavior.

Respiratory sinus arrhythmia. In regards to expressed emotion behaviors, the relations with baseline RSA and RSA change score were mixed (i.e., positive and negative), while relationships with task RSA were consistently near zero (i.e., $D = -.15$ to $.00$). Baseline RSA was weakly associated with emotion regulation behaviors ($D = .002$ to $.15$), while results were mixed (i.e., near-zero or negative relations) for task RSA and RSA change score. These results are consistent with classical theories suggesting that the SNS is most associated with fear and frustration (Fowles, 1988). As such, toddlers’ responses to fear and frustration inductions may be more relevant to sympathetic responses such as HR whereas RSA may be more related to regulatory strategies that are aligned with the PNS (e.g., self-soothing) (Kahle et al., 2018). Indeed, preliminary research by Kahle and colleagues (2018) provides credence for this assertion, but there is currently not enough research investigating the relation between task RSA and emotion regulation behaviors to confirm these hypotheses.

Pre-ejection period. Scientists have only begun assessing the increase in sympathetic influence, as indexed by PEP, during emotional challenge in toddlers. Two studies investigating task PEP and emotion expression behaviors found near-zero associations (i.e., $D = -.005$ to $.04$) between these variables. PEP may be more related to emotion dysregulation or emotion expression behaviors that are poorly matched to context (Cole & Hall, 2008; Davidson et al., 2000). Alternatively, Kahle and colleagues (2018) suggest that attempts to regulate emotions are more strongly linked with physiology than expressions of anger. Unfortunately, there was a lack of research investigating the association between PEP and emotion expression or regulatory behaviors to make any definitive claims.

Methodological Challenges

Despite some consistent patterns emerging in the literature, many associations between cardiac and behavioral indicators were inconsistent and no conclusions could be drawn outside the relationship between task HR and emotion expression behaviors. Relatedly, all studies were rated as lower quality and most studies did not report a justification for their sample size, used predictor and outcome variables with questionable reliability and validity, and did not use proper blinding procedures or covariates. Given that 60% of the studies were published before guidelines were provided for reporting observational studies (von Elm et al., 2007), these limitations are not unexpected. It may be that bias-reducing procedures were in fact undertaken, but not reported. However, to clarify the mixed relations between behavioral and physiological indicators during distress, a more in-depth discussion of the methodological differences (i.e., behavioral and cardiac measures, timing and length of measurement epochs, additional analyses) among studies is discussed below.

Behavioral indices measured during distress. Behavioral indicators of emotion expression have been integral in understanding the development of emotion regulation. However, our synthesis revealed that there are not many valid or reliable measures of emotion expression behaviors being used, and the same higher quality behavioral coding systems are not implemented consistently across studies. Indeed, only nine of 22 studies provided citations regarding the validity and reliability of their behavioral coding system. This finding is concerning because most studies included coding systems that involve a high degree of subjectivity and, as such, could

introduce a substantial amount of measurement error. Additionally, despite Cole and colleagues (2004) assertion that we need independent measurement of the activated emotion and the putative regulatory strategies to assess the dynamic nature of emotion properly, behavioral indicators are not commonly coded or categorized as such. Thus, we organized the behavioral coding schemes as either emotion expression behaviors or emotion regulation behaviors, but at times this dichotomization was challenging due to the inclusion of behaviors that were from both expression and regulatory categories. Despite these challenges, our decision to dichotomize studies based on this distinction led to a more nuanced understanding of behavioral responses during distress, as expressed emotions were most consistently related to task HR. This finding is consistent with past research that has suggested that measures of discrete negative emotions (e.g., fear, sadness) are more sensitive than global ratings of negative affect or distress (Buss & Goldsmith, 2007). Yet, emotion regulation behaviors were inconsistently related to cardiac indexes during distress. Many studies investigated behavioral inhibition, a dispositional construct defined by an inborn bias to respond to unfamiliar events by showing anxiety (Kagan, 1994). Behavioral inhibition is often reflected in regulatory behaviors (e.g., latency to approach the stranger or unfamiliar objects, time spent proximal to the mother, or physical distance from the mother), and past work has found little evidence that indexes of behavioral inhibition are consistent across situations (Rubin et al., 1997). Additionally, recent work (e.g., Kahle et al., 2018; Perry et al., 2016) has suggested that emotion regulation behaviors should be further categorized by strategy (e.g., distraction, orienting toward mother, verbalizations), as each strategy can be differentially related to the SNS and PNS. Given that most studies included in the review collapsed across behavioral regulation strategies, it was not possible to further dichotomize emotion regulation strategies. However, this may account for the inconsistent relations between emotion regulation behaviors and cardiac indicators. Moreover, initially we had wanted to more specifically distinguish between the peak reactivity phase to a stressor and the less proximal regulation phase. However, the majority of studies (16/22) did not distinguish between initial post-stressor (reactivity) and less-proximal (regulation) measurement epochs (i.e. researchers collapsed coding systems over the toddler's immediate reaction to stimuli and the regulation from the stimuli), which is an important distinction that should be made in the future.

Cardiac indices of measured during distress. Our narrative synthesis suggests that task HR is positively correlated with expressed emotion behaviors. However, baseline HR and HP, baseline and task RSA, and task PEP are not consistently related to expressed emotion or emotion regulation behaviors. Past research has suggested that there are important developmental changes in the sympathetic and PNS across toddlerhood. Thus, in the second year of life, autonomic measures such as RSA may be more reflective of physiological regulation than simply emotional reactivity or arousal (Eisenberg et al., 1995). Furthermore, few studies have investigated the stability of cardiac measures across the first years of life, which limits our ability to state that these measures represent stable individual differences in distress responses. Finally, baseline and task RSA reflect different aspects of self-regulatory functioning (Porges, 2007). While baseline RSA reflects an individual's ability to regulate internal bodily processes and temperamental characteristics (Beauchaine, 2001; Porges, 1996), task RSA (following a distressing event) is intended to reflect an individual's ability to

attend and respond to environmental demands (Beauchaine et al., 2007). As such, null or mixed findings may result from researchers choosing cardiac measures that are not aligned theoretically with their behavioral outcomes. Again, the lack of disambiguation in studies between reactivity and regulation measurement epochs would likely have an important impact.

Timing and length of measurement. As mentioned above, key methodological factors that we hypothesized to affect the relation between behavioral and cardiac indexes were the timing and length of measurement (i.e., epoch length) for behavioral and cardiac indicators. Indeed, distress reactivity and regulation have been clearly delineated in psychophysiological research (Rothbart & Derryberry, 1981) and researchers have suggested that distress may be more accurately measured through temporally sensitive techniques that capture dynamic change (Thompson et al., 2008). Regarding the timing of measurement, most of the stronger relations were characterized by concurrent measurement of behavioral and cardiac indexes during distress. As such, studies that use predictive analyses (e.g., baseline cardiac variables) may not capture salient aspects of toddlers' distress within these specific paradigms, which helps explain the weaker relations between indicators. Relatedly, the operationalization of baseline cardiac indicators differed across studies, with some studies investigating physiology at rest whereas others collected physiological data during presentation of auditory or visual stimuli. It can be posited that the degree of association may be affected by the operationalization of baseline physiology. As well, although most studies collapsed their behavioral and cardiac measures across the entire distress paradigm, studies that differentiated between phases of distress (i.e., reactivity vs. regulation measurement epochs) had stronger effects.

Additional analyses. Only approximately half of the studies (12/22) included in our review measured and adjusted for key potential confounding variables that are known to impact physiology (i.e., age, sex, baseline characteristics, time of day, activity level, and respiration; Oberlander & Saul, 2002; Waxman et al., 2016). Moreover, the covariates and moderators used were not consistent across studies and included child-level (i.e., sex, age, baseline HR, and activity) and dyad-level (i.e., attachment, early adversity) factors. Sex was the most widely examined variable across studies, and results suggest that there are not sex differences in the relation between behavioral and cardiac indicators following distress during this specific developmental stage. These results counter past research that has suggested that the association between behavioral and cardiac indexes of emotion may differ for boys and girls (Buss et al., 2005). As well, although previous research has noted that there are rapid physiological changes in infancy and toddlerhood which make it difficult to detect the relationship between behavior and biology (Stifter & Jain, 1996), little research has controlled for or examined the effect of age in its analyses. Other child-level (i.e., baseline HR/activity) and dyad factors (i.e., attachment, early adversity) were significant covariates or moderators in the relation between behavioral and cardiac indexes. However, conclusions cannot be made given that these variables were tested in only one study each.

Limitations

Despite our detailed search strategy, it is possible that we omitted relevant studies. We specifically excluded non-English to

publications, which may have eliminated relevant studies. Moreover, we collapsed across tasks (i.e., fear vs. frustration paradigms) given a lack of *qualitative* differences in effect sizes. However, it is possible that if quantitative techniques were utilized (i.e., meta-analysis), there may have been task-related differences in the association between emotion expression/regulatory behaviors and cardiac indicators of distress. Indeed, research in preschool- and school-aged children (e.g., Quigley & Stifter, 2006; Zeytinoglu et al., 2019) has found task-related (e.g., cognitive vs. negatively valenced emotional challenges) differences in cardiac reactivity and the magnitude of physiological responses. It is also important to consider that almost all studies (21/22) were rated as *lower* quality. However, this result may be partly due to reporting standards for papers that reflect journal requirements and decisions by journal reviewers and editors instead of the quality of the original study. Using quality checklists that delineate clear standards of reporting are critical. The EQUATOR network provides an excellent resource to ensure high-quality reporting scores (<http://www.equator-network.org/>). Relatedly, many studies did not report data on the relation between indicators when results were null. As such, the synthesis may underestimate the relation between indicators, given that null results may have been a result of small sample sizes and not due to a lack of relation. Additionally, most studies were from North-American researchers, which may limit the generalizability of our findings. Finally, the study focused on research that specifically analyzed the concurrent relation (i.e., within the same appointment) between behavioral and cardiac indexes measured during distress and did not include studies that investigated the longitudinal relation between these measures. As such, not all studies with informative findings pertaining to predictive relations between these measures were included.

Conclusions and Future Directions

The current study is one of the first to review the concurrent relation between behavioral and cardiac indicators measured during distress during the second and third years of life. The narrative synthesis suggests that currently only HR during the post-stimuli phase is consistently related to expressed emotion behaviors (i.e., standardized mean difference, $D = .05$ to $.54$). These results suggest that research using *distress* as a key construct could obtain divergent results depending on whether distress is measured with cardiac or behavioral indicators. Significant gaps in the literature were highlighted, with a lack of studies investigating the relation between HR indicators and emotion regulation behaviors, task HP and behavioral indicators, and PEP and emotion regulation behaviors. Furthermore, it is noteworthy that there are currently no studies on the relation between behavioral and physiological indicators in severe distress (e.g., pain) in toddlerhood. Given the particular salience of high distress experiences in development and increased dependency on parents during high distress, this will be an important avenue to investigate.

Mixed or near-zero associations may suggest that context-appropriate physiological and expressed emotion behaviors or emotion regulation behaviors can co-occur without being closely linked (Hastings et al., 2009). Relatedly, a recent review by Barrett and colleagues (2019) found that there is not yet sufficient evidence to conclude that facial movements have the sensitivity or specificity to express the instance of any emotion category. As such, it is possible

that the reliability of facial expressions relating to distressing emotions may not emerge until later childhood or adolescence.

However, it is possible that the lack of relations between behavioral and cardiac indicators of distress may differ due to methodological inconsistencies. Thus, another important contribution of this work is concrete suggestions for improving the quality and consistency of research in this area. Based on our review, we recommend the following guidelines for future work examining behavioral and cardiac measures of distress:

1. Improve study methodology (e.g., psychometrically sound baseline conditions and behavioral measures, appropriate frequency bands for measuring RSA in toddlerhood [i.e., 0.24 to 1.04 Hz], coder blinding to hypotheses, distinguish between phases of the post-stimuli responses [e.g., immediate responses vs. more distal responses]) and/or reporting practices (e.g., reporting on standardized data collection [1996 Task Force], ECG sampling rates/high frequency HRV bands, sufficient statistics, recruitment rates/missing data, participation rates, study population characteristics).
2. Behavioral coding schemes are needed that clearly distinguish specific types of behavior being measured (i.e., *emotion expression behavior* vs. *emotion regulation behavior*) and have established psychometrics.
3. Reliability coefficients for cardiac data reduction need to be reported. A review by Laborde and colleagues (2017) noted that the ease of access to cardiac indicators in psychophysiological research has been overshadowed by the difficulty of editing and interpreting cardiac findings. Our results confirm this assertion and we suggest that future research mirror reporting standards for behavioral observation tools. Specifically, authors should report number of coders editing cardiac data and report reliability coding (e.g., double coding 10–20% of the sample, reporting reliability coefficients).
4. The relative contributions of parasympathetic and sympathetic reactivity and regulation need to be accounted for in future research (e.g., collection of HR, RSA, PEP, galvanic skin response, cortisol, cortical responses).
5. Temporal measurement epochs need to be defined in a similar manner across studies to facilitate comparisons (e.g., utilizing concurrent 30-s measurement epochs for both behavioral coding and cardiac data reduction) of the impact of phase of distress (i.e., reactivity phase vs. regulation phase) on the relation between behavioral and cardiac indicators of distress.
6. Statistical approaches that capture dynamic aspects of distress reactivity and regulation (e.g., latent growth curve models) should be utilized, given that levels of parasympathetic and sympathetic activity are likely to fluctuate over the course of a distressing event. As well, a consistent set of covariates should be included for all studies investigating the relation between behavioral and cardiac indexes of distress in toddlerhood. These covariates should include but are not limited to age, sex, time since last feeding and nap, baseline physiological variables, activity, and respiration rate.

Acknowledgment

The authors would like to thank Alanna Marson at the Hospital for Sick Children for performing the data searches, and Drs. Louis Schmidt and David Flora for their consultation on this review.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by national, provincial, and internal fellowships to JW from the Canadian Institutes of Health Research, Ontario Graduate Scholarship and the Lillian Meighen and Don Wright Foundation. JW and MD are also a trainee member of Pain In Child Health (PICH), a strategic research training initiative of the Canadian Institutes of Health Research. Operating funds from the Natural Sciences and Engineering Council and infrastructure grants from the Canadian Foundation for Innovation and Ontario Research Fund to RPR supported the completion of this study.

ORCID iD

Rebecca R. Pillai Riddell  <https://orcid.org/0000-0003-3990-3680>

Supplemental material

Supplemental material for this article is available online.

Note

1. “Measurement epoch” refers to the time interval (in seconds) that behavioral and cardiac data were analyzed/coded across a distress task. The measurement epoch is the final time interval utilized to measure the association between behavioral and cardiac indicators during distress. For example, for a given study, heart rate variability may have been edited in 30-s epochs and averaged across a 2-min distress paradigm, but the final “measurement epoch” length for that study would be 120 s.

References

- Augustine, M. E., Leerkes, E. M., Smolen, A., & Calkins, S. D. (2018). Relations between early maternal sensitivity and toddler self-regulation: Exploring variation by oxytocin and dopamine D2 receptor genes. *Developmental Psychobiology*, *60*, 789–804.
- Baker, E., Baibazarova, E., Ktistaki, G., Shelton, K. H., & van Goozen, S. H. M. (2012). Development of fear and guilt in young children: Stability over time and relations with psychopathology. *Development and Psychopathology*, *24*, 833–845. <https://doi.org/10.1017/S0954579412000399>
- Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review*, *5*(4), 379–389. <https://doi.org/10.1177/1754073913489753>
- Barrett, L. F. (2015). In L. F. Barrett & J. A. Russell (Eds.), *The Psychological construction of emotion* (pp. 45–79). Guilford.
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial expressions. *Developmental Psychology*, *55*(9), 1830–1849.
- Beauchaine, T. P. (2001). Vagal tone, development, and gray’s motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. *Development and Psychopathology*, *13*, 183–214. <https://doi.org/10.1017/S0954579401002012>
- Beauchaine, T. P., Bell, Z., Knapton, E., McConough-Caplan, H., Shader, T., & Zisner, A. (2019). Respiratory sinus arrhythmia reactivity across empirically-based structural dimensions of psychopathology: A meta-analysis. *Psychophysiology*, *56*, e13329.
- Beauchaine, T. P., Gatzke-Kopp, L., & Mead, H. K. (2007). Polyvagal theory and developmental psychopathology: Emotion dysregulation and conduct problems from preschool to adolescence. *Biological Psychology*, *74*, 174–184. <https://doi.org/10.1016/j.biopsycho.2005.08.008>
- Berntson, G. G., Quigley, K. S., & Lozano, D. (2007). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 182–210). Cambridge University Press.
- Berntson, G. G., Uchino, B. N., & Cacioppo, J. T. (1994). Origins of baseline variance and the law of initial values. *Psychophysiology*, *31*, 204–210.
- Bocknek, E. L., Richardson, P. A., van den Heuvel, M. I., Qipo, T., & Brophy-Herb, H. E. (2018). Sleep moderates the association between routines and emotion regulation for toddlers in poverty. *Journal of Family Psychology*, *32*(7), 966–974.
- Bohlin, G., & Hagekull, B. (1993). Stranger wariness and sociability in the early years. *Infant Behavior and Development*, *16*, 53–67. [https://doi.org/10.1016/0163-6383\(93\)80028-7](https://doi.org/10.1016/0163-6383(93)80028-7)
- Booth, A. (2013). PROSPERO: International register of systematic reviews. <https://www.crd.york.ac.uk/PROSPERO/2013>.
- Bornstein, M. H., & Suess, P. E. (2000). Child and mother cardiac vagal tone: Continuity, stability, and concordance across the first 5 years. *Developmental Psychology*, *36*(1), 54–65.
- Brenner, S. L., Beauchaine, T. P., & Sylvers, P. D. (2005). A Comparison of psychophysiological and self-report measures of behavioral approach and behavioral inhibition. *Psychophysiology*, *42*, 108–115.
- Brooker, R. J., & Buss, K. A. (2010). Dynamic measures of RSA predict distress and regulation in toddlers. *Developmental Psychobiology*, *52*(4), 372–382. <https://doi.org/10.1002/dev.20432>
- Burgess, K. B., Marshall, P. J., Rubin, K. H., & Fox, N. A. (2003). Infant attachment and temperament as predictors of subsequent externalizing problems and cardiac physiology. *Journal of Child Psychology and Psychiatry*, *44*(6), 819–831. <https://doi.org/10.1111/1469-7610.00167>
- Buss, K. A., & Goldsmith, H. H. (2007). Biobehavioral approaches to early socioemotional development. In C. A. Brownell & C. B. Kopp (Eds.) *Socioemotional development in the toddler years* (pp. 370–395). Guilford Press.
- Buss, K. A., Goldsmith, H. H., & Davidson, R. J. (2005). Cardiac reactivity is associated with changes in negative emotion in 24-month-olds. *Developmental Psychobiology*, *46*, 118–132. <https://doi.org/10.1002/dev.20048>
- Calkins, S. D. (2010). Conceptual and methodological challenges to the study of emotion regulation and psychopathology. *Journal of Psychopathology and Behavioral Assessment*, *32*, 92–95.
- Calkins, S. D., & Dedmon, S. E. (2000). Physiological and behavioral regulation in two-year-old children with aggressive/destructive behavior problems. *Journal of Abnormal Child Psychology*, *28*, 103–118.
- Calkins, S. D., & Fox, N. A. (1992). The relations among infant temperament, security of attachment, and behavioral inhibition at twenty-four months. *Child Development*, *63*, 1456–1472. <https://doi.org/10.2307/1131568>
- Calkins, S. D., & Fox, N. A. (2002). Self-regulatory processes in early personality development: A multilevel approach to the study of childhood social withdrawal and aggression. *Development and Psychopathology*, *14*, 477–498.
- Calkins, S. D., & Hill, A. (2007). Caregiver influences on emerging emotion regulation: Biological and environmental transactions in early development. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 229–248). The Guilford Press.

- Calkins, S. D., & Johnson, M. C. (1998). Toddler regulation of distress to frustrating events: Temperamental and maternal correlates. *Infant Behavior & Development, 21*(3), 379–395. [https://doi.org/10.1016/S0163-6383\(98\)90015-7](https://doi.org/10.1016/S0163-6383(98)90015-7)
- Calkins, S. D., & Leerkes, E. M. (2004). Early attachment processes and the development of emotional self-regulation. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory and applications* (pp. 324–339). The Guilford Press.
- Campos, J. J., Campos, R. G., & Barrett, K. C. (1989). Emergent themes in the study of emotional development and emotion regulation. *Developmental Psychology, 25*(3), 394–402. <https://doi.org/10.1037/0012-1649.25.3.394>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cole, P. M., & Hall, S. E. (2008). Emotion dysregulation as a risk factor for psychology. In T. P. Beauchaine & S. P. Hinshaw (Eds.), *Child and adolescent psychopathology* (pp. 265–300). John Wiley and Sons.
- Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development, 75*(2), 317–333. <https://doi.org/10.1111/j.1467-8624.2004.00673.x>
- Crombie, I. K., & McQuay, H. J. (1998). The systematic review: A good guide rather than a guarantee. *Pain, 76*, 1–2. [https://doi.org/10.1016/S0304-3959\(98\)00038-4](https://doi.org/10.1016/S0304-3959(98)00038-4)
- Crowell, S. E., Beauchaine, T. P., Gatzke-Kopp, L., Sylvers, P., Mead, H., & Chipman-Chacon, J. (2006). Autonomic correlates of attention-deficit/hyperactivity disorder and oppositional defiant disorder in preschool children. *Journal of Abnormal Psychology, 115*(1), 174–178.
- Davidson, R. J., Jackson, D. C., & Kalin, N. H. (2000). Emotion, plasticity, context and regulation: Perspectives from affective neuroscience. *Psychological Bulletin, 126*, 890–906. <https://doi.org/10.1037/0033-2909.126.6.890>
- Domes, G., Schulze, L., Böttger, M., Grossman, A., Hauenstein, K., Wirtz, P. H., Heinrichs, M., & Herpertz, S. C. (2009). The neural correlates of sex differences in emotional reactivity and regulation. *Human Brain Mapping, 31*(5), 758–769. <https://doi.org/10.1002/hbm.20903>
- Downs, S. H., & Black, N. (1998). The Feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*, 377–384. <https://doi.org/10.1136/jech.52.6.377>
- Eisenberg, N., Fabes, R. A., Murphy, B., Maszk, P., Smith, M., & Karbon, M. (1995). The role of emotionality and regulation in children's social functioning: A longitudinal study. *Child Development, 66*, 1360–1384. <https://www.jstor.org/stable/1131652>
- Ekas, N. V., Braungart-Rieker, J. M., & Messinger, D. S. (2018). *The development of infant emotion regulation: Time is of the essence*. In P. M. Cole & T. Hollestein (Eds.), (pp. 31–51). Emotion regulation: A matter of time.
- Enebrink, P., Andershed, H., & Långström, N. (2005). Callous–unemotional traits are associated with clinical severity in referred boys with conduct problems. *Nordic Journal of Psychiatry, 59*(6), 431–440. <https://doi.org/10.1080/08039480500360690>
- Fowles, D. (1988). Psychophysiology and psychopathology: A motivational approach. *Psychophysiology, 25*, 373–391.
- Fox, N. A. (1989). Psychophysiological correlates of emotional reactivity during the first year of life. *Developmental Psychology, 25*(3), 364–372. <https://doi.org/10.1037/0012-1649.25.3.364>
- Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003). Callous-unemotional traits and conduct problems in the prediction of conduct problem severity, aggression, and self-report of delinquency. *Journal of Abnormal Child Psychology, 31*(4), 457–470. http://resolver.scholarsportal.info/resolve/00910627/v31i0004/457_ctacpisaasod.xml
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*. Advance online publication. <https://doi.org/10.1177/2515245919847202>
- Garcia Coll, C., Kagan, J., & Reznick, J. S. (1984). Behavioral inhibition in young children. *Child Development, 55*(3), 1005–1019. <https://www.jstor.org/stable/1130152>
- Gatzke-Kopp, L., & Ram, N. (2018). Developmental dynamics of autonomic function in childhood. *Psychophysiology, 55*, e13218.
- Gendron, M., & Feldman Barrett, L. (2009). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review, 1*(4), 316–339. <https://doi.org/10.1177/1754073909338877>
- Gunnar, M. R., & Vazquez, D. (2006). Stress neurobiology and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Developmental neuroscience* (pp. 533–577). John Wiley & Sons Inc.
- Hastings, P. D., Nuselovici, J. N., Klimes-Dougan, B., Kendziora, K. T., Usher, B. A., Ho, M. R., & Zahn-Waxler, C. (2009). Dysregulated coherence of subjective and cardiac emotional activation in adolescents with internalizing and externalizing problems. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 50*(11), 1348–1356.
- Hay, D. F., van Goozen, S. H. M., Mundy, L., Phillips, R., Roberts, S., Meeuwse, M., Goodyer, I., & Perra, O. (2017). If you go down to the woods today: Infants' distress during a teddy bear's picnic in relation to peer relations and later emotional problems. *Infancy, 22*(4), 552–570. <https://doi.org/10.1111/inf.12172>
- Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. www.handbook.cochrane.org
- Holzman, J. B., & Bridgett, D. J. (2017). Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: A meta-analytic review. *Neuroscience & Biobehavioral Review, 74*, 233–255.
- Kagan, J. (1994). *Galen's prophecy: Temperament in human nature*. Basic Books.
- Kagan, J., Reznick, J. S., & Gibbons, J. (1989). Inhibited and uninhibited types of children. *Child Development, 60*(4), 838–845. <https://www.jstor.org/stable/1131025>
- Kahle, S., Miller, J. G., Helm, J. L., & Hastings, P. D. (2018). Linking autonomic physiology and emotion regulation in preschoolers: The role of reactivity and recovery. *Developmental Psychobiology, 60*(7), 775–788.
- Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*, 199–214. <https://doi.org/10.1037/0012-1649.18.2.199>
- Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting. *Frontiers of Psychology, 8*, 213. <https://doi.org/10.3389/fpsyg.2017.00213>
- Liew, J., Eisenberg, N., Spinrad, T. L., Eggum, N. D., Haugen, R. G., Kupfer, A., Reiser, M. R., Smith, C. L., Lemery-Chalfant, K., & Baham, M. E. (2011). Physiological regulation and fearfulness as

- predictors of young children's empathy-related reactions. *Social Development*, 20(1), 111–113. <https://doi.org/10.1111/j.1467-9507.2010.00575.x>
- Matthews, S. G., & Phillips, D. I. (2010). Minireview: Transgenerational inheritance of the stress response: A new frontier in stress research. *Endocrinology*, 151, 7–13. <https://doi.org/10.1210/en.2009-0916>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151, 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- National Heart, Lung, and Blood Institute. (2014). Quality assessment tool for observational cohort and cross-sectional studies. *National Institute of Health*. <https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiac-risk-reduction/tools/cohort>
- Oberlander, T., & Saul, J. P. (2002). Methodological considerations for the use of heart rate variability as a measure of pain reactivity in vulnerable infants. *Clinical Perinatology*, 29, 427–443. [https://doi.org/10.1016/S0095-5108\(02\)00013-1](https://doi.org/10.1016/S0095-5108(02)00013-1)
- Obradović, J., & Finch, J. E. (2017). Linking executive function skills and physiological challenge response: Piecewise growth curve modeling. *Developmental Science*, 20(6), e12476.
- Paret, L., Bailey, H. N., Roche, J., Bureau, J. F., & Moran, G. (2015). Preschool ambivalent attachment associated with a lack of vagal withdrawal in response to stress. *Attachment & Human Development*, 17(1), 65–82. <https://doi.org/10.1080/14616734.2014.967786>
- Perry, N., & Calkins, S. (2018). *A biopsychosocial perspective on the development of emotion regulation across childhood*. In P. M. Cole & T. Hollestein (Eds.), *Emotion regulation: A matter of time* (pp. 3–30).
- Perry, N., Calkins, S., & Bell, M. A. (2016). Indirect effects of maternal sensitivity on infant emotion regulation behaviors: The role of vagal withdrawal. *Infancy*, 21, 128–153.
- Pillai Riddell, R., & Racine, N. (2009). Assessing pain in infancy: The caregiver context. *Journal of Pain Research and Management*, 14(1), 27–32. <https://doi.org/10.1155/2009/410725>
- Pillai Riddell, R., Flora, D. B., Stevens, S. A., Stevens, B., Cohen, L., Greenberg, S., & Garfield, H. (2013). Variability in infant acute pain responding meaningfully obscured by averaging pain responses. *PAIN*, 154, 714–721.
- Popay, P., Roberts, H., Sowden, A., Petticrew, M., Arai, L., & Rodgers, M. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme*. Lancaster: Institute of Health Research.
- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage: A Polyvagal theory. *Psychophysiology*, 32, 301–318.
- Porges, S. W. (1996). Physiological regulation in high-risk infants: A model for assessment and potential intervention. *Development and Psychopathology*, 8, 43–58.
- Porges, S. W. (2007). The Polyvagal perspective. *Biological Psychology*, 74, 116–143.
- Provost, M. A., & Gouin-Decarie, T. (1979). Heart rate reactivity of 9- and 12-months old infants showing specific emotions in natural setting. *International Journal of Behavioral Development*, 2, 109–120. <https://doi.org/10.1177/016502547900200201>
- Qu, J., & Leerkes, E. M. (2018). Patterns of RSA and observed distress during the still-face paradigm predict later attachment, compliance and behavior problems: A person-centered approach. *Developmental Psychobiology*, 60, 707–721.
- Quigley, K. S., & Stifter, C. A. (2006). A comparative validation of sympathetic reactivity in children and adults. *Psychophysiology*, 43, 357–365.
- Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (Vol. 1, pp. 39–85). Erlbaum.
- Rothbart, M. K., Ziaie, H., & Boyle, C. G. (1992). Self-regulation and emotion in infancy. In N. Eisenberg & R. A. Fabes (Eds.), *Emotion and its regulation in early development* (pp. 7–23). San Francisco: Jossey-Bass.
- Rubin, K. H., Hastings, P. D., Stewart, S. L., Henderson, H. A., & Chen, X. (1997). Consistency and concomitants of inhibition: Some of the children, all of the time. *Child Development*, 68(3), 467–483. <https://doi.org/10.1111/j.1467-8624.1997.tb01952.x>
- Sanderson, S., Tatt, I. D., & Higgins, J. P. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, 36, 666–676. <https://doi.org/10.1093/ije/dym018>
- Santucci, A. K., Silk, J. S., Shaw, D. S., Gentzler, A., Fox, N. A., & Kovacs, M. (2008). Vagal tone and temperament as predictors of emotion regulation strategies in young children. *Developmental Psychobiology*, 50(3), 205–216.
- Seifer, R., Schiller, M., Sameroff, A. J., Resnick, S., & Riordan, K. (1996). Attachment, maternal sensitivity, and infant temperament during the first year of life. *Developmental Psychology*, 32(1), 12–25. <https://doi.org/10.1037/0012-1649.32.1.12>
- Shader, T. M., Gatzke-Kopp, L. M., Crowell, S. E., Reid, M. J., Thayer, J. F., Vasey, M. W., Webster-Stratton, C., Bell, Z., & Beauchaine, T. P. (2018). Quantifying respiratory sinus arrhythmia: Effects of misspecifying breathing frequencies across development. *Development and Psychopathology*, 30, 351–366.
- Sherwood, A., Allen, M. T., Fahnenberg, J., Kelsey, R. M., Lovallo, W. R., & van Dooren, L. J. P. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology*, 27, 1–23.
- Shirlcliff, E. A., Vitacco, M. J., Graf, A. R., Gostisha, A. J., Merz, J. L., & Zahn-Waxler, C. (2009). Neurobiology of empathy and callousness: Implications for the development of antisocial behavior. *Behavioral Sciences & The Law*, 27(2), 137–171. <https://doi.org/10.1002/bsl.862>
- Spangler, G., & Grossman, K. E. (1993). Biobehavioral organization in securely and insecurely attached infants. *Child Development*, 64(5), 1439–1450. <https://www.jstor.org/stable/1131544>
- Stifter, C. A., & Fox, N. A. (1990). Infant reactivity: Physiological correlates of newborn and 5-month temperament. *Developmental Psychology*, 26(4), 582. <https://doi.org/10.1037/0012-1649.26.4.582>
- Stifter, C. A., & Jain, A. (1996). Psychophysiological correlates of infant temperament: Stability of behavior and autonomic patterning from 5 to 18 months. *Developmental Psychobiology*, 29(4), 379–391. [https://doi.org/10.1002/\(SICI\)1098-2302\(199605\)29:4<379:AID-DEV5>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1098-2302(199605)29:4<379:AID-DEV5>3.0.CO;2-N)
- Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task force of the European society of cardiology and the North American society of pacing and electrophysiology. *Circulation*, 93, 1043–1065. <https://doi.org/10.1161/01.CIR.93.5.1043>

- Thompson, R. A. (1994). Emotion regulation: A Theme in search of definition. In N. A. Fox (Ed.), *The Development of emotion regulation: Biological and behavioral considerations. Monographs of the Society for Research in Child Development*, 59 (Serial No. 240), 25–52.
- Thompson, R. A., Lewis, M. D., & Calkins, S. D. (2008). Reassessing emotion regulation. *Child Development Perspectives*, 2(3), 124–131. <https://doi.org/10.1111/j.1750-8606.2008.00054.x>
- Voegtline, K. M. (2010). *Physiological bases of emotion regulation in young children living in low-income, rural communities*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, Pennsylvania, USA.
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gotsche, P. C., & Vandembroucke, J. P. (2007). The Strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, 147(8), 573–577. <https://doi.org/10.1371/journal.pmed.0040296>
- Waxman, J. A., Pillai Riddell, R. R., Pinhasov, A., Schmidt, L. A., & Tablon, P. (2016). Development of cardiovascular indexes of acute pain responding in infants: A systematic review. *Pain Research and Management*, 16, 1–16. <https://doi.org/10.1155/2016/8458696>
- Zahn, D., Adams, J., Krohn, J., Wenzel, M., Mann, C. G., Gomille, L. K., Jacobi-Scherbening, V., & Kubiak, T. (2016). Heart rate variability and self-control—A meta-analysis. *Biological Psychology*, 115, 9–26.
- Zeytinoglu, S., Calkins, S. D., & Leerkes, E. M. (2019). Autonomic nervous system functioning in early childhood: Responses to cognitive and negatively valenced emotional challenges. *Developmental Psychobiology*, 00, 1–17.
- Zimmer-Gembeck, M. J., Webb, H. J., Pepping, C. A., Swan, K., Merlo, O., Skinner, E. A., Avdagic, E., & Dunbar, M. (2017). Review: Is parent-child attachment a correlate of children's emotion regulation and coping? *International Journal of Behavioral Development*, 41, 74–93. <https://doi.org/10.1177/0165025415618276>