Psychometrics of Pediatric Heart Rate Variability: Reliability and Stability

Oren Weiner

A Thesis

In

The Department

Of

Psychology

Presented in Partial Fulfillment of the Requirement

for the Degree of Master of Arts (Psychology) at

Concordia University

Montreal, Quebec, Canada

August 2014

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Oren Matthew Weiner

Entitled: Psychometrics of Pediatric Heart Rate Variability: Reliability and Stability

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Psychology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Wayne Brake

_____ Examiner
Dr. Lucie Bonneville

_____ Examiner
Dr. Jean-Philippe Gouin

_____ Supervisor
Dr. Jennifer J. McGrath

Approved by _____
Chair of Department or Graduate Program Director

_____
Dean of Faculty

_____
Date

ABSTRACT

Psychometrics of Pediatric Heart Rate Variability: Reliability and Stability

Oren Weiner

Measures of heart rate variability (HRV) are commonly used in longitudinal studies among infants, children, and adolescents as an indicator of autonomic cardiovascular control. However, the psychometric properties of pediatric HRV measures have yet to be established. This thesis examined the psychometrics of time- and frequency-domain HRV measures among infants and youth through two complementary studies. The first study was a systematic review and meta-analysis of 46 studies that evaluated how study methodology (study protocol, sample characteristics, ECG signal acquisition and pre-processing, HRV analyses) affects HRV test-retest reliability. HRV displayed moderate reliability overall across infant/toddler (M_{age} <5 yrs; Fisher's Z = 0.42) and child/adolescent groups (M_{age} 5–18 yrs; Z = 0.64); reliability among infant/toddler studies was relatively more sensitive to examined *a priori* moderator variables. The second study examined temporal stability of HRV measures obtained from children (M_{age} 9 yrs) participating in a large, longitudinal cohort study (N = 632) within Quebec. Results indicated that HRV is a moderately stable individual difference in children (ICC_{range} = 0.74–0.85; r_{range} = 0.67–0.75), and stability was robust to initial differences in, and developmental changes among demographic (age, puberty), cardiovascular (blood pressure), anthropometric (height, adiposity indices), and physical activity (e.g., weekday/end screen time) covariates. However, stability of HRV measures reflecting parasympathetic (rMSSD, pNN50, HF) activity was augmented following statistical control of time-varying heart rate measures. Together, this thesis contributes original knowledge regarding the psychometrics of pediatric HRV measures. Pertinent

methodological and intra-individual factors recommended for consideration by future researchers are discussed.

ACKNOWLEDGEMENT

I would like to thank Dr. Jennifer J. McGrath for her guidance and dedication to my research, as well as members of the Pediatric Public Health Psychology Laboratory for their helpful feedback and assistance with this project. A special thank you goes to Ms. Jinshia Ly for her assistance with coding articles for the meta-analysis. I would also like to thank the participants and families of the QUALITY Cohort. Finally, I would like to thank my family and friends in Montreal and Toronto for their continued support and encouragement.

TABLE OF CONTENTS

List of Tables……..……xi

List of Figures……..……..……...xiii

List of Appendices...……..……..……..xiv

List of Abbreviations……..xv

INTRODUCTION……...……1

Autonomic and Cardiovascular Physiology……1

Frequency-Domain HRV……4

Time-Domain HRV……..7

Autonomic Dysfunction, Health, and HRV……8

HRV Measurement, Analysis, and Reporting……10

Study Protocol……..……..11

Sample Characteristics…….……...12

ECG Signal Acquisition and Pre-Processing…….…13

HRV Analyses…….……...16

HRV Reliability……...…….……20

Study Protocol……....20

Sample Characteristics…….……...21

ECG Signal Acquisition and Pre-Processing…….……23

HRV Analyses……....28

HRV Stability…….…....30

Age…..……...……..31

Heart Rate…….……32

Blood Pressure……...33

Pubertal Development……...……...……...34

Growth and Adiposity……...35

Physical Activity……37

Summary……38

Current State of Knowledge…….……..38

HRV Reliability…….38

HRV Stability……....39

Knowledge Gaps…….……...40

Study Aims…….....43

STUDY 1

METHOD…….. 45

Literature Search Strategy……..45

Article Inclusion and Exclusion…….46

Reliability of Article Selection and Coding……...47

Article Coding and Data Extraction……...47

Study and Sample Characteristics……..47

Psychophysiological Recording……48

HRV Pre-Processing and Analysis…….49

Selection of Baseline HRV Measures……49

Statistical Analysis…….50

Effect Size Calculation……..50

Effect Size Management……...…….51

Analytic Strategy……....51

RESULTS……...54

Study Participant and Recording Characteristics……...54

Overall Summary Analyses……....54

Moderator Analyses ……...55

Study Protocol……...55

Sample Characteristics……...57

ECG Signal Acquisition and Pre-Processing…….58

HRV Analyses……...…60

Study Quality…….62

DISCUSSION……64

Study Protocol……64

Sample Characteristics……...67

ECG Signal Acquisition and Pre-Processing…….70

HRV Analyses……....75

Limitations, Strengths, and Current State of the Literature……...79

Conclusion…….86

STUDY 2

METHOD……...89

Cohort Profile……..89

Participants…….90

Measures……90

Anthropometrics……90

Puberty……...91

Blood Pressure……...91

Physical Activity……...92
Heart Rate Variability……93
ECG Signal Acquisition……93
ECG Signal Processing……..93
HRV Analyses……...94
Procedure……...94
Data Screening……...95
Outliers and Statistical Assumptions……95
Missing Data……..96
Analytic Strategy……...97
RESULTS……100
ECG Data Integrity……..101
HRV Stability……...101
Pearson Correlations……101
Intra-Class Correlations……...102
Partial Correlations……..102
Age……...103
Heart Rate……103
Blood Pressure……..104
Puberty……104
Growth and Adiposity……..105
Physical Activity……..105
Secondary *Post Hoc* Analyses……105
DISCUSSION……109
Developmentally Relevant Covariates……109

Summary……..116

Limitations, Strengths, and Future Recommendations……117

Conclusion……...123

GENERAL DISCUSSION……..124

General Limitations……..130

Conclusion……...132

REFERENCES……....134

LIST OF TABLES

Table 1: Continuous Study Descriptive Characteristics by Age Group……...168

Table 2: ECG Recording Characteristics by Age Group……170

Table 3: Categorical Study Descriptive Characteristics by Age Group……..171

Table 4: Overall Summary Analyses for HRV Reliability ……176

Table 5: HRV Reliability – Continuous Moderator Variables by Age Group…….177

Table 6: Infant/Toddler HRV Reliability – Categorical Moderator Variables……179

Table 7: Child/Adolescent HRV Reliability – Categorical Moderator Variables……...183

Table 8: Sample Demographic, Anthropometric, and Physical Activity Measures…….187

Table 9: Descriptive Statistics for Electrocardiogram (ECG) and Heart Rate Variability (HRV) Measures……..189

Table 10: HRV Stability: Intraclass (*ICC*) and Pearson Correlations (*r*) ……...190

Table 11: HRV Stability: Partial Correlations ($r_{ab.c}$), Visit 1 Covariates Only……....191

Table 12: HRV Stability: Partial Correlations ($r_{ab.c}$), Time-Varying Covariates…….192

Table 13: HRV Regression: Standardized Beta Coefficients, Visit 1 Covariates Only……193

Table 14: HRV Regression: Unstandardized Beta Coefficients, Visit 1 Covariates Only……194

Table 15: HRV Regression: Standardized Beta Coefficients, Time-Varying Covariates……...195

Table 16: HRV Regression: Unstandardized Beta Coefficients, Time-Varying Covariates……...196

Table 17: HRV Stability: Intraclass (*ICC*) and Pearson Correlations (*r*) with Imputed Data……197

Table 18: HRV Stability: Partial Correlations ($r_{ab.c}$), Visit 1 Covariates Only with Imputed Data……..198

Table 19: HRV Stability: Partial Correlations ($r_{ab.c}$), Time-Varying Covariates with Imputed Data……..199

LIST OF FIGURES

Figure 1: Flow Chart for Article Identification and Inclusion in Meta-Analysis……200

Figure 2: Forest Plot for HF Reliability……...201

Figure 3: Forest Plot for NN and SDNN Reliability……202

Figure 4: Forest Plot for LF and LF:HF Ratio Reliability……...203

LIST OF APPENDICES

Appendix A……..……...……..204

LIST OF ABBREVIATIONS

BP……...Blood Pressure

BMI……Body Mass Index

ECG……..Electrocardiogram

HRV……Heart Rate Variability

HF…….High Frequency

LF……..Low Frequency

LF:HF Ratio…….Low Frequency to High Frequency Ratio

NN…….Mean NN Interval

pNN50…….Proportion of NN Intervals Greater than 50 ms

rMSSD……...Root Mean Square of Successive Differences

SDNN……...Standard Deviation of NN Intervals

VLF……..Very Low Frequency

Heart rate variability (HRV) is an established measure of autonomic cardiovascular control. Although HRV is commonly reported in studies of health with adult and pediatric samples, the psychometric properties of pediatric HRV measures are not well-established. The overarching goal of this thesis was to conduct a comprehensive examination of test-retest reliability and temporal stability of HRV measures from pediatric samples. A thorough background introduction will first describe physiological and methodological factors pertinent to HRV measurement and interpretation; following this, key research findings relevant to the reliability and stability of HRV will be highlighted and discussed.

Autonomic and Cardiovascular Physiology

The autonomic nervous system is a highly interconnected physiological network. The autonomic nervous system contains a sympathetic and a parasympathetic branch, both of which innervate multiple bodily systems (e.g., heart, lungs, kidney; Guyton & Hall, 2006; Johnson, 2003). Sympathetic nerves originate from thoracic and lumbar spinal regions; these nerves release epinephrine and norepinepherine to target organs (Guyton & Hall, 2006) and activate the "fight or flight" response. "Fight or flight" mobilizes the body for action upon detection of a real or perceived stressor by stimulating certain physiological processes (e.g., cardiovascular activity, respiration) and suppressing others (e.g., digestion; Johnson, 2003; Rogers, 2011). Parasympathetic nerves originate from cranial (e.g., vagus nerve) and sacral spinal regions; these nerves release acetylcholine to target organs (Guyton & Hall, 2006) and activate the "rest and digest" response. "Rest and digest" promotes recovery throughout the body upon removal of a

stressor by suppressing certain physiological processes (e.g., cardiovascular activity, respiration) and stimulating others (e.g., digestion; Johnson, 2003; Rogers, 2011).

Early theories of autonomic nervous system functioning posited that sympathetic and parasympathetic branches act reciprocally, wherein activation of one autonomic branch inhibited the other; this was referred to as "autonomic reciprocity" (Berntson, Cacioppo, & Quigley, 1991). Contemporary research suggests that autonomic functioning is more complex and can exist in states of reciprocal, coactive, or independent functioning, such that sympathetic and parasympathetic activity can be negatively correlated, positively correlated, or uncorrelated; this is referred to as "autonomic determinism" or "autonomic space" (Berntson et al., 1991; Berntson, Quigley, & Lozano, 2007; Reyes Del Paso, Langewitz, Mulder, van Roon, & Duscheck, 2013).

Activity of the sympathetic and parasympathetic autonomic branches on the heart produces quantifiable modulations to cardiovascular activity. This modulation reflects autonomic regulation of homeostasis across bodily systems (Berntson et al., 2007; Guyton & Hall, 2006), and can be captured with measures of heart rate variability (HRV). Measures of HRV can be non-invasively quantified through continuous recordings of the heart's electrical activity using an electrocardiogram (ECG). While HRV can also be derived using photoplethysmography (Lu, Yang, Taylor, & Stein, 2009), blood pressure waveforms (McKinley, Shapiro, Bagiella, Myers, De Meersman, Grant, & Sloan, 2003), or Doppler ultrasound techniques (Jezewski, Kupka, & Horoba, 2008), electrocardiogram methodology is often preferred because it produces a clear and well-defined visual signal of cardiovascular activity from which HRV can be derived

(Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, 1996).

The electrocardiogram produces a characteristic visual signal that represents distinct phases of each heartbeat, or cardiac cycle. A cardiac cycle begins in the right atrium, where electrical discharges from the sino-atrial node are transmitted to the atrio-ventricular node (Johnson, 2003); this is represented on an electrocardiogram as the P-wave. Signals from the atrio-ventricular node propagate through Purkinje fibres, located in the septum wall between the heart's left and right ventricles, and stimulate contraction of the ventricles, which is referred to as ventricular depolarization (Johnson, 2003). This ventricular depolarization is represented on an electrocardiogram as the QRS-complex, with the R-wave typically the most visually prominent signal in the waveform. Finally, the ventricles become repolarised in preparation for the next contraction (Johnson, 2003); this is represented on an electrocardiogram as the T-wave.

While the sino-atrial node has an internal pacemaker that generates its own electrical activity (Cacioppo, Berntson, Binkley, Quigley, Uchino, & Fieldstone, 1994), sympathetic and parasympathetic mechanisms regulate this activity, which results in continuous adjustments to the timing of each heartbeat (Berntson, Cacioppo, & Quigley, 1993; Berntson et al., 2007). As such, HRV reflects variance in the time between successive heartbeats (i.e., R-waves; Berntson, Bigger, Jr., Eckberg, Grossman, Kaufmann, Malik, et al., 1997; Ernst, 2014). This time between beats is commonly referred to as an NN-interval (representing a "normal-to-normal" cardiac cycle), but also called an RR-interval, inter-beat interval (IBI), or heart period (HP). NN-interval data is processed by specialized algorithms, which derive measures of HRV. Several NN-

interval processing techniques are available, each of which provides a slightly unique perspective of autonomic cardiovascular control. Linear-based time- and frequency-domain measures are the most commonly reported in contemporary HRV literature (Ernst, 2014; Kleiger, Stein, & Bigger, Jr., 2005). Other non-linear HRV measures can also be derived using Poincare plot, geometric (e.g., triangular), entropy-based, and wavelet-based analyses.

Frequency-Domain HRV

Frequency-domain analyses examine HRV using specialized algorithms that decompose the NN-interval signal into distinct sinusoidal waveforms (Ernst, 2014). More specifically, spectral analyses determine the frequency amplitude of each sinusoidal waveform as a function of a pre-set bandwidth that dictates the minimum and maximum frequency range to be detected (Berntson et al., 1997). Frequency-domain bandwidths are classified as high-frequency (HF), low-frequency (LF), very-low-frequency (VLF), or ultra-low-frequency (ULF) range. LF:HF ratio and total spectral power can also be derived (Berntson et. al., 2007; Sztajzel, 2004). In the pediatric HRV literature, HF is commonly expressed using indices of vagal tone or respiratory sinus arrhythmia; these are typically defined as reflecting the rhythmic oscillations in heart rate activity occurring at the frequency of respiration mediated by neural projections of the vagus nerve, or tenth cranial nerve, to the sino-atrial node (El-Sheikh, 2005; Porges, Doussard-Roosevelt, Portales, & Suess, 1994). Notedly, measures of HF, vagal tone, and respiratory sinus arrhythmia all reflect the same underlying construct.

Frequency-domain analyses are thought to provide valid estimates of autonomic factors that regulate HRV (Karim, Hasan, & Ali, 2011; Sztajzel, 2004). Frequency-

domain methods isolate variance (i.e., power) within a series of NN intervals into distinct spectral components associated with different frequency bandwidths. Power within a frequency-domain bandwidth reflects the extent to which faster (e.g., HF) or slower (e.g., VLF) physiological mechanisms contribute to variance in heart rate. Mechanisms underlying each frequency-domain HRV measure have been examined using pharmacological blockade studies with chemical agents to partially or totally inhibit activity of one or both autonomic branches. Changes within an HRV measure following blockade provides insight into the extent to which each autonomic branch contributes to that index (Cacioppo et al. 1994). Notedly, when both autonomic branches are completely blocked, there is no variability between heartbeats; thus, HRV is abolished, resulting in a metronome-like heartbeat with equal NN intervals (Akselrod, Gordon, Ubel, Shannon, Barger, & Cohen, 1981).

Physiological mechanisms underlying high-frequency (HF) spectral components are generally well-established. Parasympathetic blockade, using muscarinic receptor antagonists (e.g., atropine), consistently reduces or abolishes power (i.e., variance) in the HF spectrum by dampening vagal and respiratory influences (Berntson et al., 1993; Yasuma & Hayano, 2004) on heart rate. Conversely, sympathetic blockade shows little or no effects on HF (Akselrod et al, 1981; Akselrod, Gordon, Madwed, Snidman, Shannon, & Cohen, 1985; Cacioppo et al., 1994; Chen & Mukkamala, 2008; Martinmaki, Rusko, Kooistra, Kettunen, & Saalasti, 2005). Together these blockade studies demonstrate that HF is chiefly mediated by parasympathetic and respiratory mechanisms.

Physiological mechanisms underlying low-frequency (LF) spectral components are less well-established and are assumed to include activity of both autonomic branches

as well as baroreceptor reflex mechanisms that regulate blood pressure homeostasis (Berntson et al., 2007; Kleiger et al., 2005). Sympathetic blockade using adrenergic receptor antagonists (e.g., propranolol, metoprolol) reduces LF power (Akselrod, 1981; Chen & Mukkamala, 2008), particularly when changing postures (e.g., sitting to standing; head-tilt) where sympathetic activity normally increases to maintain blood circulation in the upper limbs and brain (Cacioppo et al., 1994; Kleiger et al., 2005). However, results are not uniform across studies. Akselrod and colleagues (1985) found that LF increased by ~50% following sympathetic blockade, and decreased by ~90% following parasympathetic blockade; decreases in LF following parasympathetic blockade have also been reported elsewhere (Akselrod, 1981; Cacioppo et al., 1994; Chen & Mukkamala, 2008). These blockade studies provide mixed results about the relative contributions of each autonomic nervous system branch on LF. Current thinking speculates that LF is mediated by a combination of sympathetic, parasympathetic, and baroreceptor reflex activity.

Physiological mechanisms underlying LF:HF ratio are also not well understood. Pharmacological studies demonstrate that blockade of one or both autonomic branches decrease (Cacioppo et al., 1994; Chen & Mukkamala, 2008) or increase LF:HF ratio (Reyes Del Paso et al., 2013). LF:HF ratio is often reported as an index of autonomic balance; however, this interpretation is actively debated because first, LF has both sympathetic and parasympathetic influences, and second, autonomic nervous system activity is not exclusively reciprocal (Reyes Del Paso et al., 2013). As LF and HF each reflect tonic influences of the autonomic nervous system on cardiovascular activity, the LF:HF ratio is more likely an index of autonomic modulation rather than of autonomic

balance. Nevertheless, despite uncertainties in its interpretation, a reduced LF:HF ratio is frequently reported as being associated with poorer health outcomes (Ernst, 2014; Heathers, 2014).

Physiological mechanisms underlying very-low (VLF) and ultra-low-frequency (ULF) spectral components are poorly understood. Blockade studies suggest VLF is mediated by sympathetic and parasympathetic mechanisms, the angiotensin system (which regulates blood pressure and water balance), and mechanisms underlying regulation of temperature and inflammatory markers (e.g., C-reactive protein, white blood cell count; Akselrod et al., 1985; Ernst, 2014; Taylor, Carr, Myers, & Eckberg, 1998). Similar mechanisms may underlie both VLF and ULF (Berntson et al., 2007); however, underlying mechanisms potentially unique to ULF also include physical activity (Serrador, Finlayson, & Hughson, 1999) and circadian rhythmicity, as indicated by long-term ECG studies (e.g., 24 hr; Bigger, Jr., Fleiss, Rolnitzky, & Steinman , 1993; Ernst, 2014; Karim et al., 2011). While underlying mechanisms of VLF and ULF are not fully elucidated, both measures of HRV have clinical utility as predictors of cardiovascular mortality (Bigger, Fleiss, Steinman, Rolnitzky, Kleiger, & Rottman, 1992; Kleiger et al., 2005; Sztajzel, 2004).

Time-Domain HRV

Time-domain measures of HRV are also commonly reported. Time-domain measures are mathematically derived from NN-intervals. The standard deviation of NN intervals (SDNN) is derived from the entire NN-interval series and reflects overall heart rate variability (Malik, 1997; Sztajzel, 2004). Physiological mechanisms underlying SDNN include both sympathetic and parasympathetic inputs. Further, 30–40% of the

variance in SDNN is thought to be attributable to circadian components of HRV (Kleiger et al., 2005). The root mean square of successive NN-interval differences (rMSSD) and the proportion of successive NN-intervals that differ by more than 50 msec (pNN50) are two time-domain methods derived using the differences between adjacent NN intervals. Physiological mechanisms underlying rMSSD and pNN50 chiefly reflect parasympathetic components of HRV (Kleiger et al., 2005; Malik, 1997; Sztajzel, 2004). While other time-domain measures can be derived (e.g., SDNNi, SDANN) to reflect short- and longer-term (e.g., circadian) components of HRV, they are rarely reported in the pediatric literature.

Time- and frequency-domain HRV measures are complementary in that they capture the same underlying autonomic nervous system phenomenon in different ways, and are thus, highly correlated. SDNN is correlated with total spectral power, while rMSSD and pNN50 are correlated with HF (Ernst, 2014; Kleiger et al., 2005; Sztajzel, 2004). The decision to use either time-domain or frequency-domain HRV is often influenced by personal preference, study aims, and the type and quality of NN interval data available (e.g., short- vs. long-term). Examination of HRV yields an indication of important mechanisms associated with autonomic cardiovascular control, and thus, autonomic nervous system functioning; as such, time- and frequency-domain HRV measures provide a useful means of quantifying autonomic dysfunction.

Autonomic Dysfunction, Health, and HRV

Autonomic dysfunction (i.e., sympathetic hyperactivity and/or parasympathetic hypoactivity) is associated with poor health. Autonomic dysfunction occurs when one autonomic nervous system branch persistently dominates over the other (Janig, 2006),

causing a prolonged state of physiological arousal that places excessive demands on the body. This demand results in "wear and tear" of the body, referred to as allostatic load (e.g., Thayer & Sternberg, 2006), which can result in many acute and chronic illnesses. As such, autonomic dysfunction, indicated by reduced HRV, is commonly studied as an index of health in both adults and children. In studies with adults, autonomic dysfunction is evident following a myocardial infarction and in those with cardiac arrhythmias, hyper/hypotension, Type 1 and Type 2 diabetes, and obesity (Boer-Martins, Figueiredo, Demacq, Martins, Consolin-Colombo, et al., 2011; Kleiger et al., 2005; Schmid, Schonlebe, Drexler, & Mueck-Weymann, 2010; Singh, Larson, Tsuji, Evans, O'Donnell, & Levy, 1998; Sztajzel, 2004; Thayer & Sternberg, 2006). Among these studies, autonomic dysfunction is typically indicated by reduced time-domain measures and HF spectral power, as well as increased LF, LF:HF ratio, and heart rate. Autonomic dysfunction is also evident among adults diagnosed with depression, bipolar disorder, schizophrenia, and sleep problems (e.g., insomnia, sleep apnea), relative to controls (Bonnet & Arand, 1998; Bylsma, Salomon, Taylor-Clift, Morris, & Rottenberg, 2014; Henry, Minassian, & Paulus, 2010; Roche et al., 2003). In studies with children, autonomic dysfunction is evident among children with hypertension, obesity, and Type 1 and Type 2 diabetes (Akinci, Celiker, Bayka, & Tezic, 1993; Faulkner, Hathaway, & Tolley, 2005; Gui-Ling et al., 2013; Latchman, Mathur, Bartels, Axtell, & De Meersman, 2011; Martini et al., 2001; Oliveira et al., 2010; Rodriguez-Colon, Bixler, Li, Vgontzas, & Liao, 2011). Among these pediatric studies, and similar to adult studies, autonomic dysfunction is typically indicated by reduced time-domain measures and HF spectral power, as well as increased LF, LF:HF ratio, and heart rate. Autonomic dysfunction is

also evident among children diagnosed with depression, anxiety, sleep problems (e.g., sleep apnea), and internalizing and externalizing behaviour disorders (Baharav, Kotagal, Rubin, Pratt, & Akselrod, 1999; El-Sheikh & Hinnant, 2011; El-Sheikh & Whitson, 2006; Sampei, Murata, Dakeishi, & Wood, 2006).

HRV has been demonstrated as a useful physiological marker of autonomic dysfunction associated with poor or declining health (e.g., risk stratification for mortality in adults following myocardial infarction; Ernst, 2014). However, it currently remains to be established whether autonomic dysfunction is merely reflective of, or is mechanistic to, chronic illnesses such as obesity, diabetes, and cardiovascular disease. Considering the implications of HRV to the study of health, it is pertinent that HRV be accurately measured, analyzed, and interpreted to minimize the probability of erroneous conclusions about health status or treatment effectiveness (i.e., clinical utility) from being made.

HRV Measurement, Analysis, and Reporting

The Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (Task Force, 1996) previously established guidelines for the measurement, analysis, and interpretation of HRV for adults, which were extended in a subsequent committee report (Berntson et al., 1997). While these available guidelines are frequently cited in the HRV literature, wide variability in study methodology remains. Further, theoretical and methodological advances in HRV measurement within the last two decades are not reflected in these guidelines. Moreover, there remains a paucity of research establishing HRV measurement guidelines for infants, children, and adolescents. Some researchers have attempted to refine existing adult guidelines to reflect more recent findings (e.g., Cerutti, Goldberger, & Yamamoto, 2006;

Denver, Reed, & Porges, 2007), and others have aimed to extend these guidelines to pediatric samples (Bar-Haim, Marshall, & Fox, 2000).

Published measurement guidelines detail several methodological conditions which are key to obtaining HRV data that is of suitable psychometric quality for psychophysiological research. One way to conceptually organize these methodological conditions (reviewed below) is by using four overarching categories: study protocol (e.g., methodological protocol), sample characteristics (e.g., exclusion criteria, participant sex), ECG signal acquisition and pre-processing (e.g., time of ECG recording, data editing and filtering), and HRV analyses (e.g., ECG recording duration, frequency bandwidth selection).

Study Protocol

The methodological protocol in which ECG data is obtained must be controlled to minimize internal and external sources of measurement error. There are at least two ways in which the methodological protocol can be controlled to ensure HRV is measured under similar experimental conditions across each study visit. First, instructing participants to refrain from behaviours (e.g., strenuous exercise, alcohol use, eating) that affect autonomic and/or cardiovascular function, allowing all participants to acclimatize or habituate to the experimental situation prior to data collection, and ensuring that each participant maintains a consistent recording posture (to account for established moderating effects of posture on autonomic functioning; e.g., Cacioppo et al., 1994; Kleiger et al., 2005) will each help promote standardized ECG recording conditions. Second, HRV is often higher during rest compared to challenge (e.g., tasks, stressors) and recovery conditions (McNarry & Lewis, 2012; Winsley, Armstrong, Bywater, & Fawkner,

2003), warranting consideration of the ECG recording protocol. Ensuring a standardized ECG recording protocol (e.g., rest, task, recovery conditions) for all participants and across each study assessment can improve reliability in both cross-sectional and longitudinal designs. Taken together, increased methodological control will likely augment data reproducibility by improving the replicability of study conditions, and further help to promote HRV psychometrics.

Sample Characteristics

The homogeneity of a study sample is a key factor in obtaining reliable HRV data. Selection of specific *a priori* inclusion and exclusion criteria permit the researcher to optimize the homogeneity of the sample, while controlling for variables known to influence HRV. For example, medications, and acute (e.g., infection) or chronic (e.g., diabetes) illnesses known to alter cardiovascular functioning can affect resulting HRV measures and should be considered when recruiting participants. Among adults, studies often report stringent exclusion criteria (e.g., medication use; food, caffeine, alcohol; cardiovascular abnormalities; acute/chronic illness; smoking; Cipryan & Litschmannova, 2013; La Fountaine, Wecht, Spungen, & Bauman, 2010; Pinna et al., 2007). Among infants and youth, sample selection criteria are less rigorous or often unspecified (e.g., Kennedy, Rubin, Hastings, & Maisel, 2004; Liew et al., 2011; Nepal & Paudel, 2012; Rigterink, Fainsilber Katz, & Hessler, 2010). A minority of pediatric studies have excluded participants taking prescription medications, or who have cardiovascular conditions (e.g., arrhythmia, hypertension) or chronic illnesses (e.g., metabolic syndrome, obesity; Bar-Haim et al., 2000; Galland, Taylor, Bolton, & Sayers, 2006; Nepal & Paudel, 2012). Further considerations for ensuring a homogenous sample include recruiting

participants with a narrow versus wide age range (e.g., 8–10 yrs vs. 5–18 yrs), and recruiting a balanced ratio of male to female participants to account for possible sex differences in autonomic cardiovascular control. Thus, greater sample homogeneity may help to promote the psychometric quality of HRV.

ECG Signal Acquisition and Pre-Processing

Circadian factors in autonomic activity warrant consideration for the time of day when ECG is recorded. Circadian variations in HRV have been established in adults (e.g., Armstrong, Kenny, Green, & Seely, 2011) and in children (e.g., Massin, Maeyns, Withofs, Ravet, & Gerard, 2000). Sympathetic dominance is normally observed just after awakening, and becomes withdrawn throughout the day, allowing for a progressive increase in resting parasympathetic dominance that is higher at night and peaks just before awakening (Guo & Stein, 2003; Huikruri, Kessler, Terracall, Castellanos, Linnaluoto, & Myerburg, 1990). Considering this, HRV measures obtained at different times of the day, either within or across study assessments, would be more difficult to compare because resulting data may reflect considerably different patterns of autonomic cardiovascular control. Ensuring that the ECG recording time is standardized across each study assessment will minimize confounding effects of circadian variations in HRV and improve data reproducibility in both cross-sectional and longitudinal studies. HRV measurement guidelines do not specify an ideal time for ECG recording, making this a predominantly researcher-based decision.

Reliable and clean HRV data requires reliable and clean ECG signals. Optimal electrode positioning and adherence is required (Jennings, Berg, Hutcheson, Obrist, Porges, & Turpin, 1981; Mulder, 1992), which involves proper skin preparation (i.e.,

abrasive cleaning to remove dirt and dead skin cells) at a site that will be minimally influenced by movement but will also be comfortable for the participant (Shaffer & Combatalade, 2013). Electrode placement reporting should include sufficient anatomical details to facilitate study replication (Jennings et al., 1981). ECG recording hardware should be routinely calibrated and assessed for precision (Berntson et al., 1997). ECG sampling rate is another factor that promotes reliable cardiovascular measurement. Sampling rate (measured in Hertz, Hz) refers to the number of physiological "snapshots" or samples recorded per second and defines the minimal R-wave oscillation amplitude that can be detected (Berntson et al., 1997; Hejjel & Roth, 2004; Riniolo & Porges, 1997). Measurement guidelines recommend 500–1000 Hz or greater as an optimal sampling rate for HRV measurement, although 250 Hz can also be used when additional processing techniques are applied (Berntson et al., 1997; Riniolo & Porges, 1997; Task Force, 1996). Sampling rates that are too slow (e.g., < 100 Hz) or too fast (e.g., > 2500 Hz) are thought to limit the precision of R-wave identification and can contribute unwanted variance, overestimate HRV, or limit reproducibility (Ernst, 2014; Hejjel & Roth, 2004; Riniolo & Porges, 1997).

Following data collection, application of proper editing and pre-processing techniques is essential for data integrity. Artifacts and recording errors are often unavoidable in cardiovascular research and can result from poor electrode adhesion, movement during ECG recording, or aberrant heart beats (e.g., premature ventricular contractions; Shaffer & Combatalade, 2013). Correction of artifacts and recording errors is a necessary prerequisite to HRV measurement, as even a single artifact can disturb the continuity of a recording, violate statistical assumptions of HRV analysis, or falsely

increase the variance in an ECG signal (Berntson & Quigley, 1990; Berntson & Stowell, 1998; Kim, Lim, Kim, & Park, 2007; Kim, Kim, Lim, & Park, 2009; Salo, Huikuri, & Seppanen, 2001). A systematic and reproducible artifact editing procedure should be performed using visual inspection of ECG data by trained technicians. Software-based editing methods are also recommended and include pattern detection algorithms as well as linear, non-linear, and cubic spline interpolations (Berntson et al., 1990; Berntson et al., 1997; Jarrin, McGrath, Giovanniello, Poirier, & Lambert, 2012; Kim et al., 2007, 2009; Salo, Huikuri, & Seppanen, 2001; Porges & Byrne, 1992). Measurement guidelines recommend using both manual and software-based editing procedures and reporting the proportion of data excluded due to artifacts (Berntson et al., 1997; Jennings et al., 1981; Task Force, 1996). Despite these recommendations, few authors report the proportion of data excluded due to artifacts, which precludes knowledge of the integrity and quality, and thus, interpretability of the HRV data. Following artifact correction, the ECG signal must be filtered using digital smoothing procedures. Additional "noise" and overlapping frequency components (i.e., spectral leakage) can magnify the apparent HRV power spectrum and lead to erroneous conclusions regarding autonomic cardiovascular control (Singh, Vinod, & Saxena, 2004). These smoothing algorithms (e.g., triangular or Hanning windows) crop ECG recording segments and remove unwanted sources of variance (Berntson & Quigley , 1990; Jarrin et al., 2012; Kim et al., 2007, 2009; Salo et al., 2001). One software filter commonly used in pediatric HRV studies is Porges' (1985) moving polynomial window. The moving polynomial algorithm filters out additional variance in cardiovascular patterns from an ECG recording and produces a residual time-series that is corrected for the confounding effects of baseline heart rate; application of this filter

potentially eliminates the need to include heart rate as a covariate in HRV analysis (Porges, 1985).

HRV Analyses

Data reduction and analyses decisions can also influence HRV reliability. Measures of HRV are commonly derived using algorithms in commercially available software programs. These programs often permit user-adjusted default settings. Four pertinent user-defined decisions are outlined below; such adjustments must be chosen carefully, as each modification can impact the resulting measures (Berntson et al., 1997).

One important user-defined analysis decision is choosing the ECG recording duration to be analyzed for HRV. ECG recordings must be long enough to sample short- and long-term heart rate variations. Published guidelines recommend that ECG be recorded for a minimum of 1 min to assess HF, 2 min for LF, and ~50 min for VLF (Berntson et al., 1997; Sztajzel, 2004). A 5-min recording duration is a generally accepted standard in short-term studies, while 24-hr recordings are often standard in long-term studies (Kim et al., 2007, 2009; Task Force, 1996). ECG recording duration determines the extent of variability that can be captured within the signal, and thus influences resulting measures of HRV. For instance, a researcher that obtains a 2-min ECG recording may only have enough data available to derive reliable estimates of HF activity. Further, because very-low-frequency oscillations occur in intervals greater than 5 min, short-term HRV recordings are likely unsuitable to adequately capture very-low-frequency sources of variability (Berntson et al., 1997; Heathers, 2014; Kleiger et al., 2005). VLF, which in part reflects longer circadian patterns, is less reliable if the ECG recording duration is too short to adequately capture variability in this frequency range (Heathers, 2014; Kleiger et

al., 2005); a signal occurring at 0.04 cycles per second (Hz) may occur less than 3 times per min (Heathers, 2014). Thus, an ECG recording duration that is appropriate for the variable of interest must be considered *a priori* to improve both the precision and the reproducibility of HRV.

A second important user-defined analysis decision is the selection of ECG data to be analysed for HRV. Following data collection, selection of the ECG recording duration to be analyzed (i.e., data reduction) is pertinent for reliable measures of HRV. An "epoch" is a segment of the ECG recording submitted for analyses. HRV can be derived using the mean of a single analytical epoch, or by analyzing multiple (i.e., ≥ 2) short analytical epochs. For example, a 5-min recording of ECG can be used to derive HRV measures from the mean value of the entire 5-min epoch, the mean value of five 1-min epochs, or the mean value of six 30-sec epochs. Certain HRV measures may be less impacted by the use of short epochs than others (e.g., HF vs. LF) that require more ECG data to properly capture sources of variability, which oscillate at periods longer than the analyzed recording duration (Berntson et al., 2007; McNames & Aboy, 2006; Porges & Byrne, 1992). Some researchers suggest deriving HRV from the mean of several short epochs can reduce the impact of missed or uncorrected artifacts or slower heart rate trends (often more present in longer recordings) that were not accounted for in the data cleaning phase (Izard et al., 1991; Karim et al., 2011; Malik, 1997; Salo et al., 2001). However, exclusive use of shorter, artifact-free recording epochs may introduce a selection bias that would limit the reliability and generalizability of results (Berntson et al., 1997, 2007; Porges & Byrne, 1992). Similar to artifact editing, the decision to derive HRV using a full or

segmented recording epoch should be explicitly reported to enhance the interpretability of results.

A third important user-defined analysis decision is the selection of spectral frequency bands, which dictate the measurement boundaries between spectral variables. Among adults, standard frequency bands are defined as: HF 0.15–0.40 Hz; LF 0.04–0.15 Hz; VLF 0.003–0.04 Hz; ULF <0.003 Hz (Berntson et al., 1997; Ernst, 2014; Task Force, 1996). Importantly, available guidelines for frequency bands among infants and young children are limited to HF (0.2–1.20 Hz or 0.3–1.30 Hz for infants; 0.24–1.04 Hz for young children; Bar-Haim et al., 2000; Fox & Porges, 1985); no guidelines exist for LF or VLF, with adult frequency bands typically applied. If adult HF frequency bands are erroneously applied to pediatric samples, which is not uncommon, resulting HRV data will be an inaccurate estimate of parasympathetic modulation of autonomic cardiovascular control; this would distort interpretations of HF among subsequent analyses. Thus, selection and explicit reporting of frequency bandwidths is also pertinent to the interpretability of results.

A fourth important user-defined analysis decision is the selection of measurement units used to present HRV data. Time-domain HRV is commonly expressed in milliseconds (e.g., SDNN, RSA) or as a percentage (e.g., pNN50). Frequency-domain HRV can be expressed in absolute (ms^2/Hz) units, which reduce dependency on ECG recording length; normalized units (n.u.), which minimize within- and between-subjects variability across the total HRV spectrum and reduce difficulties with comparing HRV between studies or analytical methods by quantifying a measure of proportional change; or, natural log (ln) transformation units, which reduce the effects of skewed distributional

characteristics (Berntson et al., 1997; Healthers, 2014; Sztajzel, 2004). Interpretations can differ based on their respective metrics. For example, HRV in normalized units is expressed in a ratio metric and is derived using data from HRV that has already been expressed in absolute units (Burr, 2007). Thus, HRV in normalized units supplement, rather than substitute, HRV in absolute units. Guidelines suggest that spectral HRV can be more precisely interpreted if both absolute and normalized units are presented together (Task Force, 1996); however, journal space limitations mean few authors in the pediatric HRV literature comply with this guideline, which further contributes to methodological variability across HRV studies and makes results difficult to compare.

As outlined above, obtaining accurate measures of HRV requires careful consideration of multiple methodological factors by the researcher, ranging from participant selection to the metric in which HRV is reported. Guidelines for HRV measurement and analysis were established to promote methodological standardization. Despite these guidelines being widely cited in the cardiovascular literature, there is considerable variability across studies. It remains unknown to what extent these methodological differences may impact HRV's psychometric properties. Few studies to date have thoroughly examined the psychometric properties of HRV, especially in children and adolescents. Given that HRV is commonly utilized as a predictor of future health outcomes, it is important to elucidate factors that can influence the psychometrics of longitudinal HRV measures. Test-retest reliability and temporal stability of HRV are psychometric properties that both relate to the reproducibility of HRV data. There are multiple methodological conditions and user-defined measurement decisions that are likely pertinent to obtaining reliable HRV measures. Further, there are multiple biological

and intra-individual characteristics that may also affect the stability of HRV. An examination of these constructs is pertinent to understanding the extent to which HRV measures change across time.

HRV Reliability

To meaningfully measure HRV, it must be psychometrically reliable and reflect a stable individual difference. Reliability refers to the level of agreement or concordance between results obtained on at least two occasions and/or between two or more measurement contexts (Field, 2009; Weir, 2005; Thompson & Vacha-Haase, 2000; See Appendix A). Classical test theory explains that measured data is comprised of a "true score" and "measurement error"; increased true score variance increases a measure's reliability, whereas increased measurement error decreases reliability (Thompson & Vacha-Haase, 2000). Stability refers to the consistency of a score obtained over time for a given individual (Bornstein & Suess, 2000). Stability is conceptually distinct from reliability as it indicates the extent to which a phenomenon is a stable individual difference.

Study protocol. The reliability of a measure typically decreases as the time between assessments (i.e., study follow-up length) increases (Cohen & Swerdlik, 2002). This observation is consistent with examinations of HRV reliability in both adults and children. Among young adults (M_{age} = 22 yrs), test-retest reliability of HF and LF measured within a single day was higher than reliability measured after a 9-day period (ICC_{LF} = 0.72 vs. 0.48, ICC_{HF} = 0.93 vs. 0.78; Cipryan & Litschmannova, 2013) and higher within a single day than reliability measured across a 24-month period (ICC_{LF} = 0.85 vs. 0.79, ICC_{HF} = 0.89 vs. 0.79; Kowalewski & Urban, 2004). Among infants and

youth, a similar pattern has been observed, but less consistently. For example, Fracasso and colleagues (1994) observed generally decreasing test-retest reliability of HF over time in infants across 2-month (r = 0.50), 3-month (r = 0.55), 5-month (r = 0.60), 6-month (r = 0.15) and 7-month periods (r = 0.10). Perry and colleagues (2012) observed higher test-retest reliability of HF in toddlers across 1 year (4–5 yrs; r = 0.52) compared to 2 years (3–5 yrs; r = 0.34). Finally, El-Sheikh and Hinnant (2011) observed progressively declining test-retest reliability of HF in children across 1-year (r = 0.63), 2-year (r = 0.54), and 3-year follow-ups (r = 0.32). All of these pediatric studies used similar HRV acquisition protocols across their study visits. It is plausible that methodological details not described (e.g., sample heterogeneity, differences in the time of ECG recording) may have also contributed to these declines in reliability over time. These findings suggest HRV reliability decreases with longer test-retest intervals.

Sample characteristics. Reliability of HRV is also sensitive to sample heterogeneity. Sample heterogeneity can result from researchers not excluding participants who consume prescription medication or who are suffering from an illness that may impact cardiovascular functioning. Consumption of several prescription medications or the presence of an acute or chronic illness can each alter sympathovagal balance (e.g., sympathetic-mediated inflammatory responses; Lin et al., 2006). Recruiting a mixed sample of participants, wherein only some take prescription medication or have a medical illness, may augment individual differences in autonomic activity and reduce HRV test-retest reliability. Medications prescribed to adults, including beta-blocker, anticholinergic, and antiarrhythmic drugs, have been associated with mean-level decreases to LF, HF, LF:HF ratio, and SDNN (Penttila, Kuusela, & Scheinin, 2005;

Shaffer & Combatalade, 2013; Task Force, 1996). Buchhorn and colleagues (2012) demonstrated that methylphenidate, a stimulant medication commonly prescribed for children with ADHD, was associated with mean-level decreases in 24-hr HR and higher rMSSD and pNN50. Few studies have directly examined the effects of other medications on HRV or HRV reliability in children. Exposure to certain medications has been associated with changes in sympathovagal balance; thus, it is plausible that studies that measure HRV from a sample wherein only some participants consume prescription medications would yield less reliable measures of HRV. Several acute and chronic illnesses (e.g., infections, hypertension, diabetes) have also been associated with changes in sympathovagal balance (i.e., sympathetic hyperactivity/parasympathetic hypoactivity) and reduced HRV among adults (e.g., Boer-Martins et al., 2011; Ernst, 2014; Henry et al., 2010; Schmid et al., 2010) and children (e.g., El-Sheikh & Whitson, 2006; Latchman et al., 2011; Oliveira et al., 2010; Rodriguez-Colon et al., 2011; Shroff et al., 2013). Indeed, HRV reliability is typically lower among clinical, relative to non-clinical, samples of adults and children (Kochiadakis, Orfanakis, Rombola, Chrysostomakis, Chlouverakis, & Vardas, 1997; Sandercock, Bromley, & Brodie, 2005). Taken together, specifying study exclusion criteria, accounting for medication use, and ensuring that participants are free of acute and chronic illnesses or cardiovascular risk factors will improve sample homogeneity, reduce individual differences in sympathovagal balance, and in turn, promote greater reliability.

Sex differences in HRV have been observed in adults and children, although results are often inconsistent. Among adults, males typically display greater sympathetic dominance (e.g., higher LF, LF:HF ratio), whereas females typically display greater

parasympathetic dominance (e.g., higher HF, rMSSD; Antelmi et al., 2004). Ramaekers and colleagues (1998) demonstrated that adult males have significantly higher mean-level SDNN, SDANN, LF, HF, and LF:HF ratio, even after controlling for basal heart rate. Adult females have also demonstrated greater stress-related decreases in HF and rMSSD, relative to males (Li et al., 2009). In a sample of young adults (N = 44; M_{age} = 20.43 yrs), Sookan and McKune (2012) indicated good to excellent reliability (ICC_{range} 0.77–0.92) for all time- and frequency-domain variables among females, but only weak to good reliability (ICC_{range} = 0.36–0.90) among males. It is plausible that sex differences in HRV reliability also exist among children; evidence for mean-level sex differences in HRV measures among children is scarce. Boys typically display increased mean-level NN and SDNN relative to girls (Faulkner et al., 2003; Silvetti, Drago, & Ragonese, 2001), and male sex has also been associated with increased HF among 9- and 11-year-old children, before and after controlling for differences in weight (El-Sheikh, 2005). However, sex differences between other HRV measures remain to be examined. Sex-specific shifts in sympathovagal balance across childhood may relate to developmental changes in sex hormone concentrations (e.g., testosterone, estrogen), which have been shown to exert unique effects on blood pressure and heart rate between boys and girls (Spear, 2000).

ECG signal acquisition and pre-processing. A 24-hr circadian rhythm of HRV has been observed in infants as early as 7 to12 weeks old (Hoppenbrouwers, Oliveira, Sandarupa, Khoo, Neuman, & Ramanathan, 2012). Studies with adults (e.g., Burges, Trinder, Kim, Luke, & Helen, 1997) and children (e.g., Massin et al., 2000) provide compelling evidence for diurnal variations in HRV; however, few studies have specifically examined HRV reliability in relation to the time of day data were collected.

Hurikuri and colleagues (1990) observed reproducible circadian variations in SDNN using data from repeated 24-hr recordings analyzed in consecutive 1-hr epochs. While similar studies are not available in the pediatric literature, mean-level changes in HRV across day, night, and 24-hr recordings suggest circadian factors impact HRV reliability. Rabbia and colleagues (2003) demonstrated that measures of HF and VLF in 12 non-clinical adolescents were higher at night compared to the day, while the opposite result was observed for heart rate and LF:HF ratio and no differences were observed for LF. The notion that circadian factors can influence HRV psychometrics is supported by observations that suggest HRV measures differ depending on the time of data collection. Considering this, consistency in the timing of HRV recordings across test-retest observations may be pertinent to HRV reliability.

Allowing participants to acclimatize (or habituate) to their surroundings prior to ECG recording is recommended by HRV measurement guidelines (Berntson et al., 1997; Task Force, 1996); however, few studies report including a designated pre-recording acclimatization period. Resting measures of HRV provide insight into baseline autonomic cardiovascular control, and are also a staple from which to compare change. However, even if participants are instructed to rest quietly during baseline recordings, they may not actually be fully at rest (i.e., anxious, humming quietly, ruminating). Unfamiliarity and/or discomfort with psychophysiological research protocols may be stressful and alter sympathovagal balance and respiration, which would hinder the collection of a stable baseline recording; this may be especially true among infants and young children. Ensuring an adequate resting baseline or "vanilla" baseline recording (wherein participants complete a standardized, minimally-demanding comparison task; Jennings et

al., 1992) may not be enough to neutralize inter-individual differences in reactivity to the testing situation (e.g., Ottaviani, Shapiro, Davydov, & Goldstein, 2008; Ottaviani & Shapiro, 2011; Salomon, Matthews, & Allen, 2000; Sharpley, 1993). A pre-recording acclimatization period helps participants to familiarize with their surroundings and to habituate to the physiological sensors prior to data collection; as familiarity with study conditions increases, sympathetic influences on heart rate will progressively decline and reduce cardiovascular reactivity to the unfamiliar (e.g., study environment, awareness of ECG sensors). Using adult ECG data, Sharpley (1993) noted that fifteen minutes may be a sufficient duration for reducing heart rate activity to resting levels; however, he argued that a standardized acclimatization duration may be less useful than a duration unique to each participant's distinctive cardiovascular profile. Data collection that follows a sufficient acclimatization period may yield more accurate and reproducible estimates of HRV because of decreased inter-individual variability in heart rate activity within and across study assessments. Thus, it is plausible that including a pre-recording acclimatization period is relevant to producing reliable measures of HRV.

HRV is influenced by postural position (e.g., supine, seated, standing). Parasympathetic dominance is often observed in seated or supine positions, while sympathetic dominance is often observed in standing or head-up-tilt positions (Cacioppo et al, 1994; Kleiger et al., 2005). Among adults (M_{age} = 22 yrs) measured across 24 months, HF reliability, but not LF reliability, was notably higher in supine versus standing postures (ICC_{HF} = 0.89 vs. 0.79; ICC_{LF} = 0.81 vs. 0.79; Kowalewski & Urban, 2004). Among children (M_{age} = 9.5 yrs), mean-level increases in NN, SDNN, rMSSD, and pNN50 have been observed in supine relative to standing postures (Nepal & Paudel,

2012), suggesting that posture may affect HRV reliability in pediatric samples. Indeed, among children (M_{age} = 11 yrs) measured across 2 weeks, HR and SDNN reliability was higher in supine versus standing postures (ICC_{HR} = 0.78 vs. 0.65; ICC_{SDNN} = 0.79 vs. 0.69), whereas HF and LF reliability were minimally affected by posture (ICC_{HF} = 0.77 vs. 0.73; ICC_{LF} = 0.66 vs. 0.63; Dietrich et al., 2010). Conversely, among adolescent males measured across 1 year, HF reliability, but not LF reliability, was higher in supine versus standing postures (r_{HF} = 0.37 vs. 0.25; r_{LF} = 0.37 vs. 0.31; Mezzacappa et al., 1997). Sympathetic modulation may exhibit greater consistency across recording postures, whereas parasympathetic modulation may be more sensitive to postural changes. Considering the above results for HF and LF between children (Dietrich et al., 2010) and adolescents (Mezzacappa et al., 1997), developmental differences in sympathovagal balance may be relevant to the study of recording posture and HRV reliability. Taken together, HRV reliability (except for LF) appears to be influenced by posture.

ECG sampling rate defines the minimal R-wave oscillation amplitude that can be detected within a heart rate recording (Berntson et al., 1997; Hejjel & Roth, 2004). Sampling rates that are too slow (e.g., < 100 Hz) or too fast (e.g., > 2500 Hz) may have increasingly negative effects on the precision of HRV estimation (e.g., increased error variance, limited R-wave detection accuracy; Berntson et al., 1997). With decreasing sampling rates, a dose-response inflation in VLF, LF, HF, LF:HF ratio, SDNN, and rMSSD has been demonstrated using adult (Ernst, 2014) and preterm infant data (Riniolo & Porges, 1997). Selection of an appropriate ECG sampling rate for HRV may be age- and sample-dependent (Merri, Farden, Mottley, & Titlebaum, 1990; Singh, Vinod, &

Saxena, 2004); it is possible that a threshold in ECG sampling rate exists among differently-aged samples, such that obtaining reliable and accurate HRV measures may require sampling rates to be adjusted based on the characteristics specific to a study population, and based on changes among these characteristics over time. Considering this, it is plausible that ensuring an adequate sampling rate for detection of cardiovascular activity contributes to HRV reliability.

Artifact editing should be performed carefully, and by using both manual visual inspection and software-based filtering algorithms. Guidelines for HRV measurement convey the importance of editing artifacts and recording errors prior to analysis (e.g., poor electrode adherence, participant movement, background physiological activity; Berntson et al., 1997; Task Force, 1996). Disturbances in the continuity and variance of an ECG recording caused by artifacts are known, and their effects on HRV have been widely discussed and examined using real and simulated data (Berntson et al., 1990; Kim et al., 2007, 2009; Mulder, 1992; Porges & Byrne, 1992; Salo et al., 2001). Berntson and Stowell (1998) demonstrated that HF becomes inflated by almost three natural log units in the presence of a single ECG artifact. Salo and colleagues (2001) demonstrated that SDNN, LF, and HF had different sensitivities to the amount (0–50%) and type (deletion vs. interpolation) of editing, and that these editing procedures had different effects between recordings from non-clinical and clinical (myocardial infarction) participants. Among samples of adults (Kroll, Freed, & Stein, 1996) and children (Batten, Urbina, & Berenson, 2000), HRV reproducibility has been demonstrated despite inter-rater variability in the manual editing process. Following manual editing, NN-intervals must be digitally filtered and smoothed to correct for problems with stationarity and spectral

leakage (Task Force, 1996). Random "noise" (e.g., slow non-stationary trends) and overlapping frequency components (i.e., spectral leakage) contribute unwanted variance to ECG recordings, which could compromise estimation of HRV and further reduce reliability of the measure. Common filtering techniques include triangular, Hanning/Hamming, or moving polynomial windows (Berntson et al., 1997; Porges, 1985). Improved accuracy and reliability of HRV have been demonstrated using adult data following digital filtering (Lee & Yu, 2010; Singh et al., 2005). Studies examining HRV reliability in the context of different (or absent) software-based filtering algorithms among children remain to be completed; however, it is plausible that HRV reliability will be reduced if NN-interval data is not properly filtered.

HRV analyses. HRV measures are sensitive to the duration of ECG data from which they are derived (e.g., SDNN, VLF; see above section on HRV measurement). In turn, reliability of HRV is also influenced by ECG recording duration. Among adults, Hurikuri and colleagues (1990) demonstrated that SDNN can be reliably measured when using the mean of repeated 24-hr ECG recordings (coefficient of variation [CV] = 7 ± 6%). Further, Marks and Lightfoot (1999) demonstrated that LF was more reliable when derived from the mean of repeated 5-min, compared to 2.5-min recordings (*ICC* = 0.82 vs. 0.78). Among infants, Richards (1995) demonstrated a positive association between HF reliability and ECG recording duration, with durations of 3 min or longer showing the highest reliabilities (>0.8). Taken together, studies suggest that HRV reliability is influenced by the ECG recording duration; it is plausible that reliability of some measures (e.g., mean NN interval) would decrease with longer recordings, whereas reliability of other measures (e.g., VLF) would increase or remain unaffected (e.g.,

LF:HF ratio). The consistency of ECG recording duration within and across study assessments is another important consideration for HRV reliability. Dalla Pozza and colleagues (2006) reported higher correlations between pediatric HRV measures derived from two 10-min ECG recordings compared to measures derived from a 10-min and 24-hr recording (r_{NN} = 0.81 vs. 0.65; r_{HF} = 0.69 vs. 0.48; r_{LF} = 0.88 vs. 0.54). Thus, HRV reliability may also depend on whether identical ECG recording durations were analyzed across study follow-up assessments. Finally, ECG data reduction methods (i.e., selection of an ECG analysis epoch) are pertinent. Among adults, McNames and Aboy (2006) demonstrated that HF was reproducible across ECG epochs ranging from 10 sec to 10 min, while measures reflecting LF activity displayed poor reproducibility with epochs less than 10 min. Among infants, Richards (1995) analyzed HF using 75-sec recording samples and observed higher HF reliability when measures were derived using fifteen 5-sec samples compared to five 15-sec samples. Thus, it is plausible that ECG data reduction techniques (i.e., selection of single or multiple analytical epochs) are pertinent to obtaining reliable HRV measures.

Spectral frequency bandwidths dictate the measurement boundaries between spectral variables and are intended to capture specific physiological underpinnings of HRV. Recommended frequency bandwidths for common HRV variables are available for adults (Ernst, 2014), whereas a recommended bandwidth for HF is available only for samples of young children (Bar-Haim et al., 2000; Fox & Porges, 1985). The selection of an appropriate frequency band is pertinent and should be carefully considered prior to analysis. Erroneous application of frequency bandwidths (i.e., using adult frequency band cutoffs with pediatric samples), results in data that will not constitute a reliable measure

of HRV because NN-interval oscillations occurring at higher or lower frequencies than the selected frequency band can detect would distort the data and produce inaccurate estimates of variability. Thus, it is plausible that selection of appropriate frequency bandwidths is pertinent to HRV reliability, although this has yet to be examined.

Taken together, evidence suggests that HRV is a moderately reliable measure of cardiac autonomic control; however, as the majority of this evidence comes from studies with adults, relatively less is understood about HRV reliability in children. Considerable variability within and between study protocols exists, which makes it difficult to determine which methodological factors are most salient for HRV reliability. Given the available data from adult and pediatric HRV studies, it is plausible that study follow-up length; extent of sample homogeneity and standardization of study protocol; ECG recording posture, sampling rate, and recording duration; collection of stable baseline data; and, selection of appropriate frequency bandwidths are each methodological considerations that are pertinent for promoting HRV reliability.

HRV Stability

Stability examines the extent to which changes in HRV over time may be attributed to developmental differences within individuals; understanding stability is important when considering trajectories of individual differences in psychophysiological development (El Sheikh, 2005; See Appendix A). Several researchers have examined HRV stability in children due to its developmental relevance for the prediction of future health outcomes, including hypertension, obesity, and Type 1 and Type 2 diabetes (Faulkner et al., 2005; Kleiger et al., 2005; Latchman et al., 2011; Martini et al., 2001; Thayer & Sternberg, 2006). Multiple, developmentally-relevant factors have been

previously associated with HRV in children, including age, heart rate, blood pressure, pubertal status, anthropometric characteristics (e.g., height, weight), and physical activity. Consideration of these factors may be pertinent to understanding HRV stability.

Age. Parasympathetic-mediated measures of HRV increase across infancy and early childhood, and subsequently decline across adolescence into adulthood (Longin et al., 2009; Lenard, Studinger, Mersich, Kocsis, & Kollai, 2004; Silvetti et al., 2001). Considering these mean-level changes in HRV across development, several studies have examined stability of measures reflecting the mean NN-interval and HF across infancy and childhood; relatively less data is available across adolescence. Measures of the mean NN-interval ("NN") do not appear consistently stable in young children. For example, NN stability was weak when measured from 2-months-old to 5-years-old ($r = 0.08$; Bornstein & Suess, 2000), but moderate when measured from 4-months-old to 4-years-old ($r = 0.38$; Bar-Haim et al., 2000). Among younger children, stability of HF has been examined across 1- to 3-months-old ($r = 0.12$–0.24; Hsu & Porter, 2004; Porter, Bryan, & Hsu, 1995), 3- to 6-months-old ($r = 0.36$), 6- to 12-months-old ($r = 0.62$; Propper et al., 2008), 2- to 4.5-years-old ($r = 0.57$; Calkins & Keane, 2004), 5- to 6-years-old ($r = 0.53$), and 6- to 7-years-old ($r = 0.62$; Heathers, Fink, Kuhnert, & de Rosnay, 2012). Among older children, HF stability has been examined in samples aged 8- to 10-years-old ($r = 0.79$; Hinnant & El-Sheikh, 2009), 8- to 11-years-old ($r = 0.56$; Hinnant, Elmore-Staton, & El-Sheikh, 2011), 9- to 11-years-old ($r = 0.49$; El-Sheikh, 2005), and among adolescents aged 15 to 17 years across 6 months ($r_{HF} = 0.66$; $r_{LF} = 0.65$; $r_{SDNN} = 0.66$; Henje Blom, Olsson, Serlachius, Ericson, & Ingvar, 2009). Notedly, most of these studies contained small sample sizes ($N < 71$), analyzed less than 5 min of ECG, did not report

several methodological details (e.g., time of ECG recording, sampling rate), and only two (El-Sheikh, 2005; Henje Blom et al., 2009) included developmentally-relevant covariates in their analyses. Taken together, moderate HRV stability is typically observed among pediatric samples despite the mixed methodological quality of available studies. Further, reported results are consistent with previous findings of mean-level changes in HRV across development, and suggest a positive association between HRV stability and advanced age.

Heart rate. Developmentally, heart rate activity (the inverse of the mean NN interval) is typically highest during infancy, and decreases steadily with advancing age (Fleming et al., 2011; Lenard et al., 2004). Heart rate is largely influenced by activity of the sympathetic autonomic branch, whereas HRV is influenced by both sympathetic and parasympathetic autonomic branches. Significant negative associations between heart rate and HRV are typically reported among samples of adults (e.g., β_{range} = -0.17 – -0.38; Tsuji et al., 1996) and older adults (e.g., r_{range} = -0.39 – -0.55; Tsuji et al., 1994); however, results are more variable among children (e.g., β_{range} = -0.40 – 0.74; Jarrin et al., In press). As developmental declines in heart rate are typically observed between infancy and adolescence, and faster heart rate activity is typically associated with lower HRV, it is plausible that heart rate is an important consideration in examinations of HRV stability. A study by Henje Blom and colleagues (2009) measured HRV twice across 6 months and reported higher stability after heart rate was included as a covariate: HF (r = 0.54 vs. 0.66), LF (r = 0.62 vs. 0.65), SDNN (r = 0.64 vs. 0.66). Given the available evidence, researchers generally agree on the relevance of including heart rate as a covariate in HRV analyses, especially in the absence of specialized filtering algorithms

which account for differences/changes in heart rate, such as a moving polynomial filter. However, including heart rate as a statistical covariate in examinations of HRV stability is uncommon. As the association between heart rate and HRV may change across time, accounting for differences in heart rate across periods of child development may augment estimates of HRV stability.

Blood pressure. Blood pressure becomes elevated as children develop into adolescents and adults (Chen, & Wang, 2008), which likely reflects compensatory cardiovascular mechanisms that augment blood flow to meet increasing bodily demands. Changes in systolic (SBP) and diastolic (SBP) blood pressure across development are likely related to similar changes in HRV. In a cross-sectional study, Tanaka and colleagues (2000) demonstrated that measures of SBP and DBP were negatively associated with HF and positively associated with LF:HF ratio among adolescents (M_{age} = 13.5 years; HF β_{SBP} = -0.51; β_{DBP} = -0.49; LF:HF ratio β_{SBP} = 0.52; β_{DBP} = 0.33), but not among pre-adolescents (M_{age} = 9.0 years; HF β_{SBP} = 0.19; β_{DBP} = 0.06; LF:HF ratio β_{SBP} = -0.13; β_{DBP} = 0.01), after adjusting for age, weight, and height. Longitudinal studies examining the relation between blood pressure and HRV among children and adolescents are rare. Campbell and colleagues (2003) demonstrated that SBP measured at 14-years-old was positively correlated with LF (r = 0.15) and LF:HF ratio (r = 0.09) measured at 22-years-old, whereas DBP measured at 14-years-old was negatively correlated with LF measured at 22-years-old (r = -0.03); however, none of the correlations reached statistical significance, and thus must be interpreted with a degree of caution. It is likely that increasing sympathetic dominance across child development is associated with modulation of both BP and HRV. Baroreceptor reflex mechanisms, which detect

moment-to-moment increases or decreases in BP, trigger autonomic responses to modulate cardiac output. Further, decreased vagal tone (e.g., across older samples) results in vasodilation, which decreases BP and baroreceptor reflex control (Berntson et al., 2007; Tanaka, Borres, Thulesius, Tamai, Ericson, & Lindblad, 2000). Thus, as BP and HRV are both controlled by autonomic modulation of cardiovascular activity, developmental changes in BP are likely associated with HRV stability.

Pubertal development. Puberty is a critical developmental period when significant physical, biological, and psychological changes occur (Shirtcliff, Dahl, & Pollak, 2009). The timing of puberty is different between girls and boys, such that girls often begin puberty 2 to 3 years earlier (Patton & Vines, 2007; Shaffer, Wood & Willoughby, 2005). Puberty triggers the release of hormones from the pituitary gland (e.g., human growth hormone, estrogen, testosterone) that contribute to physical, biological, and sexual maturations. More data are available regarding the effects of estrogen, relative to testosterone, on autonomic functioning; however, literature reviews have been inconsistent about whether estrogen depresses (Kajantie & Phillips, 2006) or augments (Ordaz & Luna, 2012) sympathetic activity. Nevertheless, evidence suggests that sex hormones released during puberty impacts blood pressure and heart rate (McCabe, Porges, & Carter, 1981; Spear, 2000) and may also impact HRV. El-Sheikh (2005) demonstrated that parent-reported pubertal status was positively associated with cross-sectional measures of HF obtained from children at 9- (r = 0.22) and 11-years-old (r = 0.28); children in this study were mostly pre-pubertal at 9-years-old, whereas only girls were classified as early-pubertal at 11-years-old. Further, pre-pubertal children exhibit lower mean-level values of LF, SDNN, and SDANN, and higher HF, relative to

post-pubertal children (Chen et al., 2012; Kowalewsky, Alifier, Bochen, & Urban, 2007). Thus, a shift in sympathovagal balance (i.e., towards greater sympathetic dominance) appears to occur during puberty. Consistent with this, Jarrin and colleagues (In press) demonstrated that increased pubertal development (as measured by trained nurses using Tanner stages) was associated with decreased rMSSD (β = -0.08), pNN50 (β = -0.08), and HF (β = -0.06), and increased LF:HF ratio (β = 0.11). Taken together, pediatric studies demonstrate clear associations between pubertal development and HRV (although any sex-specific differences in the mechanisms and timing of this association remain to be examined). Thus, it is plausible that pubertal development is also associated with HRV stability.

Growth and adiposity. As children grow, they naturally become taller and heavier (Adair, Martorell, & Stein, 2009; De Onis & World Health Organization, 2006; Shaffer, Wood, & Willoughby, 2005). Developmental changes in height and weight are normal occurrences, and both increase cardiovascular demand because the heart is required to maintain circulation throughout a larger frame; a possible explanation for this relates to the increased production of growth hormone during pubertal development. Although studies in children have not directly examined the relation between growth hormone and HRV, adults with growth hormone deficiencies have decreased sympathetic activity relative to normal controls (Leong, Mann, Wallymahmed, MacFarlane, & Wilding, 2000). Considering this, greater bioavailability of growth hormone may further contribute to a shift towards sympathetic dominance across child development. Seppälä and colleagues (2013) reported significant correlations between height and HRV (r_{range} = -0.15 – 0.14) in a large sample of children; however, they did not report which HRV

variables were more associated with height than others, which limits the ability to draw sound conclusions. A study by Rodriguez-Colon and colleagues (2011) with 9-year-old children demonstrated that height percentile was negatively associated with HF (β = -0.11), LF (β = -0.07), SDNN (β = -3.20), and rMSSD (β = -4.88), and positively associated with LF:HF ratio (β = 0.07) and heart rate (β = 0.47; non-significant), based on a regression model including age, sex, race, sleep apnea, and sampling weight. Thus, it is plausible that height is also associated with HRV stability. Rodriguez-Colon and colleagues (2011) also demonstrated that weight percentile was negatively associated with HF (β = -0.15), LF (β = -0.11), SDNN (β = -4.50), and rMSSD (β = -5.71), and positively associated with LF:HF ratio (β = 0.05) and heart rate (β = 1.12); associations between waist circumference and HRV were similar to those between weight and HRV, but these analyses were not adjusted for similar covariates. Jarrin and colleagues (In press) also demonstrated that central adiposity (i.e., waist circumference) in children was positively associated with LF (β = 0.12) and LF:HF ratio (β = 0.18), even after adjusting for age, sex, puberty, heart rate, blood pressure, TV/computer screen time, physical activity, household income, and parental education. Longitudinal studies further support the relation between weight and HRV. Graziano and colleagues (2011) reported that reduced HF regulation in 5.5-year-olds was associated with being overweight 5 years later. Further, Campbell and colleagues (2003) reported that BMI at 14-years-old correlated negatively with HF (r = -0.10) and positively with LF:HF Ratio (r = 0.23) when participants were followed up at 22-years-old. Notedly, there is considerable intra-individual variability in the timing and rate of growth among children across different periods of development. Considering this, although weight and other measures of

adiposity have been associated with HRV, it is plausible that the *change* in these measures across time is particularly relevant to examinations of HRV stability.

Physical activity. Exercise promotes regularity and coordination among contractions of the heart's ventricles and augments balance of autonomic cardiovascular control (Stein & Kleiger, 1999). Physical activity has been positively associated with parasympathetic dominance among adults (Sandercock, Bromley, & Brodie, 2005). However, studies in children often fail to show significant differences in HRV when measured before and after exercise (e.g., Gamelin et al., 2009; Winsley et al., 2003), suggesting that acute physical activity may not be associated with HRV stability in early development. However, there is evidence for an effect of overall physical fitness and HRV. Among adults, Gregoire and colleagues (1996) demonstrated that participants who engage in regular physical activity display higher parasympathetic and lower sympathetic activity at rest, compared to age- and sex-matched sedentary controls. Among children, Hutin and colleagues (2000) demonstrated that rMSSD measured from obese children steadily increased across a 4-month period of regular (4–5 days/week) exercise, but decreased across a subsequent 4-month period when training was halted. Among adolescents, Henje Blom and colleagues (2009) demonstrated that physical activity is negatively associated with heart rate (r = -0.44), and positively associated with LF (r = 0.29), HF (r = 0.32), and SDNN (r = 0.37), based on partial correlations controlling for BMI, capillary glucose, and SBP; physical activity accounted for 13–17% of the variance among HF, LF, and SDNN in regression models. Sustained maintenance of physical activity may promote parasympathetic dominance, whereas sustained sedentary

behaviour may augment parasympathetic withdrawal. It is plausible that changes in physical fitness over time (e.g., joining team sport) could influence HRV stability.

Summary

Research from the adult (e.g., Tanaka et al., 2000; Tsuji et al., 1994, 1996) and pediatric literature (e.g., Henje Blom et al., 2009; Jarrin et al., In press; Rodriguez-Colon et al., 2011) has typically demonstrated that HRV is positively associated with more advanced age and greater physical fitness, and negatively associated with increased heart rate, blood pressure, pubertal status, and physical growth (e.g., increased weight). It is possible that these same factors influence the stability of pediatric HRV measures. Given that stability reflects consistency in a measure across development, it is plausible that accounting for individual developmental differences among measures of age, heart rate, blood pressure, pubertal status, growth, and physical activity across time would augment HRV stability. The psychometric stability of HRV can likely be improved if researchers account for developmental changes in factors associated with HRV by increasing methodological or statistical control. A poor understanding of developmental factors that affect the psychometrics of HRV may result in erroneous conclusions about the utility of predicting health outcomes using longitudinal HRV data.

Current State of Knowledge

HRV reliability. Available evidence from the adult and pediatric HRV literature suggested that HRV is a reliable measure of autonomic cardiovascular control. The majority of evidence for HRV reliability came from studies with adults. As such, evidence for HRV reliability among samples of infants, children, and adolescents was relatively less abundant. Most reviewed studies arrived at the same conclusion of

moderate HRV reliability. However, there was often considerable methodological variability across the adult and pediatric HRV literature (e.g., heterogeneous samples, distinct ECG recording protocols, different ECG recording durations [e.g., < 5-min, 10-min, 24-hr recordings]). This methodological variability makes it challenging to draw meaningful comparisons about HRV reliability reported by different studies. Moreover, there was a paucity of research across both the adult and pediatric HRV literature that sufficiently conformed to published HRV measurement guidelines. Many of the reviewed studies failed to report pertinent study protocol (e.g., participant inclusion/exclusion criteria, time of ECG recording), sample characteristics (e.g., anthropometrics, ethnicity), and technical details (e.g., excluded NN-interval data, HRV derivation method). Minimal conformity to published guidelines reduces research standardization limits the interpretability of obtained results and yields greater potential for measurement error. Nevertheless, studies examining HRV reliability demonstrated that multiple methodological factors merit consideration when measuring HRV. Evidence suggested that methodological factors relevant to sample characteristics, study protocol, ECG signal acquisition and pre-processing methods, and HRV analysis techniques are relevant to producing reliable measures of HRV. In sum, while moderate HRV reliability was typically reported among studies with adults and children, poor reporting practices and limited research standardization were noteworthy concerns. As such, it is difficult to determine whether certain methodological features of a study made more of a contribution to HRV reliability than others.

HRV stability. Available evidence from the pediatric HRV literature suggested that HRV reflects a moderately stable individual difference in children. HRV stability

was typically reported as weak-to-moderate among newborns and very young infants; moderate stability was observed by approximately 6-months-old, and was retained across childhood development into early adolescence. Most examinations of HRV stability have focused on HF; few studies have examined stability using other HRV measures. Further, a review of the extant pediatric HRV literature indicated that mean-level differences in HRV have been associated with developmentally-relevant, individual difference factors such as age, pubertal development, heart rate, blood pressure, height, weight, and physical fitness. However, given that only a few pediatric HRV researchers have examined stability while statistically controlling for such developmental factors, it remains to be tested empirically whether these same factors alter HRV stability. Studies that have included covariates in examinations of stability (El-Sheikh, 2005; Henje Blom et al., 2009) suggested that controlling for developmentally-relevant moderators would augment HRV stability. In sum, while moderate HRV stability is typically reported among children, studies were limited by a somewhat restricted focus on one measure of HRV, and a lack of statistical control of developmentally-relevant moderators, which have been previously associated with HRV. As such, it is difficult to determine whether these individual difference factors are also associated with HRV stability.

Knowledge Gaps

The current literature suggests that HRV is a reliable and stable measure of autonomic cardiovascular control among pediatric samples. However, three primary knowledge gaps regarding the psychometrics of HRV in the pediatric literature exist. First, studies examining HRV reliability have typically compared only limited methodological conditions (e.g., paced vs. spontaneous breathing; supine vs. standing

recording postures; rest vs. exercise). Mean-level differences in HRV (and heart rate) have been associated with methodological conditions such as whether or not participants were able to acclimate prior to data collection, the consistency of an ECG recording duration, and the application of editing and digital filtering techniques. It is intuitive that changing a study methodology across test-retest intervals (e.g., inconsistent ECG recording times, different ECG recording postures, unstandardized recording protocols) could compromise reliability. However, it remains to be empirically examined whether reliability of HRV is robust to methodological inconsistencies across time; this would help elucidate whether the consistency of some methodological factors are more salient to promoting reliability than others. Given that replication of initial study conditions is pertinent for ensuring reproducibility of results, it can be hypothesized that studies that maintain greater research standardization would yield more reliable measures of HRV compared to studies that use less consistent methodology. Further, the specific methodology used to derive HRV measures is relevant for deriving HRV measures that accurately reflect autonomic functioning. Using an appropriate ECG sampling rate, applying necessary editing and digital filtering to NN-interval data, and selecting appropriate frequency bandwidths are each pertinent methodological factors detailed by published HRV guidelines. Although it remains to be empirically examined whether or not a consideration of such methodological factors is relevant to improving the reliability of HRV measures, it can be hypothesized that studies that conform to published guidelines would yield more reliable measures of HRV.

Second, studies examining HRV stability typically do not examine the influence of developmentally-relevant covariates. Multiple physiological and behavioural factors

have been associated with mean-level differences in HRV among children in cross-sectional and longitudinal studies. However, few researchers control for age, heart rate, weight, or other developmentally-relevant covariates when examining HRV stability. Examining how such factors influence HRV stability is relevant to establishing whether HRV reflects a stable individual difference during childhood, and to understanding the strengths and limitations of using HRV as a predictor of future health outcomes. As such, it can be hypothesized that controlling for the influence of individual differences among factors such as age, heart rate, and weight, among others, could augment HRV stability by accounting for extraneous variance in longitudinal measures. Moreover, examinations of HRV stability chiefly focus on measures of HF relative to other time- and frequency-domain measures. As such, evidence for stability of other HRV measures (e.g., LF, rMSSD) that provide additional information regarding autonomic cardiovascular control is lacking. Given that HF has been previously demonstrated as a stable measure among children, it can be hypothesized that other measures of HRV in children, particularly those reflecting parasympathetic modulation, would exhibit a similar degree of stability.

Third, studies examining HRV stability typically use statistical methods that provide a somewhat restricted understanding of this psychometric property. Most studies that examine HRV stability use correlational techniques, which only account for rank order changes between individuals across test-retest intervals, not magnitude of mean changes. Exclusive use of these analytical methods may limit a thorough understanding of stability (i.e., stability as reflected only in the context of relative rank-order differences) because a consideration of mean-level changes is also pertinent. Statistical methods that account for mean-level changes and random error in longitudinal data (such

as the intra-class correlation) have utility in providing a more complete picture of HRV as a stable individual difference. However, examination of HRV stability using more advanced techniques is not common practice. Given the longitudinal associations observed between HRV and other biological processes (e.g., heart rate, puberty), examining whether simultaneous changes in measures of HRV and in developmentally-relevant covariates across time may be pertinent to understanding HRV stability. As such, it can be hypothesized that accounting for developmental changes among specific within-person factors would yield more meaningful information regarding stability of HRV in children. Taken together, a better understanding of the psychometric properties of HRV measured in infants, children, and adolescents requires an examination of methodological and intra-individual factors that could affect the reliability and stability of HRV, respectively.

Study Aims

The overarching aim of the present thesis was to conduct an empirical evaluation of the psychometric properties of time- and frequency-domain HRV measures in pediatric populations, from infancy through adolescence. The specific goals of this research were: (1) to synthesize available evidence for the test-retest reliability of time- and frequency-domain HRV measures from the extant literature, (2) to examine methodological factors that impact the reliability of HRV measures among samples of infants, children, and adolescents, and (3) to examine intra-individual and developmental factors that impact the temporal stability of time- and frequency-domain HRV measures among children. These aims were achieved through two complementary studies. The first study was a comprehensive meta-analysis of the existing pediatric literature to evaluate

the test-retest reliability of HRV. The goal of this study was to consider the impact of study methodology on the reliability of HRV measurement. The second study examined the 2-year stability of HRV measures obtained from children who had participated in a large, longitudinal cohort study within Quebec. The goal of this second study was to assess the impact of demographic (age, sex, pubertal status), cardiovascular (heart rate, blood pressure), anthropometric (height, weight, BMI, waist-circumference), and physical activity (self-reported weekly activity, time spent watching TV) variables on HRV measurement stability using time-varying covariates. This study evaluated HRV as a stable individual difference in youth. Taken together, these two studies provided timely information regarding the psychometric properties of HRV among infants, children, and adolescents.

Study 1: Method

Literature Search Strategy

First, a literature search was performed using PsycInfo and PubMed (MedLine) databases from January 1, 1970 to December 31, 2013, using the following keyword combinations: (“neonatal” OR “infant” OR “toddler” OR “child” OR “children” OR “adolescent” OR “adolescence” OR “teen” OR “youth” OR "young adult") AND (“heart period” OR “heart rate variability” OR “HRV” OR “R-R interval” OR “RR interval” OR “N-N interval” OR “NN interval” OR “heart interval” OR “respiratory sinus arrhythmia” OR “RSA” OR “vagal tone” OR “inter-beat interval” OR “IBI” OR “SDNN” OR “SDANN” OR “SDNNi” OR “rMSSD” OR “pNN50” OR “high frequency heart rate variability” OR “high frequency HRV” OR “low frequency heart rate variability” OR “low frequency HRV”) AND (“reliability” OR “reliable” OR “stable” OR “stability” OR “reproducible” OR “reproducibility” OR “psychometric” OR “intra-individual” OR “test-retest” OR “repeatable” OR “repeatability” OR “longitudinal”). The following filters were also applied: Species: humans; Language: English; Age Range: birth to 18 years. This search yielded 975 non-redundant articles.

Second, article titles were reviewed and abstracts were selected for follow-up if the title indicated that HRV was obtained from infants, children, or adolescents; 586 titles met this inclusion. Third, abstracts of the selected titles were reviewed and articles were selected for follow-up if the abstract suggested that HRV was measured from non-clinical samples of infants, children, or adolescents on more than one occasion; 82 articles met this criterion. Fourth, a descendancy approach was used to identify additional articles by reviewing the reference lists from selected articles; 328 non-redundant titles were identified. Following an abstract review of these references, 37 additional articles met

inclusion criteria. Fifth, an ascendancy approach was used to identify additional articles based on a forward citation search of the 82 initially selected articles using Web of Science; 191 non-redundant titles were identified. Following an abstract review of these references, 29 additional articles met inclusion criteria. Finally, letters of solicitation requesting additional or unpublished data were sent to authors who published two or more included articles; one additional article was obtained. A total of 143 non-redundant articles were identified for full review (see Figure 1).

Article Inclusion and Exclusion

Each of the 143 identified articles were read and fully reviewed for possible study inclusion. Articles were included in the meta-analysis if: (a) HRV was measured at least twice during separate, resting baseline conditions; (b) participants were non-clinical infants, children, or adolescents with a sample mean age of ≤ 18.0 years at the initial study assessment; and (c) sufficient baseline HRV data were presented for effect size computation (i.e., reporting only means was not sufficient for inclusion). Articles were excluded based on an *a priori* hierarchy: (a) not an empirical study ($k = 1$); (b) sample size less than 10 ($k = 4$); (c) sample consisted only of clinical (e.g., diabetes, anxiety) or special-population individuals (e.g., pre-term infants, athletes; $k = 9$); (d) second baseline ECG recording was obtained only on the same day as the first recording ($k = 1$); (e) second baseline ECG recording was not obtained ($k = 51$); (f) HRV data was obtained solely during periods of overnight sleep ($k = 3$); (g) HRV was derived from averaged heart rate (using beats per minute, BPM), rather than NN intervals ($k = 4$); (h) HRV data were averaged across rest and between task conditions (i.e., stress, exercise; $k = 3$); (i) data not reported for baseline HRV ($k = 20$); (j) data were redundant with another study

already included in the meta-analysis (k = 1; more complete article was included). Of the 143 potentially relevant articles, 46 met inclusion criteria (see Figure 1).

Reliability of Article Selection and Coding

To evaluate article selection fidelity, inter- and intra-rater reliability were examined. A single rater (OW) initially coded all titles, abstracts, and articles. Next, a random sample of 10% of titles and abstracts was blindly coded a second time by OW after a four-month interval; a separate independent rater (JL) also coded titles and abstracts at this time. Excellent intra-rater (kappa = 1.0) and inter-rater (kappa = 1.0) agreement for title selection were obtained. Excellent intra-rater (kappa = 0.92) and adequate inter-rater agreement (kappa = 0.8) for abstract selection were obtained. Differences in abstract selection were resolved through discussion. For the included articles, excellent intra-rater reliability was obtained for categorical (kappa = 0.93) and continuous coding decisions (*ICC* = 1.0, 95% CI [1.0, 1.0]), and for selection of data used to calculate effect sizes (*ICC* = 1.0, 95% CI [1.0, 1.0]). Good inter-rater reliability was obtained for categorical (kappa = 0.83) and continuous coding decisions (*ICC* = 0.96, 95% CI [0.927, 0.975]), and for selection of data used to calculate effect sizes (*ICC* = 1.0, 95% CI [1.0, 1.0]). Each article required approximately 60 coding decisions.

Article Coding and Data Extraction

Study and sample characteristics. If demographic information (e.g., age, weight) was reported separately for males and females, a weighted average value was calculated. Mean participant age was coded in months. Sex was coded as percent male. Participant height (centimetres), weight (kilograms), BMI (kg/m^2), systolic and diastolic blood pressure (mmHg), ethnicity (African American, European American, Hispanic, Asian, Mixed, Other, Unknown), and puberty status (Tanner stages) were coded.

Exclusion criteria for study participants as reported by each article were dichotomously coded (yes/no) within the following *a priori* categories: (a) prescription medication use; (b) abnormal pregnancy or birth (e.g., preterm birth, abnormal pediatric exam); (c) chronic medical illness (e.g., diabetes, cardiovascular pathology); (d) mental health diagnosis (e.g., cognitive, intellectual, psychological, behavioural disorder); (e) anthropometric characteristics (e.g., overweight, obese). Study follow-up length was coded in weeks. "Unmentioned" was coded when a study or sample characteristic of interest was not reported.

Psychophysiological recording. Study location was coded (university/hospital lab; school; home/ambulatory; nursery). Time of day that the ECG recording occurred was coded (morning 06:00–11:59; afternoon 12:00 –17:59; evening 18:00–23:59). ECG sampling rate was dichotomously coded (≥ 500 Hz vs. < 500 Hz). Baseline ECG recording posture was coded (seated, supine, standing); "changed" was coded when children assumed different recording postures between study assessments. An *a priori* hierarchy was applied when more than one recording posture was reported ([1] seated, [2] supine, [3] standing). Studies with infants and toddlers were often recorded in a seated or supine posture; however, when no recording posture was identified infant/toddler studies were coded as supine. Inclusion of a pre-recording acclimatization period (i.e., time spent resting quietly prior to baseline ECG recording) was dichotomously coded (yes/no). Total ECG recording duration (baseline, task, recovery conditions combined) and baseline ECG recording duration were each coded in minutes. All studies were also dichotomously coded (yes/no) to indicate whether the location, recording time, recording posture, acclimatization period, total ECG recording duration, and baseline ECG

recording duration were identical across follow-up study assessments. Studies were dichotomously coded (yes/no) to indicate whether participants were awake for the entire duration of each ECG recording; several studies with infants had mixed sleep/wake recordings. "Unmentioned" was coded when a recording characteristic of interest was not reported.

HRV pre-processing and analysis. Manual editing of ECG data and use of software-based filtering algorithms (e.g., Hanning/Hamming window, moving polynomial window) were both dichotomously coded (yes/no). The baseline ECG recording duration that was analyzed for HRV was coded in minutes; studies were dichotomously coded (yes/no) to indicate whether the baseline ECG recording duration analyzed for HRV was identical across all study follow-up assessments. The method of HRV signal processing (i.e., data reduction) was dichotomously coded (HRV processed using the mean of one, longer ECG recording epoch vs. HRV processed using the mean of multiple, shorter ECG recording epochs). "Unmentioned' was coded when a pre-processing or analysis characteristic of interest was not reported.

Selection of baseline HRV measures. Baseline HRV was defined as any HRV measure obtained from ECG data recorded either independent of, or prior to, any experimental study condition (e.g., stress, tilt, exercise) across all study follow-up assessments. Time- and frequency-domain measures of baseline HRV were coded using *a priori* categories. Measures reported as heart period, mean NN-interval, or inter-beat interval were coded as NN in milliseconds (ms); SDNN and "heart period standard deviation" were coded as SDNN in ms; HF, respiratory sinus arrhythmia, and vagal tone were coded as HF in absolute (ms^2) and normalized (n.u.) units; LF was coded in

absolute (ms^2) and normalized (n.u.) units. Frequency-domain HRV bandwidths were dichotomously coded (yes/no) to indicate whether a study selected an age-appropriate frequency band consistent with published HRV guidelines or recommendations (Bar-Haim et al., 2000; Fox & Porges, 1985; Task Force, 1996); studies with infants and toddlers often used wider frequency bands relative to studies with children.

Statistical Analysis

Effect size calculation. Effect sizes are scale-free measures of the strength of association between independent and dependent variables that can be represented in standardized (e.g., Hedges' *g*, Fisher's *Z*) or unstandardized (e.g., regression coefficient, mean difference) metrics (Rosenthal, 1994; Wilkinson & Task Force on Statistical Inference, 1999). Fisher's *Z* was selected as the standardized common effect size for this meta-analysis. Fisher's *Z* ranges from -∞ to +∞ and can be interpreted similar to a correlation; however, caution about bias is warranted when converting statistics derived from small samples (e.g., $N < 30$; Rosenberg et al., 2000; Rosenthal, 1994; Silver & Dunlap, 1987). Studies reporting effect sizes derived using different statistical methods were coded based on an *a priori* hierarchy ([1] Intra-class correlations, [2] partial correlation, [3] Pearson/zero-order correlation, [4] Spearman rank-order correlations, [5] unstandardized beta coefficients [6] F-ratios, [7] *p*-values). Intra-class correlations (*ICC*), as well as Pearson and Spearman correlations, were converted directly to Fisher's *Z* (Rosenberg et al., 2000; Rosenthal, 1994). It has been suggested that *ICC*s are more comparable to *r*-type correlations following square-root transformation (e.g., Bornman, Mutz, & Daniel, 2010); however, preliminary analyses revealed no differences between Fisher's *Z* values derived from untransformed and from square-root transformed *ICC*s.

As such, the untransformed (i.e., original) *ICC* value was converted to Fisher's *Z*. Test statistics, including unstandardized beta coefficients and *F*-ratios, were converted to *r* and then to Fisher's *Z* (Rosenthal, 1994; Rosenberg et al., 2000). When no other test statistic was available, exact *p* values were converted to the standard normal deviate (z-score), then to *r,* and then to Fisher's *Z* (Rosenberg et al., 2000).

Effect size management. Studies reporting effect sizes derived from two or more follow-up assessments were coded for both the entire study follow-up length and shorter follow-up intervals. Thus, multiple effect sizes for each reported HRV variable were extracted from each study if available; this allowed for both conservative (non-redundant; one effect size per study) and non-conservative (redundant; multiple effect sizes per study) analytical approaches. Non-redundant effect sizes were defined as those from the entire study follow-up length; multiple non-redundant effect sizes (i.e., from non-overlapping study follow-up intervals; e.g., 2–4 weeks, 4–6 weeks) were also sample weight averaged to allow for one effect size per study. Redundant effect sizes were defined as those from multiple, overlapping, study follow-up intervals (e.g., 2–4 weeks, 3–5 weeks).

Analytic strategy. Because of noted methodological differences between HRV research conducted with infants and children, analyses were age-stratified (see Williams et al., 2012). Infant/toddler studies were defined as those with mean participant age less than 5 years ($k = 31$); child/adolescent studies were defined as those with mean participant age greater than 5 years but less than or equal to18 years ($k = 15$).

To determine whether to conduct a fixed or random effects meta-analysis, we considered that the categorical variables coded are believed to capture all possible options

represented in the extant HRV literature (e.g., standardized vs. unstandardized ECG recording time; sitting vs. supine vs. standing posture; ≥ 500 Hz vs. < 500 Hz sampling rate); these coding options are not sampled from a larger population of possible options. Thus, a fixed-effects meta-analytic model, which assumes that variability in effect sizes are due only to sampling error (Rosenthal, 1994; Field & Gillett, 2010), was chosen for this study. A random-effects meta-analytic model, which assumes that variability in effect sizes is due to both random variance and sampling error, was not deemed appropriate given the data available (Rosenberg et al., 2000).

All analyses were age-group stratified. A mean cumulative effect size was calculated overall and for each HRV variable separately. The heterogeneity (Q_T) test statistic was calculated as a measure of variation of the included effect sizes. A non-significant Q_T indicates a homogeneous distribution, suggesting that variability of the effect sizes is less than that which would be expected from sampling error. A significant Q_T indicates a heterogeneous distribution, suggesting that additional moderator analyses may be warranted (Rosenberg et al., 2000). Moderator analyses were conducted using continuous and categorical variables coded from each study. Similar to regression, continuous analyses calculate a slope to test associations between an effect size and a continuous moderator (e.g., ECG recording duration). Similar to ANOVA, categorical analyses partition variance explained by the model (Q_M) and residual error variance (Q_E; Rosenberg et al., 2000). Bootstrap methods (1000 samples) were used to produce robust non-parametric estimates of confidence intervals around each effect size (Rosenberg et al., 2000). Finally, to address concerns of possible publication bias and the file-drawer problem, Orwin's fail-safe numbers were used to determine the number of missing,

unpublished, or non-significant comparisons needed to make an overall effect size not different from 0.01. Analyses were performed using MetaWin 2 (Sinauer Associates, 2000).

The primary study hypothesis was that time- and frequency-domain HRV would exhibit moderate test-retest reliability among both infant/toddler and child/adolescent studies. Considering the recommendations outlined by published HRV guidelines and results from available methodological examinations of HRV (reviewed above), it was hypothesized that higher HRV reliability would be associated with studies that (a) considered factors relevant to study protocol and sample characteristics (i.e., shorter study follow-up length, greater sample homogeneity, smaller proportion of male participants); (b) maintained greater standardization of ECG recording conditions (i.e., consistent ECG recording time, included a pre-recording acclimatization period, consistent supine or seated ECG recording postures); and, (c) measured ECG and analyzed HRV in accordance with published guidelines and recommendations (i.e., sampling rate $\geq$ 500 Hz, edited and filtered ECG data prior to HRV analysis, analyzed longer [opposed to shorter] and consistent ECG recording durations, derived HRV using the multiple, shorter ECG epochs [opposed to a single, longer ECG epoch], selected a recommended frequency bandwidth [opposed to an alternative frequency bandwidth]). In addition, this meta-analysis examined whether test-retest reliability of HRV was significantly different between infant/toddler and child/adolescent studies.

Study 1: Results

Study Participant and Recording Characteristics

A total of 46 studies (N = 4790) were included in the present meta-analysis. Thirty-one studies of infants and toddlers (N = 3257) and 15 studies of children and adolescents (N = 1533) each had a mean of approximately 100 participants per study (See Tables 1 and 3). Approximately half of all participants were male for both age groups. Overall, few studies reported details about participant characteristics, presenting little information about height, weight, BMI, ethnicity, or study attrition. Data collection typically occurred at a university or hospital research lab for both infant/toddler (71%) and child/adolescent (67%) studies. The mean ECG recording duration analyzed for HRV was 4.54 min (SD = 4.32) in infant/toddler studies, and 3.22 min (SD = 0.96) in child/adolescent studies (See Table 2). The most commonly reported HRV measures were HF and NN in infant/toddler studies, and HF and LF in child/adolescent studies. A mean of 2.0 (SD = 1.03) and 2.5 (SD = 2.12) HRV variables were reported among infant/toddler and child/adolescent studies, respectively.

Overall Summary Analyses

Consistent with the main study hypothesis, HRV exhibited moderate test-retest reliability across both age groups. Further, reliability was significantly lower for infant/toddler studies compared to child/adolescent studies (Z = 0.42 vs. 0.64; Q_M = 33.84, $p < 0.001$). Among infant/toddler studies, reliability of HF (Z = 0.42) and SDNN (Z = 0.43) were higher relative to NN (Z = 0.33). All infant/toddler analyses were heterogeneous, indicating further moderator analyses would be appropriate; HF reliability displayed the greatest relative heterogeneity among infant/toddler studies (Q_T = 71.61, p

< .001). Among child/adolescent studies, reliability of SDNN ($Z = 0.85$) was higher relative to HF ($Z = 0.65$), LF ($Z = 0.61$), and LF:HF ratio ($Z = 0.52$). All child/adolescent analyses were heterogeneous, except SDNN and LF:HF ratio; HF reliability displayed the greatest relative heterogeneity among child/adolescent studies ($Q_T = 61.15$, $p < .001$). (see Table 4). Further examination of each HRV variable was conducted using continuous and categorical moderator variables. Forest plots depicting HRV reliability effect sizes are presented in Figures 2 to 4.

Moderator Analyses

Continuous and categorical moderator analyses examined variables within the following four categories: *study protocol* (study follow-up length; consistency of study location; time of ECG recording; recording exclusively awake ECG data vs. not), s*ample characteristics* (participant exclusion criteria; sample size, age, and weight; proportion of males), *ECG signal acquisition and pre-processing* (inclusion of a pre-recording acclimatization period, ECG recording posture, ECG sampling rate, editing and filtering techniques), and *HRV analyses* (ECG recording duration, consistency of ECG recording durations, HRV derivation method, HRV frequency bandwidth selection). Results from continuous moderator analyses are presented in Table 5; results from categorical moderator analyses are presented in Tables 6 and 7. Notedly, due to the paucity of studies available, most categorical analyses were limited to NN and HF among infant/toddler studies and HF and LF among child/adolescent studies.

Study protocol. Longer study follow-up length was significantly associated with lower SDNN reliability ($B = -0.01$, $p < 0.001$), but was not associated with NN or HF reliability among infant/toddler studies. In contrast, longer study follow-up length was

significantly associated with lower HF reliability (B < -0.01, p = 0.012), but was not associated with SDNN reliability among child/adolescent studies. Further, longer study follow-up length was significantly associated with lower LF reliability (B = -0.01, p = 0.015), but was not associated with LF:HF ratio reliability among child/adolescent studies. Results provided partial support for the hypothesis that longer study follow-up length is associated with reduced HRV reliability. However, significant heterogeneity remained for SDNN (Q_T = 14.92, p < .001) and HF (Q_T = 71.61, p < .001) among infant/toddler studies and for HF (Q_T = 61.15, p < .001) and LF (Q_T = 12.36, p < .05) among child/adolescent studies, indicating that further moderator analyses are required.

No differences in reliability were observed between studies that maintained an identical study location across follow-up assessments and studies that changed study location at a subsequent follow-up assessment.

No differences in reliability were observed between studies that limited ECG recording to the morning hours, or between studies that reported a consistent time of ECG recording across follow-up assessments (i.e., same vs. different ECG recording time) and HRV reliability across both age groups. Recording ECG exclusively while participants remained awake, compared to recording ECG without restricting participants from sleeping, yielded significantly higher HF reliability among infant/toddler studies (Z = 0.44 vs. 0.21; Q_M = 6.50, p = 0.011); all child/adolescent studies recorded ECG data while participants were awake. Results from infant/toddler studies that recorded ECG while participants remained awake, compared to studies that did not restrict participants from sleeping, provided partial support for the hypothesis that accounting for circadian factors is associated with higher HRV reliability.

Sample characteristics. Sample exclusions (yes/no) based on prescription medication use ($Z = 0.61$ vs. 0.23; $Q_M = 14.06$, $p < 0.001$), chronic medical illnesses ($Z = 0.47$ vs. 0.26; $Q_M = 5.26$, $p = 0.022$), and low birth weight (e.g., < 2500 g; $Z = 0.48$ vs. 0.27; $Q_M = 4.71$, $p = 0.03$) yielded significantly higher NN reliability among infant/toddler studies. Further, sample exclusion due to atypical neonatal characteristics (e.g., preterm birth, abnormal pediatric exam; $Z = 0.34$ vs. 0.45; $Q_M = 8.34$, $p = 0.004$) and low birth weight ($Z = 0.23$ vs. 0.44; $Q_M = 6.31$, $p = 0.012$) yielded significantly lower HF reliability among infant/toddler studies. NN reliability was most influenced by sample exclusions based on medication use; HF reliability was most influenced by sample exclusions based on birth weight among infant/toddler studies. Child/adolescent studies with sample exclusion criteria yielded no differences in HF reliability, suggesting HF reliability is more robust to sample heterogeneity among children and adolescents relative to infants and toddlers. Results from infant/toddler studies provided partial support for the hypothesis that increased sample heterogeneity is associated with higher HRV reliability. However, as significant heterogeneity was observed for NN and HF among infant/toddler studies and for HF among child/adolescent studies upon examination of participant exclusions, further moderator analyses are needed.

Larger sample size was significantly associated with higher NN reliability ($B < 0.01$, $p = 0.027$) and lower SDNN reliability ($B = -0.01$, $p < 0.001$) among infant/toddler studies, and lower LF reliability ($B < -0.01$, $p = 0.049$) among child/adolescent studies. Larger sample size was not associated with HF reliability in either age group, and was not associated with SDNN or LF:HF ratio reliability among child/adolescent studies.

Older sample age was significantly associated with higher NN ($B = 0.01$, $p < 0.001$) and HF ($B = 0.01$, $p = 0.002$) reliability, but lower SDNN reliability ($B = -0.53$, $p < 0.001$) among infant/toddler studies. Conversely, older sample age was not associated with reliability of any HRV variables among child/adolescent studies. There were no associations between increased sample weight and HRV reliability across both age groups.

Having a greater proportion of male participants was significantly associated with lower NN reliability ($B = -0.02$, $p = 0.017$) among infant/toddler studies; no relation was observed for HF. Conversely, having a greater proportion of male participants was significantly associated with lower HF ($B = -0.01$, $p = 0.007$) and LF ($B = -0.01$, $p = 0.008$) reliability among child/adolescent studies; no relation was observed for LF:HF ratio. Results based on frequency-domain measures from child/adolescent studies provided partial support for the hypothesis that having a greater proportion of male participants is associated with lower HRV reliability. However, significant heterogeneity remained for NN ($Q_T = 36.46$, $p < .001$) among infant/toddler studies and for HF ($Q_T = 58.99$, $p < .001$) and LF ($Q_T = 12.36$, $p < .05$) among child/adolescent studies that had a greater proportion of male participants, indicating that further moderator analyses are required.

ECG signal acquisition and pre-processing. Including a pre-recording acclimatization period yielded significantly higher NN reliability ($Z = 0.62$ vs. 0.21; $Q_M = 18.31$, $p < 0.001$) among infant/toddler studies, and higher HF ($Z = 0.68$ vs. 0.49; $Q_M = 4.85$, $p = 0.023$) and LF reliability ($Z = 0.77$ vs. 0.39; $Q_M = 9.03$, $p = 0.003$) among child/adolescent studies; HF reliability among infant/toddler studies did not differ in the

presence or absence of an acclimatization period (Z = 0.43 vs. 0.42). Results largely supported the hypothesis that including a pre-recording acclimatization period is associated with higher HRV reliability, particularly for frequency-domain measures among child/adolescent studies. However, there remained significant heterogeneity for HF among child/adolescent studies that included a pre-recording acclimatization period (Q_T = 50.06, p < .001), indicating that further moderator analyses are required.

ECG recording posture was significantly associated with NN (Q_M = 9.42, p = 0.002) and HF (Q_M = 23.12, p < 0.001) reliability among infant/toddler studies; HF reliability was highest when ECG was recorded in the seated posture (Z = 0.50), and lowest when ECG was recorded in different postures across follow-up assessments (Z = 0.19). HF and LF reliability did not differ between ECG recording postures among child/adolescent studies; however, effect sizes for HF and LF reliability were larger among studies that recorded ECG in a seated, relative to a supine posture. No child/adolescent studies reported a change in the ECG recording posture across study follow-up assessments. Results from infant/toddler studies provided partial support for the hypothesis that ECG recording postures associated with parasympathetic dominance yields higher reliability of measures reflecting parasympathetic modulation of heart rate. However, significant heterogeneity was observed for NN (Q_T = 31.55, p < .001) and HF (Q_T = 42.13, p < .001) among infant/toddler studies and for HF (Q_T = 33.85, p < .001) among child/adolescent studies that recorded ECG in a seated posture, indicating that further moderator analyses are required.

Reporting an ECG sampling rate < 500 Hz versus no mention of ECG sampling rate was associated with significantly lower HF reliability (Z = 0.28 vs. 0.44; Q_M = 4.84,

$p = 0.028$) among infant/toddler studies. Reporting an ECG sampling rate ≥ 500 Hz versus no mention of ECG sampling rate was not associated with higher HF reliability among child/adolescent studies; this association approached significance for LF, such that studies reporting an ECG sampling rate ≥ 500 Hz yielded higher reliability relative to studies that did not report an ECG sampling rate ($Z = 0.72$ vs. 0.44; $Q_M = 3.67$, $p = 0.056$). Minimal reporting of ECG sampling rate among included studies limited this analysis to comparisons between studies that did and did not report an ECG sampling rate; as such, this study could not adequately test the hypothesis that sampling rates ≥ 500 Hz are associated with higher HRV reliability.

Reporting manual editing of ECG data versus no mention of manual editing was significantly associated with higher NN reliability ($Z = 0.38$ vs. 0.14; $Q_M = 4.99$, $p = 0.025$), while reporting a software-based filtering algorithm versus no mention of applying filtering algorithms was significantly associated with higher HF reliability ($Z = 0.45$ vs. 0.35; $Q_M = 4.85$, $p = 0.028$) among infant/toddler studies. No differences in HF or LF reliability were observed among child/adolescent studies that reported or did not report manual editing or software-based filtering of ECG data. Results from infant/toddler studies provided partial support for the hypothesis editing and filtering of ECG data is associated with higher HRV reliability.

HRV analyses. Longer ECG recording duration was significantly associated with lower NN reliability ($B = -0.03$, $p = 0.026$) among infant/toddler studies; however, results were no longer statistically significant after re-examining only those studies with ECG recording duration ≤ 5 min (excluded $k = 2$; $B = 0.06$, $p = 0.083$). No relation was observed between ECG recording duration and HF reliability among infant/toddler

studies. Conversely, longer ECG recording duration was significantly associated with higher HF (B = 0.14, p = 0.004), LF (B = 0.10, p = 0.040), and lower LF:HF ratio reliability (B = -0.35, p = 0.032) among child/adolescent studies; no relation was observed for SDNN. Results from child/adolescent studies provided partial support for the hypothesis that longer ECG recording durations are associated with higher HRV reliability. However, significant heterogeneity was observed for NN (Q_T = 45.82, p < .001) among infant/toddler studies and for HF (Q_T = 61.15, p < .001) and LF (Q_T = 12.36, p < .05) among child/adolescent studies that had longer ECG recording durations, indicating that further moderator analyses are required.

Analyzing the same ECG recording duration across study follow-up assessments, compared to different recording durations, was associated with significantly higher NN (Z = 0.38 vs. 0.12; Q_M = 5.72, p = 0.017) and HF (Z = 0.45 vs. 0.33; Q_M = 5.13, p = 0.024) reliability among infant/toddler studies. As well, infant/toddler studies that derived HF using the mean of two or more ECG analysis epochs (i.e., by segmenting a single ECG recording duration into shorter, sequential epochs) yielded significantly higher reliability compared to studies that derived HF using the mean of a single ECG epoch (Z = 0.46 vs. 0.25; Q_M = 5.41, p = 0.020). Results from infant/toddler studies provided partial support for the hypothesis that analyzing consistent ECG recording durations across follow-up assessments is associated with higher HRV reliability, and for the hypothesis that analyzing HRV from shorter ECG epochs is associated with higher HRV reliability. In the child/adolescent studies, all but one analyzed the same ECG recording duration across study follow-up assessments. As well, no difference in HF reliability was found between studies that analyzed two or more ECG epochs compared to studies that

analyzed a single epoch. Results from child/adolescent studies did not provide support for the hypothesis that analyzing consistent ECG recording durations across follow-up assessments is associated with higher HRV reliability, or for the hypothesis that analyzing HRV from shorter ECG epochs is associated with higher HRV reliability

Selection of an HRV frequency bandwidth consistent with published guidelines and recommendations, compared to an alternative frequency bandwidth, yielded significantly higher HF reliability among child/adolescent studies (Z = 0.86 vs. 0.38; Q_M = 16.49, $p < 0.001$); no association with HF was observed among infant/toddler studies. LF reliability was not significantly associated with frequency bandwidth selection among child/adolescent studies; however, effect sizes for LF reliability were larger when studies selected a frequency bandwidth consistent with published guidelines and recommendations, relative to studies that selected an alternative frequency bandwidth (Z = 0.61 vs. 0.39). Results from infant/toddler studies did not support the hypothesis that selecting a frequency bandwidth consistent with published guidelines and recommendations is associated with higher HRV reliability, and instead suggested that HF reliability among infant/toddler studies was robust to frequency bandwidth selection. Results from child/adolescent studies provided partial support for the hypothesis that frequency bandwidth selection is associated with higher HRV (i.e., HF) reliability.

Study Quality

Ten dimensions of study quality were rated dichotomously and summed to yield a quality rating. Study quality variables associated with continuous data were dichotomized based on the median value of included studies. Study quality dimensions included: (1) explicit exclusion criteria for medical illness (28.3%); (2) explicit exclusion criteria for

prescription medication use (10.9%); (3) sample size of ≥ 50 participants (43.5%); (4) specified at least one pre-data collection participant instruction (e.g., overnight fast, no exercise 12 hr prior to data collection; 19.6%); (5) included pre-recording acclimatization period (32.6%); (6) resting/baseline ECG recording duration ≥ 3.5 min (52.2%); (7) ECG recording duration analyzed for HRV ≥ 3 min (50%); (8) identified both ECG recording hardware and HRV analysis software (80.4%); (9) explicitly reported both manual editing and digital filtering of ECG data (50%); and (10) used recommended, age-appropriate HRV frequency bandwidths (50%). Study quality was correlated with reliability effect sizes pertaining to the total study follow-up length. If reliability of two or more non-redundant HRV variables were reported in a given study, Fisher's *Z* values were averaged to yield one effect size per study. Study quality was not associated with effect sizes among infant/toddler studies ($M = 3.94$, $SD = 2.06$; $r = -0.013$, $p = .945$). Conversely, study quality was positively associated with effect sizes among child/adolescent studies ($M = 4.67$, $SD = 1.76$; $r = 0.514$, $p = 0.050$).

Study 1: Discussion

The overarching aim of this meta-analysis was to examine the test-retest reliability of time- and frequency-domain HRV measures among pediatric studies. Consistent with the main study hypotheses, time- and frequency-domain HRV measures displayed moderate test-retest reliability across both age groups. Overall, HRV reliability among infant/toddler studies was significantly lower compared to child/adolescent studies. Among infant/toddler studies, all effect sizes were in the moderate range; SDNN and HF test-retest reliability were similar, while NN was relatively lower in comparison. Among child/adolescent studies, all effect sizes were also in the moderate range; HF and LF test-retest reliability were similar, while SDNN reliability was relatively higher and LF:HF ratio was relatively lower in comparison.

Several moderating variables influenced HRV reliability. The paucity of study information reported by researchers precluded examination of several moderator variables of interest. Moderator analyses examined methodological factors relevant to *study protocol* (e.g., study follow-up length, time of ECG recording), *sample characteristics* (e.g., participant exclusion criteria, sample size, proportion of males), *ECG signal acquisition and pre-processing* (e.g., inclusion of a pre-recording acclimatization period, editing and filtering techniques), and *HRV analyses* (e.g., ECG recording duration, HRV frequency bandwidth selection).

Study Protocol

Longer study follow-up length was associated with decreased reliability for SDNN among infant/toddler studies, and decreased reliability of HF and LF among child/adolescent studies. Lower test-retest reliability with longer study follow-up lengths is not unexpected given basic psychometric principles (c.f., Cohen & Swerdlik, 2002).

The impact of age-related increases in HRV across childhood, which in part reflects physiological changes in the complexity of autonomic cardiovascular control, offers a plausible explanation for decreasing reliability with longer study follow-up lengths (i.e., longer test-retest intervals). For instance, Gentzler and colleagues (2012) obtained almost yearly measures of HF from children across 8 years and yielded one of the lowest effect sizes among child/adolescent studies, whereas effect sizes among studies that spanned from 1 week to 4 years were typically more homogeneous. One reason why study follow-up length was not associated with HF reliability among infant/toddler studies could be related to study methodology in the derivation of HF. Infant/toddler studies typically derived HF using Porges' (1985) moving polynomial filter, which can account for slow periodic processes and non-stationarity in the ECG data by excluding background sources of variance. Conversely, child/adolescent studies typically derived HF using spectral analysis techniques (e.g., fast Fourier transformation) or the peak-valley method (i.e., obtaining the difference in mean NN-intervals during inspiration and expiration between consecutive respiratory cycles; El-Sheikh, 2005). Given that these analytical methods do not inherently account for heart rate, this may explain why HF among child/adolescent studies was less reproducible across longer test-retest intervals. Application of a moving polynomial filter could contribute to more reproducible measures of HF by accounting for developmental changes in heart rate activity over longer test-retest intervals (e.g., among newborns followed-up after several months or years). Heart rate is one of many physiological changes that occur as children grow, decreasing from infancy. This change in heart rate likely contributes to the reproducibility of HRV, and should thus be considered when HRV is examined longitudinally in children. Given that longer study

follow-up length was associated with lower HF reliability among studies that typically did not account for heart rate, an examination of long-term HF reproducibility using measures derived with and without a moving polynomial filter as compared to HF derived with and without statistically controlling for heart rate may be warranted.

Meta-analytic results indicated that limiting ECG recording to morning hours or using identical time-of-day of ECG recording for both study assessments was not associated with HF reliability, across either age group. However, based on limited data, reliability of SDNN and LF was higher among child/adolescent studies that used identical time-of-day of ECG recording for both study assessments relative to studies that did not. Circadian variation in HRV (i.e., greater sympathetic dominance in morning, parasympathetic dominance in evening/sleep) has been observed in infants (Hoppenbrouwers et al., 2012) and children (Massin et al., 2000; Rabbia et al., 2003). However, results provided only partial support for the hypothesis that controlling potential circadian confounds to HRV (i.e., maintaining an identical time of ECG recording for all participants across follow-up assessments) augments reliability. Other results from this meta-analysis indicated that among infant/toddler studies that recorded ECG during daytime hours, HF reliability was higher for those studies that exclusively contained awake ECG data compared to those that contained mixed awake/asleep recordings (i.e., studies that did not restrict participants from falling asleep during ECG recording). It is likely that infant/toddler studies that examined HF using combined awake/asleep data were unable to replicate the initial recording conditions at study follow-up, such that a different subset of participants may have fallen asleep at each study assessment. Long-term and overnight HRV recordings have been considered better

predictors of all-cause morbidity and mortality compared to short-term daytime measures (e.g., Ernst, 2014) that are less able to capture fluctuations in HRV between periods of wakefulness and sleep. The present meta-analysis could not thoroughly examine whether circadian factors affect HRV psychometrics. As such, future studies should consider examining repeated full day (i.e., 24-hr) and separate 12-hr (e.g., day-time vs. overnight) ECG recordings to better elucidate circadian factors, which may influence HRV reliability. Namely, considering the observed variations in autonomic activity throughout a 24-hr period (e.g., Guo & Stein, 2003; Huikruri et al., 1990), an examination of whether reliability changes based on the specific time of ECG recording could help determine whether some HRV variables (e.g., LF) are more reliable when consistently measured in the morning versus others (e.g., HF) that are measured in the evening.

Sample Characteristics

Sample exclusion criteria selected by researchers were another study protocol moderator that influenced HRV test-retest reliability. Increased sample homogeneity, attributable to the defined participant exclusion criteria, was typically associated with higher NN reliability and lower HF reliability among infant/toddler studies. HF reliability among child/adolescent studies with sample exclusion criteria did not differ from studies without exclusion criteria, suggesting that HF from children and adolescents, but not infants and toddlers, is more robust to reduced sample homogeneity. Infant/toddler studies that excluded participants based on prescription medication use, chronic illnesses, or low birth weight yielded higher NN reliability; exclusions based on medication use demonstrated the strongest effect on NN reliability ($Z_{\text{exclusion reported}} = 0.61$ vs. $Z_{\text{unmentioned}} = 0.23$) relative to other exclusion criteria. Several medications and illnesses have been

associated with changes in sympathovagal balance (i.e., sympathetic hyperactivity and/or parasympathetic hypoactivity) among adults and children (e.g., Buchhorn et al., 2012; Earnst, 2014; Task Force, 1996). Given that NN is influenced by both sympathetic and parasympathetic autonomic branches, excluding participants who consume medications or have compromised health status from an otherwise healthy sample would help eliminate an additional source of variance in measures of heart rate and improve NN reliability. Conversely, studies that excluded participants based on prescription medication use had relatively lower HF reliability across both age groups; exclusion due to presence of a medical illness was not associated with HF reliability among child/adolescent studies. This result for medication exclusions and HF reliability among infant/toddler studies was paradoxical given results for NN; however, only two studies among both age groups reported this exclusion, and effect size confidence intervals were largely overlapping. As such, results provided only partial support for the hypothesis that increased sample homogeneity augments HRV reliability. It is recommended that future researchers examine the cross-sectional and longitudinal effects on HRV attributed to illnesses common to younger individuals and prescription medications typically prescribed to children. This would help determine if excluding participants based on certain illnesses or medications have a greater impact on HRV reliability compared to others.

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, yet lower HF and LF reliability among child/adolescent studies. These child/adolescent results are seemingly consistent with

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 are commonly reported as having greater mean-level HF and yield greater HF reliability relative to males, although these findings are not consistent (e.g., El-Sheikh, 2005). Biologically, sex differences in HRV reliability may be associated with hormonal differences related to pubertal development (e.g., sympathetic alterations in response to increased concentration of estrogen; Kajantie & Phillips, 2006; Ordaz & Luna, 2012; Shirtcliff, Dahl, & Pollak, 2009). As girls typically begin puberty approximately 2–3 years earlier than boys (Patton & Vines, 2007; Shaffer, Wood & Willoughby, 2005), sex differences in HRV reliability may become more pertinent as children achieve greater sexual maturation. For example, Silvetti and colleagues (2001) demonstrated that young boys have higher mean-level values of SDNN, rMSSD, and pNN50 relative to young girls, while adolescent boys have lower mean-level values of SDNN, rMSSD, and pNN50 relative to adolescent girls. However, Jarrin and colleagues (In press) demonstrated that pubertal onset (as indicated by measures of gonadarche and adrenarche) was negatively associated with rMSSD, pNN50, and HF, and positively associated with LF:HF ratio, suggesting that a more thorough examination of the associations between pubertal status, related hormone levels, and HRV reliability among youth may be warranted. As such, future studies should specifically examine sex differences in HRV reliability among children across different stages of puberty (e.g., pre-pubertal vs. mid-pubertal vs. post-pubertal) to better determine whether pubertal development can account for observed sex differences. Further, examining sex differences using 24-hr ECG data can help elucidate any possible

interactions between sex and circadian fluctuations in HRV. One simple recommendation to evaluate the role of sex on the psychometric properties of HRV measures would be for researchers to present sex-stratified analyses.

ECG Signal Acquisition and Pre-Processing.

Inclusion of a pre-recording acclimatization (i.e., habituation) period was a significant moderator of HRV reliability. Reporting an acclimatization period was associated with significantly higher NN reliability among infant/toddler studies and significantly higher HF and LF reliability among child/adolescent studies compared to studies that did not report this detail. As well, higher SDNN reliability across both age groups and higher LF:HF ratio reliability among child/adolescent studies was observed among those studies that included a pre-recording acclimatization period relative to studies that did not. Studies that reported an acclimatization period typically yielded the largest effect sizes and associated failsafe *N*s relative to studies that did not, and relative to many of the other examined moderators. Further, while the Q_T statistic indicated that significant heterogeneity remained for HF among child/adolescent studies that reported an acclimatization period, the Q_T for LF among child/adolescent studies and for NN among infant/toddler studies indicated homogenous effect sizes. Results largely supported the hypothesis that allowing participants to habituate prior to data collection, as recommended by HRV measurement guidelines (e.g., Berntson et al., 1997), improves HRV reliability. Notably, given that the duration of the acclimatization periods reported by included studies was typically 5 min or less among infant/toddler studies and 10 min or less among child/adolescent studies, allowing even a brief habituation period prior to data collection may be sufficient to stabilize HRV and augment reliability. Obtaining

stable baseline HRV data is important for examining resting autonomic cardiovascular control and HRV reactivity to experimental stressors. Participants who are acclimated to experimental conditions prior to data collection would likely yield more reliable data because individual differences in physiological reactivity to experimental conditions are minimized (i.e., stress-related increases in sympathetic or respiratory activity; e.g., Ottaviani & Shapiro, 2011; Salomon, Matthews, & Allen, 2000; Traina, Cataldo, Galullo, & Russo, 2011). Conversely, participants who are not acclimated may yield less stable HRV measures, which in turn, would distort mean-level values and decrease test-retest reliability. Including a pre-recording acclimatization period may be especially pertinent for very young children, who may be more reactive to the unfamiliar relative to older children or adolescents. Thus, the results of this meta-analysis suggest pediatric researchers can improve HRV reliability by providing participants time to acclimate to their surroundings. Future researchers should consider examining HRV reliability among groups given different durations of time in which to acclimate prior to data collection. This would help determine if there is a minimum "acclimatization threshold" in children that could allow for the most reliable measures to be obtained; examining this between children of different ages would help to further elucidate whether the duration of optimal habituation changes across childhood development.

The posture in which HRV was recorded influenced HRV reliability. Specifically, infant/toddler NN and HF reliability was higher for measures obtained in seated postures and lowest when posture changed between study recordings. Child/adolescent HF reliability was relatively higher among studies that recorded ECG in a seated, rather than a supine or an unmentioned posture. These observations are consistent with Young and

Leicht (2011), who demonstrated that adult HF reliability was higher in seated versus standing postures. Children who are asked to maintain a standing posture and remain still for a baseline ECG recording, even for a recording of only 3- to 5-min, may become more restless/fatigued compared to maintaining a seated posture; increased restlessness during the baseline recording could reduce the stationarity of heart rate signals, which could reduce reliability. Further, given that pharmacological blockade studies suggest that sympathetic dominance is augmented with increasing orthostatic load (e.g., standing vs, supine postures, Cacioppo et al., 1994), it may be that HF is more reliably measured in seated or supine, relative to standing, postures because it is in these postures that parasympathetic dominance is more readily observed. Relatedly, LF (which in part reflects sympathetic modulation) may be more reliably measured in a posture that reflects greater sympathetic, opposed to parasympathetic dominance; indeed, although it could not be tested statistically given the available data, LF reliability was higher when measured in a seated, relative to a supine position (see Table 7). Head-up-tilt paradigms are useful for determining the influence of posture on HRV, but are more commonly employed among the adult, compared to child HRV literature. Thus, future researchers should examine reliability of pediatric HRV using postural paradigms (e.g., head-up-tilt) to help improve the collective understanding of how HRV reliability among children is impacted by postural position.

Very few studies reported ECG sampling rate, which precluded meaningful examination of whether sampling rate was associated HRV test-retest reliability. Infant/toddler HF reliability was higher among studies that did not report a sampling rate compared to studies that used a sampling rate < 500 Hz, while child/adolescent HF and

LF reliability was higher among studies that used a sampling rate ≥ 500 Hz relative to studies that did not report a sampling rate. Thus, conclusions about these results must be made with caution. Notedly, less than 20% (k = 5) of infant/toddler studies reporting HF explicitly detailed the ECG sampling rate. The minimal detailing of ECG sampling rate reflects an important gap in reporting practices among the pediatric, relative to the adult literature. HRV measurement guidelines (Berntson et al., 1997; Task Force, 1996) and findings described among the wider HRV methodology literature (e.g., Hejjel & Roth, 2004; Merri et al., 1990; Riniolo & Porges, 1997; Singh et al., 2004), articulate the importance of using an "adequate" ECG sampling rate for both interpretive and study replication purposes. ECG sampling rate defines the minimal R-wave oscillation amplitude that can be detected within a heart rate recording, and is thus pertinent to accurate measurement of NN-interval data. It is possible that an “adequate” sampling rate may be age- or sample-specific (i.e., a different sampling rate may be required for infants who have faster heart rates compared to adults or to those with cardiovascular morbidities; Singh et al., 2004). Thus, there may be a threshold in R-wave detection across different age groups that can only be accommodated through application of an appropriate ECG sampling rate. As newer ECG hardware becomes available, many with different sampling rates (e.g., 128 Hz, 500 Hz, 1000 Hz), it will be important for future studies to examine the association between ECG sampling rate and HRV reliability.

Editing of ECG data was associated with HRV reliability. Among infant/toddler studies, manual editing of ECG data was associated with higher NN reliability, while software-based filtering was associated with higher HF reliability. In contrast, child/adolescent studies that reported manual editing and software-based filtering did not

yield higher reliability compared to child/adolescent studies that did not report these details. The presence of artifacts and recording errors can significantly lengthen or shorten NN-intervals, distort variability, and vastly inflate measures of HRV (Bernston & Stowell, 1998; Mulder, 1992). Considering this, 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 attributed to differences in heart rate activity between the two groups. Given that infants and toddlers have faster heart rates compared to children and adolescents, ECG recordings from infants and toddlers may be more sensitive to aberrations in NN-interval data due to ectopic heartbeats or other artifacts (e.g., movement) that distort measures of HF and hinder reproducibility. As most infant/toddler studies applied a moving polynomial filter (Porges, 1985) to NN-interval data, extraneous variance heart rate measures attributed to artifacts (e.g., distortions in HF due to spectral leakage) or recording errors associated with poor data resolution (e.g., inadequate ECG sampling rates) may have been better accounted for among these studies compared to others that did not apply this filter. Editing and filtering of ECG data were not associated with HF reliability among child/adolescent studies, which suggests that HF reliability among children and adolescents is robust to unedited or unfiltered NN-interval data. It is possible that ECG recordings from children and adolescents exhibit greater inherent stationarity, which reduces the need for the same degree of data cleaning relative to ECG recordings from infants and toddlers. Thus, results from infant/toddler studies supported the hypothesis that carefully applied data cleaning procedures (i.e., editing/filtering) is associated with better HRV reliability. Notedly, given limited available data, nested analyses of reliability

(e.g., between studies that reported either manual/software or no editing at all; between studies that each reported different software-based filters) could not be performed. Studies that examine the effect of editing and filtering NN-interval data are available, however they are predominantly focused on either simulated or adult-derived ECG data. As such, it is recommended that future studies more thoroughly examine how different degrees of editing and filtering affects reliability of HRV measures obtained from NN-interval data collected from pediatric samples.

HRV Analysis

Longer ECG recording duration was associated with lower NN reliability among infant/toddler studies, and higher HF and LF reliability among child/adolescent studies. Further, longer ECG recording duration was associated with lower LF:HF ratio reliability among child/adolescent studies, which is a paradoxical finding given the results for HF and LF. Infant/toddler studies typically used 2- to 5-min ECG recordings for HRV analysis, whereas child/adolescent studies typically used 3-min recordings. While these recording durations do not conform to the typical 5-min standard, results of this meta-analysis suggest that HF can be reliably measured even when a recording of only 2 or 3 min is obtained. The observed LF results are consistent with Marks and Lightfoot (1999), who demonstrated greater LF reliability when measures were derived from repeated 5-min, compared to 2.5-min ECG recordings. Deriving reliable measures of LF requires longer ECG recordings than HF because low-frequency heart rate oscillations typically occur in intervals of 5-min or greater (Heathers, 2014; Kleiger et al., 2005). While longer ECG recordings can produce more reliable measures of HF and LF, the reduced reliability of LF:HF ratio with longer ECG recordings may reflect how the relative

contributions of HF and LF activity across a longer period of time (e.g., longer ECG recording) are less easily replicated upon study follow-up. However, results for LF:HF ratio must be interpreted with caution as only three studies contributed data for LF:HF ratio reliability; among these, the study with the longest ECG recording duration had the lowest effect size, and vice versa. Meta-analytic results also indicated that HF reliability among infant/toddler studies was higher when measures were derived using the mean of multiple ECG analysis epochs (relative to a single, longer ECG epoch or to studies that did not report a derivation method); similar results were not observed among child/adolescent studies. Among studies reporting this information, 30-sec epochs were typically used for both age groups. Results for studies that used multiple, shorter epochs were consistent with earlier findings from infant (Richards, 1995) and adult studies (McNames & Aboy, 2006). Shorter epochs are less contaminated by artifacts and recording errors that could distort measures and hinder reliability. As such, the use of multiple epochs for HF among infant/toddler studies may have better accounted for the presence of artifacts and recording errors, which increase variance in the measure. Given that HF reliability among infant/toddler studies was not associated with ECG recording duration, but was associated with the use of multiple epochs, it is possible that HRV reliability is influenced by an interaction between ECG recording duration and the HRV derivation method (e.g., better reliability among studies with both longer ECG recording durations and more ECG epochs). This could be tested using nested analyses; however, this was not possible in the present study due to limited available data. As such, it will be important for pediatric researchers to further examine how HRV reliability is impacted by deriving measures using multiple ECG recording durations (e.g., 5 min, 10 min, 15 min),

multiple data reduction techniques (e.g., mean of 2 epochs vs. mean of 10 epochs), and any interactions between the two. This will help inform decisions regarding optimal ECG recordings for pediatric HRV research, and extend available HRV guidelines.

Selection of an HRV frequency bandwidth consistent with available guidelines and recommendations did not impact HF reliability among infant/toddler studies, but was associated with higher HF reliability among child/adolescent studies; effect sizes for LF reliability among child/adolescent studies were higher when a recommended frequency bandwidth was used relative to when an alternative bandwidth was selected. Spectral frequency bands determine the measurement boundaries of frequency-domain variables; if variability is above or below the frequency bandwidth cut-offs, results will provide less accurate estimates of autonomic cardiovascular control. Notably, published recommendations for HRV frequency bandwidths appropriate for pediatric studies only exist for HF among infants and toddlers up to 4-years-old (i.e., HF 0.24–1.04 Hz for children under 4-yrs; HF 0.20–1.20 Hz or 0.30–1.30 Hz for infants; Bar-Haim et al., 2000; Fox & Porges, 1985); as such, most HRV studies with children and adolescents use adult frequency bandwidths (i.e., HF 0.15–0.40 Hz, LF 0.04–0.15 Hz; Berntson et al., 1997; Task Force, 1996). Frequency bandwidth selection may not have affected reliability across infant/toddler studies because the wider bandwidths used among this age group accounted for heart period variance attributed to faster and more chaotic cardiovascular and respiratory patterns. Relatedly, the prominent use of a moving polynomial filter when deriving HF among infant/toddler studies, compared to child/adolescent studies that did not typically use this same filter, may have further influenced the present results. Among child/adolescent studies, those that used the HF

frequency bandwidth recommended for adults yielded higher HF reliability compared to studies that used a narrower bandwidth to account for paced breathing (e.g., 0.20–30 Hz; Mezzacappa et al., 1997) or a wider bandwidth to account for a broader sample age-range (e.g., 0.20–1.00 Hz; Gentzler et al., 2012). Although the influence of other methodological factors across included studies must be considered, this result suggested that the HF bandwidth recommended for adults may be suitable for deriving reliable measures of HRV among samples of children and adolescents. However, limited available data among child/adolescent studies precluded an examination of whether reliability would differ between studies that applied frequency bandwidths, which were wider or narrower than those recommended for adults. Examining such nuances in HRV measurement among pediatric studies can help maximize the psychometrics and interpretability of HRV data obtained from children and adolescents. Overall, results suggested that frequency bandwidth selection is pertinent to obtaining reliable measures of spectral HRV among children and adolescents. However, although this study demonstrated that use of adult HRV bandwidths (opposed to an alternative) was associated with higher reliability, this result does not imply that HRV measures derived using adult frequency bandwidths are necessarily valid. Given that frequency bandwidth recommendations are currently available only for adults and for infants/young children, the lack of frequency bandwidth recommendations for use among children and adolescents is a notable gap in the literature. It is recommended that future research examine cross-sectional and longitudinal HRV measures using a variety of frequency bandwidths to determine whether deriving reliable and valid HRV measures from children and adolescents requires a unique frequency bandwidth, or whether adult

bandwidths are equally appropriate among child and adolescent studies; this may help establish recommendations specific to HRV studies with pediatric samples.

Limitations, Strengths, and Current State of the Literature

There are four primary limitations of the present meta-analysis that merit consideration. First, results must be interpreted in light of the coding decisions. Namely, although multiple variables were initially coded from each included study, the performed moderator analyses were restricted due to limited variability. One example is for the coding of the time of ECG recording. This moderator was intended to examine reliability between studies that recorded ECG in the morning (06:00–12:00), afternoon (12:00–18:00), and evening (18:00–24:00). However, given that most studies failed to report this detail, analyses could only be performed after collapsing this moderator into a dichotomous variable (i.e., same vs. different time of ECG recording). Another example is for the coding of ECG sampling rate. This moderator was intended to examine reliability between studies that used slower (128 Hz, 250 Hz) relative to faster (500 Hz, 1000 Hz) sampling rates. However, given that most studies also failed to report this detail, analyses could only be performed after collapsing this moderator into a dichotomous variable (< 500 Hz, ≥ 500 Hz). As such, several moderator variables were collapsed into categories that were more restricted than initially planned; this precluded a more comprehensive examination of how such factors influence test-retest reliability, and restricted testing of several specific hypotheses. Given that few included studies reported target methodological factors, the decision to collapse several moderators into dichotomous variables was warranted. Remaining coding decisions were developed in accordance with considerations for HRV measurement outlined by published guidelines

and recommendations (e.g., Bar-Haim et al., 2000; Berntson et al., 1997; Task Force, 1996), common methodology among the pediatric HRV literature, and decisions of similar methodological reviews (Heathers, 2014). However, given that most published HRV guidelines are specific to adults, other moderator variables were coded to reflect available standards. For example, while selection of HRV frequency bandwidths is key to obtaining valid and reliable frequency-domain HRV variables, bandwidths specific to children and adolescents are not available. As such, reliability among child/adolescent studies in the context of frequency bandwidth selection could only be examined in relation to adult standards. Further, as many studies did not report this detail, frequency bandwidth could only be examined using a collapsed, dichotomized (appropriate vs. alternative frequency bandwidth) variable, instead of examining reliability between studies that used wider or narrower frequency bandwidths than recommended. Collapsing of moderators into smaller, more broad categories was a noted limitation of this meta-analysis; however, doing so allowed for meaningful examinations of reliability moderators to be conducted. As such, future meta-analyses should attempt to replicate this coding strategy to examine whether the same moderator categories used here are similarly meaningful among a different sample of longitudinal HRV studies.

Second, the focus on time- and frequency-domain measures precluded an examination of several moderator variables. Time- and frequency-domain measures are only two of several analytical methods for quantifying HRV. HRV measures derived from geometric (e.g., triangular index), wavelet-based (e.g., wavelet coefficient), and entropy-based (e.g., approximate entropy) analyses provide alternative perspectives of autonomic cardiovascular control, have different analytical assumptions, and have also

been associated with health outcomes (e.g., Kleiger et al., 2005). However, it remains to be determined whether these alternative HRV analytic methods yield data with greater predictive validity for future health outcomes relative to time- and frequency-domain measures. There are many more studies across the HRV literature that measure and discuss time- and frequency-domain HRV measures relative to available alternatives. Thus, the decision to examine only time- and frequency-domain HRV measures in this meta-analysis was made to accommodate an examination of HRV variables most commonly reported in the pediatric literature. However, future meta-analytic studies could perhaps avoid this limitation by examining reliability of alternative HRV measures beyond time- and frequency-domain indices. In addition to providing more data that can be used for moderator analyses in future meta-analytic studies, including a broader selection of HRV measures, other than time- and frequency-domain indices, would allow for a suitable way to compare psychometric properties of HRV derived from different techniques (e.g., frequency-domain vs. wavelet analysis). In turn, this would also allow for more comprehensive analyses of HRV reliability

Third, examination of included studies using fixed-effects meta-analytic models may limit the generalizability of these results. A random-effects model assumes that variability in effect sizes is due to both random variance and sampling error, and also produces wider confidence intervals (Rosenberg et al., 2000). Thus, a random-effects model may have allowed for greater generalizability of these results to other time- and frequency-domain HRV studies. The present analyses used a fixed-effects model with non-redundant study effect sizes. Upon comparison of these results with those obtained using random-effects models, three overarching differences were observed between the

two sets of analyses (data not shown for parsimony). First, while moderate HRV reliability was still demonstrated, there were often slight differences in the mean effect size across many analyses; effect sizes from random-effects models were typically, but not consistently smaller than those produced using fixed-effects models. Second, heterogeneity statistics were typically not statistically significant among random-effects analyses, suggesting greater homogeneity of effect sizes across included studies. Finally, while some moderator analyses yielded significant results regardless of the meta-analytic model used (e.g., pre-recording acclimatization period for NN among infant/toddler studies, frequency bandwidth selection for HF among child/adolescent studies), several of the moderator analyses that were significant among fixed-effects models were no longer significant among random-effects models. However, methodological moderators that were not significant among fixed-effects models typically remained as such when examined using random-effects models. The differing sets of results between the two meta-analytic models suggest that caution regarding Type I error is warranted. Namely, while these results still indicated moderate test-retest reliability, results may only be applicable to the examined studies, and not a wider population of studies. However, this consideration does not necessarily nullify the observations and inferences made herein. For instance, it is possible that non-significant results observed across all analyses signify a robust lack of effect between certain moderators and HRV reliability. Given that the moderator variables chosen for this study reflect constructs that are conceptually meaningful for HRV research, it is almost equally relevant to understand which methodological factors are, and are not salient to producing reliable HRV data. The included moderators were coded to reflect a small number of methodological conditions

that were believed to capture all possible options represented in the extant HRV literature (e.g., specified vs. unspecified exclusion criteria; similar vs. dissimilar ECG recording durations). As such, a fixed-effects meta-analytic model was considered appropriate, given the nature of the data. Future meta-analyses of HRV reliability should consider replicating this comparative analytic strategy (i.e., fixed- vs. random-effects models) to determine whether these results are robust among a different sample of longitudinal HRV studies.

Finally, as indicated above, the minimal reporting of pertinent methodological details among included studies restricted the number of moderator analyses that could be performed. This precluded an examination of several study hypotheses across one or both age groups, but also speaks to the current state of the wider pediatric HRV literature. Included studies failed to report essential study protocol (e.g., time of ECG recording), sample (e.g., participant ethnicity, anthropometrics, sample attrition), or technical details (e.g., ECG sampling rate, editing/filtering). Relatedly, the majority of included studies exclusively reported HF. It is generally established that HF is a strong predictor of health and development among pediatric and adult participants; as such, it is understandable (i.e., for comparative reasons) that HF is the most commonly reported HRV variable among pediatric studies. Increased reporting of other time- and frequency-domain variables (e.g., SDANN, rMSSD, VLF) and methodological details associated with their measurement among the pediatric literature can provide additional information about physiological (i.e., autonomic, circadian, hormonal) mechanisms that influence autonomic cardiovascular control in children. Importantly, the minimal reporting practices observed among included pediatric studies precluded an examination of nested

analyses (i.e., moderators of moderators). The present study examined test-retest reliability in the context of several moderator variables, some of which were pertinent to reliability among studies in both age groups. However, due to limited available data, moderator variables were examined in isolation of one another. Thus, because many studies reported such minimal methodological information, potentially important interactions between the examined moderators could not be examined. Such methodological interactions may also be pertinent to HRV reliability. The examination of nested analyses was warranted in this meta-analysis because significant heterogeneity remained evident following many of the included moderator analyses. This heterogeneity suggests that there was additional variability among effect sizes that had yet to be explained. Several nested analyses appeared particularly relevant and would help to elucidate whether one moderator had more of an impact on reliability than another. For example, the present meta-analysis was unable to examine whether studies that included a pre-recording acclimatization period (a key moderator of HRV reliability observed in this study) yielded higher reliability of HF among longer (e.g., $\geq$ 1 year) relative to shorter (e.g., $\leq$ 1 month) studies, or whether further differences existed between the two age groups. As another example, many of the factors associated with HRV reliability pertained to the control of variance attributed to heart rate activity; in many cases among infant/toddler studies, use of Porges' (1985) moving polynomial filter was proposed as a methodological explanation for why HF reliability among infant/toddler studies was more robust to certain moderator variables (e.g., study follow-up length, use of software-based filters) relative to other HRV variables, or to HF among child/adolescent studies. As such, it would have been pertinent to conduct a nested analysis between studies that did

and did not apply this filter, and examine whether additional moderator analyses could account for remaining heterogeneity among effect sizes. In addition to the limitations posed for the present meta-analysis, minimal reporting of pertinent methodological details observed among included studies maintains previously noted gaps in the pediatric HRV literature. Taken together, observations from this research suggest that HRV studies among the pediatric literature would benefit from augmented reporting practices of key methodological and analytical details. More rigorous reporting of methodological and analytical details would increase the interpretability and replicability of published findings, and allow for more comprehensive and meaningful meta-analyses to be conducted in the future.

There are five noted strengths in the present meta-analysis. First, to the best of our knowledge, this is the only available meta-analysis of HRV test-retest reliability among pediatric samples. Thus, the present study has potential to contribute important new knowledge to the field of HRV research. Primarily, the present study identified pertinent methodological considerations for obtaining reliable HRV measures among infants, children, and adolescents. Similar examinations do not appear available even in the adult HRV literature; while a quantitative review of HRV methodology was recently published (Heathers, 2014), it was not a meta-analysis, and the studies consisted only of adult HRV data specific to frequency-domain analyses. As such, pediatric HRV researchers can use the empirical data reported herein to guide the design of future studies in a way that helps produce more reliable and stable measures of HRV. Second, examined moderator variables largely reflected methodological factors that merit consideration by researchers before collected ECG data is analyzed for HRV. Namely, examined factors related to

study design (e.g., study follow-up length), participant recruitment (e.g., exclusion criteria), and ECG signal processing (e.g., editing and filtering, frequency bandwidth selection) are factors that are relevant to most HRV research contexts, regardless of the analytical method used to actually derive HRV (e.g., frequency-domain analysis, Poincare plot analysis). Thus, it is plausible that results from this study are generalizable to the wider HRV literature (e.g., adult studies), and may not necessarily be specific to reliability of pediatric measures of time- and frequency-domain HRV. Third, the identification and selection of studies for this meta-analysis, as well as relevant moderators, were determined from a comprehensive and systematic review of the extant (pediatric) HRV literature; as such, this meta-analysis included a broad selection of relevant pediatric studies and potential moderators of HRV reliability. Fourth, excellent coding fidelity of included studies was established following the blind re-coding of articles by an independent rater; thus, coding decisions were adequately objective and reproducible. Finally, reliability was examined using both non-redundant and redundant effect sizes, which yielded largely similar results across most analyses between the two.

Conclusion

Taken together, HRV is a reliable measure of autonomic cardiovascular control among pediatric samples. Several moderator variables were found to influence HRV reliability among both infant/toddler and child/adolescent studies. Notedly, this meta-analysis demonstrated three key findings that appear most pertinent to pediatric HRV research. First, HRV reliability was higher and more robust to differences in study methodology among samples of children and adolescents relative to infants and toddlers. Second, methodological factors that control for extraneous variance attributed to heart

rate are pertinent to obtaining reproducible measures of HRV, particularly among samples of infants and toddlers who have faster and more chaotic heart rate activity relative to older children and adolescents. For instance, if participants are unable to habituate to experimental conditions or they are asked to maintain a posture that increases orthostatic stress during data collection, obtained ECG data would likely be contaminated by stress-related increases in heart rate, which in turn would distort mean-level estimates of HRV and limit reproducibility. Moreover, studies that do not properly edit or filter their ECG data for aberrations are likely to produce less reliable measures because the background influences of heart rate are not accounted for. Third, the pediatric HRV literature is notably restricted in its reporting of key methodological and technical details relative to the adult HRV literature; many of these details pertain to recommendations for improved inter-study consistency, as outlined by published HRV guidelines, and are important for interpretability and replicability of study findings. Further, as significant heterogeneity was observed across several analyses, including those wherein significant differences in reliability were identified between moderator variables (e.g., study follow-up length, ECG recording posture, data editing/filtering), the minimal reporting of methodological details among examined studies precluded the ability to conduct nested analyses, which are important to elucidate whether interactions among the moderator variables could help explain remaining heterogeneity among effect sizes.

As the current study has shown, carefully planned and maintained study methodology is an important consideration for obtaining reproducible measures of HRV from pediatric samples. However, given that HRV reflects a biological construct, methodological characteristics are likely not the only factors that influence the

psychometrics of pediatric HRV measures. HRV is mediated by multiple physiological, behavioural, and cognitive processes, which develop over time as children grow. Considering this it is plausible that internal, biological factors can also influence the reproducibility of HRV measures. Thus, in addition to the methodological considerations examined in this meta-analysis, it will be important for researchers to examine intra-individual factors that further moderate reproducibility. In sum, HRV can be reliably measured among pediatric samples, but there are likely several additional factors not examined in the current study that could explain differences in reliability among HRV studies with infants, children, and adolescents. As such, future researchers should consider replicating these meta-analytic procedures to examine whether HRV reliability is robust among similar moderators, and whether additional heterogeneity among effect sizes can be explained by the use of nested moderator analyses.

Study 2: Method

Cohort Profile

Children and their families participated in the QUebec Adipose and Lifestyle InvesTigation in Youth (QUALITY) Cohort. The QUALITY Cohort (see Lambert et al., 2011) included a large sample (N = 632) of children recruited from schools located within 75 km of Montreal or Quebec City, Quebec, Canada. The overarching aim of the QUALITY study was to investigate the natural history of excess weight, including its genetic, biological, environmental, and psychosocial determinants, and its related cardiometabolic consequences (e.g., autonomic dysfunction, inflammation). Study inclusion criteria included having at least one overweight biological parent. Parents were defined as overweight with either a self-reported height and weight that yielded a BMI > 30 kg/m^2, or a waist circumference > 102 cm in men and > 88 cm in women. To account for genetic admixture, children included in QUALITY were limited to those of Caucasian, Western European decent. Further exclusion criteria for child participants included: (a) previous diagnosis of Type 1 or 2 diabetes; (b) medical or psychological condition that compromised ability to participate (i.e., hospitalized within the last month for a serious illness; hypertension, renal failure, inflammatory bowel disease); (c) restricted low-calorie diet (≤ 600 kcal/day); or (d) medication prescribed for hypertension (e.g., thiazides, β-blockers). Informed consent and assent were provided by parents and children, respectively. The QUALITY study was approved by the ethics review board of Direction Santé Québec, Institut de la Statistique du Québec, and CHU Sainte-Justine Hospital (#2040). The present study is based on analyses of QUALITY data; only relevant methods and measures are described below.

Participants

Data for the present study are based on 632 children aged 8 to 10-years-old (*M* = 9.60 yrs, *SD* = 0.92) who participated in the first clinic visit between 2005 and 2008. Of these children, 564 (89%) returned for a second clinic visit after 2 years (*M* = 2.08 yrs, *SD* = 0.16) when they were 10 to 12-years-old (*M* = 11.66 yrs, *SD* = 0.94). Attrition was due to (a) child refused to participate at second visit (*n* = 12); (b) families were unavailable or could not be located at second visit (*n* = 21); or, (c) families were no longer interested in participating at second visit/other (n = 35). Demographic, anthropometric, and cardiovascular measures obtained from QUALITY participants were obtained at the Centre hospitalier universitaire Sainte-Justine in Montreal or Hôspital Laval in Québec City, Canada. Approximately 81% of participating families resided in the greater Montreal area.

Measures

Anthropometrics. Trained research nurses measured participant height, weight, and waist circumference. During these measurements, participants were dressed in light clothing with their shoes off. Height was measured using a stadiometer during maximal inspiration while participants stood against a wall. Weight was measured to the nearest 0.2 kg using an electronic scale calibrated regularly using a set of standard weights. Waist circumference was measured at the end of normal expiration using measuring tape that was placed around the narrowest part of the body, midway between the lowest rib cage and the iliac crest. Height and waist circumference were each measured in duplicate to the nearest 0.1 cm and the mean was obtained; if values differed by more than 0.5 cm, a third measure was taken, and the mean of the two closest measurements was obtained.

BMI was calculated as weight in kg divided by height in m^2. Age- and sex-adjusted height, weight, and BMI Z-score scores were also derived based on United States Center for Disease Control (2000) growth charts. A child was defined as overweight with a BMI $\geq 85^{th}$ to $< 95^{th}$ age- and sex-specific percentile, or as obese with a BMI $\geq 95^{th}$ percentile.

Puberty. Trained research nurses measured children's pubertal status using the Tanner puberty staging scale (Marshall & Tanner, 1969, 1970). The Tanner scale is a 5-point categorical system that classifies physical development in children and adolescents using measureable primary and secondary sex characteristics. Criteria used to classify Tanner stages were as follows: Tanner 1 (no body hair growth for either sex, no breast growth for girls or genital growth for boys); Tanner 2 (indication of body hair growth for either sex, indication of breast budding for girls or slight genital growth for boys); Tanner 3 (body hair is more abundant and more dispersed for either sex, indication of breast growth for girls or genital growth for boys); Tanner 4 (body hair has adult appearance but is not as abundant and does not extend to the inner thigh for either sex, further breast growth for girls or further genital growth for boys); Tanner 5 (complete body hair growth that extends to the inner thigh for either sex, adult-sized breasts for girls or adult-sized genitals for boys). Physician-rated measures of puberty are generally considered as a gold-standard method compared to self- and parent-reported measures (Lee, Valeria, Kochman, & Lenders, 2006; Netherton, Goodyer, Tamplin, & Herbert, 2004); given that puberty was examined by trained nurses, pubertal status of QUALITY participants was measured in accordance with a highly valid methodology.

Blood pressure. Trained research nurses measured blood pressure (BP) using an appropriate-sized occlusion cuff (determined via arm circumference at the midpoint

between the acromion [i.e., shoulder blade] and elbow). The occlusion cuff was applied to the medial surface of the right arm over the brachial artery. Following a 5-min acclimatization period, measures of systolic (SBP) and diastolic (DBP) blood pressure were obtained in a seated posture using an oscillometric instrument (model CR9340; Dinamap XL, USA). Five consecutive blood pressure readings were obtained at 1-min intervals; the mean of the last three measures was used for analysis. Measurement of blood pressure was standardized for all participants at both clinic visits to maximize stationarity. Validity (Chiolero, Paradis, & Lambert, 2010; Park & Menard, 1998) and reliability (Skirton, Chamerlain, Lawson, Ryan, & Young, 2011) of this blood pressure instrument have been previously reported.

Physical activity. As part of a larger questionnaire, children answered questions about physical activity. Children rated how much time they spent watching television (in hrs) during a typical weekday and a typical weekend. They also rated how frequently they engage in physical activity during a typical week using a Likert scale ranging from 1 (*no* or *very little* physical activity) to 5 (*frequent* or *almost daily* physical activity). These questions were adapted from the International Physical Activity Questionnaire (IPAQ; Booth, 2000), an established self-report measure of physical activity with good test-retest reliability ($r = 0.8$) and moderate criterion validity against accelerometer data ($r = 0.33$; Craig et al., 2003). While objective measures of physical activity (e.g., accelerometry) are typically preferred over self-report, systematic reviews indicate that youth self-report of physical activity demonstrates adequate test-retest reliability (e.g., $r > 0.7$) and criterion validity (e.g., $r = 0.60 – 0.84$) against objective indices of physical activity (Lubans et al., 2011).

Heart Rate Variability

ECG signal acquisition. ECG was recorded using disposable, pre-gelled snap silver/silver-chloride electrodes arranged in a modified lead II configuration. To improve signal quality, electrode resistance was minimized (< 10 kO) by pre-cleaning the skin using a rubbing alcohol swap; this method is consistent with previous guidelines (Hermens, Freriks, Disselhorst-Klug, & Rau, 2000). The active electrode was placed on the right clavicle next to the sternum over the first rib between the two collarbones. A second electrode was placed on the left mid-clavicular line at the apex of the heart over the ninth rib. A ground electrode was placed near the lowest possible right rib on the abdomen. Additional impedance (dZ) electrodes were placed over the right fourth intercostal space at the sternal edge, the fifth intercostal space at the left axillary line, and on the sixth rib in the mid-clavicular line. ECG data were acquired in a seated posture using the 8500 Marquette MARS Holter Monitor (GE Marquette Medical Systems, Milwaukee, Wisconsin, USA) at a sampling rate of 128 Hz; data were stored on frequency-modulated cassettes for later processing.

ECG signal processing. ECG data were uploaded to the MARS Holter Analysis Workstation (GE Marquette Medical Systems, Milwaukee, Wisconsin, USA) for editing and analysis. Data were manually inspected for aberrant heart beats and recording artifacts. Beat-to-beat visual inspection of the shape, trend, and length of each QRS-complex was performed using standard Marquette algorithms. ECG data were also verified for accuracy through visual inspection by a board-certified cardiologist. Artifact removal was based on a 20% change from the immediately preceding signal (Kleiger, Miller, Bigger, & Moss, 1987). A mean of 6.5% and 6.3% of ECG data was removed due

to artifacts and errors from recordings obtained at the first and second clinic visits, respectively. Artifacts or automatically excluded NN-intervals were corrected using cubic spline interpolation (via Marquette algorithms), which replaced missing data points based on a minimum of four acceptable surrounding R-peaks. Finally, ECG data were linearly detrended, mean-centered, and tapered for spectral leakage using a Hanning window.

HRV analyses. Measures of HRV were derived for time- and frequency-domain variables based on all available ECG data. Time-domain variables included: NN, SDNN, rMSSD, and pNN50. Frequency-domain variables, derived using spectral analysis with Fast Fourier Transformation, included VLF (0.0033–0.04 Hz), LF (0.04–0.15 Hz), HF (0.15–0.40 Hz), and LF:HF Ratio. Heart rate was computed from NN using the following equation: 60,000 / NN (Berntson et al., 2007).

Procedure

Parents and children arrived at the clinic at approximately 07:00; children were instructed to fast for 12 hrs and abstain from strenuous exercise for 48 hrs before participating. Parents and children were greeted by a research assistant, reviewed the study procedures, and provided their consent. First, parents completed questionnaires (e.g., demographics, socioeconomic status) and reported child medication use over the preceding 2 weeks (e.g., antibiotics, pain/fever, colds/allergies). Second, children's height, weight, and waist circumference were measured, and pubertal status was determined using the Tanner staging system; height and weight were used to calculate measures of BMI. Third, research nurses prepared children for ECG measurement, which began at 08:00 and lasted ~3 hrs. Participants were seated and remained awake for the entire ECG recording duration. Fourth, the nurse removed the ECG recording equipment

and fitted the BP occlusion cuff to record blood pressure. Finally, children completed a questionnaire that inquired about physical activity and sedentary behaviour. The study protocol was identical across both clinic visits. Children underwent additional procedures as part of the larger QUALITY study, which are not relevant to the present analyses and are described elsewhere (Lambert et al., 2011).

Data Screening

Outliers and statistical assumptions. Data were screened to identify biologically implausible values and statistical extreme outliers (±6 standard deviations). From the first clinic visit, one participant had biologically implausible VLF, LF, and HF values, which were case-wise deleted; three additional values of LF and one value of weekday screen time were also deleted as extreme outliers. From the second clinic visit, one value for each of LF:HF ratio and weekday screen time were deleted as extreme outliers. Linearity and homoscedasticity were tested using bivariate scatter plots of each HRV and covariate pair from both clinic visits, and pairs of residuals. Visual inspection indicated most bivariate pairs exhibited linear relations and at least some homoscedasticity. The pattern of results was largely identical before and after the removal of additional potential outliers. Normality of each HRV and moderator variable from both clinic visits, and their residuals, were tested using visual inspection of histograms, PP-, and QQ-plots, and examination of skewness and kurtosis statistics (Kline, 2009; Tabachnick & Fidel, 2007). Normality was indicated for all variables and residuals at both clinic visits except weight, BMI, waist circumference, weekday and weekend screen time, SDNN, and frequency-domain HRV variables, as indicated by QQ-plots. A natural log transformation of each variable corrected non-normality. As the pattern of the results was identical between

untransformed and transformed data, results using untransformed data are presented for parsimony.

Missing data. Of the 632 participants who completed the first clinic visit, 72.5% had complete HRV data, and 87.3% had complete demographic (e.g., age, sex), cardiovascular (e.g., blood pressure), anthropometric (e.g., height, weight), and physical activity (e.g., self-reported weekly activity, weekday and weekend screen time) data at both visits. Prior to imputing missing values, independent samples *t*-tests and Chi-square analyses were conducted to examine whether HRV, demographic, cardiovascular, anthropometric, or physical activity measures obtained at the first clinic visit differed between return and non-return participants. For the HRV variables, independent samples *t*-tests yielded significant differences, such that those who did not return had lower NN (676.09 vs. 703.31 ms; t (573) = 2.75, p = 0.006), rMSSD (43.18 vs. 48.41 ms; t (573) = 2.40, p = 0.017), pNN50 (25.89 vs. 21.17%; t (573) = 2.46, p = 0.014), LF (1098.96 vs. 1302.38 ms^2; t (80) = 2.35, p = 0.021), and HF (730.95 vs. 931.96 ms^2; t (82) = 2.73, p = 0.008), compared to return participants. For the moderator variables, independent samples *t*-tests yielded significant differences, such that those who did not return had higher SBP (95.96 vs. 93.54 mm/Hg; t (629) = -2.32, p = 0.021), DBP (50.30 vs. 48.35 mm/Hg; t (629) = -2.96, p = 0.003), heart rate (89.72 vs. 86.17 bpm; t (573) = -3.55, p = 0.004), and increased weekday screen time (2.69 vs. 2.21 hours; t (626) = -2.05, p = 0.040), compared to return participants. There were no significant differences between return and non-return participants on measures of SDNN, VLF, LF:HF ratio, height, weight, waist circumference, BMI, or weekend screen time. Chi-square analyses

indicated no significant differences between return and non-return participants with regards to sex, pubertal status, or self-reported physical activity.

Missing data analyses indicated that while 185 out of 632 cases (29.27%) had at least one missing value, and in total 7,470 out of 65,728 values (11.37%) were missing; missing values appeared non-monotone. Little's tests indicated missing data for all HRV variables (χ^2 [59] = 241.03, $p < 0.001$) and all moderator variables (χ^2 [372] = 419.90, p = 0.044) were not missing completely at random. This suggests that there may be a systematic relationship between missing and observed data (Little, 1988; McKnight, McKnight, Sidani, & Figueredo, 2007; Scheffer, 2002); this finding coincides with the examination of differences among measures obtained at the first clinic visit between the return and non-return participants. Missing data were imputed via multiple imputation with re-sampling techniques. All available demographic, cardiovascular, anthropometric, and HRV data were entered to inform the imputation model (Sterne et al., 2009). Five imputed datasets were created. Imputed data yielded an identical pattern of results compared to non-imputed data, although resulting imputed values were noticeably smaller relative to non-imputed data; only non-imputed results described for parsimony.

Analytic Strategy

Analyses were performed using SPSS v. 20 software (SPSS, INC., Chicago, IL). The overarching aim of the present study was to examine the 2-year stability of time- and frequency-domain HRV measures in a large sample of children. First, stability of NN, SDNN, rMSSD, pNN50, VLF, LF, HF, and LF:HF ratio was examined using Pearson (*r*) and Intra-class Correlations (*ICCs*). Average-measures *ICCs,* which range from 0 to 1, represent the ratio of between-subject to total variability ($ICC = \sigma^2_{BS} / [\sigma^2_{BS} + \sigma^2_{WS}]$), and

were calculated using a one-way random model. In contrast to Pearson correlations, which only account for rank-order differences between two variables, *ICC*s are advantageous in stability analyses as they account for rank order and mean-level differences between the two variables (Field, 2009; Weir, 2005). *ICC*s are categorized as poor (0–0.2), fair (0.3–0.4), moderate (0.5–0.6), strong (0.7–0.8), or excellent (0.9–1.0; Shrout & Fleiss, 1979).

Second, partial correlations ($r_{ab.c}$) were calculated while controlling for *a priori* moderators specific to measures obtained at the first clinic visit, and time-varying covariates: age, heart rate, blood pressure, pubertal status, height, weight, BMI, waist circumference, weekday and weekend screen time, and weekly physical activity. Residualized change scores for the time-varying covariates were derived using linear regression analyses; data from the first clinic visit predicted data from the second, and the resulting coefficient was subtracted from its respective variable obtained at the second clinic visit (Field, 2009; Llabre, Spitzer, Saab, Ironson, & Schneiderman, 1991). Partial correlation analyses partialled out: (a) the variable measured at the first clinic visit, and (b) the time-varying covariate, which included both the residualized change score and its respective variable from the first clinic visit.

The main study hypotheses were: (a) time- and frequency-domain HRV would exhibit moderate 2-year stability, as indicated using Pearson correlations and *ICC*s; (b) accounting for intra-individual differences in age, heart rate, blood pressure, pubertal status, height, weight (i.e., adiposity indices), and physical activity measured at the first clinic visit, and change among these variables across time would augment HRV stability, as indicated using partial correlations controlling for the variable measured at the first

clinic visit and for time-varying covariates; and (c) measures of heart rate, blood pressure, pubertal status, and weight would have the strongest moderating effects on HRV stability across both partial correlation analyses, particularly among HRV measures which predominantly reflect parasympathetic modulation (i.e., rMSSD, HF).

Exploratory linear regression models were produced as a secondary analysis to examine individual associations between (a) covariates measured at the first clinic and (b) time-varying covariates (i.e., the residualized change score and its associated variable measured at the first clinic visit), and HRV measured at the second clinic visit. All models were adjusted for HRV measured at the first clinic visit. Both standardized and unstandardized beta coefficients were examined.

Study 2: Results

At the first visit, 632 children and their parents participated; 564 children returned for a second visit. The majority of parents (M_{age} = 41.2, *SD* = 5.6 yrs) were overweight/ obese (78.7%, BMI ≥ 25 kg/m^2), and had a post-secondary education (79.4%). All parent and child participants were Caucasian and of European descent (due to genetic inclusion criteria). At the first visit, 54.4% of children (M_{age} = 9.60, *SD* = 0.92 yrs) were male, and 78.2% were classified in Tanner stage 1; children had ~6 hours per week of screen time (e.g., watching TV, playing computer games) and 16.1% reported no physical activity in a typical week. At the second visit, 49.5% of children (M_{age} = 11.66, *SD* = 0.94 yrs) were male, and 29.4% remained classified in the pre-pubertal stage; children had ~7.5 hours per week of screen time and 19.8% reported no physical activity in a typical week (See Table 8).

Measures of height and heart rate from the QUALITY sample were comparable to population-based normative data from similarly aged children; measures of systolic and diastolic blood pressure were lower in comparison (Fleming et al., 2011; ELDEQ, Jarrin et al., In press; QCAHS, Lambert et al., 2011; Paradis et al., 2004). Not surprisingly, due to the objectives of the study, measures of weight and adiposity (BMI, waist circumference) from this sample were higher relative to population-based normative data (ELDEQ, Jarrin et al., In press; QCAHS, Lambert et al., 2011). Moreover, the QUALITY cohort were of higher SES, and were more likely to come from two-parent homes and reside in urban regions relative to a representative sample of similarly-aged Quebec children (Lambert et al., 2011).

ECG Data Integrity

The methodological protocol for the clinic visits was standardized for all participants, such that ECG recordings began in the morning between 08:00 and 08:30. Across both visits, approximately 200 minutes of ECG data were recorded and only ~6% of data were excluded due to errors and artifacts (see Table 9). One participant from each clinic visit had less than 1 hour of ECG data available, and approximately 3% of participants at each clinic visit had ≥ 20% of ECG data removed due to errors and artifacts; results were similar before and after removal of this data (results not shown for parsimony). Time- and frequency-domain HRV measures were largely consistent with population-based normative data from similarly-aged children (Jarrin et al., In press), but were less consistent with normative data from studies with smaller samples (Acharya, Kannathal, Sing, Ping, & Chua, 2004; Martini et al., 2001; Rodriguez-Colon et al., 2011; Silvetti et al., 2001).

HRV Stability

Pearson correlations. To test the first hypothesis that time- and frequency-domain HRV measures exhibit moderate 2-year stability, Pearson product-moment correlations were calculated for each HRV measure across both QUALITY visits (see Table 10). Results yielded statistically significant correlations for all time- and frequency-domain measures and indicated that HRV was moderately stable across 2 years. Correlations among time-domain measures ranged from $r = 0.67$ to 0.72; NN was the relatively least stable measure, whereas SDNN was the relatively most stable. Correlations among frequency-domain measures ranged from $r = 0.71$ to 0.75; LF was the relatively least stable measure, whereas HF was the relatively most stable. These

results supported the first study hypothesis that HRV was moderately stable across 2 years.

Intra-class correlations. Stability of time- and frequency-domain HRV measures was also examined using intra-class correlations (*ICCs;* see Table 10). Results yielded statistically significant *ICC*s for all time- and frequency-domain measures and indicated that HRV exhibited strong stability across 2-years. *ICC*s for the time-domain measures ranged from *ICC* = 0.74 to 0.82; NN was the relatively least stable measure, while SDNN and rMSSD were the relatively most stable. *ICC*s for the frequency-domain measures similarly ranged from *ICC* = 0.79 to 0.85; VLF was the relatively least stable measure, while HF was the relatively most stable. *ICCs* suggest that within-subjects variance in HRV was low, that mean-level differences in HRV between the first and second clinic visit were minimal, and that participants' relative rank-order of HRV was maintained across both QUALITY visits. These results further supported the first hypothesis that HRV is stable across 2 years.

Partial correlations. To test the second hypothesis that controlling for developmentally-relevant moderators would augment HRV stability, partial correlations for each HRV variable were examined, adjusting for (a) individual covariates measured at the first clinic visit (age, heart rate, blood pressure, pubertal status, growth and adiposity, physical activity), and (b) time-varying developmental change in each covariate measured across both clinic visits (i.e., a residualized change score and the respective variable measured at the first clinic visit).

Results of the partial correlation analysis yielded statistically significant correlations for all time- and frequency-domain measures, indicating that all HRV

variables were moderately stable across 2 years. Stability was robust to the effects of most examined moderators when analyzed using covariates from the first clinic visit ($r_{\text{time } range}$ = 0.24–0.72; $r_{\text{frequency } range}$ = 0.65–0.75; see Table 11) and time-varying covariates ($r_{\text{time } range}$ = 0.55–0.75; $r_{\text{frequency } range}$ = 0.71–0.80; see Table 12). Specific results for each of the moderators follows.

Age. Controlling for age at the first clinic visit, and change in age across the study follow-up did not alter HRV stability (see Tables 11 and 12). This result could be explained by the minimal variability in participant age at both the first (SD_{age} = 0.92 yrs) and second (SD_{age} = 0.94 yrs) clinic visit. Results did not support the hypothesis that age augments HRV stability, and suggested that HRV stability is robust to differences in participant age at the first clinic visit and the change in age across a 2-year period.

Heart rate. Controlling for heart rate measured at the first clinic visit yielded weak to moderate stability of time-domain measures (r_{range} = 0.24 to 0.61) relative to stability examined by Pearson correlations, while moderate stability of frequency-domain measures (r_{range} = 0.65 to 0.72) was retained; heart rate did not alter stability of LF:HF ratio. Conversely, controlling for time-varying change in heart rate across the study follow-up yielded moderate stability of all time-domain (r_{range} = 0.55 to 0.75) and frequency-domain measures (r_{range} = 0.72 to 0.80). Relative to other HRV variables, NN stability was weakest after controlling for heart rate among both analyses (see Tables 11 and 12), which is not surprising considering that NN is the inverse of heart rate (Berntson et al., 2007). Moreover, stability of rMSSD (r = 0.75 vs. 0.69), pNN50 (r = 0.72 vs. 0.68), and HF (r = 0.80 vs. 0.75), all of which reflect parasympathetic modulation, was augmented relative to stability examined by Pearson correlations after controlling for

time-varying change in heart rate. Given that reduced stability was observed after controlling for heart rate measured at the first clinic visit was not expected, results provided partial support for the hypothesis that controlling for heart rate augments HRV stability.

Blood pressure. Controlling for blood pressure measured at the first clinic visit, and change in blood pressure across the study follow-up did not alter HRV stability. Results from both analyses indicated that HRV remained moderately stable; values were largely consistent with stability examined by Pearson correlations after controlling for differences in blood pressure at the first clinic visit and time-varying change in blood pressure across the study follow-up (see Tables 11 and 12). Results were identical when blood pressure Z-scores were used (data not shown for parsimony). Results did not support the hypothesis that controlling for blood pressure augments HRV stability, and suggested that HRV stability is robust to blood pressure and change in blood pressure across a 2-year period.

Puberty. Controlling for intra-individual differences in pubertal status measured at the first clinic visit and pubertal development across the study follow-up did not alter HRV stability (see Tables 11 and 12). This result could be explained by the minimal variability in pubertal status, and change in pubertal status across the study follow-up. Most QUALITY participants were classified as Tanner 1 (78.2%) at the first clinic visit, and most were classified as Tanner 2 or Tanner 3 (50%) at the second clinic visit. Results did not support the hypothesis that controlling for pubertal status augments HRV stability, and suggested that HRV stability is robust to pubertal status and pubertal development across a 2-year period.

Growth and adiposity. Controlling for measures of height, weight, BMI (CDC [2000] age- and sex-adjusted Z-score), or waist circumference obtained at the first clinic visit, and the change among these measures across the study follow-up did not alter HRV stability (see Tables 11 and 12). Results were identical when height and weight Z-scores (age- and sex-adjusted) were used (data not shown for parsimony). Controlling for BMI and waist circumference measured at the first clinic visit, and change among these measures across the study follow-up increased NN, SDNN, and LF stability relative to other HRV measures. Results did not support the hypothesis that controlling for measures of growth and adiposity augments HRV stability, and suggested that HRV stability is robust to measures of height, weight, BMI, and waist circumference, and change among these measures across a 2-year period.

Physical activity. Controlling for self-reported measures of physical activity and screen time measured at the first clinic visit, and change among these measures across the study follow-up did not alter HRV stability (See Tables 11 and 12). Results did not support the hypothesis that controlling for measures of physical activity augments HRV stability, and suggested that HRV stability is robust to measures of physical activity and change in physical activity across a 2-year period.

Secondary *Post Hoc* Analyses

To further test for relations between each proposed moderator and HRV, exploratory linear regression models examined the variance explained in each HRV variable measured at the second clinic visit by each covariate, after adjusting for HRV measured at the first clinic visit. Three models for each HRV measure were tested including (a) covariate measured at the first clinic visit, (b) residualized change scores

(variable at the first clinic visit predicting variable at the second), and (c) time-varying covariates, which included a residualized change score covariate and its respective covariate measured at the first clinic visit). The magnitude and direction of each regression coefficient was largely identical between analyses using time-varying covariates and analyses using only residualized change scores (i.e., change score not adjusted for the respective measure obtained at the first clinic visit). As such, specific results from regression models using only residualized change score covariates are not detailed for parsimony. Standardized and unstandardized coefficients for covariates measured at the first clinic visit are presented in Tables 13 and 14, respectively. Standardized and unstandardized coefficients for time-varying covariates are presented in Tables 15 and 16, respectively.

Among covariates measured at the first clinic visit, significant positive associations were observed between age and NN ($B = 0.85$, $\beta = .12$, $p < .001$) and VLF ($B = 8.52$, $\beta = .10$, $p < .001$); puberty status and NN ($B = 17.64$, $\beta = .11$, $p < .01$); systolic blood pressure and NN ($B = 0.71$, $\beta = .07$, $p < .05$) and LF ($B = 8.05$, $\beta = .07$, $p < .05$); heart rate and NN ($B = 4.54$, $\beta = .48$, $p < .05$); all cardiovascular measures and LF ($B_{\text{range}} = 8.05$ to 12.58, $\beta_{range} = .07$ to $.10$, all $p < .05$); height and all HRV measures except HF and LF:HF ratio ($B_{range} = 0.15$ to 18.66, $\beta_{range} = .08$ to $.17$, all $p < .05$); all adiposity indices and all HRV measures except LF:HF ratio ($B_{range} = 0.13$ to 124.09, $\beta_{range} = .08$ to $.16$, all $p < .05$); and weekend screen time and VLF ($B = 21.45$, $\beta = .07$, $p < .05$). Standardized coefficients indicated that heart rate measured at the first clinic visit had the largest relative influence on HRV (i.e., NN) measured at the second clinic visit. The magnitude of the effect size between heart rate measured at the first clinic visit and NN is

consistent with results from the stability analysis (see above); however, the directionality of this association was opposite to that indicated by partial correlation results.

Among time-varying covariates (i.e., both residualized change scores and the respective variable measured at the first clinic visit), significant positive associations were observed between all cardiovascular indices and LF:HF ratio (B_{range} = 0.03 to 0.07, β_{range} = .11 to .33, all $p < .01$); height and LF:HF ratio (B = 0.05, β = .57, $p < .001$); weight and LF:HF ratio (B = 0.02, β = .28, $p < .05$); weekly physical activity and all HRV variables except LF:HF ratio (B_{range} = 17.12 to 1322.14, β_{range} = .33 to .39, all $p < .01$); and weekday screen time and LF:HF ratio (B = 0.09, β = .13, $p < .01$). Significant negative associations were observed between pubertal status and NN (B = -53.41, β = -.38, $p < .01$), VLF (B = -687.41, β = -.43, $p < .01$), and LF:HF ratio (B = -0.62, β = -.47, $p < .001$); age and rMSSD (B = -1.32, β = -.90, $p < .05$), pNN50 (B = -1.28, β = -1.01, $p < .05$), and HF (B = -51.62, β = -.87, $p < .05$); all cardiovascular indices and all HRV measures except LF:HF ratio (B_{range} = -0.34 to -69.37, β_{range} = -.09 to -.1.02, all $p < .05$); weight and all HRV measures except LF:HF ratio (B_{range} = -0.35 to -25.76, β_{range} = -.32 to -.43, all $p < .05$); waist circumference and all HRV measures except LF:HF ratio (B_{range} = -0.29 to -22.28, β_{range} = -.21 to -.41, all $p < .05$); and weekday screen time and NN (B = -7.02, β = -.10, $p < .05$), rMSSD (B = -1.45, β = -.10, $p < .05$), pNN50 (B = -1.38, β = -.11, $p < .05$), and HF (B = -53.91, β = -.09, $p < .05$; see Table 14). Standardized coefficients indicated that time-varying measures of heart rate had the largest relative influence across all regression models and across all HRV variables. Results demonstrated that greater change in heart rate over time was associated with lower HRV at the second clinic visit;

this is consistent with results from partial correlation analyses, in that HRV stability was typically augmented after controlling for these time-varying changes.

Regression analyses demonstrated that many of the examined covariates measured at the first clinic visit were positively associated with HRV, while many of the examined time-varying covariates were negatively associated with HRV. Overall, although HRV stability was robust to the influence of most selected covariates, and time-varying change among these covariates, results from the secondary analysis suggested that consideration of such developmentally-relevant moderators (e.g., age, pubertal status, weekly physical activity) may be pertinent to a thorough understanding of factors that influence developmental changes in HRV over time.

Study 2: Discussion

The overarching aim of the second study was to examine the 2-year stability of time- and frequency-domain measures of HRV within a large sample of Quebec children. Overall, and consistent with the main study hypothesis, HRV exhibited moderate stability across 2 years. Further, frequency-domain measures exhibited greater stability relative to time-domain measures. These results replicate and extend previous findings among similarly-aged children (El-Sheikh, 2005; Hinnant & El-Sheikh, 2009; Hinnant et al., 2011; Keller & El-Sheikh, 2009; Leicht & Allen, 2008), and among adolescents and young adults (Li et al., 2009).

While previous cross-sectional and longitudinal pediatric studies observed negative associations between biological factors (e.g., heart rate, blood pressure, and weight) and mean-level differences in HRV, a comprehensive examination of whether such factors affected HRV *stability* had not been conducted. To address this gap, the present study examined HRV stability while controlling for *a priori* selection of developmentally-relevant, time-varying demographic, cardiovascular, anthropometric, and behavioural covariates. The second hypothesis tested whether accounting for differences in age, heart rate, blood pressure, pubertal status, growth and adiposity, and physical activity across 2 years would augment HRV stability.

Developmentally-Relevant Covariates

HRV stability was robust to changes in age. Age-related changes in HRV and HRV stability have been observed across infancy through adolescence. HRV increases between infancy and childhood, but subsequently decreases throughout adolescence and across adulthood (e.g., Finley et al., 1987; Goto et al., 1997; Longin et al., 2009; Seppala et al., 2013; Silvetti et al., 2001); this likely reflects developmental changes in autonomic

cardiovascular control across the lifespan. Associations between age and HRV have typically been examined in cross-sectional studies, many of which compare groups of small samples ($N < 40$) that included wide age ranges (e.g., 6–10 years, 5–11 years); however, cross-sectional methodology precludes consideration of the temporal relation between age and HRV. Results from previous longitudinal studies indicate that HRV stability is weak-to-moderate across infancy, and increases to moderate stability across early and middle childhood; this stability is maintained through adolescence (Bar-Haim et al., 2000; Bornstein & Suess, 2000; El-Sheikh, 2005; Henje Blom et al., 2009; Hinnant & El-Sheikh, 2009; Hinnant et al. 2011). Methodological features of the QUALITY study, namely, inclusion of 9-year-old children followed-up after 2 years, may have limited variability in participant age such that the change in age across 2 years was too narrow to observe the proposed effect. This restriction in age among QUALITY participants can explain why results did not support the hypothesis that controlling for age would augment HRV stability. To better examine whether controlling for age augments HRV stability, future studies should replicate these analyses using a sample of children with a wider age range.

Controlling for heart rate influenced HRV stability. After controlling for heart rate measured at the first clinic visit, HRV stability was weak to moderate relative to results from Pearson correlations. However, after controlling for the change (i.e., decrease) in heart rate over 2 years (while also accounting for heart rate measured at the first clinic visit), HRV stability was moderate across all variables. Further, controlling for change in heart rate noticeably augmented stability of rMSSD, pNN50 and HF; given that these HRV measures all reflect parasympathetic modulation, it is not unexpected that there

would be higher stability after partialling out the change in heart rate (which is highly influenced by sympathetic mechanisms). The relatively higher (yet still moderate) HF stability after controlling for time-varying heart rate measures is consistent with results from a 6-month study with a smaller sample (N = 77) of healthy adolescents (M_{age} = 16.5 years; Henje Blom et al., 2009); however, reduced stability of SDNN and LF found in this study is inconsistent with Henje Blom and colleagues (2009), who reported relatively higher SDNN and LF stability after controlling for continuous heart rate measures. Importantly, Henje Blom made no indication of whether analyses controlled for a single time-point heart rate measure from their first or second study visit, or if a change score was used; thus, methodological differences may account for these discrepancies. Results suggested that accounting for extraneous variance in HRV measures attributed to heart rate, and developmental changes in heart rate over time, is pertinent to longitudinal examinations of HRV among children; controlling for this variance can augment estimates of stability. While HRV is derived from heart rate, the two measures reflect different constructs with somewhat distinct implications; indeed, HRV measures are typically more sensitive to detecting improvements or declines in health relative to heart rate measures (Ernst, 2014). However, not considering the effect of heart rate on HRV measures may lead to erroneous conclusions about relations between HRV and health. The QUALITY study, along with many other studies with child and adolescent participants, did not apply a moving polynomial filter in the HRV data reduction phase; as the polynomial filter is designed to isolate variance in the HF band from extraneous background variability, application of this filter reduces the need to statistically control heart rate when studying HRV. As such, future researchers should examine whether

controlling for developmental changes in heart rate would result in similar augmentation of HRV stability after a moving polynomial filter is applied to the data

HRV stability was robust to measures of systolic and diastolic blood pressure obtained at the first clinic visit, and changes in blood pressure across the 2-year follow-up after also controlling for blood pressure measured at the first clinic visit. Consistent with developmental patterns, QUALITY participants' systolic and diastolic blood pressure increased across the study follow-up. Baroreflex receptors, located along the walls of major blood vessels (e.g., aortic arch, carotid sinuses), respond to increases and decreases in blood pressure by triggering autonomic mechanisms that decrease and increase heart rate, respectively, to maintain homeostasis. Thus, moment-to-moment changes in blood pressure can indirectly modulate heart rate, and thus influence HRV. Baroreflex sensitivity has been observed to decline across childhood and adolescence (e.g., Lenard et al., 2004); however, higher mean-level measures of HRV have been associated with higher blood pressure among previous cross-sectional studies (e.g., between pre-pubertal and post-pubertal children; Tanaka et al., 2000). Further, moderate associations between developmental increases in blood pressure and HRV have been observed among previous longitudinal studies (e.g., 14-year-olds followed-up after 8 years; Campbell et al., 2003). Given that QUALITY participants were about 9-years-old and followed-up after 2 years, it is possible that too little time had elapsed between clinic visits to detect an effect of blood pressure change on HRV stability in the present study. Considering this, future research could examine whether HRV stability is impacted by changes in blood pressure that occur over longer periods of time (e.g., 5+ years). Taken together, the present results

suggest that HRV stability is robust to intra-individual differences and 2-year change in blood pressure.

Puberty status, measured using the Tanner staging system, did not alter HRV stability. Significant physical and biological (e.g., metabolic, hormonal) changes occur during puberty (Shirtcliff et al., 2009), and include developments in autonomic and cardiovascular functioning. Circulation of sex hormones (e.g., testosterone, estrogen) increases with more advanced pubertal development, and has been associated with sex-specific alterations in parasympathetic functioning, heart rate, and blood pressure regulation (e.g., Spear, 2000). However, previous associations between pubertal status and HRV have been largely inconsistent. For example, lower mean-level measures of LF and SDNN, and higher HF have been observed in pre-pubertal, relative to post-pubertal children (Chen et al., 2012; Kowalewsky et al., 2007). Further, while El-Sheikh (2005) observed a positive association between pubertal status and HF, Jarrin and colleagues (In press) observed negative associations between pubertal status and rMSSD, pNN50 and HF. Inconsistent results between studies may relate to how puberty was measured (i.e., Tanner staging vs. measure of pubertal onset [i.e., gonadarche, menarche] vs. self-report). Most children in the QUALITY sample were classified in Tanner 1 at the first clinic visit, while approximately half of children were classified in Tanner 2 or 3 by the second clinic visit; only a minority of participants had reached a more advanced pubertal stage (i.e., Tanner 4, Tanner 5) by the 2-year follow-up. As such, limited variability among pubertal status and pubertal development in the QUALITY cohort may explain why there was no association with HRV stability. To address this shortcoming, and more comprehensively examine whether physiological changes associated with puberty influences HRV

stability, future researchers could obtain repeated measures of HRV from a youth followed-up across each Tanner stage. Further, an experimental study that involves administration of sex hormones to children and adolescents may allow for a controlled examination of how physiological changes associated with puberty affect HRV and HRV stability. Given the present findings, results did not support the hypothesis that pubertal status augments HRV stability, and suggested that HRV stability was robust to pubertal development across 2 years.

Growth, measured by increases in height, weight, BMI, and waist circumference over time, did not alter HRV stability. Physical growth during child development can affect HRV because cardiovascular demand is increasing to maintain adequate circulation throughout the body. The present findings for HRV stability are inconsistent with previous cross-sectional (e.g., Rodriguez-Colon et al., 2011; Seppala et al., 2013) and longitudinal (e.g., Campbell et al., 2003; Graziano et al., 2011) studies which have reported negative associations between increased height, weight, BMI, and waist circumference and mean level measures of SDNN, HF, and LF, and positive associations with mean-level measures of LF:HF ratio and heart rate. Graziano and colleagues (2011) demonstrated that children with lower, compared to higher NN and HF activity at 5.5-years-old demonstrated significantly greater BMI growth and were more likely to be overweight or at-risk for overweight at 10.5-years old. Campbell and colleagues (2003) demonstrated that BMI measured at 14-years-old was negatively associated with mean-level measures of HF and positively associated with mean-level LF:HF ratio measured at 22-years-old, after adjusting for systolic blood pressure measured at 14-years-old. While this study provides support for the longitudinal relation between HRV and BMI,

Campbell and colleagues (2003) did not examine time-varying associations using BMI change scores. Given the 2-year follow-up of QUALITY, change in anthropometric measures among QUALITY participants was not as large relative previous studies with longer test-retest intervals. Graziano and colleagues (2011) and Campbell and colleagues (2003) demonstrated significant negative associations between anthropometric and HRV measures across 5 and 8 years, respectively. Thus, it is possible that the extent of change in anthropometric measures between the first and second QUALITY visit ($SD_{\text{weight, first clinic visit}}$ = 11.47 kg; $SD_{\text{weight, second clinic visit}}$ = 14.84 kg) was not substantial enough to affect HRV stability relative to studies with longer test-retest intervals. Future pediatric studies with longer test-retest intervals (e.g., 5+ years) would allow for a more thorough examination of the association between anthropometric change and HRV stability. Considering the present findings, results suggested that HRV stability is robust to developmental increases in measures of growth and adiposity across 2 years.

Measures of physical activity and weekday/weekend screen time did not affect HRV stability. Overall, relative to the first clinic visit, most QUALITY participants at the second visit reported less frequent regular physical activity, and watched more television on both weekdays and weekends. This suggests that participants may have become more sedentary across time. A meta-analysis of adult HRV studies (Sandercock, Bromley, & Brodie, 2005) demonstrated a positive moderate relation between physical activity training and parasympathetic dominance as measured by HRV (i.e., higher NN and HF); further, effect sizes for change in HF after physical activity training were higher among active, relative to sedentary, study participants. Associations between physical fitness and HRV among children are less consistent. Overall physical fitness has been associated

with higher parasympathetic dominance (e.g., Chen et al., 2012) and higher measures of LF, HF, and SDNN, even after controlling for BMI, systolic blood pressure, and blood glucose level (Henje Blom et al., 2009). Sedentary behaviour may augment sympathovagal imbalance across time, which could compromise HRV stability. Importantly, in the QUALITY study, only self-report measures of activity were available. Objective measurement of physical activity has potential to reduce response biases (e.g., recall errors), a psychometric concern which may affect child-reported data. As such, future longitudinal research should consider obtaining objective indices of physical activity (e.g., accelerometry, maximal oxygen consumption [VO_2 max]), to examine whether time-varying change among these measures can provide a better understanding of how physical fitness influences HRV stability in children compared to self-report measures. Overall, results did not support the hypothesis that controlling for change in physical activity augments HRV stability; results indicated that HRV stability was robust to changes among self-reported physical activity measures across 2 years.

Summary

The present study demonstrated that time- and frequency-domain HRV measures are moderately stable. Stability was reduced after controlling for heart rate measured at the first QUALITY visit, whereas stability of measures reflecting parasympathetic functioning (rMSSD, pNN50, HF) was augmented after controlling for the 2-year change in heart rate. However, HRV stability was robust to intra-individual differences in developmentally-relevant demographic (age, puberty), cardiovascular (blood pressure), growth and adiposity (height, weight, BMI, waist circumference) and self-reported

physical activity (weekday/weekend screen time, frequency of weekly physical activity) covariates.

Limitations, Strengths, and Future Recommendations

There are five primary limitations of the present study that warrant acknowledgement. First, QUALITY data were from a large, homogeneous sample of healthy children aged 9- to 11-years old. In turn, this sample homogeneity may pose a credible limitation in that variability among some examined covariates (e.g., age, pubertal status, height) was too limited to influence HRV stability. However, the 2-year follow-up improves upon shorter follow-up intervals used in the prior literature (e.g., 1 week, 2 weeks); this longer follow-up permits better assessment of the contribution of development and growth on HRV stability. Future studies should examine HRV stability using a more diverse sample (e.g., racial diversity, wider age range, normally distributed weight) to determine the generalizability of these results.

Second, the methodological protocol of the present study was highly controlled and standardized for all participants across both study assessments, which may influence the ecological validity of our stability findings. More specifically, the QUALITY study included detailed inclusion and exclusion criteria, ECG recordings lasted approximately 3 hrs and started between 08:00 and 09:00, identical hardware was used for all participants, and identical editing and scoring procedures were used for all HRV data. Further, most covariates examined in this study were measured by trained staff under standardized conditions. Observing HRV stability that was robust to several developmentally-relevant covariates may be partly attributable to the methodological rigor of the QUALITY study; clinical or practical uses of HRV may yield lower stability estimates. In other words,

clinic-based HRV stability measures may be less ecologically valid compared to measures obtained under less controlled or standardized conditions. However, QUALITY's examination of short-term (i.e., 3 hr) daytime time- and frequency-domain HRV measures obtained in a clinic setting is consistent with many other pediatric studies. Further, the moderate stability of HRV obtained in this study is consistent with results presented among other studies of similarly-aged children (e.g., Hinnant & El-Sheikh, 2009; Keller & El-Sheikh, 2009). Future studies should replicate the present analyses using 24-hr ambulatory recordings to address potential concerns about the overall ecological validity of these findings, and examine relations between HRV stability and objective measures of physical activity (e.g., via accelerometers, pedometers, or concurrent measures of heart rate and oxygen consumption; VO^2; Sirard & Pate, 2001) to address noted psychometric concerns. Moreover, other intra-individual factors not included in the present study that have been previously associated with HRV in children, such as ethnicity (e.g., Graziano et al., 2011; Li et al., 2009), socio-economic status (e.g., Campbell et al., 2013), and multiple sleep-related parameters (e.g., Burr, 2007; Jarrin et al., In press), should be examined in future studies.

Third, missing data analysis indicated that ~11% of individual data points among the examined HRV variables and covariates were missing; this was primarily due to sample attrition. Analyses using imputed data revealed significantly reduced stability relative to analyses using the original data. Participants who did not return for a second assessment appeared more at-risk, as indicated by significantly higher SBP, DBP, heart rate, and weekday screen time, and significantly lower NN, rMSSD, pNN50, LF, and HF relative to participants who did return. The significant differences between return and

non-return participants on the above-mentioned measures suggest data were not missing at random (i.e., there may have been a systematic relationship between missing and observed data; Little, 1988; McKnight et al., 2007; Scheffer, 2002), and the reduction in HRV stability among imputed data may reflect this increased health risk. However, despite this missing data, analyses were based on a large sample of participants, allowing for sufficient statistical power.

Fourth, only two assessments of HRV (and covariates) were obtained in this study. Although this typically allows for an adequate examination of stability, having only two measures of the included variables limits an adequate examination of growth over time (i.e., growth trajectory curves), and how such growth may impact HRV stability. Contrary to initial expectations, HRV stability was higher when measured using *ICC*s compared to Pearson correlations despite mean-level increases in HRV between the first and second study visit (Shrout & Fleiss, 1979). It is possible that mean-level differences (i.e., increases) in HRV among QUALITY participants across 2 years were too small to produce lower *ICC*s relative to Pearson correlations. Thus, using *ICC*s as a conservative measure of stability may be more meaningful when examining data collected over longer test-retest intervals to allow for more change in the measure to occur. Such an examination may require more advanced statistical techniques relative to the partial correlations employed in this study. The moderator variables examined in the present study were chosen based on findings from previous HRV research with adults and children. Most of the available research indicated that the selected moderators are associated with mean-level differences in HRV. However, the present study demonstrated that differences among moderator variables measured at the first clinic visit, and

developmental change in these measures across 2 years, did not alter HRV stability. Considering this, exploratory linear regression models were tested to examine the role of the covariates on HRV at the second clinic visit. Regression models indicated that several moderator variables measured at the first clinic visit had significant positive associations with HRV measures obtained at the second clinic visit; the magnitude of each association was generally weak, with the only exception being the association between heart rate and NN, which was moderate. In most cases, measures of growth and adiposity from the first clinic visit were positively associated with HRV, suggesting that greater initial mass was a significant predictor of greater HRV at the study follow-up. However, regression models with time-varying covariates (i.e., residualized change scores and the associated covariate from the first clinic visit) demonstrated the opposite effect. Namely, most significant time-varying predictors were negatively associated with HRV at the second clinic visit; the magnitude of each association was weak to moderate, with time-varying measures of age and heart rate typically revealing the largest associations. Among time-varying regression models, the strongest associations were between change in age and pNN50, and change in heart rate and NN, suggesting that greater change (i.e., increase) in age, and change (i.e., decrease) in heart rate across 2 years was associated with greater reductions in pNN50 and NN at the second clinic visit, respectively. Another notable finding across regression models was that time-varying change (i.e., decrease) in regular physical activity across 2 years was significantly associated with HRV at the second clinic visit (whereas no associations were found among regression models with covariates measured at the first clinic visit, or among examinations of HRV stability). These results suggested that change (relative to initial differences) in the moderator variables

accounted for more variance in HRV measures obtained at the second clinic visit. Further, unstandardized coefficients revealed that time-varying change in the covariates (particularly heart rate, diastolic blood pressure, and weekly physical activity) had the strongest negative associations with VLF and LF. This suggests that change in these moderators across 2 years may have had a particularly salient effect on lower-frequency sources of autonomic cardiovascular control. Considering these findings in the larger context of this study, although the present results indicate that HRV stability was robust to many of the proposed moderator variables, these same moderators were significant predictors of HRV measured at the second clinic visit, particularly when the change in the moderator across 2 years was included in the model. Although it was beyond the scope of this thesis, taken together, regression analysis results suggested that more comprehensive statistical techniques may be required to better elucidate factors that influence stability or change in HRV measures across time. As such, it is plausible that a more comprehensive analytical technique could better examine relations between developmentally-relevant covariates and stability of HRV measures. Namely, generalizability theory, which can parse measurement variance into different components, or facets, may better explain key sources of variation in examinations of stability. By applying generalizability theory to the study of stability, a more statistically comprehensive examination can be performed. Similar to the proposed advantage of *ICC*s relative to Pearson correlations, "partial *ICC*s" can account for mean-level and rank-order change in HRV data across time, while partialling out additional variance attributed to a particular covariate. However, given that an analytical strategy such as this was beyond the scope of the present thesis, future studies of HRV stability should

consider applying generalizability theory to elucidate factors (or facets) that may be pertinent to a better understanding of moderators that influence HRV stability.

Finally, some HRV measures used in the present study were not derived in accordance with recommendations outlined by established HRV guidelines (Berntson et al., 1997; Task Force, 1996). Namely, the sampling rate used during data collection (128 Hz) may have influenced resulting HRV measures and the observed results in turn. The QUALITY study also did not apply a moving polynomial filter during the data reduction phase; not using this filter may have further impacted the fidelity of resulting HRV values. Further, the focus on time- and frequency-domain HRV measures in this study precluded an examination of stability among other HRV measures recommended by published guidelines, such as normalized frequency-domain indices, and measures derived from geometric (e.g., triangular index), Poincare plot (e.g., SD1, SD2), wavelet-based (e.g., wavelet coefficient), or entropy-based (e.g., approximate entropy) analyses; these measures have also been associated with health outcomes (e.g., Kleiger et al., 2005), but are less commonly reported among the pediatric HRV literature. However, (as previously noted) ECG recordings were of sufficient duration to examine the derived HRV measures, the methodological protocol was highly standardized for all study participants, validated hardware and software platforms were used, and ECG data were carefully screened and edited using established procedures. Thus, the time- and frequency-domain measures used in this study were obtained under largely ideal methodological conditions, and the specific variables examined herein are consistent with those most often reported among the pediatric HRV literature.

Conclusion

Taken together, the present results demonstrated that autonomic cardiovascular control, as measured by time- and frequency-domain HRV, is moderately stable among children. Overall, the present study demonstrated three key findings that appear pertinent to pediatric HRV research. First, time- and frequency-domain measures reflect a stable individual difference in childhood. Second, the moderate stability of HRV measures in children is robust to developmental changes and growth related to time-varying demographic, biological, and behavioural covariates. Third, results further demonstrated that HRV stability is influenced by variance attributed to changes in heart rate across time; this was indicated by reduced stability after controlling for heart rate measured at the first clinic visit, and augmented stability of HRV measures after controlling for change in heart rate across 2 years, particularly among measures reflecting parasympathetic modulation of heart rate (rMSSD, pNN50, HF).

General Discussion

HRV is an established measure of autonomic cardiovascular control in adults and children. While the psychometric properties of HRV have been extensively examined among adult samples, they are less well established among pediatric samples. Establishing test-retest reliability and temporal stability of HRV is an important requisite in determining whether the association between HRV and health is robust. If HRV cannot be reliably measured or does not reflect a stable individual difference, erroneous conclusions may be drawn about relations between autonomic cardiovascular control and long-term health outcomes. Among the extant pediatric literature there is a paucity of evidence for the reliability and stability of HRV measures among children. Existing studies are limited by small sample sizes, inconsistent methodological protocols, few methodological conditions considered, or scarce statistical controlling for covariates. Considering these gaps in the literature, the overarching goal of this thesis was to conduct a comprehensive examination of test-retest reliability and temporal stability of time- and frequency-domain HRV measures among pediatric samples. This goal was achieved using two complementary studies.

The first study was a comprehensive meta-analytic review of HRV test-retest reliability among infants, children, and adolescents. Results indicated that HRV exhibited moderate test-retest reliability overall, while reliability among infants and toddlers was lower compared to reliability among children and adolescents. Moreover, there was less methodological consistency across infant/toddler studies relative to child/adolescent studies. Results indicated several methodological factors were pertinent to HRV reliability across both age groups. Namely, greater reliability of NN and HF among infant/toddler studies was associated with participant exclusion criteria based on

prescription medications, medical illnesses, or low birth weights; inclusion of a pre-recording acclimatization period; recording ECG while participants were awake and seated; and, deriving HRV using the mean of several short recording epochs. Greater reliability of HF and LF among child/adolescent studies was associated with a lower proportion of male participants; longer ECG recording duration; inclusion of a pre-recording acclimatization period; and, selection of HRV frequency bandwidths consistent with published guidelines and recommendations. Notedly, several moderator analyses were restricted as several studies across both age groups did not specify target methodological factors. Many of these methodological factors pertain to technical details that are relevant to data replication and interpretation, as outlined by established HRV guidelines. In other words, several factors that may be associated with HRV reliability (e.g., time of ECG recording, ECG sampling rate) could not be examined. Other discussed limitations of this meta-analysis included reduction of some categorical moderators to dichotomized variables, focus on short-term daytime HRV measures, and use of fixed- vs. random-effects models. Despite these shortcomings, the meta-analytic findings have important implications for the methodological design of future HRV studies with infants and youth.

The second study was an examination of temporal stability of time- and frequency-domain HRV measures among children aged 9- to 11-years-old. Study results indicated that HRV exhibits moderate stability overall, and that stability was robust to developmental changes among measures of age, blood pressure, pubertal status, growth and adiposity, and self-reported physical activity across 2 years. Heart rate demonstrated notable influences on HRV stability across all variables except LF:HF ratio; controlling

for heart rate measured at the first clinic visit reduced HRV stability, while controlling for change in heart rate (i.e., decrease) augmented stability of rMSSD, pNN50, and HF, all of which reflect parasympathetic contributions to HRV. Taken together, this study demonstrated that HRV stability is robust to intra-individual differences across 2-year development of several physiological processes among children. To the best of our knowledge, this study was the first comprehensive examination of the relation between HRV stability and developmentally-relevant, time-varying covariates among children.

This thesis provided strong evidence that time- and frequency-domain HRV measures can be reliably measured and are a stable individual difference among children. Reproducibility of pediatric HRV measures appears more sensitive to methodological differences during data collection and analysis across study assessments relative to intra-individual differences within a sample across time. Standardizing the methodological protocol across study assessments, allowing participants to acclimate to experimental conditions prior to collection of baseline ECG data, and obtaining an ECG recording of adequate duration for the variable of interest are all methodological design issues which can improve test-retest reliability of HRV. Including a pre-recording acclimatization period and recording ECG in a seated posture were each substantial moderators of HRV reliability across both age groups. Attention to these methodological factors can minimize extraneous variance in HRV measures that contribute to poorer reproducibility.

One consistent finding across both studies was the role of heart rate in moderating HRV psychometrics. The psychometric quality of pediatric HRV measures declined if extraneous variance attributed to the influence of heart rate was not accounted for across test-retest intervals. The meta-analysis demonstrated that HF reliability was reduced

among child/adolescent studies, but not infant/toddler studies, that had longer study follow-up length, did not include a pre-recording acclimatization period, and did not apply software-based filtering. The examination of stability indicated that HRV was more stable after controlling for change in heart rate across the 2-year study follow-up, particularly among measures reflecting parasympathetic modulation. Results of both studies indicate that the influence of heart rate contributes notable variance to longitudinal measures of HRV in children, and suggest that changes in heart rate activity over time is likely a key contributor to mean-level differences in HRV across test-retest intervals. Multiple factors contribute to heart rate, and would in turn affect HRV and its longitudinal psychometric properties. It has been demonstrated that a moving polynomial filter (Porges, 1985) extracts heart rate variance within a specified frequency band from slower periodic processes and background physiology, and can eliminate the need to include heart rate as a covariate. Application of a moving polynomial filter could help account for natural developmental changes and sudden aberrations in background heart rate activity better than alternative methods that do not account for heart rate changes. As discussed, young children may be especially sensitive to unfamiliar experimental conditions and stress-related increases in heart rate and respiration, which could distort estimates of HRV and limit the measure's reproducibility. The lack of association between a pre-recording acclimatization period and HF reliability among infant/toddler studies in the meta-analysis might be explained by the more common use of a moving polynomial filter relative to child/adolescent studies. As such, although reliability was lower among infant/toddler, relative to child/adolescent studies, differences in heart rate activity among infants and toddlers across study assessments may have been accounted

for by the polynomial filter; this would explain why HF reliability among infant/toddler studies was not moderated by study follow-up length or inclusion of a pre-recording acclimatization period as it was among child/adolescent studies. Moreover, this suggests that use of a moving polynomial filter may be a methodological factor that is pertinent to improving pediatric HRV psychometrics. Considering this, although the QUALITY ECG data were linearly detrended, mean-centered, and tapered for spectral leakage using a Hanning window, a moving polynomial filter was not applied; thus, it remains to be determined whether including heart rate as a covariate would influence HRV stability in the same way had a moving polynomial filter been implemented. Taken together, results suggest that minimizing extraneous variance associated with intra-individual differences in heart rate activity, and the developmental change in heart rate across time, can produce more reliable and stable measures of HRV.

Published guidelines (Berntson et al., 1997; Task Force, 1996) offer recommendations for improved HRV measurement quality and inter-study consistency, such as standardized data collection, analysis, and reporting practices. Current guidelines for HRV methodology, which are almost 2 decades old, are principally focused on adult HRV assessment. This thesis provides empirical support regarding the application of several guidelines recommended among adult HRV studies (e.g., inclusion of a pre-recording acclimatization period, analyzing adequately-long and identical ECG recording durations, careful editing and filtering of NN-interval data, frequency bandwidth selection), which were pertinent for HRV reliability among children. This thesis also provides empirical support for how statistically controlling developmental changes in heart rate can augment estimates of HRV stability, particularly among measures

reflecting parasympathetic modulation. Moreover, exploratory regression analyses indicated that change over time across multiple developmentally-relevant moderators other than heart rate (pubertal development, blood pressure, growth and adiposity, physical activity) were significant predictors of HRV measured at the second QUALITY visit. As such, results of this thesis can help to inform future research into factors that are pertinent to HRV measurement among children. For instance, with the exception of studies with very young children (i.e., infants and toddlers), pediatric HRV measures in the frequency-domain are almost consistently derived using frequency bandwidths noted in adult measurement guidelines. An explicit examination of whether adult frequency bandwidths are actually appropriate for use among children and adolescents does not appear available. As such, an examination of the upper and lower limits of HF, LF, and VLF frequency bandwidths in children across various ages may help determine the point at which researchers should consider adjusting HRV frequency bandwidths to account for autonomic development throughout childhood. Relatedly, across the developmental period, children's heart rate decreases, blood pressure increases, they naturally become taller and heavier, and they experience increasing hormonal, behavioural, and psychological changes with more advanced pubertal development; these changes have each been associated with shifts towards sympathetic dominance and should be considered when examining pediatric HRV. However, while age, heart rate, blood pressure, mass, and physical fitness appear pertinent to HRV studies with participants of any age, factors associated with pubertal development are a consideration that are unique to pediatric samples. Alterations to HRV associated with pubertal development have not been frequently examined (c.f., the number of studies in the meta-analysis that included

adolescent participants). Considering the results of this thesis, an extension of current HRV guidelines to include greater consideration of factors associated with HRV specific to children may be warranted.

As previously described, the meta-analysis was restricted due to minimal reporting of target methodological factors intended for use in moderator analyses. Despite this drawback, moderator analyses indicated that reliability was generally robust across examined study and recording conditions. Several of these moderators pertain to important methodological considerations outlined by published guidelines for the interpretation and replication of HRV data. It is paramount that pediatric researchers adequately report pertinent methodological details, including participant characteristics (e.g., anthropometric data, sample ethnicity), ECG signal acquisition and pre-processing techniques (e.g., ECG sampling rate, editing and filtering procedures), and HRV analysis procedures (e.g., data reduction, presentation of absolute and normalized HRV units). The minimal methodological reporting practices among pediatric HRV studies contrasts with the more rigorous reporting practices often observed among adult HRV studies. More precise methodological reporting may help improve standardization across HRV studies, increase the interpretability and replication of findings, and allow for more comprehensive meta-analyses to be conducted in the future.

General Limitations

There are two common limitations across both studies in the present thesis that merit consideration. First, this thesis focused on short-term (i.e., < 5-hrs) daytime measures of time- and frequency-domain HRV. While these measures are frequently reported in the adult and pediatric HRV literature, other HRV measures derived from

longer recordings or alternative analytical techniques have also demonstrated clinically utility. Thus, as previously indicated, efforts to replicate these analyses using long-term HRV measures (i.e. ≥ 24-hr, 12-hr overnight) or measures derived using alternative analytical techniques (e.g., non-linear, geometrical) is warranted to determine the generalizability of these findings to alternative measures of HRV. Relatedly, HRV data from most studies included in the meta-analysis as well as data collected from the QUALITY cohort were obtained in a laboratory/clinic setting. While this methodology allows for increased standardization of ECG recordings, it potentially limits the ecological validity of these findings. Future examinations using ambulatory ECG data could help address this concern. However, short-term ECG measures obtained in laboratory or clinical settings are often more common among the HRV literature, due in part to the ease and reduced cost of collecting and analyzing shorter, relative to longer ambulatory HRV recordings. As such, considering the methodology commonly applied in contemporary research, the results of this thesis are largely applicable to many future HRV studies.

Second, this thesis focused on HRV measures that were obtained during study conditions not associated with experimental stressors or manipulations. Although HRV measures obtained during passive (relative to task or challenge) conditions are prominently reported among adult and pediatric HRV studies, HRV reactivity is an important construct that was not examined here. Measures of reactivity are important for determining how the autonomic nervous system responds to cognitive (e.g., mental workload, attentional demands), emotional (e.g., negative affect, social stress), and physical (e.g., tilting, exercise) demands. As such, efforts to replicate these psychometric

analyses using data reflecting HRV reactivity, and more specifically, reactivity to different stressors, are warranted. This would help determine if the present findings generalize beyond the context of HRV measures obtained during states of passivity, and extend to measures of HRV reactivity.

Conclusion

Overall, the meta-analysis demonstrated that among infant/toddler studies, recording exclusively awake ECG data, careful editing and filtering of ECG data, and analyzing shorter ECG recording duration epochs were each associated with higher reliability; however, sample exclusions based on prescription medication use or the presence of a chronic illness were associated with higher NN, but lower HF reliability. Among child/adolescent studies, moderator variables pertaining to sample exclusions based on prescription medication use or the presence of a chronic illness, and to the explicit reporting of manual editing or digital filtering of ECG data prior to HRV analysis, were not associated with HRV reliability; however, higher reliability was yielded among studies that obtained longer (vs. shorter) ECG recording durations and used frequency-bandwidths consistent with published HRV guidelines and recommendations. Recording ECG in a seated posture and allowing participants an opportunity to acclimate to experimental conditions prior to data collection were both associated with higher reliability across both age groups. Analyses using data collected from the QUALITY cohort indicated that stability of time- and frequency-domain HRV was robust to developmental changes in measures of blood pressure, pubertal status, growth and adiposity, and physical activity, all variables which have been previously associated with pediatric HRV in cross-sectional and longitudinal studies. Further, stability of measures reflecting parasympathetic activity increased after controlling for

change in heart rate. Altogether, this research indicated that time- and frequency-domain HRV are moderately reliable measures of autonomic cardiovascular control, and reflect a moderately stable individual difference among children. In conclusion, this thesis provides comprehensive empirical evidence for the psychometric properties of pediatric HRV measures, and demonstrates that time- and frequency-domain HRV exhibit adequate and generally robust reproducibility over time. Evidence for the psychometric quality of pediatric HRV measures is particularly relevant for future studies that examine prospective relations between HRV and health outcomes across development.

References

Acharya U, R., Kannathal, N., Sing, O. W., Ping, L. Y., & Chua, T. (2004). Heart rate analysis in normal subjects of various age groups. *Biomedical Engineering Online*, *3*(1), 24. doi:10.1186/1475-925X-3-24

Adair, L., Martorell, R., & Stein, A. (2009). Size at birth, weight gain in infancy and childhood, and adult blood pressure in 5 low-and middle-income-country cohorts: when does weight gain matter? *The American Journal of Clinical Nutrition*, *89*, 1383–1392. doi:10.3945/ajcn.2008.27139

Akinci, A., Celiker, A., Baykal, E., & Tezic, T. (1993). Heart rate variability in diabetic children: Sensitivity of the time- and frequency-domain methods. *Pediatric Cardiology*, *14*(3), 140–146.

Akselrod, S., Eliash, S., Oz, O., & Cohen, S. (1985). Hemodynamic regulation in SHR: investigation by spectral analysis. *The American Journal of Physiology*, *253*(1 Pt 2), H176–83.

Akselrod, S., Gordon, D., Ubel, F., Shannon, D., Barger, A., & Cohen, R. (1981). Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science*, *213*(4504), 220–222.

*Alkon, A., Boyce, W. T., Davis, N. B., & Eskanazi, B. (2011). Developmental changes in autonomic nervous system resting and reactivity measures in Latino children from 6 to 60 months of age. *Journal of Developmental and Behavioural Pediatrics*, *32*(9), 668–677.

*Alkon, A., Lippert, S., Vujan, N., & Boyce, W. T. (2006). The ontogeny of autonomic measures in 6- and 12-month-old infants. *Developmental Psychobiology*, *48*, 197–208. doi:10.1002/dev

Antelmi, I., de Paula, R. S., Shinzato, A. R., Peres, C. A., Mansur, A. J., & Grupi, C. J. (2004). 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*, *93*(3), 381–385. doi:10.1016/j.amjcard.2003.09.065

*Arendt, R. E., Halpern, L. F., MacLean, W. E., & Youngquist, G. a. (1991). The properties of V in newborns across repeated measures. *Developmental Psychobiology*, *24*(2), 91–101. doi:10.1002/dev.420240203

Armstrong, R. G., Kenny, G. P., Green, G., & Seely, A. J. E. (2011). Diurnal variation in heart rate variability before and after maximal exercise testing. *Chronobiology International*, *28*(4), 344–351. doi:10.3109/07420528.2011.559674

Baharav, A., Kotagal, S., Rubin, B. K., Pratt, J., & Akselrod, S. (1999). Autonomic cardiovascular control in children with obstructive sleep apnea. *Clinical Autonomic Research*, *9*(6), 345–351.

*Bar-Haim, Y., Marshall, P. J., & Fox, N. a. (2000). Developmental changes in heart period and high-frequency heart period variability from 4 months to 4 years of age. *Developmental Psychobiology*, *37*(1), 44–56.

Batten, L. A, Urbina, E. M., & Berenson, G. S. (2000). Interobserver reproducibility of heart rate variability in children (the Bogalusa Heart Study). *The American Journal of Cardiology*, *86*(11), 1264–1266, A9.

Berntson, G. G., & Stowell, J. R. (1998). ECG artifacts and heart period variability: Don't miss a beat! *Psychophysiology*, *35*(1), 127–132.

Berntson, G. G., Bigger, Jr., J. R., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., … Van Der Molen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*, 623–648.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: the modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*, *98*(4), 459–487.

Berntson, G. G., Quigley, K. S., & Lozano, D., in Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). Cardiovascular Psychophysiology. In Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.) *The Handbook of Psychophysiology* (3rd ed.). Cambridge University Press, New York, NY.

Berntson, G. G., & Quigley, K. (1990). An approach to artifact identification: Application to heart period data. *Psychophysiology*, *27*(5), 586–598.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. (1993). Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology*, *30*, 183–196.

Bigger Jr, J. T., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E., & Rottman, J. N. (1992). Correlations among time and frequency domain measures of heart period variability two weeks after acute myocardial infarction. *The American Journal of Cardiology*, *69*(9), 891–898.

Bigger, J. T., Fleiss, J. L., Rolnitzky, L. M., & Steinman, R. C. (1993). The ability of several short-term measures of RR variability to predict mortality after myocardial infarction. *Circulation*, *88*(3), 927–934. doi:10.1161/01.CIR.88.3.927

*Blandon, A. Y., Calkins, S. D., Keane, S. P., & O'Brien, M. (2010). Contributions of child's physiology and maternal behavior to children's trajectories of temperamental reactivity. *Developmental Psychology*, *46*(5), 1089–1102. doi:10.1037/a0020678.Contributions

Boer-Martins, L., Figueiredo, V. N., Demacq, C., Martins, L. C., Consolin-Colombo, F., Figueiredo, M. J., … Moreno, H. (2011). Relationship of autonomic imbalance and circadian disruption with obesity and type 2 diabetes in resistant hypertensive patients. *Cardiovascular Diabetology*, *10*(24). doi:10.1186/1475-2840-10-24

Bonnet, M. H., & Arand, D. L. (1998). Heart rate variability in insomniacs and matched normal sleepers. *Psychosomatic Medicine*, *60*(5), 610–615.

Booth, M. (2000). Assessment of physical activity: An international perspective. *Research Quarterly for Exercise and Sport,* 71(2 Suppl), S114-20.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PloS One*, *5*(12), e14331. doi:10.1371/journal.pone.0014331

*Bornstein, M. H., & Suess, P. E. (2000). Child and mother cardiac vagal tone: Continuity, stability, and concordance across the first 5 years. *Developmental Psychology*, *36*(1), 54–65. doi:10.1037//0012-1649.36.1.54

*Bornstein, M. H., & Suess, P. E. (2000). Physiological self-regulation and information processing in infancy: Cardiac vagal tone and habituation. *Child Development*, *71*(2), 273–287.

Buchhorn, R., Conzelmann, A., Willaschek, C., Störk, D., Taurines, R., & Renner, T. J. (2012). Heart rate variability and methylphenidate in children with ADHD. *Attention Deficit and Hyperactivity Disorders*, *4*(2), 85–91. doi:10.1007/s12402-012-0072-8

Burgess, H., & Trinder, J. (1997). Sleep and circadian influences on cardiac autonomic nervous system activity. *American Journal of Physiology*, *273*(42), H1761–H1768.

Burr, R. L. (2007). Interpretation of normalized spectral heart rate variability indices in sleep research: A critical review. *Sleep*, *30*(7), 913–919.

Bylsma, L. M., Salomon, K., Taylor-Clift, A., Morris, B. H., & Rottenberg, J. (2014). Respiratory sinus arrhythmia reactivity in current and remitted major depressive disorder. *Psychosomatic Medicine*, *76*(1), 66–73. doi:10.1097/PSY.0000000000000019

Cacioppo J. T., Berntson, G. G., Binkley, P. f., Quigley, K. S., Uchino, B. N., & Fieldstone, A. (1994). Autonomic cardiac control. II. Noninvasive indices and basal response as revealed by autonomic blockades. *Psychophysiology*, *31*, 586–598.

*Calkins, S. D., & Keane, S. P. (2004). Cardiac vagal regulation across the preschool period: Stability, continuity, and implications for childhood adjustment. *Developmental Psychobiology*, *45*(3), 101–112. doi:10.1002/dev.20020

Campbell, T. S., Ditto, B., Séguin, J. R., Sinray, S., & Tremblay, R. E. (2003). Adolescent pain sensitivity is associated with cardiac autonomic function and blood

pressure over 8 years. *Hypertension*, *41*(6), 1228–33. doi:10.1161/01.HYP.0000072802.84202.86

*Cavaliere, C. M. (2011). Role of the vagal system in the regulatory sensory processing patterns in children with autistic spectrum disorder: Physiologic underpinnings and reliability of measurement (Doctoral dissertation). Retrieved from www.proquest.com

Cerutti, S., Goldberger, a. L., & Yamamoto, Y. (2006). Recent advances in heart rate variability signal processing and interpretation. *IEEE Transactions on Biomedical Engineering*, *53*(1), 1–3. doi:10.1109/TBME.2005.859777

Chang, J. S., Yoo, C. S., Yi, S. H., Hong, K. H., Oh, H. S., Hwang, J. Y., … Kim, Y. S. (2009). Differential pattern of heart rate variability in patients with schizophrenia. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *33*(6), 991–995. doi:10.1016/j.pnpbp.2009.05.004

Chen, J.-Y., Lee, Y. L., Tsai, W.-C., Lee, C.-H., Chen, P.-S., Li, Y.-H., … Lin, L.-J. (2011). Cardiac autonomic functions derived from short-term heart rate variability recordings associated with heart rate recovery after treadmill exercise test in young individuals. *Heart and Vessels*, *26*(3), 282–288. doi:10.1007/s00380-010-0048-6

Chen, S.-R., Chiu, H.-W., Lee, Y.-J., Sheen, T.-C., & Jeng, C. (2012). Impact of pubertal development and physical activity on heart rate variability in overweight and obese children in Taiwan. *The Journal of School Nursing*, *28*(4), 284–290. doi:10.1177/1059840511435248

Chen, X., & Mukkamala, R. (2008). Selective quantification of the cardiac sympathetic and parasympathetic nervous systems by multisignal analysis of cardiorespiratory

variability. *American Journal of Physiology*, *294*, 362–371. doi:10.1152/ajpheart.01061.2007.

Chen, X., & Wang, Y. (2008). Tracking of blood pressure from childhood to adulthood: A systematic review and meta-regression analysis. *Circulation*, *117*(25), 3171–3180. doi:10.1161/CIRCULATIONAHA.107.730366

Chiolero, A., Paradis, G., & Lambert, M. (2010). Accuracy of oscillometric devices in children and adults. *Blood Pressure*, *19*(4), 254–259. doi:10.3109/08037051003606439

Cipryan, L., & Litschmannova, M. (2013). Intra-day and inter-day reliability of heart rate variability measurement. *Journal of Sports Sciences*, *31*(2), 150–158.

Cohen, R. J., & Swerdlik, M. E. (2002). *Psychological testing and assessment: An introduction to tests and measurement* (5th ed.). New York: McGraw-Hill.

Coleman, L., & Coleman, J. (2002). The measurement of puberty: A review. *Journal of Adolescence*, *25*(5), 535–550. doi:10.1006/jado.2002.0494

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., … Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*, *35*(8), 1381–1395. doi:10.1249/01.MSS.0000078924.61453.FB

Dalla Pozza, R., Fuchs, a, Bechtold, S., Kozlik-Feldmann, R., Daebritz, S., & Netz, H. (2006). Short-term testing of heart rate variability in heart-transplanted children: Equal to 24-h ECG recordings? *Clinical Transplantation*, *20*(4), 438–442. doi:10.1111/j.1399-0012.2006.00502.x

De Onis, M., & Organization, W. H. (2006). WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development.

Denver, J., Reed, S., & Porges, S. (2007). Methodological issues in the quantification of respiratory sinus arrhythmia. *Biological Psychology*, *74*(2), 286–294.

*Dietrich, A., Rosmalen, J. G. M., Althaus, M., van Roon, A. M., Mulder, L. J. M., Minderaa, R. B., … Riese, H. (2010). Reproducibility of heart rate variability and baroreflex sensitivity measurements in children. *Biological Psychology*, *85*(1), 71–78. doi:10.1016/j.biopsycho.2010.05.005

*Doussard-Roosevelt, J. a, Montgomery, L. A., & Porges, S. W. (2003). Short-term stability of physiological measures in kindergarten children: Respiratory sinus arrhythmia, heart period, and cortisol. *Developmental Psychobiology*, *43*(3), 230–242. doi:10.1002/dev.10136

Eckberg, D. L. (2009). Point:counterpoint: Respiratory sinus arrhythmia is due to a central mechanism vs. respiratory sinus arrhythmia is due to the baroreflex mechanism. *Journal of Applied Physiology, 106*(5), 1740–1742; discussion 1744. doi:10.1152/japplphysiol.91107.2008

Elghozi, J.-L., & Julien, C. (2007). Sympathetic control of short-term heart rate variability and its pharmacological modulation. *Fundamental & Clinical Pharmacology*, *21*(4), 337–347. doi:10.1111/j.1472-8206.2007.00502.x

*Elmore-staton, L. D. (2011). Sleep and cognitive functioning: The moderating role of vagal indices (Doctoral dissertation). Retrieved from www.proquest.com.

*El-Sheikh, M. (2005). Stability of respiratory sinus arrhythmia in children and young adolescents: A longitudinal examination. *Developmental Psychobiology*, *46*(1), 66–74. doi:10.1002/dev.20036

El-Sheikh, M., & Hinnant, J. B. (2011). Marital conflict, respiratory sinus arrhythmia, and allostatic load: Interrelations and associations with the development of children's externalizing behavior. *Development and Psychopathology*, *23*(3), 815–829. doi:10.1017/S0954579411000320

El-Sheikh, M., & Whitson, S. A. (2006). Longitudinal relations between marital conflict and child adjustment: Vagal regulation as a protective factor. *Journal of Family Psychology*, *20*(1), 30–39. doi:10.1037/0893-3200.20.1.30

El-sheikh, M., Hinnant, J. B., & Erath, S. (2012). Developmental trajectories of delinquency symptoms in childhood: The role of marital conflict and autonomic nervous system activity. *Journal of Abnormal Psychology*, *120*(1), 16–32. doi:10.1037/a0020626.Developmental

Ernst, G. (2014). *Heart Rate Variability*. Springer-Verlag, London, UK.

Faulkner, M. S., Hathaway, D., & Tolley, B. (2003). Cardiovascular autonomic function in healthy adolescents. *Heart & Lung: The Journal of Critical Care*, *32*(1), 10–22. doi:10.1067/mhl.2003.6

Field, A. P. (2009). *Discovering Statistics Using SPSS: And Sex and Drugs and Rock "n" Roll* (3rd ed.). London: Sage publications.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *The British Journal of Mathematical and Statistical Psychology*, *63*(Pt 3), 665–694. doi:10.1348/000711010X502733

Finley, J. P., Nugent, S. T., & Hellenbrand, W. (1987). Heart-rate variability in children. Spectral analysis of developmental changes between 5 and 24 years. *Canadian Journal of Physiology and Pharmacology*, *65*(10), 2048–2052.

Fleming, S., Thompson, M., Stevens, R., Heneghan, C., Pluddemann, A., Maconochie, I., … Mant, D. (2011). Normal ranges of heart rate and respiratory rate in children from birth to 18 years: A systematic review of observational studies. *Lancet*, *377*(9770), 1011–1018. doi:10.1016/S0140-6736(10)62226-X.Normal

*Fox, N. A. (1983). Maturation of autonomic control in preterm infants. *Developmental Psychobiology*, *16*(6), 495–504. doi:10.1002/dev.420160605

*Fox, N. A., & Field, T. M. (1989). Individual differences in preschool entry behavior. *Journal of Applied Developmental Psychology*, *10*(4), 527–540. doi:10.1016/0193-3973(89)90025-7

Fox, N. A., & Porges, S. W. (1985). The relation between neonatal heart period patterns and developmental outcome. *Child Development*, *56*(1), 28–37.

*Fracasso, M. P., Porges, S. W., Lamb, M. E., & Rosenberg, A. A. (1994). Cardiac activity in infancy: Reliability and stability of individual differences. *Infant Behavior and Development*, *284*(17), 277–284.

*Galland, B. C., Hayman, R. M., Taylor, B. J., Bolton, D. P., Sayers, R. M., & Williams, S. M. (2000). Factors affecting heart rate variability and heart rate responses to tilting in infants aged 1 and 3 months. *Pediatric Research*, *48*(3), 360–368. doi:10.1203/00006450-200009000-00017

*Galland, B. C., Taylor, B. J., Bolton, D. P. G., & Sayers, R. M. (2006). Heart rate variability and cardiac reflexes in small for gestational age infants. *Journal of*

Applied Physiology (Bethesda, Md. : 1985), *100*(3), 933–939. doi:10.1152/japplphysiol.01275.2005

*Gamelin, F.-X., Baquet, G., Berthoin, S., Thevenet, D., Nourry, C., Nottin, S., & Bosquet, L. (2009). Effect of high intensity intermittent training on heart rate variability in prepubescent children. *European Journal of Applied Physiology*, *105*(5), 731–738. doi:10.1007/s00421-008-0955-8

García-González, M. A, Fernández-Chimeno, M., & Ramos-Castro, J. (2004). Bias and uncertainty in heart rate variability spectral indices due to the finite ECG sampling frequency. *Physiological Measurement*, *25*(2), 489–504. doi:10.1088/0967-3334/25/2/008

*Gentzler, A. L., Rottenberg, J., Kovacs, M., George, C. J., & Morey, J. N. (2012). Atypical development of resting respiratory sinus arrhythmia in children at high risk for depression. *Developmental Psychobiology*, *54*(5), 556–567. doi:10.1002/dev.20614

Goto, M., Nagashima, M., Baba, R., Nagano, Y., Yokota, M., Nishibata, K., & Tsuji, A. (1997). Analysis of heart rate variability demonstrates effects of development on vagal modulation of heart rate in healthy children. *The Journal of Pediatrics*, *130*(5), 725–729.

Graziano, P. A, Calkins, S. D., Keane, S. P., & O'Brien, M. (2011). Cardiovascular regulation profile predicts developmental trajectory of BMI and pediatric obesity. *Obesity*, *19*(9), 1818–1825. doi:10.1038/oby.2011.98

Gregoire, J., Tuck, S., Yamamoto, Y., & Hughson, R. L. (1996). Heart rate variability at rest and exercise: Influence of age, gender, and physical training. *Canadian Journal of Applied Physiology*, *21*(6), 455–470.

Gui-Ling, X., Jing-Hua, W., Yan, Z., Hui, X., Jing-Hui, S., & Si-Rui, Y. (2013). Association of high blood pressure with heart rate variability in children. *Iranian Journal of Pediatrics*, *23*(1), 37–44.

Gunnar, M., & Wewerka, S. (2009). Developmental changes in hypothalamus–pituitary–adrenal activity over the transition to adolescence: normative changes and associations with puberty. *Development and Psychopathology*, *21*(1), 69–85. doi:10.1017/S0954579409000054.Developmental

Guo, Y., & Stein, P. (2003). Circadian rhythm in the cardiovascular system: Chronocardiology. *American Heart Journal*, *145*, 779–786. doi:10.1016/S0002-8703(02)00088-1

Gutin, B., Barbeau, P., Litaker, M. S., Ferguson, M., & Owens, S. (2000). Heart rate variability in obese children: Relations to total body and visceral adiposity, and changes with physical training and detraining. *Obesity Research*, *8*(1), 12–9. doi:10.1038/oby.2000.3

Guyton, A. C. & Hall, J. E. (2006). *Textbook of Medical Physiology*. Philadelphia, PA: Saunders, Elsevier.

Hardy, L. L., Booth, M. L., & Okely, A. D. (2007). The reliability of the Adolescent Sedentary Activity Questionnaire (ASAQ). *Preventive Medicine*, *45*(1), 71–74. doi:10.1016/j.ypmed.2007.03.014

*Hastings, P. D., Nuselovici, J. N., Utendale, W. T., Coutya, J., McShane, K. E., & Sullivan, C. (2008). Applying the polyvagal theory to children's emotion regulation: Social context, socialization, and adjustment. *Biological Psychology*, *79*(3), 299–306. doi:10.1016/j.biopsycho.2008.07.005

Heathers, J. A. J. (2014). Everything Hertz: Methodological issues in short-term frequency-domain HRV. *Frontiers in Physiology*, *5*(May), 177. doi:10.3389/fphys.2014.00177

Hejjel, L., & Roth, E. (2004). What is the adequate sampling interval of the ECG signal for heart rate variability analysis in the time domain? *Physiological Measurement*, *25*(6), 1405–1411. doi:10.1088/0967-3334/25/6/006

*Henje Blom, E., Olsson, E. M. G., Serlachius, E., Ericson, M., & Ingvar, M. (2009). Heart rate variability is related to self-reported physical activity in a healthy adolescent population. *European Journal of Applied Physiology*, *106*(6), 877–883. doi:10.1007/s00421-009-1089-3

Henry, B., Minassian, A., & Paulus, M. (2010). Heart rate variability in bipolar mania and schizophrenia. *Journal of Psychiatric Research*, *44*(3), 1–19. doi:10.1016/j.jpsychires.2009.07.011.Heart

Hermens, H. J., Freriks, B., Disselhorst-Klug, C., & Rau, G. (2000). Development of recommendations for SEMG sensors and sensor placement procedures. *Journal of Electromyography and Kinesiology*, *10*(5), 361–374.

*Hinnant, J. B., & El-Sheikh, M. (2009). Children's externalizing and internalizing symptoms over time: The role of individual differences in patterns of RSA

responding. *Journal of Abnormal Child Psychology*, *37*(8), 1049–1061. doi:10.1007/s10802-009-9341-1

*Hinnant, J. B., Elmore-Staton, L., & El-Sheikh, M. (2011). Developmental trajectories of respiratory sinus arrhythmia and preejection period in middle childhood. *Developmental Psychobiology*, *53*(1), 59–68. doi:10.1002/dev.20487

Hoppenbrouwers, T., Oliveira, F., Sandarupa, S., Khoo, M., Neuman, M., & Ramanathan, R. (2012). The development of the circadian heart rate rhythm (CHR) in Asian infants. *Early Human Development*, *88*(7), 555–561. doi:10.1016/j.earlhumdev.2011.12.031

*Hsu, H.-C., & Porter, C. L. (2004). Young infants' behavioral reactivity to mild perturbation: Developmental continuity, stability, and organization. *Infancy*, *6*(1), 95–120. doi:10.1207/s15327078in0601_5

Huikuri, H. V., Kessler, K. M., Terracall, E., Castellanos, A., Linnaluoto, M. K., & Myerburg, R. J. (1990). Reproducibility and circadian rhythm of heart rate variability in healthy subjects. *American Journal of Cardiology*, *65*, 391–393.

*Izard, C. E., Porges, S. W., Simons, R. F., Haynes, O. M., Hyde, C., Parisi, M., & Cohen, B. (1991). Infant cardiac activity: Developmental changes and relations with attachment. *Developmental Psychology*, *27*(3), 432–439.

Janig, W. (2006). *The Integrative Action of the Autonomic Nervous System: Neurobiology of Homeostasis*. Cambridge, UK: Camebridge University Press.

Janz, K. F., Dawson, J. D., & Mahoney, L. T. (2000). Predicting heart growth during puberty: The Muscatine Study. *Pediatrics*, *105*(5), e63–e63. doi:10.1542/peds.105.5.e63

Jarrin, D. C., McGrath, J. J., Giovanniello, S., Poirier, P., & Lambert, M. (2012). Measurement fidelity of heart rate variability signal processing: the devil is in the details. *International Journal of Psychophysiology*, *86*(1), 88–97. doi:10.1016/j.ijpsycho.2012.07.004

Jarrin, D. C., McGrath, J. J., Poirier, P., Seguin, L., Tremblay, R. E., Montplaisir, J. Y., Paradis, G. & Seguin, J. R., (In press). Short-term heart rate variability in a population-based sample of 10-year-old children. *Pediatric* Cardiology. doi:10.1007/s00246-014-0962-y

Jennings, J. R., Kamarck, T., Stewart, C., Eddy, M., & Johnson, P. (1992). Alternate cardiovascular baseline assessment techniques: Vanilla or resting baseline. *Psychophysiology*, *29*(6), 742–750.

Jezewski, J., Kupka, T., & Horoba, K. (2009). Extraction of fetal heart-rate signal as the time event series from evenly sampled data acquired using doppler ultrasound technique. *IEEE Transactions on Biomedical Engineering*, *55*(2), 805–810.

Johnson, L. R. (ed. . (2003). *Essential Medical Physiology* (3rd ed.) London, UK: Academic Press.

Julien, C. (2006). Cardiovascular variability is/is not an index of autonomic control of circulation. *Journal of Applied Physiology*, *101*(2), 684–688. doi:10.1152/japplphysiol.00562.2006

Kajantie, E., & Phillips, D. I. W. (2006). The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology*, *31*(2), 151–178. doi:10.1016/j.psyneuen.2005.07.002

Karim, N., Hasan, J., & Ali, S. (2011). Heart rate variability - A review. *Journals of Basic and Applied Sciences*, *7*(1), 71–77.

*Keller, P. S., & El-Sheikh, M. (2009). Salivary alpha-amylase as a longitudinal predictor of children's externalizing symptoms: Respiratory sinus arrhythmia as a moderator of effects. *Psychoneuroendocrinology*, *34*(5), 633–643. doi:10.1016/j.psyneuen.2008.12.016

*Kennedy, A. E., Rubin, K. H., Hastings, P. D., & Maisel, B. (2004). Longitudinal relations between child vagal tone and parenting behavior: 2 to 4 years. *Developmental Psychobiology*, *45*(1), 10–21. doi:10.1002/dev.20013

Kim, K. K., Kim, J. S., Lim, Y. G., & Park, K. S. (2009). The effect of missing RR-interval data on heart rate variability analysis in the frequency domain. *Physiological Measurement*, *30*(10), 1039–1050. doi:10.1088/0967-3334/30/10/005

Kim, K. K., Lim, Y. G., Kim, J. S., & Park, K. S. (2007). Effect of missing RR-interval data on heart rate variability analysis in the time domain. *Physiological Measurement*, *28*(12), 1485–1494. doi:10.1088/0967-3334/28/12/003

Kleiger, R. E., Miller, J. P., Bigger, J. T., & Moss, A J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American Journal of Cardiology*, *59*(4), 256–262.

Kleiger, R. E., Stein, P. K., & Bigger, J. T. (2005). Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrocardiology*, *10*(1), 88–101. doi:10.1111/j.1542-474X.2005.10101.x

Kline, R. B. (2009). *Becoming a behavioral science researcher: A guide to producing research that matters*. New York, NY: The Guilford Press.

Kochiadakis, G. E., Orfanakis, A. E., Rombola, A. T., Chrysostomakis, S. I., Chlouverakis, G. I., & Vardas, P. E. (1997). Reproducibility of time-domain indexes of heart rate variability in patients with vasovagal syncope. *The American Journal of Cardiology*, *79*(2), 160–165.

Kowalewski, M. A, & Urban, M. (2004). Short- and long-term reproducibility of autonomic measures in supine and standing positions. *Clinical Science*, *106*(1), 61–66. doi:10.1042/CS20030119

Kowalewski, M., Alifier, M., Bochen, D., & Urban, M. (2007). Heart rate turbulence in children–age and heart rate relationships. *Pediatric Research*, *62*(6), 710–714. doi:10.1203/PDR.0b013e3181598836

Kroll, D., Freed, L., & Stein, K. (1996). Rhythm annotation and interobserver reproducibility of measures of heart rate variability. *The American Journal of Cardiology*, *78*(1), 1055–1057.

La Fountaine, M. F., Wecht, J. M., Spungen, A. M., & Bauman, W. A. (2010). Intra-inter visit reproducibility of short-term linear and nonlinear measurement of heart rate variability in tetraplegia and neurologically intact controls. *Physiological Measurement*, *31*(3), 363–374. doi:10.1088/0967-3334/31/3/006

Lambert, M., Van Hulst, A., O'Loughlin, J., Tremblay, A., Barnett, T. A., Charron, H., … Paradis, G. (2012). Cohort profile: The Quebec adipose and lifestyle investigation in youth cohort. *International Journal of Epidemiology*, *41*(6), 1533–1544. doi:10.1093/ije/dyr111

Latchman, P. L., Mathur, M., Bartels, M. N., Axtell, R. S., & De Meersman, R. E. (2011). Impaired autonomic function in normotensive obese children. *Clinical Autonomic Research*, *21*(5), 319–323. doi:10.1007/s10286-011-0116-8

Lee, K., Valeria, B., Kochman, C., & Lenders, C. M. (2006). Self-assessment of height, weight, and sexual maturation: Validity in overweight children and adolescents. *The Journal of Adolescent Health*, *39*(3), 346–352. doi:10.1016/j.jadohealth.2005.12.016

*Leicht, A. S., & Allen, G. D. (2008). Moderate-term reproducibility of heart rate variability during rest and light to moderate exercise in children. *Brazilian Journal of Medical and Biological Research*, *41*(7), 627–633.

Lenard, Z., Studinger, P., Mersich, B., Kocsis, L., & Kollai, M. (2004). Maturation of cardiovagal autonomic function from childhood to young adult age. *Circulation*, *110*(16), 2307–12. doi:10.1161/01.CIR.0000145157.07881.A3

Leong, K. S., Mann, P., Wallymahmed, M., MacFarlane, I. a, & Wilding, J. P. (2000). Abnormal heart rate variability in adults with growth hormone deficiency. *The Journal of Clinical Endocrinology and Metabolism*, *85*(2), 628–633. doi:10.1210/jcem.85.2.6396

Li, Z., Snieder, H., Su, S., & Ding, X. (2009). A longitudinal study in youth of heart rate variability at rest and in response to stress. *International Journal of Psychophysiology*, *73*(3), 212–217. doi:10.1016/j.ijpsycho.2009.03.002.A

*Liew, J., Eisenberg, N., Spinrad, T. L., Eggum, N. D., Haugen, R. G., Kupfer, A., … Baham, M. E. (2011). Physiological Regulation and Fearfulness as Predictors of Young Children's Empathy-related Reactions. *Social Development*, *20*(1), 111–113. doi:10.1111/j.1467-9507.2010.00575.x

Lin, M.-T., Wang, J.-K., Lu, F. L., Wu, E.-T., Yeh, S.-J., Lee, W.-L., … Wu, M.-H. (2006). Heart rate variability monitoring in the detection of central nervous system complications in children with enterovirus infection. *Journal of Critical Care*, *21*(3), 280–286. doi:10.1016/j.jcrc.2006.02.005

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Llabre, M., Spitzer, S., & Saab, P. (1991). The reliability and specificity of delta versus residualized change as measures of cardiovascular reactivity to behavioral challenges. *Psychophysiology*, *28*(6), 701–711.

Longin, E., Dimitriadis, C., Shazi, S., Gerstner, T., Lenz, T., & König, S. (2009). Autonomic nervous system function in infants and adolescents: Impact of autonomic tests on heart rate variability. *Pediatric Cardiology*, *30*(3), 311–324. doi:10.1007/s00246-008-9327-8

Lu, G., Yang, F., Taylor, J. A, & Stein, J. F. (2009). A comparison of photoplethysmography and ECG recording to analyze heart rate variability in healthy subjects. *Journal of Medical Engineering & Technology*, *33*(8), 634–641. doi:10.3109/03091900903150998

Lubans, D. R., Hesketh, K., Cliff, D. P., Barnett, L. M., Salmon, J., Dollman, J., … Hardy, L. L. (2011). A systematic review of the validity and reliability of sedentary behaviour measures used with children and adolescents. *Obesity Reviews*, *12*(10), 781–99. doi:10.1111/j.1467-789X.2011.00896.x

Maestri, R., Pinna, G. D., Porta, A., Balocchi, R., Sassi, R., Signorini, M. G., … Raczak, G. (2007). Assessing nonlinear properties of heart rate variability from short-term recordings: are these measurements reliable? *Physiological Measurement*, *28*(9), 1067–1077. doi:10.1088/0967-3334/28/9/008

Magagnin, V., Bassani, T., Bari, V., Turiel, M., Maestri, R., Pinna, G. D., & Porta, A. (2011). Non-stationarities significantly distort short-term spectral, symbolic and entropy heart rate variability indices. *Physiological Measurement*, *32*(11), 1775–1786. doi:10.1088/0967-3334/32/11/S05

Malik, M. (1997). Time domain measurements of heart rate variability. *Cardiac Electrophysiology Review*, *3*, 329–334.

Malpas, S. C., Leonard, B. L., Guild, S., Ringwood, J. V, Navakatikyan, M., Austin, P. C., … Burgess, D. E. (2001). The sympathetic nervous system's role in regulating blood pressure variability. *IEEE Engineering in Medicine and Biology,* (March/April).

Mandigout, S., Melin, A, Fauchier, L., N'Guyen, L. D., Courteix, D., & Obert, P. (2002). Physical training increases heart rate variability in healthy prepubertal children. *European Journal of Clinical Investigation*, *32*(7), 479–487.

Marks, B. L., & Lightfoot, J. T. (1999). Reproducibility of resting heart rate variability with short sampling periods. *Canadian Journal of Applied Physiology*, *24*(4), 337–348.

Marshall, W. A. & Tanner, J. M. (1969). Variations in pattern of pubertal changes in girls. *Archives of Disease in Childhood*, 44, 291–303.

Marshall, W. A. & Tanner, J. M. (1970). Variations in pattern of pubertal changes in boys. *Archives of Disease in Childhood*, 45, 13–23.

*Marshall, P. J., & Stevenson-Hinde, J. (1998). Behavioral inhibition, heart period, and respiratory sinus arrhythmia in young children. *Developmental Psychobiology*, *33*(3), 283–292.

Martini, G., Riva, P., Rabbia, F., Molini, V., Ferrero, G. B., Cerutti, F., … Veglio, F. (2001). Heart rate variability in childhood obesity. *Clinical Autonomic Research*, *11*(2), 87–91.

Martinmaki, K., Rusko, H., Kooistra, L., Kettunen, J., & Saalasti, S. (2006). Intraindividual validation of heart rate variability indexes to measure vagal effects on hearts. *American Journal of Physiology. Heart and Circulatory Physiology*, *290*, 640–647. doi:10.1152/ajpheart.00054.2005.

Massin, M., Maeyns, K., Withofs, N., Ravet, F., & P, G. (2000). Circadian rhythm of heart rate and heart rate variability. *Archives of Disease in Childhood*, *83*(2), 179–182.

McCabe, P. M., Porges, S. W., & Carter, C. S. (1981). Heart period variability during estrogen exposure and withdrawal in female rats. *Physiology & Behavior*, *26*(3), 535–538.

McKinley, P. S., Shapiro, P. a, Bagiella, E., Myers, M. M., De Meersman, R. E., Grant, I., & Sloan, R. P. (2003). Deriving heart period variability from blood pressure waveforms. *Journal of Applied Physiology*, *95*(4), 1431–1438. doi:10.1152/japplphysiol.01110.2002

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: The Guilford Press.

McNames, J., & Aboy, M. (2006). Reliability and accuracy of heart rate variability metrics versus ECG segment duration. *Medical & Biological Engineering & Computing*, *44*(9), 747–756. doi:10.1007/s11517-006-0097-2

McNarry, M. A, & Lewis, M. J. (2012). Heart rate variability reproducibility during exercise. *Physiological Measurement*, *33*(7), 1123–1133. doi:10.1088/0967-3334/33/7/1123

Merri, M., Farden, D. C., Mottley, J. G., & Titlebaum, E. L. (1990). Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Transactions on Bio-Medical Engineering*, *37*(1), 99–106. doi:10.1109/10.43621

*Mezzacappa, E., Tremblay, R. E., Kindlon, D., Saul, J. P., Arseneault, L., Seguin, J., … Earls, F. (1997). Anxiety, antisocial behavior, and heart rate regulation in adolescent males. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(4), 457–469.

Mulder, L. J. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology*, *34*(2-3), 205–236.

Nepal, G. B., & Paudel, B. H. (2012). Effect of posture on heart rate variability in school children. *Nepal Medical College Journal : NMCJ*, *14*(4), 298–302.

Netherton, C., Goodyer, I., Tamplin, A., & Herbert, J. (2004). Salivary cortisol and dehydroepiandrosterone in relation to puberty and gender. *Psychoneuroendocrinology*, *29*(2), 125–140. doi:10.1016/S0306-4530(02)00150-6

Oliveira, F. M., Tran, W. H., Lesser, D., Bhatia, R., Ortega, R., Mittelman, S. D., … Khoo, M. C. (2010). Autonomic and metabolic effects of OSA in childhood obesity. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, (4), 6134–6137. doi:10.1109/IEMBS.2010.5627789

Ordaz, S., & Luna, B. (2012). Sex differences in physiological reactivity to acute psychosocial stress in adolescence. *Psychoneuroendocrinology*, *37*(8), 1135–1157. doi:10.1016/j.psyneuen.2012.01.002.Sex

Ottaviani, C., & Shapiro, D. (2011). Do we need a stressor to be stressed? Insights from cardiac regulation. *Japanese Psychological Research*, *53*(2), 155–162. doi:10.1111/j.1468-5884.2011.00462.x

Ottaviani, C., Shapiro, D., Davydov, D. M., & Goldstein, I. B. (2008). Autonomic stress response modes and ambulatory heart rate level and variability. *Journal of Psychophysiology*, *22*(1), 28–40. doi:10.1027/0269-8803.22.1.23

Paradis, G., Lambert, M., O'Loughlin, J., Lavallée, C., Aubin, J., Delvin, E., … Hanley, J. A. (2004). Blood pressure and adiposity in children and adolescents. *Circulation*, *110*(13), 1832–1838. doi:10.1161/01.CIR.0000143100.31752.B7

Park, M. K., & Menard, S. W. (1998). Blood pressure measurement in pediatric patients: State of the art. *Bahrain Medical Bulletin*, 20(2), 44–51.

Park, G., & Thayer, J. F. (2014). From the heart to the mind: Cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Frontiers in Psychology*, *5*(May), 278. doi:10.3389/fpsyg.2014.00278

Patton, G. C., & Viner, R. (2007). Pubertal transitions in health. *Lancet*, *369*(9567), 1130–1139. doi:10.1016/S0140-6736(07)60366-3

Penttilä, J., Kuusela, T., & Scheinin, H. (2005). Analysis of rapid heart rate variability in the assessment of anticholinergic drug effects in humans. *European Journal of Clinical Pharmacology*, *61*(8), 559–565. doi:10.1007/s00228-005-0953-2

*Perry, N. B., Nelson, J. A, Swingler, M. M., Leerkes, E. M., Calkins, S. D., Marcovitch, S., & O'Brien, M. (2013). The relation between maternal emotional support and child physiological regulation across the preschool years. *Developmental Psychobiology*, *55*(4), 382–394. doi:10.1002/dev.21042

Pinna, G. D., Maestri, R., Torunski, A., Danilowicz-Szymanowicz, L., Szwoch, M., La Rovere, M. T., & Raczak, G. (2007). Heart rate variability measures: A fresh look at reliability. *Clinical Science (London, England : 1979)*, *113*(3), 131–140. doi:10.1042/CS20070055

Porges, S. (1985). Method and apparatus for evaluating rhythmic oscillations in aperiodic physiological response systems. *US Patent 4,510,944*.

*Porges, S. W., Doussard-Roosevelt, J. A, Portales, A. L., & Suess, P. E. (1994). Cardiac vagal tone: Stability and relation to difficultness in infants and 3-year-olds. *Developmental Psychobiology*, *27*(5), 289–300. doi:10.1002/dev.420270504

Porges, W., & Byrne, E. A. (1992). Research methods for measurement of heart rate & respiration. *Biological Psychology*, *34*, 93–130.

*Porges. S. W. (2014). Unpublished raw data.

*Porter, C. L., Bryan, Y. E., & Hsu, H.-C. (1995). Physiological markers in early infancy: Stability of 1-to 6-month vagal tone. *Infant Behavior and Development*, *18*(3), 363–367. doi:10.1016/0163-6383(95)90025-X

*Propper, C., Moore, G. A., Mills-Koonce, W. R., Halpern, C. T., Hill-Soderlund, A. L., Calkins, S. D., … Cox, M. (2008). Gene-environment contributions to the development of infant vagal reactivity: The interaction of dopamine and maternal sensitivity. *Child Development*, *79*(5), 1377–1394. doi:10.1111/j.1467-8624.2008.01194.x

Rabbia, F., Silke, B., Conterno, A., Grosso, T., De Vito, B., Rabbone, I., … Veglio, F. (2003). Assessment of cardiac autonomic modulation during adolescent obesity. *Obesity Research*, *11*(4), 541–548. doi:10.1038/oby.2003.76

Ramaekers, D., Ector, H., Aubert, A. E., Rubens, A., & Van de Werf, F. (1998). Heart rate variability and heart rate in healthy volunteers. Is the female autonomic nervous system cardioprotective? *European Heart Journal*, *19*(9), 1334–1341.

Reyes del Paso, G. A, Langewitz, W., Mulder, L. J. M., van Roon, A., & Duschek, S. (2013). 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*, *50*(5), 477–487. doi:10.1111/psyp.12027

Richards, J. (1995). Reliability of respiratory sinus arrhythmia in R-R intervals, in 14-, 20-, and 26-week-old infants. *Infant Behavior and Development*, *18*, 155–161.

*Richards, J. E. (1989). Development and stability in visual sustained attention in 14, 20, and 26 week old infants. *Psychophysiology*, *26*(4), 422–430.

*Rigterink, T., Fainsilber Katz, L., & Hessler, D. M. (2010). Domestic violence and longitudinal associations with children's physiological regulation abilities. *Journal of Interpersonal Violence*, *25*(9), 1669–1683. doi:10.1177/0886260509354589

Riihimaa, P. H., Suominen, K., Knip, M., Tapanainen, P., & Tolonen, U. (2002). Cardiovascular autonomic reactivity is decreased in adolescents with Type 1 diabetes. *Diabetic Medicine*, *19*(11), 932–938.

Riniolo, T., & Porges, S. (1997). Inferential and descriptive influences on measures of respiratory sinus arrhythmia: Sampling rate, R-wave trigger accuracy, and variance estimates. *Psychophysiology*, *34*, 613–621.

Roche, F., Pichot, V., Sforza, E., Court-Fortune, I., Duverney, D., Costes, F., … Barthelemy, J.-C. (2003). Predicting sleep apnoea syndrome from heart period: A time-frequency wavelet analysis. *European Respiratory Journal*, *22*(6), 937–942. doi:10.1183/09031936.03.00104902

Rodríguez-Colón, S. M., Bixler, E. O., Li, X., Vgontzas, A. N., & Liao, D. (2011). Obesity is associated with impaired cardiac autonomic modulation in children. *International Journal of Pediatric Obesity*, *6*(2), 128–34. doi:10.3109/17477166.2010.490265

Roeser, K., Obergfell, F., Meule, A., Vögele, C., Schlarb, A. a, & Kübler, A. (2012). Of larks and hearts - morningness/eveningness, heart rate variability and cardiovascular stress response at different times of day. *Physiology & Behavior*, *106*(2), 151–157. doi:10.1016/j.physbeh.2012.01.023

Rogers, K., (ed.) (2011). *The Brain and the Nervous System*. New York, NY: Britannica Educational Publishing.

Rosenberg, M. S., Adams, D., & Gurevitch, J. (2000). *MetaWin. Statistical software for meta analysis, version 2.0*. Sunderland, Massachusetts: Sinauer Associates.

Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, *59*(2), 464–468. doi:10.1111/j.0014-3820.2005.tb01004.x

Rosenthal, R. (1994). Parametric Measures of Effect Size. In Cooper, H., & Hedges, L. V. (Eds.), *The Handbook of Research Synthesis* (pp 231-244). New York, NY: Russel Sage Foundation.

Salo, M. A, Huikuri, H. V, & Seppänen, T. (2001). Ectopic beats in heart rate variability analysis: Effects of editing on time and frequency domain measures. *Annals of Noninvasive Electrocardiology*, *6*(1), 5–17.

*Salomon, K. (2005). Respiratory sinus arrhythmia during stress predicts resting respiratory sinus arrhythmia 3 years later in a pediatric sample. *Health Psychology*, *24*(1), 68–76. doi:10.1037/0278-6133.24.1.68

Salomon, K., Matthews, K. a, & Allen, M. T. (2000). Patterns of sympathetic and parasympathetic reactivity in a sample of children and adolescents. *Psychophysiology*, *37*(6), 842–849.

Salvadori, M., Sontrop, J. M., Garg, A. X., Truong, J., Suri, R. S., Mahmud, F. H., … Clark, W. F. (2008). Elevated blood pressure in relation to overweight and obesity among children in a rural Canadian community. *Pediatrics*, *122*(4), e821–7. doi:10.1542/peds.2008-0951

Sampei, M., Murata, K., Dakeishi, M., & Wood, D. C. (2006). Cardiac autonomic hypofunction in preschool children with short nocturnal sleep. *The Tohoku Journal of Experimental Medicine*, *208*(3), 235–242.

Sandercock, G. R. H., Bromley, P. D., & Brodie, D. A. (2005). Effects of exercise on heart rate variability: Inferences from meta-analysis. *Medicine & Science in Sports & Exercise*, *37*(3), 433–439. doi:10.1249/01.MSS.0000155388.39002.9D

Sandercock, G. R. H., Bromley, P. D., & Brodie, D. A. (2005). The reliability of short-term measurements of heart rate variability. *International Journal of Cardiology*, *103*(3), 238–247. doi:10.1016/j.ijcard.2004.09.013

Scheffer, J. (2002). Dealing with missing data. *Res. Lett. Inf. Math. Sci.*, *3*, 153–160. Retrieved from http://muir.massey.ac.nz/handle/10179/4355

Schmid, K., Schönlebe, J., Drexler, H., & Mueck-Weymann, M. (2010). Associations between being overweight, variability in heart rate, and well-being in the young men. *Cardiology in the Young*, *20*(1), 54–59. doi:10.1017/S1047951109992009

Schmitz, K. E. (2004). A validation study of early adolescents' pubertal self-assessments. *The Journal of Early Adolescence*, *24*(4), 357–384. doi:10.1177/0272431604268531

*Schuetze, P., Eiden, R. D., & Edwards, E. P. (2009). A longitudinal examination of physiological regulation in cocaine-exposed infants across the first 7 months of life. *Infancy*, *14*(1), 19–43. doi:10.1080/15250000802569660

Seppälä, S., Laitinen, T., Tarvainen, M. P., Tompuri, T., Veijalainen, A., Savonen, K., & Lakka, T. (2013). Normal values for heart rate variability parameters in children 6-8 years of age: The PANIC Study. *Clinical Physiology and Functional Imaging*, *34*(4), 290–296. doi:10.1111/cpf.12096

Shaffer, D. R., Wood, E., & Willoughby, T. (2005). *Developmental Psychology: Childhood and Adolescence* (2nd Canadian ed.) Toronto: Nelson.

Shaffer, F., & Combatalade, D. C. (2013). Don't add or miss a beat: A guide to cleaner heart rate variability recordings. *Biofeedback*, *41*(3), 121–130. doi:10.5298/1081-5937-41.3.04

Sharpley, C. F. (1993). Effects of brief rest periods upon heart rate in multiple baseline studies of heart rate reactivity. *Biofeedback and Self-Regulation*, *18*(4), 225–235.

Shirtcliff, E. A., Dahl, R. E., & Pollak, S. D. (2009). Pubertal development: Correspondence between hormonal and physical development. *Child Development*, *80*(2), 327–337. doi:10.1111/j.1467-8624.2009.01263.x.Pubertal

Shroff, R., Dégi, A., Kerti, A., Kis, E., Cseprekál, O., Tory, K., … Reusz, G. S. (2013). Cardiovascular risk assessment in children with chronic kidney disease. *Pediatric Nephrology*, *28*(6), 875–884. doi:10.1007/s00467-012-2325-3

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Silver, N., & Dunlap, W. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, *72*(1), 1979–1982. Retrieved from http://psycnet.apa.org/journals/apl/72/1/146/

Silvetti, M. S., Drago, F., & Ragonese, P. (2001). Heart rate variability in healthy children and adolescents is partially related to age and gender. *International Journal of Cardiology*, *81*(2-3), 169–174.

Singh, D., Vinod, K., & Saxena, S. C. (2004). Sampling frequency of the RR interval time series for spectral analysis of heart rate variability. *Journal of Medical Engineering & Technology*, *28*(6), 263–272. doi:10.1080/03091900410001662350

Singh, J. P., Larson, M. G., Tsuji, H., Evans, J. C., O'Donnell, C. J., & Levy, D. (1998). Reduced heart rate variability and new-onset hypertension: Insights into pathogenesis of hypertension: The Framingham Heart Study. *Hypertension*, *32*(2), 293–297. doi:10.1161/01.HYP.32.2.293

Sirard, J. R., & Pate, R. R. (2001). Physical activity assessment in children and adolescents. *Sports Medicine*, *31*(6), 439–454.

Skirton, H., Chamberlain, W., Lawson, C., Ryan, H., & Young, E. (2011). A systematic review of variability and reliability of manual and automated blood pressure readings. *Journal of Clinical Nursing*, *20*(5-6), 602–614. doi:10.1111/j.1365-2702.2010.03528.x

*Snidman, N., Kagan, J., Riordan, L., & Shannon, D. C. (1995). Cardiac function and behavioral reactivity during infancy. *Psychophysiology*, *32*(3), 199–207. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7784528

Sookan, T., & McKune, A. J. (2012). Heart rate variability in physically active individuals: Reliability and gender characteristics. *Cardiovascular Journal of Africa*, *23*(2), 67–72. doi:10.5830/CVJA-2011.108

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience and Biobehavioral Reviews*, *24*(4), 417–463.

Stein, P., & Kleiger, R. (1998). Insights from the study of heart rate variability. *Annual Review of Medicine*, *50*, 249–261.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and

clinical research: Potential and pitfalls. *British Medical Journal*, *338*. doi:10.1136/bmj.b2393

*Stifter, C. A., & Anju, J. (1996). Psychophysiological correlates of infant temperament: Stability of behavior and autonomic patterning from 5 to 18 months. *Developmental Psychobiology*, *29* (January 1995), 379–391.

*Stifter, C. A., & Fox, N. A. (1990). Infant reactivity: Physiological correlates of newborn and 5-month temperament. *Developmental Psychology*, *26*(4), 582–588. doi:10.1037//0012-1649.26.4.582

Sztajzel, J. (2004). Heart rate variability: A noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Medical Weekly*, *134*(35-36), 514–522. doi:2004/35/smw-10321

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Pearson.

Tanaka, H., Borres, M., Thulesius, O., Tamai, H., Ericson, M. O., & Lindblad, L. E. (2000). Blood pressure and cardiovascular autonomic function in healthy children and adolescents. *The Journal of Pediatrics*, *137*(1), 63–67. doi:10.1067/mpd.2000.108098

Task Force of the European Society of Cardiology, & Task Force of the European Society of Cardiology (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*, 354–381.

Taylor, J. A., Carr, D. L., Myers, C. W., & Eckberg, D. L. (1998). Mechanisms underlying very-low-frequency RR-interval oscillations in humans. *Circulation*, *98*(6), 547–555. doi:10.1161/01.CIR.98.6.547

Thayer, J. F., & Sternberg, E. (2006). Beyond heart rate variability: Vagal regulation of allostatic systems. *Annals of the New York Academy of Sciences*, *1088*, 361–372. doi:10.1196/annals.1366.014

Thayer, J. F., Ahs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews*, *36*(2), 747–756. doi:10.1016/j.neubiorev.2011.11.009

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, *60*, 174–195.

Thuraisingham, R. A. (2006). Preprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals. *Computer Methods and Programs in Biomedicine*, *83*(1), 78–82. doi:10.1016/j.cmpb.2006.05.002

Traina, M., Gataldo, A., Galullo, F., & Russo, G. (2011). Effects of anxiety due to mental stress on heart rate variability in healthy subjects. *Minerva Psichiatrica*, *52*(4), 227–231.

Tremblay, M. S., & Willms, J. D. (2000). Secular trends in the body mass index of Canadian children. *CMAJ : Canadian Medical Association Journal*, *163*(11), 1429–1433.

Tsuji, H., Venditti, F. J., Manders, E. S., Evans, J. C., Larson, M. G., Feldman, C. L., & Levy, D. (1994). Reduced heart rate variability and mortality risk in an elderly cohort. The Framingham Heart Study. *Circulation*, *90*(2), 878–883. doi:10.1161/01.CIR.90.2.878

Tsuji, H., Venditti, F. J., Manders, E. S., Evans, J. C., Larson, M. G., Feldman, C. L., & Levy, D. (1996). Determinants of heart rate variability. *Journal of the American College of Cardiology*, *28*(6), 1539–1546.

Tzeng, Y. C., Sin, P. Y. W., Lucas, S. J. E., & Ainslie, P. N. (2009). Respiratory modulation of cardiovagal baroreflex sensitivity. *Journal of Applied Physiology*, *107*(3), 718–724. doi:10.1152/japplphysiol.00548.2009

Weir, J. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, *19*(1), 231–240.

Weise, M., Eisenhofer, G., & Merke, D. P. (2002). Pubertal and gender-related changes in the sympathoadrenal system in healthy children. *The Journal of Clinical Endocrinology and Metabolism*, *87*(11), 5038–5043. doi:10.1210/jc.2002-020590

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604.

Williams, K., Thomson, D., Seto, I., Contopoulos-Ioannidis, D. G., Ioannidis, J. P. A, Curtis, S., … Klassen, T. (2012). Standard 6: Age groups for pediatric trials. *Pediatrics*, *129 Suppl 3*, S153–160. doi:10.1542/peds.2012-0055I

*Winsley, R. J., Armstrong, N., Bywater, K., & Fawkner, S. G. (2003). Reliability of heart rate variability measures at rest and during light exercise in children. *British Journal of Sports Medicine*, *37*(6), 550–552.

Yasuma, F., & Hayano, J. (2004). Respiratory sinus arrhythmia: Why does the heartbeat synchronize with respiratory rhythm? *CHEST*, *125*(2), 683–690.

*Yiallourou, S. R., Sands, S. A., Walker, A. M., & Horne, R. S. C. (2012). Maturation of heart rate and blood pressure variability during sleep in term-born infants. *Sleep*, *35*(2), 177–186. doi:10.5665/sleep.1616

Young, F., & Leicht, A. (2011). Short-term stability of resting heart rate variability: influence of position and gender. *Applied Physiology, Nutrition, and Metabolism*, *218*(September 2010), 210–218. doi:10.1139/H10-103

Table 1

Continuous Study Descriptive Characteristics by Age Group

	Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
	k	*N*	*M* (*SD*)	*k*	*N*	*M* (*SD*)
Study follow-up length (wk)	31	3257	64.27 (63.18)	15	1533	85.84 (111.04)
Sample size (*N*)						
First study visit	31	3257	105.06 (89.48)	15	1533	102.2 (84.95)
Final study visit	31	2202	71.03 (74.94)	15	1123	74.87 (90.93)
Percent (%) of sample attrition						
First to second study visit	12	1865	21.67 (17.14)	7	1151	19.57 (7.70)
Second to third study visit	3	794	9.33 (2.52)	1	251	16.00 (0.00)
Sample mean age (mo)						
First study visit	31	3257	14.31 (17.87)	14	1261	122.43 (33.36)
Final study visit	31	3257	31.77 (29.92)	13	1465	133.61 (34.48)
Percent (%) of male participants						
First study visit	17	1337	45.47 (8.23)	12	1060	56.85 (21.66)
Second study visit	16	1806	46.76 (6.48)	12	1383	57.45 (21.29)
Height (cm)						
First study visit	2	108	52.30 (0.58)	2	38	147.35 (12.24)
Second study visit	0	–	–	1	26	141.20 (0.00)
Weight (kg)						
First study visit	10	499	3.61 (0.30)	4	264	83.93 (13.27)
Second study visit	2	61	5.96 (0.01)	2	36	37.30 (0.85)

	Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
	k	*N*	*M* (*SD*)	*k*	*N*	*M* (*SD*)
Body mass index (BMI; kg/m^2)						
First study visit	0	–	–	3	483	19.20 (1.15)
Second study visit	0	–	–	1	77	22.30 (0.00)
Effect sizes per study	31	3257	3.06 (4.70)	15	1533	2.60 (1.81)

Note. ECG = electrocardiogram; wk = week; mo = month; cm = centimetre; kg = kilograms; kg/m^2 = kilograms per meters squared

Table 2

ECG Recording Characteristics by Age Group

	Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
	k	*N*	*M* (*SD*)	*k*	*N*	*M* (*SD*)
Acclimatization period duration (min) [a]	5	500	3.46 (2.11)	10	1244	8.10 (8.79)
ECG recording duration (min) [a]						
Total recording duration	30	3230	24.98 (46.65)	14	1477	18.59 (20.77)
Baseline recording duration	29	3196	4.76 (4.25)	15	1533	4.73 (3.63)
Analyzed recording duration	29	3017	4.54 (4.32)	15	1533	3.22 (0.96)
ECG recording epoch duration (sec) [a]	18	2161	30.28 (13.32)	4	179	48.75 (48.02)

Note. ECG = electrocardiogram; HRV = heart rate variability; min = minutes; sec = seconds

[a] Based on initial study assessment

Table 3

Categorical Study Descriptive Characteristics by Age Group

	Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
	k	*N*	% [a]	*k*	*N*	% [a]
Study recording location						
Identical study location	27	2750	87.1%	14	1517	93%
Changed/multiple location(s)	4	507	12.9%	1	16	7%
Sample exclusion criteria						
Prescription medications						
Exclusion reported	3	165	9.7%	2	217	13.3%
Unmentioned	28	3092	90.3%	13	1316	86.7%
Abnormal pregnancy/birth						
Exclusion reported	4	316	12.9%		–	
Unmentioned	27	2941	87.1%	15	1533	100%
Single parent home/low SES						
Exclusion reported	3	114	9.7%	4	599	26.7%
Unmentioned	28	3143	90.3%	11	934	73.3%
Chronic medical illness						
Exclusion reported	6	376	19.4%	7	951	46.7%
Unmentioned	25	2881	80.6%	8	582	53.3%
Mental health condition						
Exclusion reported		–		6	874	40%
Unmentioned	31	3257	100%	9	659	60%

	Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
	k	N	% [a]	k	N	% [a]
Weight characteristics						
Exclusion reported	2	202	6.5%	1	201	6.7%
Unmentioned	29	3055	93.5%	14	1332	93.3%
Time of ECG recording						
06:00-12:00	3	92	9.7%	5	324	33.3%
12:00-18:00	1	20	3.2%	1	64	6.7%
Unmentioned	27	3145	87.1%	9	1145	60%
Pre-recording acclimatization period duration (min)						
1–2	2	132	4%	4	622	40.5%
5–10	3	368	11%	4	535	35%
15–30		–		2	103	7%
No acclimatization period	25	2631	80.6%	5	289	19%
Daytime ECG recorded exclusively while awake						
Exclusively awake ECG data	25	2847	80.6%	15	1533	100%
Participant sleep unrestricted	6	410	19.4%		–	
Additional measure of respiration						
Study measured respiration	8	757	25.8%	6	1048	40%
Not measured/unmentioned	23	2500	74.2%	9	485	60%
ECG recording posture						
Seated posture	19	2214	61.3%	5	700	33.3%
Supine posture	6	744	19.4%	10	833	66.7%

	Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
	k	N	% [a]	k	N	% [a]
Changed recording posture	6	299	19.4%		–	
ECG sampling rate						
≥ 500 Hz	3	85	9.7%	9	1233	60%
< 500 Hz	4	295	12.9%	1	57	6.7%
Unmentioned	24	2877	77.4%	5	243	33.3%
ECG recording duration (min) [b]						
Total recording duration						
1–3	6	357	11%	1	251	16.4%
3.1–4.99	3	468	14.4%	1	77	5%
5–10	9	850	26.1%	6	652	42.5%
10.1–25	8	1207	37.1%	4	286	18.7%
50+	4	526	16%	2	211	13.8%
Baseline recording duration						
1–2	10	1067	32.8%	1	16	1%
2.1–4.99	6	662	20.3%	10	1412	92%
5–10	11	1348	41.4%	3	95	6%
10.1–25	2	119	4%	1	10	1%
Manual editing of ECG data						
Yes/reported	23	2638	74.2%	13	1511	86.7%
No/unmentioned	8	619	25.8%	2	22	13.3%
Software-based filter algorithm						

	Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
	k	*N*	% [a]	*k*	*N*	% [a]
Yes/reported	21	2091	67.7%	8	433	53.3%
No/unmentioned	10	1166	32.3%	7	1100	46.7%
Analyzed ECG recording duration (min) [b]						
0.5–1.9	4	369	11.3%		–	
2–2.9	7	539	16.5%	3	201	13.1%
3–3.9	4	441	13.5%	8	1160	75.7%
4–4.9	3	229	7%	2	103	6.2%
5–10	9	1300	40%	2	69	5%
10.1–25	2	119	4.7%		–	
HRV derivation method						
Mean of 1 ECG epoch	4	387	12.9%	5	873	33.3%
Mean of ≥2 ECG epochs	18	2161	58.1%	4	179	26.7%
Unmentioned	9	709	29%	6	481	40%
ECG recording epoch duration (sec) [b]						
≤ 15	2	234	7.2%	1	30	2%
20	1	104	3.1%		–	
30	13	1537	47.2%	2	72	4.6%
60	2	286	8.8%		–	
120		–		1	77	5%
Frequency bandwidth selection						
LF						

		Infant/toddler ($k = 31$)			Child/adolescent ($k = 15$)		
		k	*N*	% [a]	*k*	*N*	% [a]
	Recommended bandwidth	1	31	3.2%	4	335	26.7%
	Alternative bandwidth		–		2	22	13.3%
HF							
	Recommended bandwidth	17	1968	54.8%	6	216	40%
	Alternative bandwidth	7	539	22.6%	3	243	20%

Note. HRV = Heart Rate Variability; ECG = Electrocardiogram; LF = Low Frequency; HF = High Frequency; min = minutes; sec = seconds

[a] Percent of total *N*. [b] Based on initial study assessment

Table 4

Overall Summary Analyses for HRV Reliability

	k	*Z* [95% CI][a]	Q_T	*p*	N_{FS}
Infant/toddler studies					
Averaged HRV indices [b]	31	0.42 [0.35, 0.49]	80.86	<.001	1279
NN	14	0.33 [0.15, 0.48]	45.82	<.001	443
SDNN	3	0.43 [0.15, 0.91]	14.92	<.001	126
HF	27	0.42 [0.35, 0.49]	71.61	<.001	1115
Child/adolescent studies					
Averaged HRV indices [b]	15	0.64 [0.51, 0.79]	59.86	<.001	949
SDNN	4	0.85 [0.46, 1.04]	6.07	.108	334
HF	15	0.65 [0.52, 0.82]	61.15	<.001	955
LF	6	0.61 [0.41, 0.81]	12.36	.030	357
LF:HF Ratio	3	0.52 [0.14, 1.19]	4.60	.101	151

Note. *k* = Number of Studies; *Z* = Fisher's *Z*; Q_T = Heterogeneity Statistic; N_{FS} = Orwin's Failsafe N; ES = effect size; HRV = Heart Rate Variability; HF = High Frequency; HP = Heart Period; SDNN = Standard Deviation of NN Intervals; LF = Low Frequency

[a] Bootstrap 95% Confidence Interval. [b] Average of non-redundant effect sizes for each reported HRV variable

Table 5

HRV Reliability – Continuous Moderator Variables by Age Group

	NN			SDNN			HF			LF			LF:HF Ratio		
	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
Study follow-up length															
Infant/Toddler	14	.00(<.01)	45.82***	3	**-.01**(**<.01**)***	14.92***	27	<.01(<.01)	71.61***		–			–	
Child/Adolescent		–		3	-.01(.01)	.409	15	**<-.01**(**<.01**)**	61.15***	6	**-.01**(**>.01**)*	12.36*	3	.10(.06)	4.59
Sample size (*N*)[a]															
Infant/Toddler	14	**<.01**(**<.01**)	45.82***	3	**-.01**(**<.01**)***	14.92***	27	<.01(<.01)	71.61***		–			–	
Child/Adolescent		–		3	.01(<.01)	6.07	15	<-.**01**(<.**01**)*	61.15***	6	**<-.01**(**<.01**)**	12.36*	3	-.02(.03)	4.59
Age (T0 mean; mo)															
Infant/Toddler	14	**.01**(**<.01**)***	45.82***	3	**-.53**(**.14**)***	14.92***	27	**<.01**(**<.01**)**	71.61***		–			–	
Child/Adolescent		–		3	<-.01(<.01)	.685	14	<-.01(<.01)	58.99***	6	<-.01(<.01)	12.36*	3	-.01(.01)	4.59
Male (percent)															
Infant/Toddler	9	**.02**(**.01**)**	36.46***		–		18	<-.01(.01)	61.67***		–			–	

	NN			SDNN			HF			LF			LF:HF Ratio		
	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
Child/Adolescent		–		3	.01(.01)	.421	14	**-.01(<.01)****	58.99***	6	**-.01(<.01)****	12.36*	3	.05(.03)	4.59
Weight (lbs)															
Infant/Toddler	5	.41(.43)	4.38		–		7	.12(.26)	10.76		–			–	
Child/Adolescent		–			–		4	<.01(.01)	3.45	3	-.01(.02)	3.89	3	-.02(.02)	4.59
Analyzed ECG recording duration (min)[b]															
Infant/Toddler	14	**-.03(.01)***	45.82***		–		27	**-**.01(.01)	71.61***		–			–	
Child/Adolescent		–		3	.24(.18)	.181	15	**.14(.05)****	61.15***	6	**.10(.05)***	12.36*	3	**-.35(.16)***	4.59

Note. *B* = slope; *SE* = standard error; Q_T = total heterogeneity; HRV = heart rate variability; ECG = electrocardiogram; wk = week; mo = month; lbs = pounds; min = minutes; HP = heart period; SDNN = standard deviation of NN intervals; HF = high frequency; LF = low frequency

[a] Sample size at first study visit. [b] Baseline recording duration submitted for HRV analysis * $p < .05$. ** $p \leq .01$. *** $p \leq .001$.

Table 6

Infant/Toddler HRV Reliability – Categorical Moderator Variables

	NN				SDNN				HF			
	k	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}
Sample exclusion criteria												
Prescription medications		Q_M = **14.06**, p < .**001**				N/A				Q_M = 0.47, p = .491		
Exclusion reported	2	0.61 [.40, .65]	1.10	120	1	0.91 [-,-]			2	0.37 [.34, .50]	0.45	71
Unmentioned	12	0.23 [.11, .39]	30.67***	269	2	0.33 [.15, .76]	8.27**	63	25	0.43 [.33, .51]	70.68***	1043
Chronic medical illness		Q_M = **5.26**, p = .**022**				N/A				Q_M = 2.97, p = .085		
Exclusion reported	3	0.47 [.04, .65]	11.74**	137	2	0.82 [.76, .91]	0.30	163	4	0.32 [.20, .40]	2.00	123
Unmentioned	11	0.26 [.12, .44]	28.82***	273	1	0.15 [-,-]			23	0.44 [.35, .52]	66.64***	984
Abnormal pregnancy/birth		Q_M = 1.07, p = .302				N/A				Q_M = **8.34**, p = .**004**		
Exclusion reported	4	0.38 [.02, .61]	18.84***	148		–			4	0.34 [-.01, .37]	6.88	91
Unmentioned	10	0.29 [.15, .48]	25.92**	280	3	0.43 [.15, .91]	14.92***	126	23	0.45 [.37, .52]	56.39***	1004
Low birth weight (<2500 g)		Q_M = **4.71**, p = .**030**				N/A				Q_M = **6.31**, p = .**012**		

	NN				SDNN				HF			
	k	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}
Exclusion reported	2	0.48 [.01, .65]	12.41***	93		–			2	0.23 [-.07, .34]	5.09*	43
Unmentioned	12	0.27 [.15, .45]	28.70**	309	3	0.43 [.15, .91]	14.92***	126	25	0.44 [.36, .51]	60.20***	1075
Study recording location		N/A				N/A				$Q_M = 0.06$, $p = .803$		
Identical study location	13	0.35 [.18, .52]	41.29***	446	3	0.43 [.15, .91]	14.92***	126	23	0.42 [.33, .50]	58.09***	944
Changed/multiple locations	1	0.01 [-,-]				–			4	0.44 [.20, .55]	13.45**	170
Time of ECG recording		N/A				N/A				$Q_M = 0.37$, $p = .541$		
Identical recording time		–			2	0.82 [.76, .91]	0.30	163	2	0.31 [.06, .33]	14.00	60
Unmentioned	14	0.33 [-.-]	45.82***	443	1	0.15 [-,-]			25	0.42 [.34, .49]	71.10**	1037
Pre-ECG acclimatization period		$Q_M = \mathbf{18.31}$, $p < .\mathbf{001}$				N/A				$Q_M = 0.02$, $p = .880$		
Yes/included	2	0.62 [0.55,0.65]	0.29	122	1	0.76 [-,-]			7	0.43 [.36, .49]	3.19	293
No/unmentioned	12	0.21 [0.10,0.38]	27.22**	243	2	0.33 [0.15,0.91]	10.50***	63	20	0.42 [.31, .51]	68.40***	822
ECG recorded while awake[a]		N/A				N/A				$Q_M = \mathbf{6.50}$, $p = .\mathbf{011}$		
Yes	13	0.35 [0.17,0.52]	41.29***	446	1	0.15 [-,-]			24	0.44 [.36, .51]	60.31***	1030

	NN				SDNN				HF			
	k	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}
No	1	0.01 [-,-]			2	0.82 [0.76,0.91]	0.30	163	3	0.21 [.03, .34]	4.79	60
ECG recording posture[a]		Q_M = **9.42**, p = **.002**				N/A				Q_M = **23.12**, p < **.001**		
Sitting	10	0.39 [0.20,0.56]	31.55***	384	1	0.15 [-,-]			18	0.50 [.41, .57]	42.13***	883
Supine/unmentioned	1	0.16 [-,-]				–			6	0.30 [.21, .34]	2.04	173
Changed recording posture	3	0.01 [-0.21,0.59]	3.79	0	2	0.82 [0.76,0.91]	0.30	163	3	0.19 [-.07, .34]	4.32	53
ECG sampling rate		N/A				N/A				Q_M = **4.84**, p = **.028**		
≥ 500 Hz	1	0.40 [-,-]			2	0.82 [0.76,0.91]	0.30	163	1	0.50 [-,-]		
< 500 Hz	1	0.65 [-,-]				–			4	0.28 [.16, .34]	1.53	107
Unmentioned	12	0.23 [.11, .40]	30.67***	269	1	0.15 [-,-]			22	0.44 [.35, .51]	65.11***	946
Manual editing of ECG data		Q_M = **4.99**, p = **.025**				N/A				Q_M = 1.92, p = .166		
Yes	12	0.38 [.16, .55]	40.80***	439	1	0.91 [-,-]			21	0.44 [.36, .51]	63.64***	894
No/unmentioned	2	0.14 [.13, .16]	0.02	25	2	0.33 [.15, .76]	**8.27****	63	6	0.34 [.23, .47]	6.04	200
Software-based filter algorithm		Q_M = 1.96, p = .161				N/A				Q_M = **4.00**, p = **.045**		

	NN				SDNN				HF			
	k	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}
Yes	11	0.36 [.14, .54]	41.87***	382		–			21	0.45 [.34, .54]	61.05***	888
No/unmentioned	3	0.21 [.13, .59]	1.98	59	3	0.43 [.15, .91]	14.92***	126	6	0.35 [.27, .45]	6.14	237
Identical ECG recording duration[a]		Q_M = **5.75**, *p* = **.017**				N/A				Q_M = **5.13**, *p* = **.024**		
Yes	11	0.38 [.16, .56]	38.43***	404	3	0.43 [.15, .85]	14.92***	126	20	0.45 [.36, .53]	53.40***	886
No	3	0.12 [.01, .32]	1.67	31		–			7	0.33 [.21, .47]	13.07*	227
HRV derivation method		Q_M = 2.14, *p* = .144				N/A				Q_M = **7.85**, *p* = **.020**		
Mean of 1 ECG epoch		–			1	0.91 [-,-]			2	0.25 [.16, .34]	1.08	48
Mean of ≥2 ECG epochs	9	0.37 [.12, .56]	35.34***	326		–			18	0.46 [.38, .53]	47.97***	811
Unmentioned	5	0.24 [0.09,0.52]	8.34	115	2	0.33 [0.15,0.76]	**8.27****	63	7	0.36 [.19, .54]	14.70*	242
Age-appropriate frequency bands		N/A				N/A				Q_M = 4.45, *p* = .108		
Yes		–				–			17	0.43 [.33, .50]	36.24**	718
No		–				–			8	0.48 [.32, .72]	30.84***	372

Note. *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} = failsafe *N*; SDNN = standard deviation of NN intervals; HF = high frequency; ECG = electrocardiogram

[a] Specific to baseline recording condition. * $p < .05$. ** $p \leq .01$. *** $p \leq .001$.

Table 7

Child/Adolescent HRV Reliability – Categorical Moderator Variables

	SDNN				HF				LF				LF:HF Ratio			
	k	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}
Sample exclusion criteria																
Prescription medications		N/A				$Q_M = 1.82$, $p = .178$				N/A				N/A		
Exclusion reported		–			2	0.55[.52, .87]	1.36	107		–				–		
Unmentioned	4	0.85[.46, 1.02]	6.07	334	13	0.66[.51,.84]	57.97***	850	6	0.61[.41, .81]	12.36*	357	3	0.52[.14, 1.19]	4.59	151
Chronic medical illness		N/A				$Q_M = 2.25$, $p = .134$				N/A				N/A		
Exclusion reported	1	0.79[-,-]			7	0.68[.50, .91]	38.68***	469	1	0.78 [-,-]				–		
Unmentioned	3	0.89[.41, 1.07]	5.69	265	8	0.58[.43, .85]	20.22**	459	5	0.54[.39, .81]	9.58*	265	3	0.52[.14, 1.19]	4.59	151
Mental health condition		N/A				$Q_M = 0.73$, $p = .392$				N/A				N/A		
Exclusion reported		–			6	0.67[.48, .93]	37.80***	395		–				–		
Unmentioned	4	0.85[.46, 1.02]	6.07	334	9	0.62[.46, .83]	22.62**	545	6	0.61[.41, .81]	12.36*	357	3	0.52[.14,1.19]	4.59	151
Time of ECG recording		N/A				$Q_M = 0.51$, $p = .474$				$Q_M = 1.80$, $p = .180$				N/A		
Identical recording time	3	0.89[.44, 1.07]	5.69	265	7	0.68[.54, .94]	13.93*	471	4	0.73[.40, .94]	4.47	287	3	0.52[.14, 1.19]	4.59	151

	SDNN				HF				LF				LF:HF Ratio			
	k	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}
Unmentioned	1	0.79[-,-]			8	0.63[.46, .84]	46.71***	499	2	0.55[.39, .78]	6.08*	107		–		
Pre-ECG acclimatization period		N/A				Q_M = **4.85**, p = **.028**				Q_M = **9.03**, p = **.003**				N/A		
Yes/included	3	0.87[.41, 1.07]	5.17	256	10	0.68[.54, .88]	50.06***	666	4	0.77[.63, .92]	3.32	305	2	0.69[.41, 1.19]	2.73	136
No/unmentioned	1	0.54[-,-]			5	0.49[.38, .79]	6.23	239	2	0.39[.39, .39]	0.00	76	1	0.14[-,-]		
ECG recording posture[a]		N/A				Q_M = **5.78**, p = **.056**				N/A				N/A		
Sitting	1	0.79[-,-]			5	0.72[.49, .97]	33.85***	354	1	0.78[-,-]				–		
Supine	3	0.89[.41, 1.07]	5.69	265	5	0.61[.40, 1.00]	16.67**	300	5	0.54[.39, .81]	9.58*	265	3	0.52[.14, 1.19]	4.59	151
Unmentioned		–			5	0.56[.50, .73]	4.56	274		–				–		
ECG sampling rate		N/A				Q_M = 1.33, p = .249				Q_M = **3.67**, p = **.056**				N/A		
≥ 500 Hz	2	0.73[.41, .79]	1.52	144	9	0.64[.49, .83]	43.32***	570	2	0.72[.41, .78]	1.44	142	1	0.41[-,-]		
Unmentioned	1	0.54[-,-]			5	0.54[.42, .92]	8.81	266	3	0.44[0.39,1.26]	4.97	129	2	0.60[.14, 1.19]	4.34*	117
Manual editing of ECG data		N/A				Q_M = 0.37, p = .544				Q_M = 0.46, p = .496				N/A		
Yes	3	0.87[.41, 1.07]	5.17	256	13	0.65[.52, .81]	58.57***	825	4	0.59[.39, .78]	8.91*	233	1	0.41[-,-]		
No/unmentioned	1	0.54[-,-]			2	0.80[.47, 1.22]	2.21	157	2	0.77[.39, 1.26]	2.98	152	2	0.60[.14, 1.19]	4.34*	117

	SDNN				HF				LF				LF:HF Ratio			
	k	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}	*k*	*Z* [95% CI]	Q_T	N_{FS}
Software-based filter algorithm		N/A				$Q_M = 0.25$, $p = .617$				N/A				N/A		
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	1	0.41[-,-]			7	0.64[.48, .86]	39.78***	439	1	0.41[-,-]			1	0.41[-,-]		
Identical ECG recording duration[a]																
Yes	4	0.84[.46, 1.02]	6.07	334	14	0.64[.51, .81]	60.55***	888	6	0.61[.43, .81]	12.36*	357	3	0.52[.14, 1.19]	4.59	151
No		–			1	0.87[-,-]				–				–		
HRV derivation method		N/A				$Q_M = 3.22$, $p = .200$				N/A				N/A		
Mean of 1 ECG epoch		–			5	0.67[.48, .93]	36.52***	329		–				–		
Mean of ≥2 ECG epochs	1	0.79[-,-]			4	0.74[.40, .83]	3.80	291	1	0.78[-,-]				–		
Unmentioned	3	0.89[.41, 1.07]	5.69	265	6	0.57[.42, .93]	17.62**	336	5	0.54[.39, .85]	9.58*	265	3	0.52[.14, 1.19]	4.59	151
Age-appropriate frequency bands		N/A				$Q_M =$ **16.49**, $p <$ **.001**				N/A				N/A		
Yes		–			6	0.86[.70, .99]	5.18	508	5	0.61[.42, .85]	11.92*	301	2	0.69[.14, 1.19]	2.73	136
No		–			3	0.38[.27, .47]	0.26	111	1	0.39[-,-]			1	0.33[-,-]		

Note. 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} = Failsafe N; SDNN = standard deviation of NN intervals; HF = high frequency; LF = low frequency; ECG = electrocardiogram; HRV = heart rate variability

[a] Specific to baseline recording condition. * $p < .05$. ** $p \leq .01$. *** $p \leq .001$.

Table 8

Sample Demographic, Anthropometric, and Physical Activity Measures

	Visit 1 (N = 632)		Visit 2 (N = 564)	
	M (SD)	Range	*M (SD)*	Range
Age (months)	115.16 (11.07)	96.00–132.72	139.88 (11.23)	119.04–162.84
Male Sex (*n*, %)	344	54.4%	313	49.5%
Tanner puberty status (*n*, %)				
Tanner 1	494[c]	78.2%	186[d]	29.4%
Tanner 2	118	18.7%	181	28.6%
Tanner 3	16	2.5%	135	21.4%
Tanner 4	2	0.3%	48	7.6%
Tanner 5	0	0.0%	10	1.6%
Cardiovascular indices[a]				
Systolic BP (mm/Hg)	93.80 (8.11)[e]	70.33–123.33	97.39 (8.75)	73.33–128.67
Diastolic BP (mm/Hg)	48.56 (5.12)[e]	32.33–67.33	50.40 (5.30)	35.00–69.00
Heart rate (beats per minute)	86.51 (8.82)	61.73–121.95	82.39 (9.31)	54.3–116.73
Growth and adiposity indices				
Height (meters)	138.86 (8.10)	116.50–162.10	151.43 (9.29)	124.50–179.60
Weight (kilograms)	38.16 (11.47)	20.90–89.95	49.05 (14.84)	24.00–105.20
BMI (Z-score)[b]	0.71 (1.08)	-3.16–2.67	0.68 (1.09)	-2.50–2.66
Waist (centemeters)	67.46 (12.13)	49.00–114.00	72.42 (13.26)	50.95–121.00
TV/computer screen time				
Weekday (hours)	2.26 (1.81)[f]	0.00–10.50	2.79 (1.98)[g]	0.00–12.00

	Visit 1 (N = 632)		Visit 2 (N = 564)	
	M (SD)	Range	*M (SD)*	Range
Weekend (hours)	3.95 (2.92)[h]	0.00–15.00	4.96 (3.25)[i]	0.00–21.00
Weekly physical activity (*n*, %)				
None (< 1 times)	102[e]	16.1%	125[j]	19.8%
Some (1 – 2 times)	206	32.6%	117	18.5%
Often (3 – 4 times)	156	24.7%	236	37.3%
Quite often (5 – 6 times)	84	13.3%	47	7.4%
Very often (> 7 times)	83	13.1%	38	6.0%

Note. BP = blood pressure; mm/Hg = millimeters of mercury; min = minute; BMI = body mass index

[a] Recorded in a seated posture. [b] CDC (2000) age- and sex-normed Z-score (2000). c N = 630. d N = 560. e N = 631. f N = 628. g N = 561. h N = 629. i N = 562 j N = 563

Table 9

Descriptive Statistics for Electrocardiogram (ECG) and Heart Rate Variability (HRV) Measures

	Visit 1 (N = 575)		Visit 2 (N = 503)	
	M (SD)	Range	*M (SD)*	Range
ECG recording				
Start time (hrs:min)	8:29 (0:20)	7:26–9:57	8:17 (0:20)	7:10–9:35
Total duration (min)	200.62 (18.60)	50–245	198.63 (19.10)	58–227
Artifacts (% removed)	6.51 (5.79)%	1–49%	6.27 (5.79)%	1–60%
Time-domain HRV				
NN (ms)	700.66 (70.89)	492–972	737.47 (83.35)	514–1105
SDNN (ms)	85.07 (23.32)	39–183	91.09 (25.44)	42–195
rMSSD (ms)	47.90 (15.56)	15–102	50.24 (16.58)	15–100
pNN50 (%)	25.43 (13.70)	0.60–68.90	27.25 (14.27)	0.90–63.70
Frequency-domain HRV				
VLF (ms^2)	1305.56 (700.51)[a]	230.47–5050.99	1654.89 (986.87)	302.67–7094.18
LF (ms^2)	1282.43 (794.90)[b]	234.52–5807.82	1443.19 (996.17)	209.33–7392.63
HF (ms^2)	912.35 (700.43)[a]	63.03–4989.17	920.75 (675.50)	69.25–4307.29
LF:HF Ratio	1.70 (0.70)	0.54–4.70	1.83 (0.80)[c]	0.54–6.07

Note. VLF = very low frequency; LF = low frequency; HF = high frequency; NN = mean NN interval; SDNN = standard deviation of NN intervals; rMSSD = root mean square of successive differences of NN intervals; pNN50 = proportion of successive NN intervals differing by >50 ms; ms = millisecond; ms^2 = millisecond squared

[a] n = 574. [b] n = 571. [c] n = 502

Table 10

HRV Stability: Intra-class (ICC) and Pearson Correlations (r)

	Intra-class Correlation	Pearson Correlation
	ICC [95% CI]	*r* [95% CI]
Time-domain HRV		
NN (ms)	0.74 [0.68, 0.78]	0.67 [0.62, 0.72]
SDNN (ms)	0.82 [0.78, 0.85]	0.72 [0.68, 0.76]
rMSSD (ms)	0.82 [0.78, 0.85]	0.69 [0.64, 0.74]
pNN50 (%)	0.80 [0.76, 0.84]	0.68 [0.63, 0.73]
Frequency-domain HRV		
VLF (ms^2)	0.79 [0.74, 0.82]	0.73 [0.69, 0.77]
LF (ms^2)	0.82 [0.79, 0.85]	0.71 [0.66, 0.75]
HF (ms^2)	0.85 [0.83, 0.88]	0.75 [0.71, 0.79]
LF:HF ratio	0.83 [0.80, 0.86]	0.73 [0.69, 0.77]

Note. All $p < 0.001$. VLF = very low frequency; LF = low frequency; HF = high frequency; NN = mean NN interval; SDNN = standard deviation of NN intervals; rMSSD = root mean square of successive differences of NN intervals; pNN50 = proportion of successive NN intervals differing by >50 ms; ms = millisecond; ms^2 = millisecond squared

Table 11

HRV Stability: Partial Correlations ($r_{ab.c}$), Visit 1 Covariates Only

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	0.67	0.71	0.69	0.68	0.73	0.71	0.75	0.73
Puberty status (Tanner)	0.68	0.72	0.69	0.68	0.73	0.71	0.75	0.73
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	0.68	0.72	0.69	0.67	0.73	0.71	0.75	0.72
Diastolic BP (mm/Hg)	0.65	0.71	0.68	0.66	0.73	0.71	0.75	0.73
Heart rate (beats per minute)	0.24	0.61	0.52	0.48	0.66	0.65	0.69	0.72
Growth and adiposity indices								
Height (meters)	0.67	0.71	0.70	0.68	0.73	0.71	0.75	0.72
Weight (kilograms)	0.69	0.72	0.70	0.69	0.73	0.72	0.75	0.72
BMI (Z-score)[b]	0.69	0.72	0.70	0.68	0.73	0.72	0.75	0.72
Waist (centemeters)	0.69	0.72	0.70	0.68	0.73	0.72	0.75	0.72
Physical activity indices								
Weekly physical activity	0.67	0.71	0.69	0.67	0.73	0.71	0.75	0.73
Weekday screen time (hrs)	0.67	0.72	0.69	0.67	0.73	0.71	0.75	0.72
Weekend screen time (hrs)	0.67	0.72	0.69	0.67	0.74	0.71	0.75	0.72

Note. All $p < 0.001$; BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

Table 12

HRV Stability: Partial Correlations ($r_{ab.c}$), Time-Varying Covariates

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	0.67	0.72	0.70	0.68	0.73	0.71	0.75	0.73
Puberty status (Tanner)	0.68	0.72	0.70	0.68	0.74	0.71	0.75	0.73
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	0.68	0.72	0.70	0.68	0.74	0.72	0.75	0.72
Diastolic BP (mm/Hg)	0.65	0.72	0.68	0.65	0.74	0.71	0.75	0.73
Heart rate (beats per minute)	0.55	0.68	0.75	0.72	0.73	0.72	0.80	0.75
Growth and adiposity indices								
Height (meters)	0.67	0.71	0.70	0.68	0.73	0.71	0.75	0.72
Weight (kilograms)	0.69	0.72	0.70	0.68	0.73	0.72	0.76	0.72
BMI (Z-score)[b]	0.69	0.72	0.70	0.68	0.73	0.72	0.75	0.72
Waist (centemeters)	0.69	0.72	0.71	0.69	0.74	0.72	0.76	0.72
Physical activity indices								
Weekly physical activity	0.68	0.72	0.70	0.68	0.74	0.71	0.75	0.73
Weekday screen time (hrs)	0.67	0.72	0.70	0.67	0.74	0.72	0.75	0.73
Weekend screen time (hrs)	0.67	0.72	0.69	0.67	0.74	0.71	0.75	0.73

Note. Time-varying covariates = visit 1 covariate and residualized change score; All $p < 0.001$; BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

Table 13

HRV Regression: Standardized Beta Coefficients, Visit 1 Covariates Only

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	**.12*****	.03	.05	.05	**.10*****	.03	.00	.05
Puberty status (Tanner)	**.11****	.03	.03	.03	.05	.02	.01	-.04
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	**.07***	.04	.03	.02	.06	**.07***	.01	**.10****
Diastolic BP (mm/Hg)	.02	.01	.00	-.02	.04	**.07***	.02	.06
Heart rate (beats per minute)	**.48***	.03	.06	-.01	.04	**.10***	.07	-.04
Growth and adiposity indices								
Height (meters)	**.17*****	**.08***	**.09***	**.08***	**.16*****	**.09****	.03	.06
Weight (kilograms)	**.19*****	**.10****	**.13*****	**.12*****	**.16*****	**.14*****	**.09****	.02
BMI (Z-score)[b]	**.14*****	**.09****	**.11*****	**.10****	**.10*****	**.14*****	**.10*****	.01
Waist (centemeters)	**.15*****	**.09****	**.12*****	**.11****	**.12*****	**.15*****	**.09****	.00
Physical activity indices								
Weekly physical activity	.03	.02	.01	.00	.02	.00	.00	-.02
Weekday screen time (hrs)	.00	.00	.01	.05	.00	.00	.00	-.04
Weekend screen time (hrs)	.01	.02	.02	.02	**.07***	.02	.01	-.03

Note. BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

* $p < .05$. ** $p \leq .01$. *** $p \leq .001$.

Table 14

HRV Regression: Unstandardized Beta Coefficients, Visit 1 Covariates Only

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	**0.85**	0.07	0.07	0.07	**8.52**	2.06	0.14	0.00
Puberty status (Tanner)	**17.64**	1.29	1.08	0.95	92.92	34.74	10.98	-0.06
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	**0.71**	0.12	0.06	0.03	7.29	**8.05**	1.05	**0.01**
Diastolic BP (mm/Hg)	0.30	0.03	-0.01	-0.05	8.24	**12.58**	3.21	0.01
Heart rate (beats per minute)	**4.54**	0.08	0.11	-0.01	4.90	**10.47**	5.36	0.00
Growth and adiposity indices								
Height (meters)	**1.77**	**0.25**	**0.17**	**0.15**	**18.66**	**10.03**	2.84	0.01
Weight (kilograms)	**1.39**	**0.23**	**0.19**	**0.15**	**13.91**	**12.16**	**5.26**	0.00
BMI (Z-score)[b]	**11.06**	**2.09**	**1.76**	**1.39**	**94.02**	**124.09**	**65.51**	0.01
Waist (centemeters)	**1.06**	**0.18**	**0.16**	**0.13**	**9.90**	**11.47**	**5.24**	0.00
Physical activity indices								
Weekly physical activity	2.17	0.37	0.09	0.06	12.02	3.02	1.59	-0.01
Weekday screen time (hrs)	0.16	-0.04	0.11	0.01	26.79	-0.59	0.70	-0.02
Weekend screen time (hrs)	0.31	0.15	0.11	0.08	**21.45**	4.32	2.23	--0.01

Note. BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

Table 15

HRV Regression: Standardized Beta Coefficients, Time-Varying Covariates

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	-.35	-.62	**-.90***	**-1.01***	-.23	.-.46	**-.87***	.72
Puberty status (Tanner)	**-.38****	-.06	.05	-.09	**-.43****	.06	.11	**-.47*****
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	**-.15*****	**-.14*****	**-.18*****	**-.19*****	**-.09***	**-.11***	**-.14*****	**.11****
Diastolic BP (mm/Hg)	**-.30*****	**-.23*****	**-.29*****	**-.30*****	**-.18*****	**-.18*****	**-.20*****	**.15*****
Heart rate (beats per minute)	**-1.02*****	**-.64*****	**-.84*****	**-.85*****	**-.62*****	**-.60*****	**-.62*****	**.33*****
Growth and adiposity indices								
Height (meters)	.05	-.15	-.25	-..22	.17	-.26	**-.43****	**.57*****
Weight (kilograms)	**-.38****	**-.43*****	**-.37****	**-..36****	**-.32***	**-.40****	**-.42*****	**.28***
BMI (Z-score)[b]	**.14***	.11	.06	.06	.08	.11	.10	.02
Waist (centemeters)	**-.41*****	**-.22***	**-.23***	**-..26***	**-.30****	**-.23****	**-.21***	.13
Physical activity indices								
Weekly physical activity	**.38*****	**.39*****	**.38*****	**.34****	**.39*****	**.38*****	**.33*****	-.09
Weekday screen time (hrs)	**-.10***	-.06	**-.10***	**-.11***	-.02	-.06	**-.09***	**.13****
Weekend screen time (hrs)	-.06	-.02	-.04	-.06	.01	.01	-.01	.06

Note. Time-varying covariates = visit 1 covariate and residualized change score; BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

* $p < .05$. ** $p \leq .01$. *** $p \leq .001$.

Table 16

HRV Regression: Unstandardized Beta Coefficients, Time-Varying Covariates

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	-2.56	-1.40	**-1.32**	**-1.28**	-23.81	-38.28	**-51.62**	0.05
Puberty status (Tanner)	**-53.41**	-2.71	-1.27	0.31	**-687.41**	88.96	122.38	**-.62**
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	**-1.52**	**-0.46**	**-0.38**	**-0.34**	**-10.74**	**-12.38**	**-11.79**	**0.07**
Diastolic BP (mm/Hg)	**-5.61**	**-1.30**	**-1.08**	**-0.96**	**-38.72**	**-38.96**	**-30.58**	**0.03**
Heart rate (beats per minute)	**-9.80**	**-1.89**	**-1.61**	**-1.41**	**-69.37**	**-64.85**	**-47.86**	**0.03**
Growth and adiposity indices								
Height (meters)	0.45	-0.41	-0.46	-0.34	17.37	-26.56	**-31.61**	**0.05**
Weight (kilograms)	**-2.31**	**-0.74**	**-0.42**	**-0.35**	**-21.00**	**-25.76**	**-19.18**	**0.02**
BMI (Z-score)[b]	**10.38**	2.46	0.89	0.77	63.45	92.05	54.72	0.01
Waist (centemeters)	**-2.60**	**-0.43**	**-0.29**	**-0.29**	**-22.28**	**-18.60**	**-10.77**	0.01
Physical activity indices								
Weekly physical activity	**111.42**	**35.05**	**22.27**	**17.12**	**1322.14**	**1236.91**	**782.05**	-0.24
Weekday screen time (hrs)	**-7.02**	-1.44	**-1.45**	**-1.38**	-20.32	-52.29	**-53.91**	**0.09**
Weekend screen time (hrs)	-2.02	-0.16	-0.29	-0.39	5.67	2.62	-3.01	0.02

Note. Time-varying covariates = visit 1 covariate and residualized change score; BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

Table 17

HRV Stability: Intra-class (ICC) and Pearson Correlations (r) with Imputed Data

	ICC [95% CI]	*r* [95% CI]
Time-domain HRV		
NN (ms)	0.59 [0.52, 0.65]	0.54 [0.48, 0.59]
SDNN (ms)	0.78 [0.74, 0.81]	0.64 [0.59, 0.69]
rMSSD (ms)	0.72 [0.68, 0.76]	0.59 [0.54, 0.64]
pNN50 (%)	0.72 [0.68, 0.76]	0.58 [0.52, 0.63]
Frequency-domain HRV		
VLF (ms2)	0.76 [0.72, 0.79]	0.67 [0.62, 0.71]
LF (ms2)	0.82 [0.79, 0.85]	0.68 [0.64, 0.72]
HF (ms2)	0.82 [0.79, 0.85]	0.68 [0.64, 0.72]
LF:HF ratio	0.80 [0.76, 0.83]	0.62 [0.57, 0.67]

Note. All $p < 0.001$. VLF = very low frequency; LF = low frequency; HF = high frequency; NN = mean NN interval; SDNN = standard deviation of NN intervals; rMSSD = root mean square of successive differences of NN intervals; pNN50 = proportion of successive NN intervals differing by >50 ms; ms = millisecond; ms^2 = millisecond squared

Table 18

HRV Stability: Partial Correlations ($r_{ab.c}$), Visit 1 Covariates Only with Imputed Data

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	0.52	0.64	0.59	0.58	0.66	0.68	0.68	0.61
Puberty status (Tanner)	0.54	0.64	0.59	0.58	0.67	0.68	0.67	0.62
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	0.54	0.64	0.59	0.57	0.67	0.68	0.67	0.60
Diastolic BP (mm/Hg)	0.51	0.64	0.58	0.56	0.67	0.68	0.67	0.61
Heart rate (beats per minute)	0.14	0.56	0.44	0.41	0.62	0.64	0.63	0.61
Growth and adiposity indices								
Height (meters)	0.52	0.63	0.59	0.58	0.66	0.68	0.67	0.61
Weight (kilograms)	0.54	0.64	0.60	0.58	0.66	0.68	0.68	0.60
BMI (Z-score)[b]	0.54	0.64	0.60	0.58	0.66	0.68	0.68	0.61
Waist (centemeters)	0.54	0.64	0.60	0.58	0.66	0.68	0.68	0.60
Physical activity indices								
Weekly physical activity	0.53	0.64	0.59	0.58	0.66	0.67	0.67	0.62
Weekday screen time (hrs)	0.54	0.64	0.59	0.58	0.67	0.68	0.67	0.62
Weekend screen time (hrs)	0.54	0.64	0.59	0.58	0.67	0.68	0.67	0.62

Note. All $p < 0.001$; BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

Table 19

HRV Stability: Partial Correlations ($r_{ab.c}$), Time-Varying Covariates with Imputed Data

	Time-domain				Frequency-domain			
Time-varying covariates	NN	SDNN	rMSSD	pNN50	VLF	LF	HF	LF:HF
Age (months)	0.52	0.64	0.59	0.58	0.66	0.68	0.68	0.61
Puberty status (Tanner)	0.61	0.68	0.63	0.61	0.72	0.72	0.71	0.70
Cardiovascular indices[a]								
Systolic BP (mm/Hg)	0.55	0.64	0.59	0.58	0.67	0.68	0.68	0.61
Diastolic BP (mm/Hg)	0.50	0.64	0.57	0.55	0.67	0.68	0.67	0.61
Heart rate (beats per minute)	0.21	0.61	0.63	0.60	0.65	0.69	0.72	0.63
Growth and adiposity indices								
Height (meters)	0.52	0.63	0.59	0.58	0.66	0.68	0.68	0.61
Weight (kilograms)	0.54	0.64	0.60	0.58	0.66	0.68	0.68	0.61
BMI (Z-score)[b]	0.54	0.64	0.60	0.58	0.66	0.68	0.68	0.61
Waist (centemeters)	0.54	0.64	0.60	0.58	0.67	0.68	0.68	0.60
Physical activity indices								
Weekly physical activity	0.60	0.68	0.63	0.61	0.72	0.72	0.71	0.70
Weekday screen time (hrs)	0.53	0.64	0.59	0.58	0.67	0.68	0.68	0.62
Weekend screen time (hrs)	0.54	0.64	0.59	0.58	0.67	0.68	0.67	0.62

Note. Time-varying covariates = visit 1 covariate and residualized change score; All $p < 0.001$; BP = blood pressure; mm/Hg = millimeters of mercury; BMI = body mass index; VLF = very low frequency (ms^2); LF = low frequency (ms^2); HF = high frequency (ms^2); NN = mean NN interval (ms); SDNN = standard deviation of NN intervals (ms); rMSSD = root mean square of successive differences of NN intervals (ms); pNN50 = proportion of successive NN intervals differing by >50 ms (%); ms = millisecond; ms^2 = millisecond squared

[a] = Recorded in a seated posture. [b] = CDC (2000) age- and sex-normed Z-score (2000)

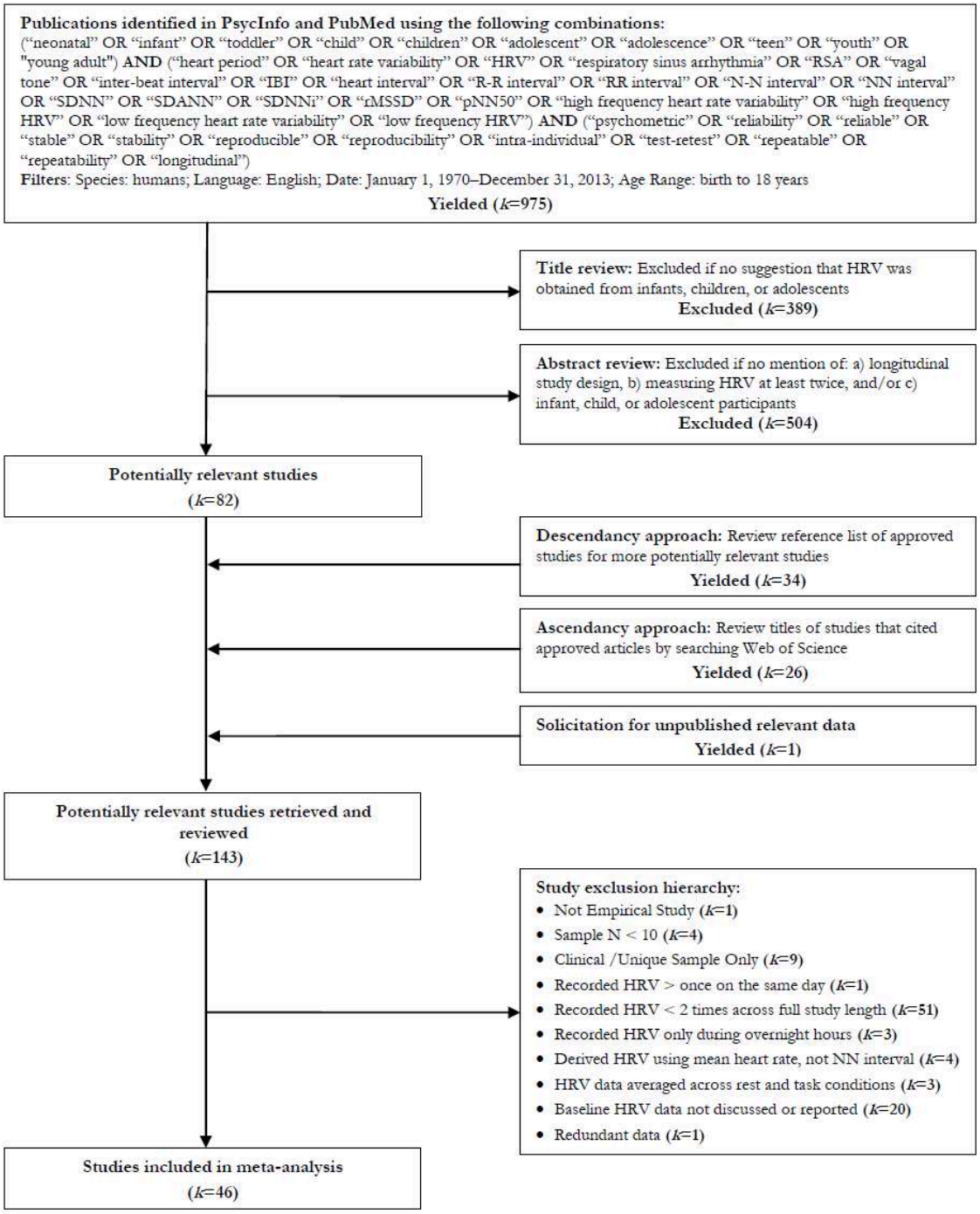

Figure 1: Flow Chart for Article Identification and Inclusion in Meta-Analysis

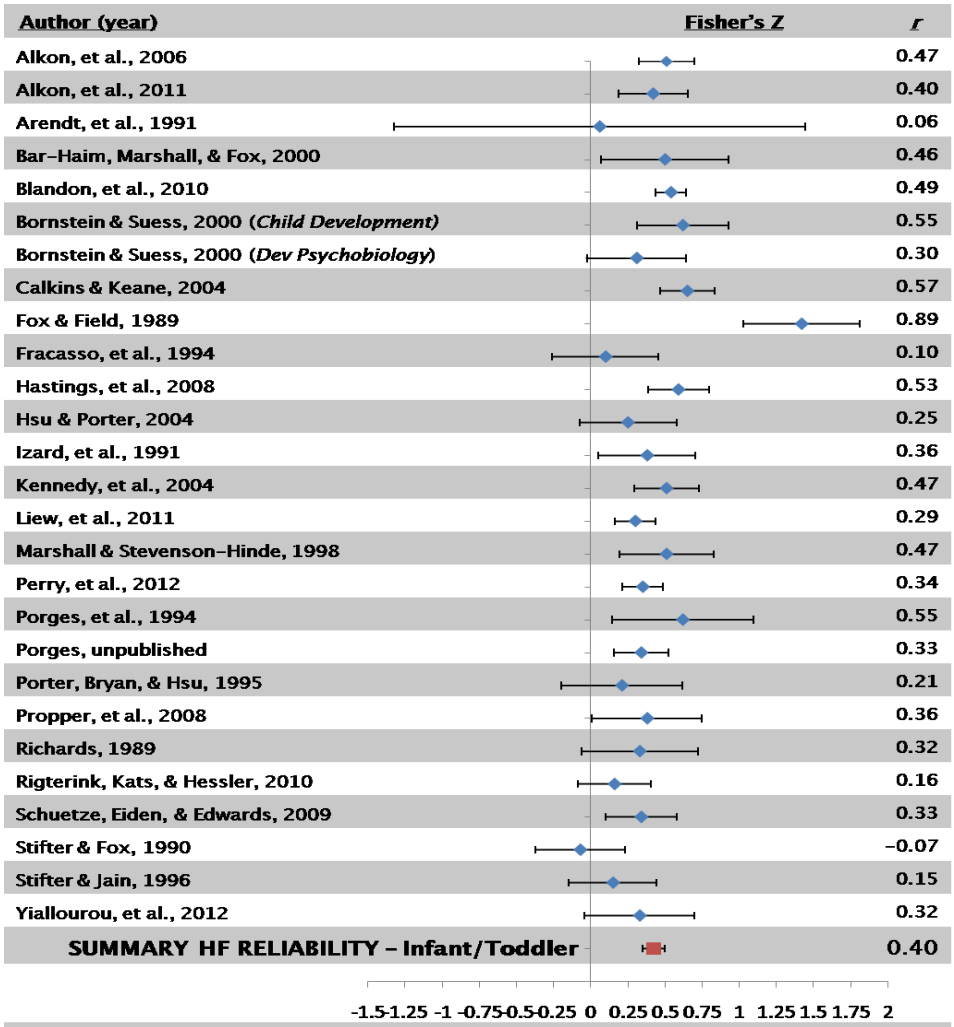

Figure 2: Forest Plot for HF Reliability (Fisher's $Z \pm$ Standard Errorn) and r

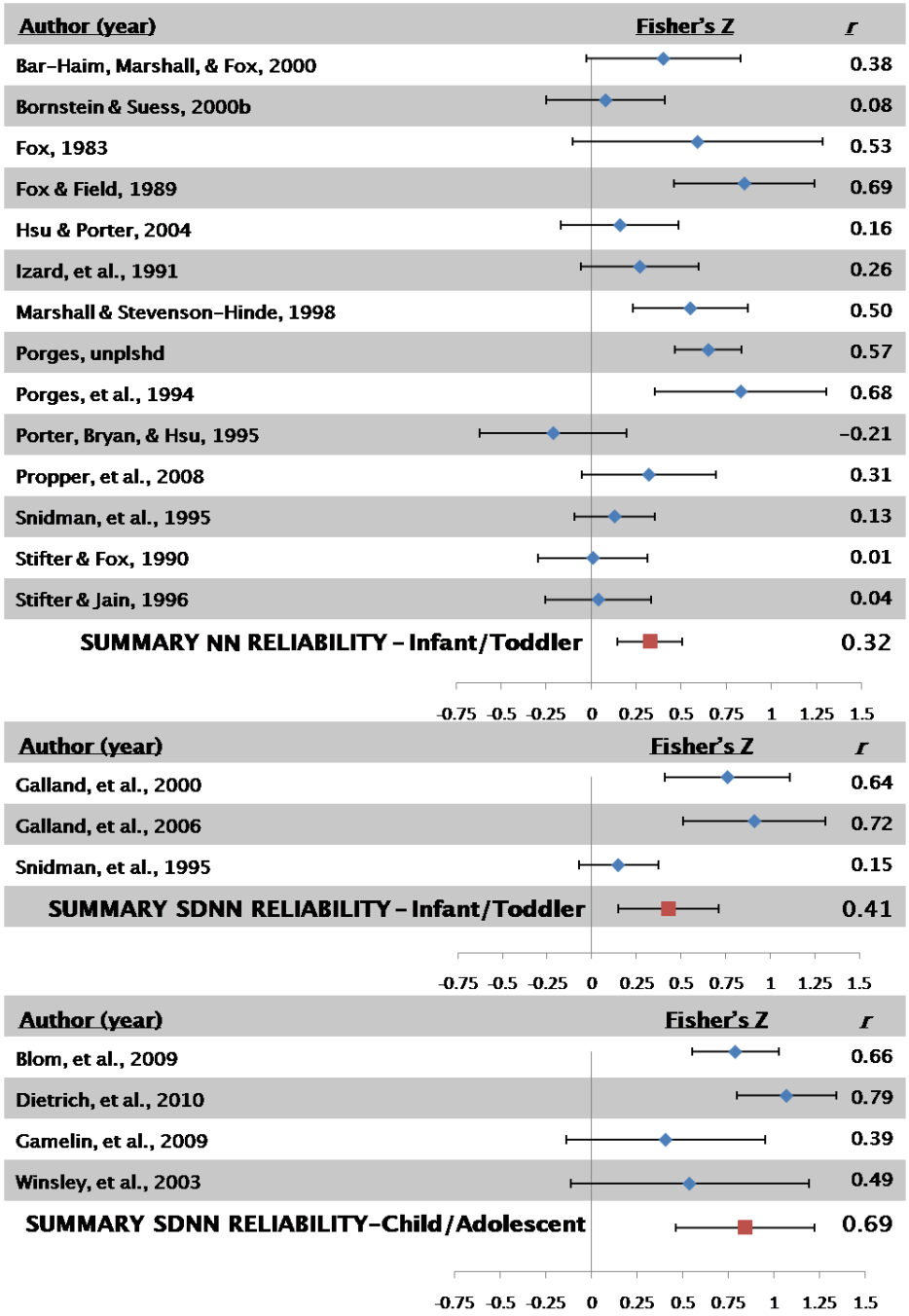

Figure 3: Forest Plot for NN and SDNN Reliability (Fisher's $Z \pm$ Standard Error) and r

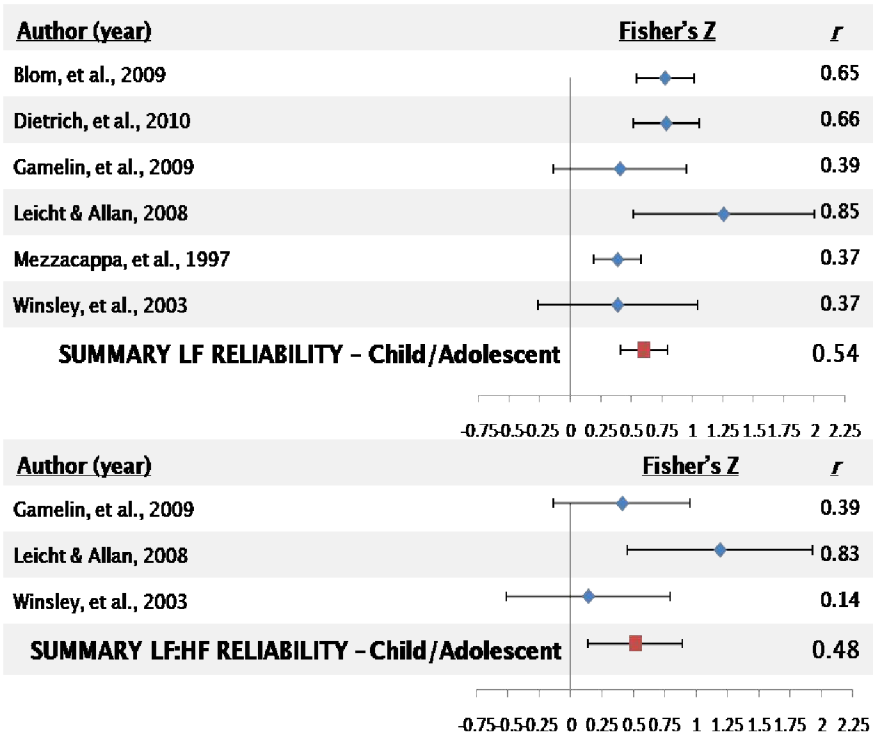

Figure 4: Forest Plot for LF and LF:HF Ratio Reliability (Fisher's $Z \pm$ Standard Error) and r

Appendix A

To determine whether a physiological variable is being accurately measured and interpreted, evidence must exist to support its validity (i.e., the accuracy with which a variable is measured), reliability (i.e., the repeatability of data within and between study contexts), and stability (i.e., the within-person consistency of a characteristic or attribute across time).

Validity

Validity represents both the extent to which a variable (e.g., heart rate) accurately quantifies what it intends to quantify (e.g., ventricular contractions), and the agreement between independent methodological procedures (e.g., stethoscope, ambulatory monitor; Campbell & Fisk, 1959; Campbell, 1960; Cohen & Swerdlik, 2010). Greater measurement validity in specific, controlled contexts (e.g., standardized protocols, homogenous samples) permits the researcher to make assertions and inferences related to the comparability and generalizability to other contexts (e.g., uncontrolled settings, heterogeneous samples). Evidence for validity is obtained when data (a) represent an adequate sample of the specific variable of interest (i.e., content validity); (b) correlate positively with data obtained *concurrently* using measures of an equivalent or related variable, or with data obtained in the future postulated as *predictable* from the initially-obtained data (i.e., criterion-related validity); and (c) behave in a way consistent with theory (i.e., correlates positively, or *converges* with established/theoretically related measures, and correlates negatively, or can be *discriminated* from established/theoretically unrelated measures). A single psychological process (e.g., anxiety) can affect several physiological variables (e.g., heart rate, respiration, blood

pressure; Tomarken, 1995); thus, specificity of a variable is paramount for content validity. Further, evidence for the appropriateness of conclusions about physiological data related to an underlying construct is an ongoing and iterative process (i.e., construct validity; Cronbach & Meehl, 1955; Smith, 2005). Evidence for construct validity is especially important in psychophysiological research given the oftentimes unavailability of suitable criterion measures with which a variable used to represent a construct can be compared (Cohen & Swerdlik, 2010). A useful way to demonstrate which variable is the most valid representation of an underlying construct is by comparing theory-driven expectations about the strength and direction of associations across both related and unrelated variables using a multitrait-multimethod matrix (Campbell & Fisk, 1959; Smith, 2005). Together, this evidence allows one to infer that a measurement is valid.

Reliability

Reliability describes the extent to which data can be measured and interpreted consistently across time and methodological contexts (Campbell & Fisk, 1959; Campbell, 1960; Cohen & Swerdlik, 2010; Kaplan & Saccuzzo, 1997). Classical test theory states that data are comprised of a "true score" and "measurement error", with greater measurement error reflecting poorer reliability (Nunnally & Bernstein, 1994). Multiple sources of measurement error can reduce a variable's reliability. It is pertinent to distinguish true score variance from random or transient fluctuations, both within- and between-assessments, by examining how different sources of error variance affect a variable's reliability (Polit, 2014; Tomarken, 1995). Evidence for reliability is obtained when data a) demonstrate minimal variation across content within a given assessment (i.e., internal consistency); b) can be replicated across time and methodologically-

equivalent contexts (i.e., test-retest reliability); c) are consistent when measured using two or more equivalent procedures (i.e., parallel-forms reliability); and, d) are measured and interpreted consistently by different individuals, such that error variance due to researcher subjectivity is minimized (i.e., inter-rater reliability; Nunnally & Bernstein, 1994; Tomarken, 1995). Generalizability theory is one analytical tool to statistically determine which source of error variance has the biggest impact on a variable's reliability when considered separately or in comparison to other sources of error (e.g., Di Nocera, Ferlazzo, & Borghi, 2001). Together, this evidence allows one to infer that the measurement of a variable is reliable.

Stability

Stability refers to the consistency of a variable, and is directly concerned with individual differences and within-person factors that contribute to observed changes in a characteristic or attribute across time (Tomarken, 1995). Stability reflects the extent of "true variation" in a variable or underlying construct across time. Once reliability has been established, observed changes in a variable obtained between test-retest intervals can be best interpreted in the context of stability (Polit, 2014; Tomarken, 1995). Evidence for stability is obtained when data a) demonstrate consistency in the absolute quantitative value of a variable measured on multiple occasions, and b) allow for inferences regarding the developmental trajectory of an underlying construct, as represented by a measured variable. It is pertinent to consider both time-invariant (i.e., static, stable; sex, ethnicity) and time-variant (i.e., dynamic, temporal; mood, growth) within-person factors when examining stability (Anastasi, 1953, Fracasso, Porges, Lamb, & Rosenberg, 1994). Together, this evidence allows one to infer that a variable is stable.

References

Anastasi, A. (1953). Individual differences. *Annual Review of Psychology*, 4, 137-156.

Cohen, R. J. & Swerdlik M. E. (2010) *Psychological Testing and Assessment: An Introduction to Tests and Measurements (*7th ed.*).* New York, NY: McGraw-Hill.

Cronbach, L.J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

Campbell, D.T., & Fiske, D. W. (1959).Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or disciminant validity. *American Psychologist*, 15, 546-553.

Fracasso, M. P., Porges, S. W., Lamb, M. E., & Rosenberg, A. A. (1994). Cardiac activity in infancy: Reliability and stability of individual differences. *Infant Behaviour and Development*, 17, 277-284.

Kaplan, R. M. & Saccuzzo, D. P. (1997) *Psychological Testing: Principles, Applications, and issues* (4th ed.). California: Brooks/Cole Publishing Co.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw Hill, Inc.

Polit, D. F. (2014). Getting serious about test-retest reliability: A critique of retest research and some recommendations. *Quality of Life Research*, 23, 1713-1720.

Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment*, 17(*4*), 396-408.

Tomarken, A. J. (1995). A psychometric perspective on psychophysiological measures. *Psychological Assessment*, 7(*3*), 387-395.