Test-Retest Reliability of Pediatric Heart Rate Variability:

A Meta-Analysis

Oren M. Weiner and Jennifer J. McGrath

Pediatric Public Health Psychology Laboratory, Concordia University, Montreal, Quebec, Canada

Abstract

Heart rate variability (HRV), an established index of autonomic cardiovascular modulation, is associated with health outcomes (e.g., obesity, diabetes) and mortality risk. Time- and frequency-domain HRV measures are commonly reported in longitudinal adult and pediatric studies of health. While test-retest reliability has been established among adults, less is known about the psychometric properties of HRV among infants, children, and adolescents. The objective was to conduct a meta-analysis of the test-retest reliability of time- and frequency-domain HRV measures from infancy to adolescence. Electronic searches (PubMed, PsycINFO; January 1970–December 2014) identified studies with nonclinical samples aged 18 years; 2 baseline HRV recordings separated by 1 day; and sufficient data for effect size computation. Forty-nine studies ($N = 5,170$) met inclusion criteria. Methodological variables coded included factors relevant to study protocol, sample characteristics, electrocardiogram (ECG) signal acquisition and preprocessing, and HRV analytical decisions. Fisher's Z was derived as the common effect size. Analyses were age-stratified (infant/toddler < 5 years, $n = 3,329$; child/adolescent 5–18 years, $n = 1,841$) due to marked methodological differences across the pediatric literature. Meta-analytic results revealed HRV demonstrated moderate reliability; child/adolescent studies ($Z = 0.62$, $r = 0.55$) had significantly higher reliability than infant/toddler studies ($Z = 0.42$, $r = 0.40$). Relative to other reported measures, HF exhibited the highest reliability among infant/toddler studies ($Z = 0.42$, $r = 0.40$), while rMSSD exhibited the highest reliability among child/adolescent studies ($Z = 1.00$, $r = 0.76$). Moderator analyses indicated greater reliability with shorter test-retest interval length, reported exclusion criteria based on medical illness/condition, lower proportion of males, prerecording acclimatization period, and longer recording duration; differences were noted across age groups. HRV is reliable among pediatric samples. Reliability is sensitive to pertinent methodological decisions that require careful consideration by the researcher. Limited methodological reporting precluded several *a priori* moderator analyses. Suggestions for future research, including standards specified by Task Force Guidelines, are discussed.

Keywords

pediatric; heart rate variability; reliability; respiratory sinus arrhythmia; psychometric

---

Jennifer J. McGrath, Pediatric Public Health Psychology Laboratory, Concordia University, 7141 Sherbrooke Street West, SP244, Montreal, Quebec, H4B 1R6, Canada, Tel. (514) 848-2424 ext. 5287, Jennifer.McGrath@concordia.ca.

Disclosure Statements

All authors disclose no actual or potential conflicts of interest including any financial, personal, or other relationships with other people or organizations that could inappropriately influence (bias) their work.

Heart rate variability (HRV) reflects the variance in time between consecutive sinoatrial depolarizations (i.e., NN-intervals) and is an established index of autonomic cardiovascular modulation. Commonly reported time- and frequency-domain HRV measures, typically derived using continuous electrocardiogram (ECG) recordings and specialized analysis software programs, each provide a unique and nuanced perspective of autonomic functioning (Ernst, 2014; Kleiger, Stein, & Bigger, 2005).

Mathematically derived time-domain measures include the standard deviation of NN-intervals (SDNN; reflecting parasympathetic, sympathetic, and circadian influences), the root mean square of successive NN-interval differences (rMSSD), and percentage of successive NN-intervals that differ by > 50 ms (pNN50); these latter indices both reflect parasympathetic influences (Malik, 1997). Frequency-domain analyses decompose NN-intervals into sinusoidal waveforms based on preestablished frequency bandwidths (Berntson, Quigley, & Lozano, 2007). Standard adult HRV frequency bandwidths are defined as High Frequency (HF 0.15–0.40 Hz), Low Frequency (LF 0.04–0.15 Hz), Very Low Frequency (VLF 0.0033–0.04 Hz), and Ultra-Low Frequency (ULF < 0.0033 Hz; Berntson et al., 1997; Task Force, 1996). Studies using pharmacological blockade demonstrate that HF (or respiratory sinus arrhythmia, RSA) chiefly reflects parasympathetic and respiratory influences (Akselrod et al., 1981; Berntson, Cacioppo, & Quigley, 1993; Cacioppo et al., 1994; Chen & Mukkamala, 2008). Physiological mechanisms underlying LF, VLF, and ULF have also been studied using pharmacological blockade (e.g., Akselrod, Eliash, Oz, & Cohen, 1985; Akselrod et al., 1981; Cacioppo et al., 1994), but are less well established. VLF and ULF are rarely reported in pediatric studies. LF:HF ratio can also be derived; however, its interpretation as reflecting autonomic balance is frequently debated, largely because both autonomic branches contribute to LF, and autonomic activity is not exclusively reciprocal (e.g., Berntson, Cacioppo, & Quigley, 1991; Reyes del Paso, Langewitz, Mulder, van Roon, & Duschek, 2013). As such, LF:HF more likely reflects overall autonomic modulation.

Decreased SDNN, HF, LF, VLF, and LF:HF are indicative of poor health and worse outcomes among conditions such as cardiac arrhythmia, obesity, hypertension, Type 1 and Type 2 diabetes, and psychological disorders in both adult (e.g., Boer-Martins et al., 2011; Henry, Minassian, & Paulus, 2010; Schmid, Schönlebe, Drexler, & Mueck-Weymann, 2010; Thayer & Sternberg, 2006) and child samples (e.g., Akinci, Celiker, Baykal, & Tezic, 1993; Baharav, Kotagal, Rubin, Pratt, & Akselrod, 1999; El-Sheikh & Hinnant, 2011; Martini et al., 2001; Xie et al., 2013). Considering its implications for health, HRV must be accurately measured, analyzed, and interpreted to minimize erroneous and potentially detrimental conclusions. Meaningful inferences about a relation between temporal changes in HRV and health require evidence for test-retest reliability in adults and children; however, the psychometrics of pediatric HRV are not well established.

HRV Measurement, Analysis, and Reporting: Implications for Test-Retest Reliability

Two committee reports have provided guidelines for HRV measurement and analysis (Berntson et al., 1997; Task Force, 1996). While these guidelines are frequently cited in the HRV literature, wide variability in study methodology remains. Further, available guidelines do not reflect theoretical and methodological advances in HRV measurement within the last two decades (e.g., Cerutti, Goldberger, & Yamamoto, 2006; Denver, Reed, & Porges, 2007), and do not account for distinctions in pediatric HRV measurement (e.g., Bar-Haim, Marshall, & Fox, 2000). Pertinent methodological considerations for HRV test-retest reliability are reviewed below and are conceptually organized into four categories: (i) study protocol; (ii) sample characteristics; (iii) ECG signal acquisition and preliminary processing; and (iv) HRV analyses. These categories coincide with milestones of salient methodological decisions.

Study Protocol

Study protocol methodological decisions include length of the follow-up period or test-retest interval; use of a standardized recording protocol; and, recording time of day and related waking state. Declining reliability with longer test-retest intervals is not unexpected (Cohen & Swerdlik, 2002), and has been observed among young adults (e.g., Cipryan & Litschmannova, 2013) and among infants and youth. For example, Perry and colleagues (2013) observed declining HF reliability in toddlers across 1 to 2 year follow-up intervals ($r = 0.52$; $r = 0.34$, respectively). El-Sheikh and Hinnant (2011) similarly observed declining HF reliability in children across 1 to 2 to 3 year follow-up intervals ($r = 0.63$; $r = 0.54$; $r = 0.32$, respectively). Another factor that can influence HRV reliability is the extent of protocol standardization. Namely, test-retest reliability can be optimized by restricting participant behaviors that affect autonomic physiology (e.g., exercise, caffeine intake, postural changes), and by standardizing the ECG recording protocol across all participants and study assessments (cf., Jennings et al., 1981). Notedly, researchers strive to balance test-retest reliability with ecological validity, which may diminish with more rigorous study protocols.

Guidelines do not specify an ideal time of day for recording ECG, leaving this study protocol decision to the discretion of researchers. Circadian variations in HRV have been observed in adults (e.g., Armstrong, Kenny, Green, & Seely, 2011), children (e.g., Massin, Maeyns, Withofs, Ravet, & Gerard, 2000), and infants as early as 7–12 weeks old (Hoppenbrouwers et al., 2012). Sympathetic dominance peaks just after awakening and withdraws during the day, while parasympathetic, or vagal, dominance becomes augmented throughout the night, reaching its peak before awakening (Guo & Stein, 2003; Huikuri et al., 1990). Pediatric studies report mean-level changes in HF, VLF, and LF:HF across day, night, and 24-hr recordings (Faulkner, Hathaway, & Tolley, 2003); however, few studies have examined how recording time influences HRV reliability. Further, waking state (awake vs. napping) during daytime hours is a relevant issue for studies conducted with infants and toddlers.

Sample Characteristics

Sample characteristics relevant to HRV reliability include age or developmental span; study exclusion criteria; and, participant biological sex. Sample heterogeneity can result from recruiting participants across a wide age range (e.g., 8–18 years), or with poorly defined, or lack of fidelity to, exclusion criteria (e.g., medical illness/condition, medication use). Evidence for HRV reliability in a homogenous sample is necessary to establish the validity of HRV measures obtained from that sample. Heterogeneous samples that span wider age ranges may actually limit reliability attributable to normal HRV changes across developmental periods. For example, lower test-retest HF reliability was observed for samples spanning from 2 months to 5 years of age ($r = .30$, Bornstein & Suess, 2000b) and from 5 to 14 years of age (Spearman's $\rho = .26$; Gentzler, Rottenberg, Kovacs, George, & Morey, 2012). Relatedly, lower 2-year test-retest HRV reliability has been reported among toddlers, compared to children (e.g., 3–5 years: $r = 0.34$, Perry et al., 2013; 8–10 years: $r = 0.54$, El-Sheikh & Hinnant, 2011). Study exclusion criteria are often less rigorous or unspecified among the pediatric HRV literature, relative to the adult literature (e.g., Kennedy, Rubin, Hastings, & Maisel, 2004; Rigterink, Fainsilber Katz, & Hessler, 2010). Medications (e.g., antiarrhythmic agents, beta-blockers) that have been associated with lower mean-level LF, HF, LF:HF, and SDNN in adults (Penttilä, Kuusela, & Scheinin, 2005; Shaffer & Combatalade, 2013; Task Force, 1996) have been rarely examined in relation to HRV, or HRV reliability, among children (e.g., Buchhorn et al., 2012). Conversely, several acute and chronic illnesses (e.g., obesity, diabetes) are associated with declines in sympathovagal balance among clinical samples of adults and children, relative to nonclinical samples (e.g., Kochiadakis et al., 1997; Latchman et al., 2011; Martini et al., 2001; Sandercock, Bromley, & Brodie, 2005). Thus, sample homogeneity is relevant for reliability.

Sex differences in HRV have been observed in adults and children, although results are often inconsistent. Adult males typically display greater sympathetic dominance (e.g., higher LF, LF:HF), while females typically display greater parasympathetic dominance (e.g., higher HF, rMSSD; Antelmi et al., 2004) and higher HRV reliability (Sookan & McKune, 2012). Conversely, while boys have displayed increased mean-level NN and SDNN relative to girls (Faulkner et al., 2003; Silvetti, Drago, & Ragonese, 2001), male sex has also been associated with higher HF among 9- and 11-year-olds (El-Sheikh, 2005). There may be sex-specific shifts in sympathovagal balance during childhood development related to changes in sex hormone concentrations (e.g., testosterone, estrogen), which in turn, exert distinct, sex-specific effects on blood pressure and heart rate (Spear, 2000). Evidence suggests that sex differences in HRV reliability may exist among both adults and children; however, pediatric studies examining biological sex differences in HRV reliability are rare.

Electrocardiogram (ECG) Signal Acquisition and Preprocessing

Electrocardiogram signal acquisition and preprocessing decisions include length of the acclimatization period; recording posture; ECG sampling rate; and, signal filtering. Allowing participants to acclimate to their surroundings prior to ECG recording is recommended by HRV measurement guidelines (Berntson et al., 1997; Task Force, 1996). Resting HRV measures index baseline autonomic cardiovascular control and serve as a reference from which to compare change. However, participants instructed to rest quietly

during the baseline recording may not actually be fully at rest (i. e., anxious, humming quietly, ruminating) if they did not have time to adequately habituate. A prerecording acclimatization period following the application of electrodes and prior to baseline recording helps participants to familiarize with their surroundings and the physiological sensors, and reduces stress-related changes in physiological activity prior to data collection; this may be especially true among infants and young children. Sharpley (1993) noted that a 15 min acclimatization period duration may be sufficient among adults, but advocated for durations that are optimal for each participant. Including a prerecording acclimatization period may augment HRV reliability by reducing unsystematic variation in resting heart rate activity; however, acclimatization periods distinct from baseline are uncommon, or are not routinely reported, among pediatric studies.

Parasympathetic dominance is often observed in seated or supine postures, while sympathetic dominance is often observed in standing or head-up-tilt postures (Cacioppo et al., 1994; Kleiger et al., 2005); thus, HRV, and HRV reliability, are influenced by postural position. HF reliability, but not LF reliability, was notably higher among adults measured across 24 months in supine versus standing postures ($ICC_{HF} = 0.89$ vs. 0.79 ; $ICC_{LF} = 0.81$ vs. 0.79 ; Kowalewski & Urban, 2004); similar results were obtained among children for heart rate and SDNN measured across 2 weeks ($ICC_{HR} = 0.78$ vs. 0.65 ; $ICC_{SDNN} = 0.79$ vs. 0.69 ; Dietrich et al., 2010), and among adolescents for HF and LF measured across 1 year ($r_{HF} = 0.37$ vs. 0.25 ; $r_{LF} = 0.37$ vs. 0.31 ; Mezzacappa et al., 1997). Considering these findings, sympathetic modulation may be less affected by postural changes than parasympathetic modulation. Shifts in sympathovagal balance across childhood development (Spear, 2000) highlight a salient consideration when examining the influence of recording posture on pediatric HRV reliability.

Electrocardiogram sampling rate may be pertinent to the reliability and accuracy of HRV data. Guidelines recommend 500–1,000 Hz as optimal for HRV sampling, although 250 Hz may also be acceptable (Berntson et al., 1997; Task Force, 1996). Studies that have empirically examined error variation attributable to ECG sampling rate are largely based on small sample sizes (e.g., $N = 1-5$), simulated comparisons, or evaluation of lower sampling rates (e.g., 64 Hz, 128 Hz, 256 Hz; cf., Merri, Farden, Mottley, & Titlebaum, 1990; Singh, Vinod, & Saxena, 2004; Wittling & Wittling, 2012). Other studies have suggested that sampling rates above 500 Hz may not have incremental utility (cf., Riniolo & Porges, 1997; Singh, Singh, & Banga, 2014). Sampling rates that are too slow, or even too fast, may reduce R-wave timing precision, contribute unwanted variance, and limit reproducibility. The ideal sampling rate also may be sample-specific (Merri et al., 1990; Singh et al., 2004) and require adjustment to account for temporal changes in age and weight, for example.

Few studies report the proportion of data excluded due to artifacts. This sharply contrasts with guideline recommendations and restricts knowledge about the integrity, quality, and interpretability of HRV data. The effects of poor continuity and stationarity of ECG signals due to artifacts (e.g., poor electrode adhesion, movement, aberrant heartbeats) are known and have been widely considered using real and simulated HRV data (Berntson & Quigley, 1990; Berntson & Stowell, 1998; Kim, Lim, Kim, & Park, 2007; Salo, Huikuri, & Seppänen, 2001). Thus, manually editing ECG data prior to HRV analysis is highly relevant for HRV

reliability. Digital automated filtering of noise and overlapping frequency components (i.e., spectral leakage) is also pertinent. Porges' (1985) moving polynomial filter is commonly used in pediatric HRV studies to remove background heart rate trends and may eliminate the need to statistically control for heart rate. HRV reliability in the context of different (or absent) automated filtering algorithms has been examined among adults (Lee & Chiu, 2010; Singh et al., 2004), but rarely in children.

Heart Rate Variability Analyses

Heart rate variability analyses include data reduction decisions pertaining to length of the recording duration analyzed; epoch length; and, frequency bandwidth selection. These HRV analyses occur post-ECG signal collection and prior to commencing statistical analyses. Choosing the ECG recording duration to be analyzed for HRV is pertinent for capturing the extent of short- and long-term variability within a signal, and for the reliability of resulting measures. Guidelines specify that ECG be recorded for at least 1 min to assess HF, 2 min for LF, and ~ 50 min for VLF. However, as 2 min may not be enough to derive reliable estimates of LF (Heathers, 2014; Kleiger et al., 2005), 5 min recording durations are generally accepted as a standard in short-term studies (Berntson et al., 1997; Task Force, 1996). Among adults, LF reliability was higher when derived from 5 min, compared to 2.5 min recordings (ICC = 0.82 vs. 0.78; Marks & Lightfoot, 1999). Among infants, recordings 3 min produced more reliable measures of HF (Richards, 1995). It is plausible that reliability of some measures may decrease with longer recording durations (e.g., NN), while others increase (e.g., VLF) or remain unaffected (e.g., LF:HF). Comparing identical length recording durations is relevant when examining HRV reliability (Dalla Pozza et al., 2006).

An ECG recording can be analyzed for HRV using the mean of a single analytical epoch or by analyzing the mean of multiple shorter epochs. For example, a 30 min recording can be analyzed as thirty 1 min epochs, six 5 min epochs, or one 30 min epoch, among several other epoch lengths. Deriving HRV using multiple epochs may reduce the impact of uncorrected artifacts or slower heart rate trends (Izard et al., 1991; Salo et al., 2001), but may introduce a selection bias, or a loss of variability that oscillates at periods longer than the analyzed epoch (Berntson et al., 1997, 2007; Porges & Byrne, 1992). McNamara and Aboy (2006) demonstrated that adult HF was reproducible across epochs ranging from 10 s to 10 min, while LF required at least 10 min epochs. Richards (1995) observed that infant HF was more reliable when a 75 s recording was analyzed in fifteen 5 s epochs, rather than five 15 s epochs. Thus, epoch selection is germane to pediatric HRV reliability and should be explicitly reported to enhance the interpretability of results.

Heart rate variability frequency bandwidths in pediatric samples differ from those defined for adults, due to young children's higher respiration rate. Namely, the HF bandwidth should be modified to include the respiratory frequency of the pediatric sample (0.3–1.30 Hz infants; 0.24–1.04 Hz young children; Bar-Haim et al., 2000; Fox & Porges, 1985). If adult bandwidths are erroneously applied to pediatric frequency-domain measures, which is not uncommon, resulting data will inaccurately estimate autonomic cardiovascular control and distort subsequent interpretations. Spectral frequency bands should be carefully considered and explicitly reported to enhance the interpretability of results. Frequency bandwidth

selection is likely pertinent to HRV test-retest reliability; however, few studies have specifically examined this, especially among children and adolescents.

Current State of Knowledge and Study Aims

There is considerable evidence for moderate to excellent reliability of HRV measures during controlled resting conditions in healthy adult participants (cf., Sandercock et al., 2005, for systematic review). Relatively less evidence is available from studies with infants, children, and adolescents. Reliable HRV measurement requires careful consideration of multiple methodological factors (e.g., Jarrin, McGrath, Giovannello, Poirier, & Lambert, 2012), many of which are detailed in frequently cited HRV measurement guidelines. There is a notable lack of conformity to these guidelines, which increases potential for error upon replication, limits the interpretability of results, and makes it challenging to draw meaningful comparisons about HRV reliability across different studies. Considering the relative lack of knowledge regarding pediatric HRV psychometrics reviewed above, the overarching aim of this meta-analysis was to systematically examine and empirically synthesize available evidence for test-retest reliability of time- and frequency-domain HRV measures from infants, children, and adolescents. Methodological decisions that could influence HRV reliability were examined as possible moderators, including (a) study protocol decisions (i.e., study follow-up length; consistent time of day recording; waking status), (b) sample characteristics (i.e., explicit exclusion criteria; proportion of male participants); (c) ECG signal acquisition and preprocessing settings (i.e., prerecording acclimatization period; recording posture; sampling rate; filtering algorithm); and (d) HRV analytical decisions (i.e., analyzed recording duration; epoch lengths; frequency bandwidth selection). Finally, given reported developmental differences, HRV reliability was compared between younger versus older children.

Method

Literature Search Strategy

An electronic literature search was performed within PsycINFO and PubMed (MedLine) databases from January 1, 1970 to December 31, 2014, using terms regarding pediatric populations (e.g., infants, children, adolescent), HRV measures (e.g., HF, RSA, SDNN), and psychometrics (e.g., reliability, test-retest, reproducibility). This search yielded 579 nonredundant journal articles; review of their titles indicated 347 were possibly relevant due to mention of HRV measures. Next, the abstracts of these 347 articles were reviewed and selected for follow-up if they suggested that HRV was measured at least twice from nonclinical pediatric or age-unspecified samples; 88 articles met this criterion. Then, ascendancy and descendancy approaches (i.e., backward and forward citation searches of 88 initial articles) identified 60 additional articles. One other article was obtained after sending solicitation letters to prominent authors for additional or unpublished data. Thus, a total of 149 nonredundant articles were retrieved for full review (see Figure 1).

Article Inclusion and Exclusion

Articles were included in the meta-analysis if: (a) time- and frequency-domain HRV measures were obtained at least twice during separate, resting baseline conditions; (b) participants were nonclinical infants, children, or adolescents with a sample mean age less than or equal to 18.0 years; and (c) sufficient data were presented for effect size computation (i.e., reporting only means was insufficient). Articles were excluded based on an *a priori* hierarchy: (i) not an empirical study ($k = 1$); (ii) sample size less than 10 ($k = 4$); (iii) sample of only clinical (e.g., diabetes, anxiety disorder) or special-population individuals (e.g., preterm infants, athletes; $k = 9$); (iv) fewer than two baseline HRV recordings were obtained ($k = 53$); (v) HRV data were obtained solely during overnight sleep ($k = 3$); (vi) NN-interval data not recorded from continuous ECG ($k = 4$); (vii) HRV data were averaged across rest and task conditions ($k = 3$); (viii) baseline HRV data not reported ($k = 22$); (ix) data were redundant with another included study ($k = 1$; more complete article retained). In total, 49 articles met final inclusion criteria (see Appendix and Figure 1).

Reliability of Article Selection and Coding

A random sample of 10% of titles, abstracts, and articles were blindly re-coded after a 4-month period by the original (OW) and an independent (JL) rater. Excellent intra- and inter-rater agreement for title selection were obtained (Cohen's kappa, $\kappa = 1.0$, for both). Excellent intra-rater ($\kappa = 0.92$) and good inter-rater agreement ($\kappa = 0.8$) for abstract selection were obtained. Abstract selection discrepancies were resolved through discussion. Excellent intrarater reliability was obtained for categorical ($\kappa = 0.93$) and continuous coding decisions (Intraclass Correlation, ICC = 1.0 [1.0, 1.0]). Good inter-rater reliability was obtained for categorical ($\kappa = 0.83$) and continuous coding decisions (ICC = 0.96, [0.93, 0.98]). Excellent intra- and inter-rater agreement for effect size selection were obtained (ICC = 1.0, [1.0, 1.0]).

Article Coding and Data Extraction

Study follow-up length was coded in months. If demographic information (e.g., age, weight) was sex-stratified, a weighted mean was calculated. Demographic variables (mean participant age [years], sex [percent male], ethnicity [African American, European American, Hispanic, Asian, Mixed, Other]), and anthropometrics were coded (height [in], weight [lbs], BMI [kg/m^2], blood pressure [mmHg], puberty [Tanner stage]). Explicit participant exclusion criteria were coded dichotomously (yes/no) within the following *a priori* categories: (a) prescription medication use; (b) medical illness or condition (e.g., chronic illness, diabetes); (c) mental health diagnosis (e.g., cognitive, intellectual, behavioral disorder); and (d) anthropometric characteristics (e.g., low birth weight, being obese). Study location was coded (university/hospital laboratory; school; home; nursery). Time of ECG recording was coded (morning 06:00–11:59; afternoon 12:00–17:59; evening 18:00–23:59). ECG sampling rate was coded in Hz. ECG recording posture was coded ([1] seated, [2] supine, [3] standing) and “changed” was coded when participants assumed different postures between study assessments; supine was assumed among studies with infants, unless otherwise specified. Durations for prerecording acclimatization period, total ECG recording, baseline ECG recording, analyzed recording, and epoch were coded in minutes. A

dichotomous code (yes/no) indicated whether: (a) participants were required to remain awake for the entire recording duration; (b) study location, recording time, posture, acclimatization period, and ECG recording durations were identical across all follow-up assessments; (c) manual editing and digital automated filtering algorithms (e.g., Hanning/Hamming window, polynomial window) were applied to ECG data; (d) HRV frequency bands were age-appropriate (e.g., HF: 0.24–1.3 Hz infant/toddler, 0.15–0.40 Hz child/adolescent), as per available guidelines and recommendations (Bar-Haim et al., 2000; Fox & Porges, 1985; Task Force, 1996); and (e) a single epoch was used to define the entire analyzed signal. “Unmentioned” was coded if a target variable was not reported.

Baseline HRV was defined as ECG-derived HRV measures that were prior to and distinct from any experimental study condition (e.g., stressor task, tilt, exercise). Baseline time- and frequency-domain HRV measures were coded using *a priori* categories. Heart period and mean NN-interval (ms) were coded as NN. SDNN and “heart period standard deviation” (ms) were coded as SDNN. rMSSD was coded (ms). HF in absolute (ms^2) and normalized units (n.u.), respiratory sinus arrhythmia (s, ms), and “vagal tone” (lnms^2) were coded as HF. LF in absolute (ms^2) and normalized units (n.u.) was coded. LF:HF ratio was also coded.

Statistical Analysis

Effect Size Calculation and Management—Fisher’s Z , which ranges from $-\infty$ to $+\infty$ and can be interpreted similar to a correlation coefficient, was selected as the standardized common effect size. Intra-class correlations (ICC), as well as Pearson and Spearman correlations, were converted to Fisher’s Z using Fisher’s variance stabilizing transformation (Rosenberg, Adams, & Gurevitch, 2000; Rosenthal, 1994). F -ratios and unstandardized beta coefficients were converted to r and then to Fisher’s Z (Rosenberg et al., 2000; Rosenthal, 1994). Exact p values, when no other test statistic was available, were converted to a standard normal deviate (Z -score), then to r , and then to Fisher’s Z (Rosenberg et al., 2000). Studies reporting effect sizes derived from > 2 follow-up assessments were coded for both the entire study follow-up length (i.e., one effect size per HRV variable, per study) and shorter, nonoverlapping follow-up intervals (e.g., 2–4 weeks, 4–6 weeks; see section Selection of Effect Sizes below).

Selection of Effect Sizes: Effect sizes were coded for available data reported, thus yielding multiple effect sizes per study. Multiple effect sizes were reported due to HRV variables (e.g., HF, LF), follow-up intervals (e.g., 2–4 wks, 4–6 wks, 2–6 wks), and postures (e.g., supine, seated). We employed a conservative approach including only one effect size per HRV variable per study (Total: 93 effect sizes, $M = 1.90$ effect sizes per study; Infants/Toddlers: 47, $M = 1.52$; Children/Adolescents: 46; $M = 2.56$); the longest follow-up interval, in one posture (selected hierarchically: seated, supine, standing), was retained. To maximize power, redundant effect sizes were permitted only for the corresponding moderator analyses for follow-up interval and recording posture.

Meta-Analytic Strategy: Analyses were age-stratified due to noted methodological differences between studies with younger and older children (cf., Williams et al., 2012). Infant/toddler studies were those with a mean sample age less than 5 years ($k = 31$); child/

adolescent studies were those with a mean sample age 5 years or greater, but less than or equal to 18 years ($k = 18$). A fixed-effects meta-analytic model was chosen because the categorical variables coded essentially capture all possible options represented in the extant HRV literature (e.g., sitting vs. supine vs. standing posture; consistent vs. inconsistent recording durations), rather than categories sampled from a larger population of possible options (Rosenberg et al., 2000). A fixed-effects model is also consistent with an earlier meta-analysis of heart rate and blood pressure reproducibility in adults (Swain & Suls, 1996). Thus, a random-effects model was not deemed appropriate given the examined data.

An effect size was calculated for each HRV variable separately. An analysis of the heterogeneity statistic (Q_T), which measured the variation for the included effect sizes, was conducted for each meta-analytic model. Significant Q_T indicated that the included effect sizes had a heterogeneous distribution and informed whether additional moderator analyses were warranted (Rosenberg et al., 2000). Continuous (i.e., slope) and categorical (i.e., Q_M) methodological variables were examined in subsequent moderator analyses (Rosenberg et al., 2000). Bootstrap methods (1,000 samples) produced robust nonparametric confidence interval estimates around each effect size (Rosenberg et al., 2000). Orwin's Fail-Safe numbers addressed possible publication bias by estimating the number of missing, unpublished, or nonsignificant studies needed to make the overall effect size negligible or not different from zero. Analyses were performed using MetaWin 2 (Sinauer Associates, 2000) and forest plots were graphed using Forest Plot Viewer (Boyles, Harris, Rooney, & Thayer, 2011).

Results

Study Participant and Recording Characteristics

A total of 49 studies ($N = 5,170$) were included in the present meta-analysis. Infant/toddler studies ($k = 31$; $N = 3,329$) and child/adolescent studies ($k = 18$; $N = 1,841$) had sample sizes ranging from 10 to 441 ($Mdn = 90, 60$, respectively), and about half of participants in both groups were male (see Table 1). Overall, few studies reported *a priori* sample exclusion criteria or participant characteristics (e.g., height, weight, study attrition), which precluded certain planned moderator analyses. Mean ECG recording duration analyzed for HRV in infant/toddler studies was 4.42 min ($SD = 4.29$), and in child/adolescent studies it was 3.74 min ($SD = 1.91$; see Table 1).

Overall Summary Analyses

Baseline HRV exhibited moderate test-retest reliability across both age groups. To compare the overall reliability between the age groups, a mean effect size was calculated across HRV variables to yield one effect size per study. Overall reliability was significantly lower for infant/toddler studies compared to child/adolescent studies ($Z = 0.42$ vs. 0.62 ; $Q_M = 32.49$, $p < .001$). All effect sizes, except LF:HF, were significantly heterogeneous, indicating that further moderator analyses were warranted. Mean summary effect sizes are presented in Table 2. Reliability forest plots are presented in Figures 2–4.

Moderator Analyses

Continuous (Table 2) and categorical (Tables 3 and 4) moderator variables were examined within the following categories: Study protocol (*test-retest follow-up interval length; standardized recording protocol; time of day, awake*), sample characteristics (*age, developmental period; participant exclusion criteria; sex, proportion male*), ECG signal acquisition and preprocessing (*acclimatization period; posture; sampling rate; artifact editing, filtering*), and HRV analyses (*recording duration analyzed; epoch length; frequency bandwidth selection*). Moderator analyses were largely restricted to NN and HF among infant/toddler studies, and HF and LF among child/adolescent studies, due to limited reported data.

Study Protocol—Longer study follow-up length was not associated with NN or HF reliability among infant/toddler studies. In contrast, longer study follow-up length was significantly associated with lower SDNN ($B = -0.057, p = .052$), rMSSD ($B = -0.553, p = .011$), HF ($B = -0.005, p = .008$), and LF ($B = -0.031, p = .007$), but not LF:HF reliability, among child/adolescent studies. Significant heterogeneity remained for all HRV variables in both age groups, except LF:HF which was homogenous, indicating that further or nested moderator analyses are merited. Moderator results were consistent for conservative and redundant effect size selection approaches; only conservative approach statistics reported for parsimony.

Across both age groups, minimal reporting of ECG recording time precluded an examination of whether time of recording ECG (e.g., morning vs. afternoon) moderated HRV reliability. However, HF reliability was significantly higher among infant/toddler studies that recorded exclusively awake ECG data, compared to recordings that contained some sleep (i.e., napping; $Z = 0.44$ vs. 0.21 ; $Q_M = 6.35, p = .012$).

Sample Characteristics—Significantly higher NN reliability was found among infant/toddler studies that excluded participants based on prescription medication use ($Z = 0.61$ vs. 0.23 ; $Q_M = 14.06, p < .001$) and medical illness/condition ($Z = 0.47$ vs. 0.26 ; $Q_M = 5.26, p = .022$); similar results were not found for HF. Significantly higher HF ($Z = 0.67$ vs. 0.55 ; $Q_M = 6.63, p = .010$) and LF reliability ($Z = 0.80$ vs. 0.54 ; $Q_M = 4.17, p = .041$) were found among child/adolescent studies that excluded participants based on medical illness/condition. Nested moderator analyses may further address remaining heterogeneity across most effect sizes.

Larger sample size was significantly associated with higher rMSSD reliability ($B = 0.051, p = .043$), but lower HF ($B = -0.001, p < .001$) and LF reliability ($B = -.003, p < .001$) among child/adolescent studies. Older sample age was significantly associated with higher NN ($B = 0.153, p < .001$) and HF reliability ($B = 0.055, p < .001$) among infant/toddler studies, and higher rMSSD ($B = 0.173, p < .001$) among child/adolescent studies.

Having a greater proportion of male participants was significantly associated with higher NN reliability ($B = 2.141, p = .017$) among infant/toddler studies; no relation was observed for HF reliability. Conversely, a greater proportion of males was significantly associated with higher SDNN ($B = 1.088, p < .002$) and rMSSD reliability ($B = 2.085, p < .001$), and

lower HF ($B = -0.334$, $p = .039$) and LF reliability ($B = -0.430$, $p = .036$) among child/adolescent studies; no relation was observed for LF:HF reliability. Further nested analyses are needed to address remaining heterogeneity.

Electrocardiogram Signal Acquisition and Preprocessing—Prerecording acclimatization period was associated with significantly higher NN reliability ($Z = 0.62$ vs. 0.21 ; $Q_M = 18.31$, $p < .001$) among infant/toddler studies, and higher LF reliability ($Z = 0.79$ vs. 0.39 ; $Q_M = 10.32$, $p = .001$) among child/adolescent studies. No significant relations were observed for HF reliability in either age group. The duration of the acclimatization period was not associated with reliability of any HRV measures. Significant heterogeneity remained for NN reliability among infant/toddler studies and HF reliability in both age groups; LF reliability was homogenous among child/adolescent studies.

Electrocardiogram recording posture was significantly associated with NN ($Q_M = 9.42$, $p = .002$) and HF reliability ($Q_M = 18.69$, $p < .001$) among infant/toddler studies; reliability was highest in seated recordings and lowest when recording posture changed across follow-up assessments. Recording posture was not associated with HRV reliability among child/adolescent studies; however, HF and LF reliability were relatively lower when recorded in standing posture. Significant heterogeneity among seated posture measures suggests further or nested moderated analyses are warranted in both age groups. Moderator analyses for posture were only conducted using the redundant effect size selection approach, to permit sufficient data for comparison.

The paucity of studies reporting ECG sampling rate precluded adequate examination as a moderator of HRV reliability. Gross binary coding (> 500 Hz vs. unmentioned) was used to salvage available information to conduct cursory categorical moderator analyses. Child/adolescent studies reporting the sampling rate > 500 Hz yielded significantly higher HF reliability ($Z = 0.64$ vs. 0.53 ; $Q_M = 3.65$, $p = .056$) and lower SDNN reliability ($Z = 0.73$ vs. 1.26 ; $Q_M = 6.53$, $p = .011$), than studies that did not report the sampling rate. Reliability did not differ among infant/toddler studies when comparing the sampling rate > 500 Hz, versus not reported or < 500 Hz.

Infant/toddler studies that reported manual editing of ECG data, versus not reporting this detail, yielded significantly higher NN reliability ($Z = 0.38$ vs. 0.14 ; $Q_M = 4.99$, $p = .025$); while those reporting an automated filtering algorithm, versus not reporting this detail, yielded significantly higher HF reliability ($Z = 0.45$ vs. 0.35 ; $Q_M = 4.00$, $p = .045$). Reliability did not differ among child/adolescent studies that reported editing and filtering of ECG data, with the exception of SDNN where manual editing was associated with lower reliability ($Z = 0.86$ vs. 1.26 ; $Q_M = 4.08$, $p = .043$; see Table 4).

Heart Rate Variability Analyses—Longer recording duration analyzed was significantly associated with lower NN reliability ($B = -0.028$, $p = .026$), but was not associated with HF reliability among infant/toddler studies. Conversely, longer recording duration analyzed was significantly associated with higher SDNN ($B = 0.130$, $p < .001$), rMSSD ($B = 0.201$, $p < .001$), HF ($B = 0.066$, $p = .004$), and LF reliability ($B = 0.059$, $p = .040$) among child/adolescent studies; no relation was observed for LF:HF reliability. Remaining effect size

heterogeneity, mostly among measures from child/adolescent studies, suggests that further or nested moderator analyses are required. Analyzing identical ECG recording durations, versus differing durations, across both study assessments was associated with significantly higher NN ($Z = 0.38$ vs. 0.12 ; $Q_M = 5.75$, $p = .017$) and HF reliability ($Z = 0.45$ vs. 0.34 ; $Q_M = 5.35$, $p = .021$) among infant/toddler studies. As well, infant/toddler HF reliability was significantly higher when derived using the mean of two or more ECG epochs, compared to a single epoch ($Z = 0.46$ vs. 0.25 ; $Q_M = 5.41$, $p = .020$). These associations were not observed among child/adolescent studies. Duration of the analytical epoch (e.g., 1-min, 5-min) was not associated with HRV reliability.

Selecting an age-specific HRV frequency bandwidth consistent with guidelines and recommendations, compared to an alternative bandwidth, yielded no significant differences. Effect sizes for HF, LF, and LF:HF reliability were relatively larger among studies that selected an age-specific frequency bandwidth, compared to an alternative bandwidth ($Z_{HF} = 0.64$ vs. 0.49 , $Z_{LF} = 0.64$ vs. 0.39 , $Z_{LF:HF} = 0.74$ vs. 0.39); yet, significant heterogeneity remained.

Study Quality

An index of study quality was coded based on 10 study characteristics: (1) having exclusion criteria for medical illness/condition (26.5% of studies had this characteristic); (2) having exclusion criteria for medication use (12.2%); (3) having sample size ≥ 50 (44.9%); (4) specified at least one prerecording participant instruction (e.g., overnight fast, no exercise for 12 hr; 18.4%); (5) included a prerecording acclimatization period (36.7%); (6) analyzed baseline ECG recording duration ≥ 3.5 min (51%); (7) analyzed recording duration ≥ 3.5 min (50%); (8) reported both ECG recording hardware and HRV analysis software (81.6%); (9) reported both manual editing and digital automated filtering of ECG data (49%); and, (10) reported using recommended HRV frequency bands for sample age (49.0%). Correlation analyses with one effect size per study (i.e., mean effect size averaged across HRV variables) revealed study quality was not associated with HRV reliability among infant/toddler studies ($M = 3.97$ out of possible 10 characteristics, $SD = 2.07$; $r = -.033$, $p = .861$), but was positively associated among child/adolescent studies ($M = 4.67$, $SD = 1.68$; $r = .489$, $p = .040$).

Discussion

The overarching aim of this systematic review was to comprehensively examine test-retest reliability of pediatric HRV using meta-analytic techniques. Effect sizes for reliability of time- and frequency-domain HRV measures were typically in the moderate range across both age groups. Further, overall test-retest reliability was significantly higher in child/adolescent studies, compared to infant/toddler studies. While several moderating variables influenced HRV reliability, limited methodological reporting precluded several *a priori* planned moderator analyses.

Study Protocol

Longer study follow-up length was associated with decreased reliability for SDNN among infant/toddler studies, and decreased reliability of SDNN, rMSSD, HF, and LF among child/adolescent studies. This result is not unexpected given basic psychometric principles (cf., Cohen & Swerdlik, 2002). As example, studies lasting 3 years or longer had lower test-retest reliability than those only 2 weeks or shorter in duration. Age-related increases in HRV partly reflect increased autonomic complexity and offer a plausible explanation for decreasing reliability with increased follow-up length. The prominent use of a moving polynomial filter (Porges, 1985) to derive HF among infant/toddler studies may augment reliability by accounting for variance attributed to normal developmental decreases in heart rate activity over longer test-retest intervals (e.g., newborns studied across several months or years). Conversely, the predominant use of spectral analysis techniques (e.g., Fourier transformation) or the peak-valley method (e.g., El-Sheikh, 2005) to derive HF among child/adolescent studies, neither of which inherently accounts for heart rate variance, may explain the lower reliability of HF across longer test-retest intervals in this age group. These results suggest that an examination of long-term reproducibility of HF derived with and without a moving polynomial filter, compared to HF derived with and without statistical control of heart rate, should be conducted.

Due to limited reported data, this meta-analysis could not examine whether consistent ECG recording time of day across follow-up assessments (i.e., to account for potential circadian effects) improved reliability. However, other results from this meta-analysis indicated that HF reliability among infant/toddler studies, which analyzed exclusively awake ECG data, was higher compared to HF derived from daytime data containing some sleep (i.e., naps; unrestricted sleep). It is unknown whether the same infants/toddlers fell asleep at both assessments, which would reduce the replicability of initial recording conditions. Circadian variation in HRV across different age groups is important to consider when interpreting HRV (Ernst, 2014). As such, more research using 24-hr ECG data is necessary to examine whether HRV reliability varies depending on certain times of day (e.g., morning, afternoon, evening).

Sample Characteristics

Increased sample homogeneity, attributable to specified participant exclusion criteria, also influenced HRV. Higher NN reliability was observed among infant/toddler studies that reported exclusion criteria for medical illness/condition and prescription medication use, while higher HF and LF reliability was observed among child/adolescent studies that excluded for medical illness/condition. Results suggest that reducing variation attributable to prescription medication use or compromised health improves reliability. Considering these findings, future studies should examine the cross-sectional and longitudinal effects on HRV attributable to illnesses and medications common in pediatric populations to determine whether employing stringent exclusion criteria (e.g., specific diagnosis or medication) would augment HRV reliability.

Biological sex was a significant moderator of HRV reliability across both age groups, albeit in different ways. Studies with a greater proportion of male participants yielded higher NN

reliability among infant/toddler studies, and yielded higher SDNN and rMSSD, but lower HF and LF reliability among child/adolescent studies. These child/adolescent results are similar to adult findings of lower HRV reliability among males relative to females (e.g., Sookan & McKune, 2012). Current thinking is that sex differences in HRV result from greater sympathetic dominance among males relative to females; indeed, females often, but not always (e.g., El-Sheikh, 2005), produce higher mean-level HF and yield greater HF reliability relative to males. Biologically, sex differences in HRV reliability may be associated with hormonal differences related to pubertal development (e.g., sympathetic alterations in response to increased estrogen concentration; Kajantie & Phillips, 2006; Ordaz & Luna, 2012; Shirtcliff, Dahl, & Pollak, 2009). Sex differences in HRV reliability may become more pertinent as children advance toward sexual maturation (e.g., Patton & Vines, 2007; Silvetti et al., 2001). Jarrin and colleagues (2015) demonstrated that measures of gonadarche and adrenarche were negatively associated with rMSSD, pNN50, and HF, and positively associated with LF:HF. As such, more research into the association between pubertal status, related hormone levels, and HRV reliability among youth is needed to address this noted knowledge gap. Sex differences in pediatric HRV reliability could be clarified if authors report sex-stratified analyses.

Electrocardiogram Signal Acquisition and Preprocessing

Including a prerecording acclimatization period was associated with significantly higher NN reliability among infant/toddler studies, and significantly higher LF reliability among child/adolescent studies; further, effect sizes for each measure were homogenous. Studies that allowed participants to habituate yielded some of the largest effect sizes associated with a particular HRV measure, relative to studies without an acclimatization period. Results largely supported HRV guideline recommendations to include a prerecording acclimatization period (Berntson et al., 1997). Infant/toddler studies typically reported < 4 min of acclimatization, while child/adolescent studies typically reported < 10 min. No relationship was found between the acclimatization period duration and HRV reliability, suggesting even a brief habituation period prior to data collection may be sufficient to stabilize HRV and augment reliability. Examination of HRV reliability among groups of younger and older children may determine whether the “optimal” acclimatization duration changes across development.

Recording posture influenced HRV reliability. Specifically, infant/toddler NN and HF reliability were higher for measures obtained in seated postures and lowest when posture changed between study recordings. These observations are consistent with Young and Leicht (2011), who reported higher HF reliability among adults in seated versus standing postures. Child/adolescent HF and LF reliability were relatively lower among studies that recorded ECG in a standing posture, rather than seated or supine. Children who are asked to maintain a standing posture for a baseline ECG recording, even of only 3- to 5-min, may become more restless/fatigued compared to maintaining a seated posture; with increasing restlessness, the stationarity of heart rate signals will decline and compromise reproducibility. Given the results from blockade studies examining HRV and posture (e.g., standing vs. supine, Cacioppo et al., 1994), HF may be more reliably measured in seated or supine, relative to standing, postures when parasympathetic dominance is more readily

observed; while, LF may be more reliably measured in seated or standing postures when greater sympathetic, as opposed to parasympathetic dominance, is more readily observed. Further exploration of head-up-tilt paradigms in pediatric studies may improve our understanding of how postural position is related to HRV and its reliability.

The paucity of reported ECG sampling rates precluded a meaningful examination of its relation with HRV reliability. Child/adolescent studies that reported a sampling rate 500 Hz had higher HF reliability than studies that did not report this detail; however, only relative differences in reliability were found across the remaining analyses. Thus, conclusions about these results must be made with caution. Notably, only 23% ($k = 7$) of infant/toddler studies reporting HF explicitly detailed the ECG sampling rate. Lack of reporting ECG sampling rate illustrates a noted limitation in HRV reporting practices among the pediatric, relative to the adult, literature. Published guidelines (Berntson et al., 1997; Task Force, 1996) and recommendations from the wider HRV methodology literature (e.g., Hejjet & Roth, 2004; Merri et al., 1990; Riniolo & Porges, 1997; Singh et al., 2004) emphasize the importance of using an “adequate” ECG sampling rate for both interpretive and replication purposes. It is possible that an adequate sampling rate may be age- (e.g., infant vs. adult) or sample-specific (e.g., nonclinical vs. clinical; Singh et al., 2004). Thus, a threshold in R-wave detection may exist across different groups that require the sampling rate to be adjusted for a given characteristic. As newer ECG recording hardware becomes available with assorted sampling rates (e.g., 128 Hz, 500 Hz, 1,000 Hz), it will be pertinent that future studies examine the relation between sampling rate and HRV reliability.

Electrocardiogram signal preprocessing was associated with HRV reliability among infant/toddler studies. Manual artifact editing of ECG data was associated with significantly higher NN reliability, while digital automated filtering was associated with significantly higher HF reliability. Similar results were not found among child/adolescent studies. Artifacts and recording errors can significantly lengthen or shorten NN-intervals, inflate estimates of variability, and decrease reliability (Berntson & Stowell, 1998; Mulder, 1992). As such, lower reliability of NN with unedited data is not unexpected. The higher reliability of HF associated with filtering ECG data among infant/toddler, but not child/adolescent studies could be attributable to differences in the frequency of heart rate artifacts between the two groups, and/or the prominent use of a moving polynomial filter in infant/toddler studies (Porges, 1985), which may better account for extraneous variance attributed to artifacts (e.g., spectral leakage) or errors from poor data resolution (e.g., inadequate sampling rate). Compared to infant/toddler studies, child/adolescent studies reported using digital filters less frequently, but had greater variability among the filter types used. Results from infant/toddler studies demonstrated that carefully applied editing and filtering is associated with better HRV reliability in pediatric samples, and suggest that further examination of how editing and filtering NN-interval data affects pediatric HRV reliability is warranted.

Heart Rate Variability Analyses

Longer analyzed ECG recording duration was associated with lower NN reliability among infant/toddler studies, and higher SDNN, rMSSD, HF, and LF reliability among child/adolescent studies. Infant/toddler studies typically analyzed 2–5 min ECG recordings for

HRV, while child/adolescent studies typically analyzed 3 min recordings. Although these recording durations are less than the typical 5 min standard, results of this meta-analysis suggest that baseline HF can be reliably measured with recordings as short as 2 min. Increased LF reliability with longer analyzed durations is consistent with findings reported by Marks and Lightfoot (1999). Meta-analytic results also indicated that HF reliability among infant/toddler studies, but not child/adolescent studies, was relatively higher when measures were derived from the mean of multiple epochs. These results are consistent with earlier findings from infant (Richards, 1995) and adult studies (McNames & Aboy, 2006). Multiple 30 s epochs were typically reported across both age groups. HRV reliability may be influenced by the interaction between longer analyzed ECG recording durations and deriving HRV using multiple epochs; this could not be tested using nested moderator analyses due to limited available data. It will be important for pediatric researchers to further examine how HRV reliability is impacted by the direct and interactive effects of analyzed recording duration (e.g., 5 min, 10 min, 15 min) with epoch length (e.g., one 10 min epoch vs. two 5 min epochs). Careful consideration of recording duration and epoch length will help inform HRV analytical decisions for pediatric research and extend recommended HRV guidelines.

Selection of an age-specific HRV frequency bandwidth consistent with available guidelines and recommendations, as opposed to alternative bandwidths or unmentioned, did not alter HF reliability among infant/toddler studies, but was associated with relatively higher HF, LF, and LF:HF reliability, among child/adolescent studies. For example, child/adolescent studies that used the HF frequency bandwidth recommended for adults (i.e., 0.15–0.40 Hz; Berntson et al., 1997; Task Force, 1996) yielded comparatively higher reliability than studies that used wider (e.g., 0.20–1.00 Hz; Gentzler et al., 2012) or narrower (e.g., 0.20–0.30 Hz; Mezzacappa et al., 1997) bandwidths. Recommendations for pediatric HRV bandwidths only exist for HF among infants and toddlers up to age 4 years (i.e., 0.30–1.30 Hz for infants, Fox & Porges, 1985; 0.24–1.04 Hz for children under 4 years, Bar-Haim et al., 2000); these wider frequency bandwidths capture the influence of faster and more chaotic heart rate and respiratory patterns. Results suggested that the frequency bandwidths recommended for adults may not be unsuitable for deriving reliable measures of HF and LF among children and adolescents. Importantly, although the use of adult frequency bandwidths was associated with higher reliability, this result does not imply these HRV measures are necessarily valid. The lack of frequency bandwidth guidelines for use among children and adolescents is a noted gap in the literature. Thus, it is recommended that future research compare HRV measures across a range of frequency bands to determine an optimal bandwidth for frequency-domain measures obtained from children and youth. Systematically examining these measurement and methodological nuances will facilitate maximizing the psychometrics and interpretability of pediatric HRV data, and may culminate in the establishment of pediatric HRV recommendation guidelines.

Limitations

There are three limitations that merit consideration. First, failure to report key methodological details was pervasive in the pediatric HRV literature. Thus, the results must be interpreted in light of the coding decisions. The initial selection of which moderator

variables to code was informed by recommendations in HRV guidelines (Berntson et al., 1997; Task Force, 1996), the wider HRV methodology literature, and decisions used in a similar methodological review (Heathers, 2014). Due to the available data, several moderator variables of interest had to be collapsed into broader categories than initially planned. For example, rather than being able to precisely code the time of the ECG recording, reported information limited the coding to merely “same” versus “different” recording time categories. As another example, exact ECG sampling rates were often not reported, so coding was constrained to the categories of <500Hz, 500Hz, or unmentioned. These reporting limitations precluded examination of several moderators, as well as nested moderator analyses, that may have explained remaining heterogeneity or elucidated potential moderator interactions. Relatedly, the preponderance of studies only reported HF, thus, the reliability of time-versus frequency-domain measures could not be meaningfully compared.

Second, the present meta-analysis focused on short-term, daytime measures of time- and frequency-domain HRV obtained from baseline recordings. Future studies should examine pediatric HRV reliability among measures derived using longer ambulatory recordings (e.g., 24 hr), alternative analytical methods (e.g., geometric, wavelet, entropy), and reactivity protocols, each of which have been associated with health outcomes. Future efforts to replicate these findings using HRV data derived under different analytical contexts are warranted to determine the generalizability of these findings.

Third, examinations using fixed-effects meta-analytic models may limit the generalizability of these results compared to random-effects models, which assume that effect size variability is due to both random variance and sampling error (Rosenberg et al., 2000). *Post hoc* random effects analyses (not shown for parsimony) generally yielded smaller effect sizes, non significant heterogeneity analyses, and fewer significant moderators. However, the difference in effect sizes magnitude was typically small, moderate reliability was still generally observed, and most moderator variables remained significant across fixed- and random-effects models. Given that the coded variables captured the range of possible methodological options, a fixed-effects model was considered appropriate.

Conclusion

Taken together, this meta-analysis demonstrates that time- and frequency-domain HRV measures exhibit moderate and generally robust reproducibility over time, and provides initial empirical support for the application of certain adult guidelines to pediatric studies. Reliability among child/adolescent studies was significantly higher compared to infant/toddler studies. Greater reliability among infant/toddler studies was largely associated with sample exclusion criteria (medical illness/condition, prescription medications); prerecording acclimatization periods; recording ECG while participants are awake and seated; and, analyzing HRV using multiple epochs. Greater reliability among child/adolescent studies was largely associated with sample exclusion criteria (medical illness/condition); smaller proportion of males; longer recording durations; prerecording acclimatization periods; and, age-specific frequency bandwidth selection. Consistent with HRV guidelines, it is recommended that pediatric HRV study authors report precise protocol details (e.g., recording time of day), sample characteristics (e.g., anthropometrics, pubertal status), ECG

signal recording and preprocessing details (e.g., ECG sampling rate, artifact editing), and HRV analytical decisions (e.g., length of recording duration analyzed, epoch length). More rigorous reporting would facilitate research standardization, improve the interpretability and replicability of study findings, and permit more comprehensive meta-analytic comparisons in the future. Overall, these meta-analytic results have potential to make important empirical contributions by informing future researchers about the salient factors relevant to pediatric HRV methodology, extending current guidelines to include consideration of changes in HRV across childhood development (e.g., pubertal growth), and ultimately improving HRV psychometrics across studies of adult and pediatric health.

Acknowledgments

This work was partly supported by the Canadian Institutes of Health Research (J. McGrath OCO79897, MOP89886, MSH95353). Jinshia Ly independently coded articles to enable estimation of inter-rater agreement; the authors express their sincere gratitude. The authors also wish to thank Jean-Philippe Gouin and Lucie Bonneville for their insightful comments on earlier drafts of this manuscript, as part of O.W.'s Master's thesis. Preliminary findings were presented during an oral presentation at the 2015 Annual Scientific Meeting of the American Psychosomatic Society in Savannah, Georgia.

References

- Akinci A, Celiker A, Baykal E, Tezic T. Heart rate variability in diabetic children: Sensitivity of the time- and frequency-domain methods. *Pediatric Cardiology*. 1993; 14:140–146. DOI: 10.1007/BF00795641 [PubMed: 8415215]
- Akselrod S, Eliash S, Oz O, Cohen S. Hemodynamic regulation in SHR: Investigation by spectral analysis. *American Journal of Physiology*. 1985; 253:H176–83.
- Akselrod S, Gordon D, Ubel F, Shannon D, Barger A, Cohen R. Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science*. 1981; 213:220–222. DOI: 10.1126/science.6166045 [PubMed: 6166045]
- Antelmi I, de Paula RS, Shinzato AR, Peres CA, Mansur AJ, Grupi CJ. Influence of age, gender, body mass index, and functional capacity on heart rate variability in a cohort of subjects without heart disease. *The American Journal of Cardiology*. 2004; 93:381–385. DOI: 10.1016/j.amjcard.2003.09.065 [PubMed: 14759400]
- Armstrong RG, Kenny GP, Green G, Seely AJE. Diurnal variation in heart rate variability before and after maximal exercise testing. *Chronobiology International*. 2011; 28:344–351. DOI: 10.3109/07420528.2011.559674 [PubMed: 21539426]
- Baharav A, Kotagal S, Rubin BK, Pratt J, Akselrod S. Autonomic cardiovascular control in children with obstructive sleep apnea. *Clinical Autonomic Research*. 1999; 9:345–351. DOI: 10.1007/BF02318382 [PubMed: 10638809]
- Berntson GG, Bigger JR Jr, Eckberg DL, Grossman P, Kaufmann PG, Malik M, ... Van Der Molen MW. Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*. 1997; 34:623–648. DOI: 10.1111/j.1469-8986.1997.tb02140.x [PubMed: 9401419]
- Berntson GG, Cacioppo JT, Quigley KS. Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*. 1991; 98:459–487. DOI: 10.1037/0033-295X.98.4.459 [PubMed: 1660159]
- Berntson GG, Cacioppo JT, Quigley KS. Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology*. 1993; 30:183–196. DOI: 10.1111/j.1469-8986.1993.tb01731.x [PubMed: 8434081]
- Berntson GG, Quigley KS. An approach to artifact identification: Application to heart period data. *Psychophysiology*. 1990; 27:586–598. DOI: 10.1111/j.1469-8986.1990.tb01982.x [PubMed: 2274622]

- Berntson, GG., Quigley, KS., Lozano, D. Cardiovascular Psychophysiology. In: Cacioppo, JT., Tassinari, LG., Berntson, GG., editors. *The Handbook of Psychophysiology*. 3. New York, NY: Cambridge University Press; 2007. p. 182-210.
- Berntson GG, Stowell JR. ECG artifacts and heart period variability: Don't miss a beat! *Psychophysiology*. 1998; 35:127–132. DOI: 10.1017/S0048577298001541 [PubMed: 9499713]
- Boer-Martins L, Figueiredo VN, Demacq C, Martins LC, Consolin-Colombo F, Figueiredo MJ, ... Moreno HJr. Relationship of autonomic imbalance and circadian disruption with obesity and type 2 diabetes in resistant hypertensive patients. *Cardiovascular Diabetology*. 2011; 10:1–12. DOI: 10.1186/1475-2840-10-24 [PubMed: 21214916]
- Boyles AL, Harris SF, Rooney AA, Thayer KA. Forest plot viewer: A new graphing tool. *Epidemiology*. 2011; 22:746–747. DOI: 10.1097/EDE.0b013e318225ba48 [PubMed: 21811115]
- Buchhorn R, Conzelmann A, Willaschek C, Störk D, Taurines R, Renner TJ. Heart rate variability and methylphenidate in children with ADHD. *ADHD Attention Deficit Hyperactivity Disorders*. 2012; 4:85–91. DOI: 10.1007/s12402-012-0072-8 [PubMed: 22328340]
- Cacioppo JT, Berntson GG, Binkley PF, Quigley KS, Uchino BN, Fieldstone A. Autonomic cardiac control. II. Noninvasive indices and basal response as revealed by autonomic blockades. *Psychophysiology*. 1994; 31:586–598. DOI: 10.1111/j.1469-8986.1994.tb02351.x [PubMed: 7846219]
- Cerutti S, Goldberger AL, Yamamoto Y. Recent advances in heart rate variability signal processing and interpretation. *IEEE Transactions on Biomedical Engineering*. 2006; 53:1–3. DOI: 10.1109/TBME.2005.859777
- Chen X, Mukkamala R. Selective quantification of the cardiac sympathetic and parasympathetic nervous systems by multisignal analysis of cardiorespiratory variability. *American Journal of Physiology*. 2008; 294:362–371. DOI: 10.1152/ajpheart.01061.2007
- Cipryan L, Litschmannova M. Intra-day and inter-day reliability of heart rate variability measurement. *Journal of Sports Sciences*. 2013; 31:150–158. DOI: 10.1080/02640414.2012.721931 [PubMed: 22963409]
- Cohen, RJ., Swerdlik, ME. *Psychological testing and assessment: An introduction to tests and measurement*. 5. New York, NY: McGraw-Hill; 2002.
- Dalla Pozza R, Fuchs A, Bechtold S, Kozlik-Feldmann R, Daebritz S, Netz H. Short-term testing of heart rate variability in heart-transplanted children: Equal to 24-h ECG recordings? *Clinical Transplantation*. 2006; 20:438–442. DOI: 10.1111/j.1399-0012.2006.00502.x0 [PubMed: 16842518]
- Denver J, Reed S, Porges S. Methodological issues in the quantification of respiratory sinus arrhythmia. *Biological Psychology*. 2007; 74:286–294. DOI: 10.1016/j.biopsycho.2005.09.005 [PubMed: 17067734]
- El-Sheikh M, Hinnant JB. Marital conflict, respiratory sinus arrhythmia, and allostatic load: Interrelations and associations with the development of children's externalizing behavior. *Development and Psychopathology*. 2011; 23:815–829. DOI: 10.1017/S0954579411000320 [PubMed: 21756434]
- Ernst, G. *Heart Rate Variability*. London, UK: Springer-Verlag; 2014.
- Faulkner MS, Hathaway D, Tolley B. Cardiovascular autonomic function in healthy adolescents. *Heart & Lung: Journal of Critical Care*. 2003; 32:10–22. DOI: 10.1067/mhl.2003.6
- Fox NA, Porges SW. The relation between neonatal heart period patterns and developmental outcome. *Child Development*. 1985; 56:28–37. DOI: 10.1111/j.1467-8624.1985.tb00082.x [PubMed: 3987407]
- Guo Y, Stein P. Circadian rhythm in the cardiovascular system: Chronocardiology. *American Heart Journal*. 2003; 145:779–786. DOI: 10.1016/S0002-8703(02)00088-1 [PubMed: 12766733]
- Heathers JAJ. Everything hertz: Methodological issues in short-term frequency-domain HRV. *Frontiers in Physiology*. 2014; 5:177–177. DOI: 10.3389/fphys.2014.00177 [PubMed: 24847279]
- Hejjel L, Roth E. What is the adequate sampling interval of the ECG signal for heart rate variability analysis in the time domain? *Physiological Measurement*. 2004; 25:1405–1411. DOI: 10.1088/0967-3334/25/6/006 [PubMed: 15712719]

- Henry B, Minassian A, Paulus M. Heart rate variability in bipolar mania and schizophrenia. *Journal of Psychiatric Research*. 2010; 44:1–19. DOI: 10.1016/j.jpsychires.2009.07.011.Heart [PubMed: 19656527]
- Hoppenbrouwers T, Oliveira F, Sandarupa S, Khoo M, Neuman M, Ramanathan R. The development of the circadian heart rate rhythm (CHR) in Asian infants. *Early Human Development*. 2012; 88:555–561. DOI: 10.1016/j.earlhumdev.2011.12.031 [PubMed: 22289413]
- Huikuri HV, Kessler KM, Terracall E, Castellanos A, Linnaluoto MK, Myerburg RJ. Reproducibility and circadian rhythm of heart rate variability in healthy subjects. *The American Journal of Cardiology*. 1990; 65:391–393. DOI: 10.1016/0002-9149(90)90308-N [PubMed: 2301268]
- Jarrin DC, McGrath JJ, Giovanniello S, Poirier P, Lambert M. Measurement fidelity of heart rate variability signal processing: The devil is in the details. *International Journal of Psychophysiology*. 2012; 86:88–97. DOI: 10.1016/j.ijpsycho.2012.07.004 [PubMed: 22820268]
- Jarrin DC, McGrath JJ, Poirier P, Seguin L, Tremblay RE, Montplaisir JY, ... Seguin JR. Short-term heart rate variability in a population-based sample of 10-year-old children. *Pediatric Cardiology*. 2015; 36:41–48. DOI: 10.1007/s00246-014-0962-y [PubMed: 25056158]
- Jennings JR, Berg WK, Hutcheson J, Obrist P, Porges S, Turpin G. Publication guidelines for heart rate studies in man. *Psychophysiology*. 1981; 18:226–231. DOI: 10.1111/j.1469-8986.1981.tb03023.x [PubMed: 7291437]
- Kajantie E, Phillips DIW. The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology*. 2006; 31:151–178. DOI: 10.1016/j.psyneuen.2005.07.002 [PubMed: 16139959]
- Kim KK, Lim YG, Kim JS, Park KS. Effect of missing RR-interval data on heart rate variability analysis in the time domain. *Physiological Measurement*. 2007; 28:1485–1494. DOI: 10.1088/0967-3334/28/12/003 [PubMed: 18057513]
- Kleiger RE, Stein PK, Bigger JT. Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrocardiology*. 2005; 10:88–101. DOI: 10.1111/j.1542-474X.2005.10101.x [PubMed: 15649244]
- Kochiadakis GE, Orfanakis AE, Rombola AT, Chrysostomakis SI, Chlouverakis GI, Vardas PE. Reproducibility of time-domain indexes of heart rate variability in patients with vasovagal syncope. *The American Journal of Cardiology*. 1997; 79:160–165. DOI: 10.1016/S0002-9149(96)00704-7 [PubMed: 9193016]
- Kowalewski MA, Urban M. Short- and long-term reproducibility of autonomic measures in supine and standing positions. *Clinical Science*. 2004; 106:61–66. DOI: 10.1042/CS20030119 [PubMed: 12889989]
- Latchman PL, Mathur M, Bartels MN, Axtell RS, De Meersman RE. Impaired autonomic function in normotensive obese children. *Clinical Autonomic Research*. 2011; 21:319–323. DOI: 10.1007/s10286-011-0116-8 [PubMed: 21318460]
- Lee TC, Chiu HW. Frequency-domain heart rate variability analysis performed by digital filters. *Computing in Cardiology*. 2010; 37:589–592.
- Malik M. Time domain measurements of heart rate variability. *Cardiac Electrophysiology Review*. 1997; 3:329–334. DOI: 10.1023/A:1009912905325
- Marks BL, Lightfoot JT. Reproducibility of resting heart rate variability with short sampling periods. *Canadian Journal of Applied Physiology*. 1999; 24:337–348. DOI: 10.1139/h99-026 [PubMed: 10470450]
- Martini G, Riva P, Rabbia F, Molini V, Ferrero GB, Cerutti F, ... Veglio F. Heart rate variability in childhood obesity. *Clinical Autonomic Research*. 2001; 11:87–91. DOI: 10.1007/BF02322051 [PubMed: 11570608]
- Massin M, Maeyns K, Withofs N, Ravet F, Gerard P. Circadian rhythm of heart rate and heart rate variability. *Archives of Disease in Childhood*. 2000; 83:179–182. DOI: 10.1136/adc.83.2.179 [PubMed: 10906034]
- McNames J, Aboy M. Reliability and accuracy of heart rate variability metrics versus ECG segment duration. *Medical and Biological Engineering and Computing*. 2006; 44:747–756. DOI: 10.1007/s11517-006-0097-2 [PubMed: 16960742]

- Merri M, Farden DC, Mottley JG, Titlebaum EL. Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Transactions on Bio-Medical Engineering*. 1990; 37:99–106. DOI: 10.1109/10.43621 [PubMed: 2303276]
- Mulder LJ. Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*. 1992; 34:205–236. DOI: 10.1016/0301-0511(92)90016-N [PubMed: 1467394]
- Ordaz S, Luna B. Sex differences in physiological reactivity to acute psychosocial stress in adolescence. *Psychoneuroendocrinology*. 2012; 37:1135–1157. DOI: 10.1016/j.psyneuen.2012.01.002.Sex [PubMed: 22281210]
- Patton GC, Vines R. Pubertal transitions in health. *Lancet*. 2007; 369:1130–1139. DOI: 10.1016/0140-6736(07)60366-3 [PubMed: 17398312]
- Penttilä J, Kuusela T, Scheinin H. Analysis of rapid heart rate variability in the assessment of anticholinergic drug effects in humans. *European Journal of Clinical Pharmacology*. 2005; 61:559–565. DOI: 10.1007/s00228-005-0953-2 [PubMed: 16133553]
- Porges, SW. Method and apparatus for evaluating rhythmic oscillations in aperiodic physiological response systems. US Patent. 4,510,944. 1985.
- Porges SW, Byrne EA. Research methods for measurement of heart rate & respiration. *Biological Psychology*. 1992; 34:93–130. DOI: 10.1016/0301-0511(92)90012-J [PubMed: 1467397]
- Reyes del Paso GA, Langewitz W, Mulder LJM, van Roon A, Duschek S. The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: A review with emphasis on a reanalysis of previous studies. *Psychophysiology*. 2013; 50:477–487. DOI: 10.1111/psyp.12027 [PubMed: 23445494]
- Richards J. Reliability of respiratory sinus arrhythmia in R-R intervals in 14-, 20-, and 26-week-old infants. *Infant Behavior and Development*. 1995; 18:155–161. DOI: 10.1016/0163-6383(95)90045-4
- Riniolo T, Porges S. Inferential and descriptive influences on measures of respiratory sinus arrhythmia: Sampling rate, R-wave trigger accuracy, and variance estimates. *Psychophysiology*. 1997; 34:613–621. DOI: 10.1111/j.1469-8986.1997.tb01748.x [PubMed: 9299916]
- Rosenberg, MS., Adams, D., Gurevitch, J. MetaWin. Statistical software for meta-analysis, version 2.0. Sunderland, MA: Sinauer Associates; 2000.
- Rosenthal, R. Parametric Measures of Effect Size. In: Cooper, H., Hedges, LV., editors. *The Handbook of Research Synthesis*. New York, NY: Russel Sage Foundation; 1994. p. 231-244.
- Salo MA, Huikuri HV, Seppänen T. Ectopic beats in heart rate variability analysis: Effects of editing on time and frequency domain measures. *Annals of Noninvasive Electrocardiology*. 2001; 6:5–17. DOI: 10.1111/j.1542-474X.2001.tb00080.x [PubMed: 11174857]
- Sandercock GRH, Bromley PD, Brodie DA. The reliability of short-term measurements of heart rate variability. *International Journal of Cardiology*. 2005; 103:238–247. DOI: 10.1016/j.ijcard.2004.09.013 [PubMed: 16098384]
- Schmid K, Schönlebe J, Drexler H, Mueck-Weymann M. Associations between being overweight, variability in heart rate, and well-being in the young men. *Cardiology in the Young*. 2010; 20:54–59. DOI: 10.1017/S1047951109992009 [PubMed: 20144255]
- Shaffer F, Combatalade DC. Don't add or miss a beat: A guide to cleaner heart rate variability recordings. *Biofeedback*. 2013; 41:121–130. DOI: 10.5298/1081-5937-41.3.04
- Sharpley CF. Effects of brief rest periods upon heart rate in multiple baseline studies of heart rate reactivity. *Biofeedback and Self-Regulation*. 1993; 18:225–235. DOI: 10.1007/BF00999081 [PubMed: 8130295]
- Shirtcliff EA, Dahl RE, Pollak SD. Pubertal development: Correspondence between hormonal and physical development. *Child Development*. 2009; 80:327–337. DOI: 10.1111/j.1467-8624.2009.01263.x [PubMed: 19466995]
- Silvetti MS, Drago F, Ragonese P. Heart rate variability in healthy children and adolescents is partially related to age and gender. *International Journal of Cardiology*. 2001; 81:169–174. DOI: 10.1016/S0167-5273(01)00537-X [PubMed: 11744133]
- Singh B, Singh M, Banga VK. Sample entropy based HRV: Effect of ECG sampling frequency. *Biomedical Science and Engineering*. 2014; 2:68–72. doi:12.12691/bse-2-3-3.

- Singh D, Vinod K, Saxena SC. Sampling frequency of the RR interval time series for spectral analysis of heart rate variability. *Journal of Medical Engineering & Technology*. 2004; 28:263–272. DOI: 10.1080/03091900410001662350 [PubMed: 15513744]
- Sookan T, McKune AJ. Heart rate variability in physically active individuals: Reliability and gender characteristics. *Cardiovascular Journal of Africa*. 2012; 23:67–72. DOI: 10.5830/CVJA-2011-018 [PubMed: 22447474]
- Spear LP. The adolescent brain and age-related behavioral manifestations. *Neuroscience & Biobehavioral Review*. 2000; 24:417–463. DOI: 10.1016/S0149-7634(00)00014-2
- Swain A, Suls J. Reproducibility of blood pressure and heart rate reactivity: A meta-analysis. *Psychophysiology*. 1996; 33:162–174. DOI: 10.1111/j.1469-8986.1996.tb02120.x [PubMed: 8851244]
- Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*. 1996; 17:354–381. [PubMed: 8737210]
- Thayer JF, Sternberg E. Beyond heart rate variability: Vagal regulation of allostatic systems. *Annals of the New York Academy of Science*. 2006; 1088:361–372. DOI: 10.1196/annals.1366.014
- Williams K, Thomson D, Seto I, Contopoulos-Ioannidis DG, Ioannidis JPA, Curtis S, ... Klassen T. Standard 6: Age groups for pediatric trials. *Pediatrics*. 2012; 129:S153–S160. DOI: 10.1542/peds.2012-0055I [PubMed: 22661762]
- Wittling, W., Wittling, RA. *Herzschlagvariabilität: Frühwarnsystem, Stress-und Fitnessindikator* [Heart rate variability: Early warning system, stress and fitness indicator]. Heiligenstadt: Eichsfeld-Verlag; 2012.
- Xie GL, Wang JH, Zhou Y, Xu H, Sun JH, Yang SR. Association of high blood pressure with heart rate variability in children. *Iranian Journal of Pediatrics*. 2013; 23:37–44. [PubMed: 23550151]
- Young F, Leicht A. Short-term stability of resting heart rate variability: Influence of position and gender. *Applied Physiology, Nutrition, and Metabolism*. 2011; 218:210–218.

Appendix

References Included in the Meta-Analysis

- Alkon A, Boyce WT, Davis NB, Eskanazi B. Developmental changes in autonomic nervous system resting and reactivity measures in Latino children from 6 to 60 months of age. *Journal of Developmental and Behavioral Pediatrics*. 2011; 32(9):668–677. DOI: 10.1097/DBP.0b013e3182331fa6 [PubMed: 22008788]
- Alkon A, Lippert S, Vujan N, Boyce WT. The ontogeny of autonomic measures in 6- and 12-month-old infants. *Developmental Psychobiology*. 2006; 48:197–208. DOI: 10.1002/dev [PubMed: 16568414]
- Arendt RE, Halpern LF, MacLean WE, Youngquist GA. The properties of V in newborns across repeated measures. *Developmental Psychobiology*. 1991; 24(2):91–101. DOI: 10.1002/dev.420240203 [PubMed: 2044850]
- Bar-Haim Y, Marshall PJ, Fox NA. Developmental changes in heart period and high-frequency heart period variability from 4 months to 4 years of age. *Developmental Psychobiology*. 2000; 37(1):44–56. DOI: 10.1002/1098-2302(200007)37:1<44::AID-DEV6>3.0.CO;2-7 [PubMed: 10937660]
- Blandon AY, Calkins SD, Keane SP, O'Brien M. Contributions of child's physiology and maternal behavior to children's trajectories of temperamental reactivity. *Developmental Psychology*. 2010; 46(5):1089–1102. DOI: 10.1037/a0020678. Contributions [PubMed: 20822225]
- Bornstein MH, Suess PE. Physiological self-regulation and information processing in infancy: Cardiac vagal tone and habituation. *Child Development*. 2000a; 71(2):273–287. DOI: 10.1111/1467-8624.00143 [PubMed: 10834463]
- Bornstein MH, Suess PE. Child and mother cardiac vagal tone: Continuity, stability, and concordance across the first 5 years. *Developmental Psychology*. 2000b; 36(1):54–65. DOI: 10.1037/0012-1649.36.1.54 [PubMed: 10645744]

- Calkins SD, Keane SP. Cardiac vagal regulation across the preschool period: Stability, continuity, and implications for childhood adjustment. *Developmental Psychobiology*. 2004; 45(3):101–112. DOI: 10.1002/dev.20020 [PubMed: 15505799]
- Cavaliere, CM. Unpublished doctoral dissertation. Seton Hall University; South Orange, New Jersey: 2011. Role of the vagal system in the regulatory sensory processing patterns in children with autistic spectrum disorder: Physiologic underpinnings and reliability of measurement. Retrieved from <http://www.proquest.com>
- Dietrich A, Rosmalen JGM, Althaus M, van Roon AM, Mulder LJM, Minderaa RB, ... Riese H. Reproducibility of heart rate variability and baroreflex sensitivity measurements in children. *Biological Psychology*. 2010; 85(1):71–78. DOI: 10.1016/j.biopsycho.2010.05.005 [PubMed: 20553793]
- Doussard-Roosevelt JA, Montgomery LA, Porges SW. Short-term stability of physiological measures in kindergarten children: Respiratory sinus arrhythmia, heart period, and cortisol. *Developmental Psychobiology*. 2003; 43(3):230–242. DOI: 10.1002/dev.10136 [PubMed: 14558045]
- El-Sheikh M. Stability of respiratory sinus arrhythmia in children and young adolescents: A longitudinal examination. *Developmental Psychobiology*. 2005; 46(1):66–74. DOI: 10.1002/dev.20036 [PubMed: 15690389]
- Elmore-Staton, LD. Unpublished doctoral dissertation. Auburn University; Auburn, Alabama: 2011. Sleep and cognitive functioning: The moderating role of vagal indices. Retrieved from <http://www.proquest.com>
- Farah BQ, Lima AHRDA, Cavalcante BR, Oliveira LMFT, Brito ALDS, Barros MVG, Ritti-Dias RM. Intra-individuals and inter- and intra-observer reliability of short-term heart rate variability in adolescents. *Clinical Physiology and Functional Imaging*. 2014; 36(1):33–39. DOI: 10.1111/cpf.12190 [PubMed: 25216444]
- Fox NA. Maturation of autonomic control in preterm infants. *Developmental Psychobiology*. 1983; 16(6):495–504. DOI: 10.1002/dev.420160605 [PubMed: 6642080]
- Fox NA, Field TM. Individual differences in preschool entry behavior. *Journal of Applied Developmental Psychology*. 1989; 10(4):527–540. DOI: 10.1016/0193-3973(89)90025-7
- Fracasso MP, Porges SW, Lamb ME, Rosenberg AA. Cardiac activity in infancy: Reliability and stability of individual differences. *Infant Behavior and Development*. 1994; 284(17):277–284. DOI: 10.1016/0163-6383(94)90006-X
- Galland BC, Taylor BJ, Bolton DPG, Sayers RM. Heart rate variability and cardiac reflexes in small for gestational age infants. *Journal of Applied Physiology*. 2006; 100(3):933–939. DOI: 10.1152/jappphysiol.01275.2005 [PubMed: 16306252]
- Gamelin FX, Baquet G, Berthoin S, Thevenet D, Nourry C, Nottin S, Bosquet L. Effect of high intensity intermittent training on heart rate variability in prepubescent children. *Euro Journal of Applied Physiology*. 2009; 105(5):731–738. DOI: 10.1007/s00421-008-0955-8
- Gentzler AL, Rottenberg J, Kovacs M, George CJ, Morey JN. Atypical development of resting respiratory sinus arrhythmia in children at high risk for depression. *Developmental Psychobiology*. 2012; 54(5):556–567. DOI: 10.1002/dev.20614 [PubMed: 21960203]
- Hastings PD, Nuselovici JN, Utendale WT, Coutya J, McShane KE, Sullivan C. Applying the polyvagal theory to children's emotion regulation: Social context, socialization, and adjustment. *Biological Psychology*. 2008; 79(3):299–306. DOI: 10.1016/j.biopsycho.2008.07.005 [PubMed: 18722499]
- Henje Blom E, Olsson EMG, Serlachius E, Ericson M, Ingvar M. Heart rate variability is related to self-reported physical activity in a healthy adolescent population. *European Journal of Applied Physiology*. 2009; 106(6):877–883. DOI: 10.1007/s00421-009-1089-3 [PubMed: 19479275]
- Hinnant JB, El-Sheikh M. Children's externalizing and internalizing symptoms over time: The role of individual differences in patterns of RSA responding. *Journal of Abnormal Child Psychology*. 2009; 37(8):1049–1061. DOI: 10.1007/s10802-009-9341-1 [PubMed: 19711181]
- Hinnant JB, Elmore-Staton L, El-Sheikh M. Developmental trajectories of respiratory sinus arrhythmia and preejection period in middle childhood. *Developmental Psychobiology*. 2011; 53(1):59–68. DOI: 10.1002/dev.20487 [PubMed: 20882584]

- Hsu HC, Porter CL. Young infants' behavioral reactivity to mild perturbation: Developmental continuity, stability, and organization. *Infancy*. 2004; 6(1):95–120. DOI: 10.1207/s15327078in0601_5
- Izard CE, Porges SW, Simons RF, Haynes OM, Hyde C, Parisi M, Cohen B. Infant cardiac activity: Developmental changes and relations with attachment. *Developmental Psychology*. 1991; 27(3): 432–439. DOI: 10.1037/0012-1649.27.3.432
- Keller PS, El-Sheikh M. Salivary alpha-amylase as a longitudinal predictor of children's externalizing symptoms: Respiratory sinus arrhythmia as a moderator; of effects. *Psychoneuroendocrinology*. 2009; 34(5):633–643. DOI: 10.1016/j.psyneuen.2008.12.016 [PubMed: 19195789]
- Keller PS, Kouros CD, Erath SA, Dahl RE, El-Sheikh M. Longitudinal relations between maternal depressive symptoms and child sleep problems: The role of parasympathetic nervous system reactivity. *Journal of Child Psychology and Psychiatry*. 2014; 55(2):172–179. DOI: 10.1111/jcpp.12151 [PubMed: 24117807]
- Kennedy AE, Rubin KH, Hastings PD, Maisel B. Longitudinal relations between child vagal tone and parenting behavior: 2 to 4 years. *Developmental Psychobiology*. 2004; 45(1):10–21. DOI: 10.1002/dev.20013 [PubMed: 15229872]
- Leicht AS, Allen GD. Moderate-term reproducibility of heart rate variability during rest and light to moderate exercise in children. *Brazilian Journal of Medical and Biological Research*. 2008; 41(7):627–633. DOI: 10.1590/S0100-879X2008000700013 [PubMed: 18719745]
- Liew J, Eisenberg N, Spinrad TL, Eggum ND, Haugen RG, Kupfer A, ... Baham ME. Physiological regulation and fearfulness as predictors of young children's empathy-related reactions. *Social Development*. 2011; 20(1):111–113. DOI: 10.1111/j.1467-9507.2010.00575.x [PubMed: 22573929]
- Marshall PJ, Stevenson-Hinde J. Behavioral inhibition, heart period, and respiratory sinus arrhythmia in young children. *Developmental Psychobiology*. 1998; 33(3):283–292. DOI: 10.1002/(SICI)1098-2302(199811)33:3 [PubMed: 9810478]
- Mezzacappa E, Tremblay RE, Kindlon D, Saul JP, Arseneault L, Seguin J, ... Earls F. Anxiety, antisocial behavior, and heart rate regulation in adolescent males. *Journal of Child Psychology and Psychiatry*. 1997; 38(4):457–469. DOI: 10.1111/j.1469-7610.1997.tb01531.x [PubMed: 9232491]
- Patriquin MA, Lorenzi J, Scarpa A, Bell MA. Developmental trajectories of respiratory sinus arrhythmia: Associations with social responsiveness. *Developmental Psychobiology*. 2014; 56:317–326. DOI: 10.1002/dev.21100 [PubMed: 23341170]
- Perry NB, Nelson JA, Swingle MM, Leerkes EM, Calkins SD, Marcovitch S, O'Brien M. The relation between maternal emotional support and child physiological regulation across the preschool years. *Developmental Psychobiology*. 2013; 55(4):382–394. DOI: 10.1002/dev.21042 [PubMed: 22573287]
- Porges SW. Unpublished raw data. 2014
- Porges SW, Doussard-Roosevelt JA, Portales AL, Suess PE. Cardiac vagal tone: Stability and relation to difficultness in infants and 3-year-olds. *Developmental Psychobiology*. 1994; 27(5):289–300. DOI: 10.1002/dev.420270504 [PubMed: 7926281]
- Porter CL, Bryan YE, Hsu HC. Physiological markers in early infancy: Stability of 1-to 6-month vagal tone. *Infant Behavior and Development*. 1995; 18(3):363–367. DOI: 10.1016/0163-6383(95)90025-X
- Propper C, Moore GA, Mills-Koonce WR, Halpern CT, Hill-Soderlund AL, Calkins SD, ... Cox M. Gene-environment contributions to the development of infant vagal reactivity: The interaction of dopamine and maternal sensitivity. *Child Development*. 2008; 79(5):1377–1394. DOI: 10.1111/j.1467-8624.2008.01194.x [PubMed: 18826531]
- Richards JE. Development and stability in visual sustained attention in 14, 20, and 26 week old infants. *Psychophysiology*. 1989; 26(4):422–430. DOI: 10.1111/j.1469-8986.1989.tb01944.x [PubMed: 2798691]
- Rigterink T, Fainsilber Katz L, Hessler DM. Domestic violence and longitudinal associations with children's physiological regulation abilities. *Journal of Interpersonal Violence*. 2010; 25(9): 1669–1683. DOI: 10.1177/0886260509354589 [PubMed: 20587477]

- Salomon K. Respiratory sinus arrhythmia during stress predicts resting respiratory sinus arrhythmia 3 years later in a pediatric sample. *Health Psychology*. 2005; 24(1):68–76. DOI: 10.1037/0278-6133.24.1.68 [PubMed: 15631564]
- Schuetze P, Eiden RD, Edwards EP. A longitudinal examination of physiological regulation in cocaine-exposed infants across the first 7 months of life. *Infancy*. 2009; 14(1):19–43. DOI: 10.1080/15250000802569660 [PubMed: 20047018]
- Snidman N, Kagan J, Riordan L, Shannon DC. Cardiac function and behavioral reactivity during infancy. *Psychophysiology*. 1995; 32(3):199–207. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7784528>. [PubMed: 7784528]
- Stifter CA, Jain A. Psychophysiological correlates of infant temperament: Stability of behavior and autonomic patterning from 5 to 18 months. *Developmental Psychobiology*. 1996; 29:379–391. DOI: 10.1002/(SICI)1098-2302(199605)29:4<379::AID-DEV5>3.0.CO;2-N [PubMed: 8732809]
- Stifter CA, Fox NA. Infant reactivity: Physiological correlates of newborn and 5-month temperament. *Developmental Psychology*. 1990; 26(4):582–588. DOI: 10.1037/0012-1649.26.4.582
- Winegust AK, Mathewson KJ, Schmidt LA. Test-retest reliability of frontal alpha electroencephalogram (EEG) and electrocardiogram (ECG) measures in adolescents: A pilot study. *International Journal of Neuroscience*. 2014; 124(12):908–911. DOI: 10.3109/00207454.2014.895003 [PubMed: 24617292]
- Winsley RJ, Armstrong N, Bywater K, Fawcner SG. Reliability of heart rate variability measures at rest and during light exercise in children. *British Journal of Sports Medicine*. 2003; 37(6):550–552. DOI: 10.1136/bjsm.37.6.550 [PubMed: 14665601]
- Yiallourou SR, Sands SA, Walker AM, Horne RSC. Maturation of heart rate and blood pressure variability during sleep in term-born infants. *Sleep*. 2012; 35(2):177–186. DOI: 10.5665/sleep.1616 [PubMed: 22294807]

Figure 1.
 Flow chart for article identification and inclusion in meta-analysis.

Figure 2.
HRV Reliability Forest Plots for NN, SDNN, and rMSSD.

Figure 3.
HRV Reliability Forest Plots for HF.

Figure 4.
HRV Reliability Forest Plots for LF and LF:HF.

Table 1

Study descriptive characteristics and ECG recording characteristics by age group

	Infant/toddler (<i>k</i> = 31)			Child/adolescent (<i>k</i> = 18)		
	<i>k</i>	<i>N</i>	<i>M</i> (<i>SD</i>)	<i>k</i>	<i>N</i>	<i>M</i> (<i>SD</i>)
Continuous Study Variables						
Follow-up interval length (mos)	31	3,329	17.55 (16.57)	18	1,841	19.39 (26.13)
Sample size (<i>N</i>) ^a	31	3,329	107.39 (88.50)	18	1,841	102.28 (91.93)
Sample mean age (yrs)						
First study visit	31	3,329	1.2 (1.48)	17	1,785	10.84 (3.18)
Final study visit	31	3,329	2.77 (2.46)	16	1,773	11.78 (3.10)
Percent (%) of male participants ^a	17	1,409	46.18 (7.78)	15	1,368	56.38 (25.67)
Percent (%) of sample attrition						
First to second study visit	12	1,865	21.67 (17.14)	10	1,422	15.65 (10.37)
Second to third study visit	3	794	9.33 (2.52)	1	251	16.00 (0.00)
Height (in) ^{a,b}	2	108	20.59 (0.23)	3	65	61.38 (6.76)
Weight (lb) ^{a,b}	9	465	7.81 (0.54)	5	291	94.08 (25.45)
Acclimatization duration (min) ^a	5	500	3.46 (2.11)	12	1,542	9.5 (10.34)
ECG recording duration (min) ^a						
Total recording duration	30	3,302	20.01 (40.39)	17	1,785	16.61 (19.26)
Baseline recording duration	30	3,302	4.63 (4.24)	18	1,841	5.00 (3.56)
Analyzed recording duration	30	3,123	4.42 (4.29)	18	1,841	3.74 (1.91)
ECG epoch duration (min) ^a	18	2,161	0.51 (0.22)	4	179	0.81 (0.80)
Number of effect sizes per study	31	3,329	4.81 (8.15)	18	1,841	3.56 (2.79)
	<i>k</i>	<i>N</i>	% ^c	<i>k</i>	<i>N</i>	% ^c
Categorical Study Variables						
Sample exclusion criteria						
Prescription medications	3	165	9.7%	3	227	16.7%
Medical illness/condition	5	359	16.1%	8	978	44.4%
Time of ECG recording						
06:00–12:00	2	58	6.5%	5	324	27.8%
12:00–18:00	1	20	3.2%	1	64	5.6%
Awake vs. asleep ECG recording ^a						
Exclusively awake data	27	3,079	87.1%	18	1,841	100%
Sleep unrestricted	4	250	12.9%	–		
Additional measure of respiration	7	723	22.6%	7	1,319	38.9%
Recording posture ^d						
Seated posture	20	2,320	64.5%	7	987	38.9%
Supine posture	5	614	16.1%	6	307	33.3%
Standing posture	1	112	3.2%	2	232	6.5%

	Infant/toddler (<i>k</i> = 31)			Child/adolescent (<i>k</i> = 18)		
	<i>k</i>	<i>N</i>	<i>M</i> (<i>SD</i>)	<i>k</i>	<i>N</i>	<i>M</i> (<i>SD</i>)
Changed recording posture	5	265	16.1%	–		
ECG sampling rate						
500 Hz	3	157	9.7%	10	1,243	55.6%
<500 Hz	4	295	12.9%	1	57	5.6%
Manual editing of ECG data	24	2,744	77.4%	15	1,792	83.3%
Digital filtering algorithm	22	2,197	71%	8	433	44.4%
HRV derivation method						
Mean of 1 epoch	5	493	16.1%	7	1,154	38.9%
Mean of 2 epochs	18	2,161	58.1%	4	179	22.2%
Recommended frequency band						
High Frequency (HF)	18	2,074	58.1%	8	514	44.4%
Low Frequency (LF)	1	31	3.2%	5	362	27.8%

Notes. HRV = Heart Rate Variability; ECG = Electrocardiogram; LF = Low Frequency; HF = High Frequency;

^aBased on initial study assessment. Unmentioned categories not included (i.e., target variable not reported; % may not total 100%).

^bLength, birthweight for Infant/toddler group. Of 9 studies reporting birthweight, mean age = 1.43 mos.

^cPercent of corresponding category group total *k*.

^dRedundant effect sizes included (i.e., multiple postures reported per study).

Table 2

HRV reliability – summary effect sizes and continuous moderator analyses by age group

		NN				SDNN				rMSSD				HF				LF				LF:HF			
		k	Z	[95% CI] ^a	Q _T	k	Z	[95% CI] ^a	Q _T	k	Z	[95% CI] ^a	Q _T	k	Z	[95% CI] ^a	Q _T	k	Z	[95% CI] ^a	Q _T	k	Z	[95% CI] ^a	Q _T
Overall summary analyses																									
Infant/Toddler	14	0.33 [15, 49]	45.8 ^{**}	2	0.33 [15, 91]	14.9 ^{**}			28	0.42 [35, 49]		72.1 ^{**}													
Child/Adolescent	2	0.75 [68, 100]	14.9 ^{**}	5	0.94 [68, 125]	15.8 ^{**}	3	1.00 [0.41, 1.53]	13.7 ^{***}	0.62 [51, 78]		67.7 ^{**}	7	0.63 [46, 82]	19.9 [*]	4	0.63 [30, 89]	5.5							
	k	B (SE)	Q _T	k	B (SE)	Q _T	k	B (SE)	Q _T	k	B (SE)	Q _T	k	B (SE)	Q _T	k	B (SE)	Q _T							
Continuous moderator analyses																									
Follow-up interval length (mos) ^b																									
Infant/Toddler	14	.004(.00)	47.5 ^{***}		–				28	.002(.00)		101.4 ^{***}													
Child/Adolescent	5	–	–.057(.03) [*]	15.8 ^{**}	3	–.553(.22) [*]	13.7 ^{***}	17	–.005(.00) ^{**}	84.3 ^{***}	7	–.031(.01) ^{**}	13.9 [*]	4	.077(.18)	5.5									
Sample size (N) ^c																									
Infant/Toddler	14	.001(.00)	45.8 ^{***}		–				28	.000(.00)		72.1 ^{***}													
Child/Adolescent	5	–	–.003(.00)	15.8 ^{**}	3	.051(.03) [*]	13.7 ^{***}	16	–.001(.01) ^{***}	67.7 ^{***}	7	–.003(.00) ^{**}	13.9 [*]	4	.001(.02)	5.5									
Mean age (yrs) ^c																									
Infant/Toddler	14	.153(.03) ^{**}	45.8 ^{***}		–				28	.055(.02) ^{***}		72.1 ^{***}													
Child/Adolescent	5	–	.018(.00)	15.8 ^{**}	3	.173(.05) ^{***}	13.7 ^{***}	16	–.005(.00)	65.8 ^{***}	7	.007(.03)	13.9 [*]	4	.018(.05)	5.5									
Male (percent) ^c																									
Infant/Toddler	9	2.141(.90) ^{**}	36.5 ^{***}		–				19	–.262(.55)		62.5 ^{**}													
Child/Adolescent	5	–	1.088(.35) ^{**}	15.8 ^{**}	3	2.085(.57) ^{***}	13.7 ^{***}	16	–.344(.17) [*]	65.8 ^{***}	7	–.430(.21) [*]	13.9 [*]	4	.742(.56)	5.5									
Acclimatization duration (min) ^c																									
Infant/Toddler		–			–				6	–.035(.03)		2.91													
Child/Adolescent	4	–	.017(.01)	14.2 ^{**}		–			11	.007(.01)		51.0 ^{***}	5	.008(.01)	3.6										
Analyzed recording duration (min) ^c																									
Infant/Toddler	14	–.028(.01) [*]	45.8 ^{***}		–				27	–.010(.01)		72.1 ^{***}													
Child/Adolescent	5		.130(.04) ^{***}	15.8 ^{**}	3	.201(.05) ^{***}	13.7 ^{***}	17	.066(.03) ^{**}	67.7 ^{***}	7	.059(.03) [*]	13.9 [*]	4	.015(.04)	5.5									

Epoch duration (min) ^c																							
NN				SDNN				rMSSD				HF				LF				LF:HF			
<i>k</i>	<i>Z</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>k</i>	<i>Z</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>k</i>	<i>Z</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>k</i>	<i>Z</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>k</i>	<i>Z</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>k</i>	<i>Z</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>
Infant/Toddler	8	.228(.79)	22.9 ^{**}	—	—	—	—	18	.052(.14)	—	48.0 ^{***}	—	—	—	—	—	—	—	—	—	—	—	—
Child/Adolescent	—	—	—	—	—	—	—	4	.059(.11)	—	3.62	—	—	—	—	—	—	—	—	—	—	—	—

Notes. *B* = slope; *SE* = standard error; *Z* = Fisher's *Z*; *Q_T* = Heterogeneity; *NFS* = Orwin's Failsafe *N*; NN = mean NN interval; SDNN = standard deviation of NN; rMSSD = root mean square of successive differences; HF = high frequency; LF = low frequency.

^a Bootstrap 95% Confidence Interval.

^b Longest follow-up interval assessment only (i.e., only non-redundant effect sizes).

^c Initial study assessment.

* *p* < .05.

** *p* .01.

*** *p* .001.

Table 3

HRV reliability – infant/toddler studies, categorical moderator variables

	<i>k</i>	NN			<i>k</i>	HF		
		<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>N_{FS}</i>		<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>N_{FS}</i>
Sample exclusion criteria								
Prescription medications		<i>Q_M</i> = 14.06, <i>p</i> < .001				<i>Q_M</i> = 0.44, <i>p</i> = .509		
Exclusion reported	2	0.61 [.40, .65]	1.10	120	2	0.37 [.34, .50]	0.45	71
Unmentioned	12	0.23 [.11, .40]	30.67***	269	26	0.42 [.34, .50]	71.25***	1078
Medical illness/condition		<i>Q_M</i> = 5.26, <i>p</i> = .022				<i>Q_M</i> = 2.82, <i>p</i> = .093		
Exclusion reported	3	0.47 [.04, .65]	11.74**	137	4	0.32 [.21, .42]	2.00	124
Unmentioned	11	0.26 [.12, .44]	28.82***	273	24	0.44 [.35, .51]	67.31***	1020
Acclimatization period		<i>Q_M</i> = 18.31, <i>p</i> < .001				<i>Q_M</i> < 0.01, <i>p</i> = .986		
Yes/included	2	0.62 [.55, .65]	0.29	122	8	0.42 [.36, .48]	3.48	329
No/unmentioned	12	0.21 [.10, .38]	27.22**	243	20	0.42 [.31, .51]	68.66***	821
Awake vs. asleep ECG recording ^b		N/A				<i>Q_M</i> = 6.35, <i>p</i> = .012		
Exclusively awake data	13	0.35 [.17, .52]	41.29***	446	25	0.44 [.36, .50]	61.00***	1065
Sleep unrestricted	1	0.01 [-, -]	–	–	3	0.21 [.03, .34]	4.79	60
Recording posture ^c		<i>Q_M</i> = 9.42, <i>p</i> = .002				<i>Q_M</i> = 18.69, <i>p</i> < .001		
Sitting	10	0.39 [.21, .58]	31.55***	384	19	0.49 [.42, .56]	43.28***	921
Supine	1	0.16 [-, -]	–	–	5	0.32 [.25, .35]	0.65	155
Standing	1	0.38 [-, -]	–	–		–		
Changed recording posture	3	0.01 [-.21, .59]	3.79	0	3	0.19 [-.01, .34]	4.32	53
Unmentioned		–			1	0.38 [-, -]	–	–
ECG sampling rate		N/A				<i>Q_M</i> = 5.07, <i>p</i> = .079		
500 Hz		–			2	0.39 [.35, .50]	0.34	76
<500 Hz		–			4	0.28 [.15, .34]	1.62	106
Unmentioned		–			22	0.44 [.36, .52]	65.11	946
Manual editing of ECG data		<i>Q_M</i> = 4.99, <i>p</i> = .025				<i>Q_M</i> = 1.80, <i>p</i> = .180		
Yes	12	0.38 [.16, .55]	40.80***	439	22	0.43 [.35, .51]	64.60***	930
No/unmentioned	2	0.14 [.13, .16]	0.02	25	6	0.34 [.21, .48]	6.04	200
Digital filtering algorithm		<i>Q_M</i> = 1.96, <i>p</i> = .161				<i>Q_M</i> = 4.00, <i>p</i> = .045		
Yes	11	0.36 [.14, .53]	41.87***	382	22	0.45 [.35, .53]	62.30***	962
No/unmentioned	3	0.21 [.13, .59]	1.98	59	6	0.35 [.27, .46]	6.13	204
Identical ECG recording durations		<i>Q_M</i> = 5.75, <i>p</i> = .017				<i>Q_M</i> = 5.35, <i>p</i> = .021		
Yes	11	0.38 [.16, .56]	38.43***	404	20	0.45 [.36, .52]	53.71***	886
No/unmentioned	3	0.12 [.01, .32]	1.67	31	8	0.34 [.23, .46]	13.08*	261
HRV derivation method		N/A				<i>Q_M</i> = 5.21, <i>p</i> = .017		
Mean of 1 epoch		–			2	0.25 [.16, .34]	1.08	48
Mean of 2 epochs	9	0.37 [.11, .57]	35.34***	326	18	0.46 [.38, .53]	47.97***	811

	NN				HF			
	<i>k</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>N_{FS}</i>	<i>k</i>	<i>Z</i> [95% CI] ^a	<i>Q_T</i>	<i>N_{FS}</i>
Recommended frequency band		N/A				$Q_M = 0.58, p = .448$		
Yes		–		18	0.43 [.33, .50]	36.86***	754	
No/unmentioned		–		8	0.48 [.31, .73]	30.84***	372	

Notes. *k* = number of studies; *Z* = Fisher's *Z*; *Q_M* = between-groups heterogeneity; *Q_T* = total heterogeneity; *N_{FS}* = Orwin's Failsafe *N*; SDNN = standard deviation of NN intervals; HF = high frequency; LF = low frequency; ECG = electrocardiogram; HRV = heart rate variability.

^aBootstrap 95% Confidence Interval.

^bInitial study assessment recording condition.

^cRedundant effect sizes included (i.e., multiple postures reported per study).

*
p < .05.

**
p < .01.

***
p < .001.

Table 4

HRV reliability – child/adolescent studies, categorical moderator variables

	SDNN				HF				LF				LF:HF			
	k	Z [95% CI] ^a	Q _T	N _{rs}	k	Z [95% CI] ^a	Q _T	N _{rs}	k	Z [95% CI] ^a	Q _T	N _{rs}	k	Z [95% CI] ^a	Q _T	N _{rs}
Sample exclusion criteria																
Prescription medications		N/A				$Q_M = 0.97, p = .324$				N/A				N/A		
Exclusion reported		-			2	0.55[.52, .87]	1.36	107		-				-		
Unmentioned	5	0.94[.66, 1.25]	15.77**	466	15	0.63[.51, .82]	63.34***	932	7	0.63[.46, .82]	13.90*	432	4	0.63[.30, .89]	5.52	249
Medical illness/condition						$Q_M = 6.63, p = .010$				$Q_M = 4.17, p = .041$				N/A		
Exclusion reported	2	0.98[.79, 1.53]	9.61***	194	8	0.67[.52, .93]	39.50***	541	2	0.80[.78, .87]	0.14	158	1	0.78[-, -]	-	-
Unmentioned	3	0.89[.41, 1.07]	5.80	265	9	0.55[.46, .73]	21.54**	482	5	0.54[.39, .81]	9.59*	265	3	0.51[.14, 1.19]	4.60	151
Acclimatization period		N/A				$Q_M = 3.29, p = .070$				$Q_M = 10.32, p = .001$				N/A		
Yes/included	4	0.97[.73, 1.33]	14.22**	382	12	0.64[.53, .82]	58.14***	756	5	0.79[.67, .91]	3.58	388	3	0.74[.41, 1.19]	2.87	217
No/unmentioned	1		-	-	5	0.49[.38, .79]	6.23	239	2	0.39[.39, .39]	0.00	76	1	0.14[-, -]	-	-
Awake vs. asleep ECG recording ^b		N/A				N/A				N/A				N/A		
Exclusively awake data	5	0.94[.66, 1.25]	15.77**	466	17	0.62 [.51, .78]	67.67***	1,039	7	0.63[.46, .82]	13.90*	432	4	0.63[.30, .89]	5.52	249
Sleep unrestricted		-				-				-				-		
ECG recording posture ^c		N/A				$Q_M = 4.89, p = .180$				$Q_M = 1.11, p = .293$				N/A		
Sitting	1	0.79[-, -]	-	-	7	0.65[.50, .90]	43.36***	447	1	0.78[-, -]	-	-		-		
Supine	4	1.05[.46, 1.36]	13.17**	414	6	0.64[.41, .99]	18.41**	378	6	0.58[.42, .85]	11.87*	341	4	0.63[.30, .89]	5.52	249
Standing	1	1.02[-, -]	-	-	2	0.49[.26, .93]	15.48***	96	2	0.47[.32, .75]	6.23*	91		-		
Unmentioned		-			4	0.55[.50, .73]	3.34	214		-				-		
ECG sampling rate						$Q_M = 6.53, p = .011$				$Q_M = 2.12, p = .145$				N/A		
500 Hz	2	0.73[.41, .79]	1.57	143	9	0.64[.49, .81]	43.34***	569	2	0.72[.41, .78]	1.49	142	1	0.41[-, -]	-	-
Unmentioned	2	1.26 [.54, 1.53]	6.41*	249	7	0.53[.46, .87]	11.99	364	4	0.52[.39, .99]	8.59*	202	3	0.71[.14, 1.19]	4.65	209
Manual editing of ECG data						$Q_M = 1.97, p = .160$				$Q_M = 1.90, p = .169$				N/A		
Yes	3	0.87[.41, 1.07]	5.17	256	14	0.62[.51, .77]	63.43***	847	4	0.59[.39, .78]	8.91*	233	1	0.41[-, -]	-	-
No/unmentioned	2	1.26[.54, 1.53]	6.41*	249	3	0.84[.47, 1.22]	2.27	249	3	0.83[.39, 1.26]	3.08	246	3	0.70[.14, 1.19]	4.65	209

	SDNN				HF				LF				LF:HF							
	<i>k</i>	<i>Z</i>	[95% CI] ^a	<i>Q_T</i>	<i>N_{FS}</i>	<i>k</i>	<i>Z</i>	[95% CI] ^a	<i>Q_T</i>	<i>N_{FS}</i>	<i>k</i>	<i>Z</i>	[95% CI] ^a	<i>Q_T</i>	<i>N_{FS}</i>					
Digital filtering algorithm																				
Yes	3	0.89	[.54, 1.07]	3.47	263	8	0.67	[.46, .91]	21.12 ^{**}	529	5	0.62	[.41, .81]	11.85 [*]	302	2	0.60	[.14, 1.19]	4.34 [*]	117
No/unmentioned	2	1.13	[.41, 1.53]	10.54 ^{**}	224	9	0.61	[.49, .78]	45.60 ^{***}	537	2	0.71	[.41, .87]	1.78	139	2	0.65	[.41, .78]	1.15	127
HRV derivation method																				
Mean of 1 epoch						6	0.62	[.49, .84]	42.27 ^{***}	364				–						
Mean of 2 epochs	1	0.79	[-, -]	–	–	4	0.74	[.40, .83]	3.80	291	1	0.78	[-, -]	–	–					
Recommended frequency band																				
Yes						6	0.64	[.52, .95]	19.02 ^{**}	376	6	0.64	[.46, .85]	13.37 [*]	375	3	0.74	[.14, 1.19]	2.87	217
No/unmentioned						5	0.49	[.37, .79]	6.23	239	1	0.39	[-, -]	–	–	1	0.39	[-, -]	–	–

Notes. *k* = number of studies; *Z* = Fisher's *Z*; 95% CI = bootstrap 95% confidence interval; *Q_M* = between-groups heterogeneity; *Q_T* = total heterogeneity; *N_{FS}* = Orwin's Failsafe *N*; SDNN = standard deviation of NN intervals; HF = high frequency; LF = low frequency; ECG = electrocardiogram.

^a Bootstrap 95% Confidence Interval.

^b Initial study assessment recording condition.

^c Redundant effect sizes included (i.e., multiple postures per study).

* *p* < .05.

** *p* .01.

*** *p* .001.