Clinical Psychology Review 33 (2013) 795–811

Contents lists available at SciVerse ScienceDirect

Clinical Psychology Review

Reaction time variability in ADHD: A meta-analytic review of 319 studies

Michael J. Kofler [a,*], Mark D. Rapport [b], Dustin E. Sarver [b], Joseph S. Raiker [b], Sarah A. Orban [b], Lauren M. Friedman [b], Ellen G. Kolomeyer [b]

[a] Department of Human Services, University of Virginia, Charlottesville, VA, USA
[b] Department of Psychology, University of Central Florida, Orlando, FL, USA

HIGHLIGHTS

- Children & adults with ADHD exhibit increased RT variability (RTV) relative to nonclinical groups.
- This RTV was attenuated by stimulants, but unaffected by psychosocial & other medical treatments.
- Individuals with ADHD did not evince slower processing speed (mean RT) after accounting for RTV.
- Comparison with clinical control groups reveals that RT variability is not specific to ADHD.
- RTV is a stable feature of ADHD & other clinical disorders observed across diverse tasks and methods.

ARTICLE INFO

Article history:
Received 5 November 2012
Revised 31 May 2013
Accepted 3 June 2013
Available online 13 June 2013

Keywords:
ADHD
Variability
Reaction time
Meta-analysis

ABSTRACT

Individuals with ADHD are characterized as ubiquitously slower and more variable than their unaffected peers, and increased reaction time (RT) variability is considered by many to reflect an etiologically important characteristic of ADHD. The present review critically evaluates these claims through meta-analysis of 319 studies of RT variability in children, adolescents, and adults with ADHD relative to typically developing (TD) groups, clinical control groups, and themselves (subtype comparisons, treatment and motivation effects). Random effects models corrected for measurement unreliability and publication bias revealed that children/adolescents (Hedges' $g = 0.76$) and adults ($g = 0.46$) with ADHD demonstrated greater RT variability relative to TD groups. This increased variability was attenuated by psychostimulant treatment ($g = -0.74$), but unaffected by non-stimulant medical and psychosocial interventions. Individuals with ADHD did not evince slower processing speed (mean RT) after accounting for RT variability, whereas large magnitude RT variability deficits remained after accounting for mean RT. Adolescents and adults with ADHD were indistinguishable from clinical control groups, and children with ADHD were only minimally more variable than clinical control children ($g = 0.25$). Collectively, results of the meta-analysis indicate that RT variability reflects a stable feature of ADHD and other clinical disorders that is robust to systematic differences across studies.

Contents

* Corresponding author at: Department of Human Services, University of Virginia, 417 Emmet Street, PO Box 400267, Charlottesville, VA 22904–4267, USA. Tel.: +1 434 243 8716.
E-mail address: kofler@virginia.edu (M.J. Kofler).

0272-7358/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cpr.2013.06.001

1. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a complex, chronic, and potentially debilitating disorder of brain, behavior, and development that affects approximately 2.8 to 3.9 million U.S. school children at an annual cost of illness of over $14,000 per child (APA, 2000; Pelham, Foster, & Robb, 2007; U.S. Census Bureau, 2010). Longitudinal studies reveal that functionally impairing ADHD symptoms continue into adolescence and adulthood for most individuals, and are associated with a host of adverse outcomes. These include scholastic underachievement and school failure, increased high school/college dropout rates, earlier/riskier sexual activity, dysfunctional interpersonal relationships, negative driving-related outcomes, lower overall socioeconomic status, poor work histories, and less secure employment (Barkley, Fischer, Smallish, & Fletcher, 2006; Mannuzza, Klein, Bessler, Malloy, & LaPadula, 1993).

Anecdotal and controlled observations suggest that individuals with ADHD are *consistently inconsistent* (Rapport, Chung, Shore, & Isaacs, 2001) both behaviorally and in their performance on neurocognitive tests (Klein, Wendling, Huettner, Ruder, & Peper, 2006; Kofler, Rapport, & Alderson, 2008; Russell et al., 2006; Willcutt, Sonuga-Barke, Nigg, & Sergeant, 2008). Although long treated as a nuisance variable in the search for underlying neurocognitive deficits in ADHD (for an exception, see Cohen & Douglas, 1972), *intraindividual variability* is currently considered by many to be a core and stable feature of the disorder, and referred to frequently as a ubiquitous and etiologically important characteristic of ADHD (Table 1). Intraindividual variability refers to moment-to-moment (within-subject) fluctuations in behavior and task performance occurring over a period of seconds or milliseconds rather than hours or days (Castellanos et al., 2005; Russell et al., 2006; Tamm et al., 2012). It is distinguished from systematic changes in behavior or performance related to practice, learning, development, treatment, or variations in the clinical condition (Buzy, Medoff, & Schweitzer, 2009; Russell et al., 2006). Neurological correlates of intraindividual variability include regions implicated consistently in ADHD, including dorsolateral prefrontal, orbital frontal, and anterior cingulate cortices (Bellgrove, Hester, & Garavan, 2004; MacDonald, Nyberg, & Backman, 2006). In addition, dysfunctional dopaminergic and noradrenergic neurotransmission, as well as reduced myelination and inadequate lactate transport, have been hypothesized as neurobiological mechanisms responsible for increased intraindividual variability in ADHD (Biederman & Spencer, 1999; Castellanos et al., 2005; Russell et al., 2006).

Intraindividual (i.e., within-subject) variability is indexed conventionally by reaction time (RT) dispersion during laboratory tasks. RT variability refers to inconsistency in an individual's speed of responding, measured in seconds or milliseconds, and has been argued to reflect a subset of abnormally slow responses during laboratory tasks (Klein et

Table 1
Description of attention-deficit/hyperactivity disorder (ADHD) etiological models with predictions regarding reaction time variability.

Model	Model description of ADHD	Role of RT Variability	Model account of variability	Representative publications
Attentional Lapse Models	Models vary from DSM-IV Clinical Model (core attention deficit in ADHD) to attention deficits attributable to alternate processes/mechanisms (see models below)	Outcome	RTV viewed as a neuropsychological indicator of lapses of attention	Leth-Steensen et al. (2000)
Behavioral Inhibition Model	A core deficit model wherein deficits in BI (stopping pre-potent/ongoing responses and interference control) result in four areas of executive dysfunction that collectively result in ADHD behavioral symptoms	Outcome	Attributable to the direct effects of BI dysfunction and indirect effects of BI deficits through executive dysfunction resulting in sustained attention/vigilance deficits	Barkley (1997)
Cognitive Neuroenergetic Model	Decreased ATP production and inadequate lactate supply from deficient astrocyte functioning causes the behavioral features of inefficient and inconsistent performance in individuals with ADHD	Causal	Attentional lapse model; RTV arises from state regulation deficits attributable to impairments in rapid, sustained neural firing	Russell et al. (2006); Sergeant (2005)
Default Mode Network Model	A multiple pathway model that hypothesizes that disruptions in cortico-striato-thalamo-cortical neuroanatomical circuitry—consisting of 'hot' and 'cool' regions—contribute to functional behavioral and cognitive differences in ADHD	Causal	Predictable oscillations in default mode (i.e., resting state) neural networks interfere with task-oriented neural processing, producing periodic lapses of attention	Castellanos et al. (2005); Castellanos & Tannock (2002); Sonuga-Barke and Castellanos (2007)
Dynamic Developmental Model	A core deficit model that hypothesizes that reduced dopaminergic functioning causes narrower reinforcement gradients and altered extinction processes in normal behavior–consequence relationships. These deficient dual processes contribute to core ADHD symptoms and behavioral variability, which vary based on context, task, and function	Correlate	Reflects a pattern of inconsistent behavior-response associations affected by deficient reinforcement/extinction mechanisms, which in turn, disrupt the accumulation of simple behavioral response units into more complex and functional response chains	Sagvolden et al. (2005)
Variability Trait Model	Childhood Hyperactivity attributed to excessive variability, both in rate and magnitude of change, in arousal level and reactivity; excessively inconsistent arousal and reactivity result in problems in sustained attention, performance, and social behavior	Causal	Excessive variability in autonomic, electrocortical and behavioral response underlies impairments in attention, performance and social behavior	Hicks et al. (1989)
Subcortical Deficit Model	A developmental model that hypothesizes that ADHD is caused by subcortical neural dysfunction that manifests early in ontogeny, remains relatively static throughout life, and is not associated with the remission of symptomatology	Causal/Core	Reflects unconsciously (i.e., non-prefrontally) mediated deficits in arousal and activation similar to those described by the Cognitive Energetic Model	Halperin & Schulz (2006); Halperin et al. (2008)
Tripartite Pathway Model	A multiple pathway/equifinality model in which ADHD symptoms are caused by deficits in one or more dissociable cognitive (behavioral inhibition, temporal processing) and/or motivational (delay aversion) processes	Outcome	RTV attributable to temporal processing deficits	Sonuga-Barke et al. (2010)
Working Memory Model	A core deficit model that views inattention, hyperactivity, and impulsivity as phenotypic/behavioral expressions of the interaction between neurobiological vulnerability and environmental demands that overwhelm these children's impaired working memory. Associated features of ADHD arise through direct effects of impaired WM, or indirect effects of impaired WM through its impact on core behavioral symptoms	Outcome	Attributable to the direct effects of CE dysfunction and indirect effects of CE deficits through CE's impact on increased motor activity, mind wandering, visual inattention and impulsive responding that temporarily disrupt task performance	Rapport et al. (2001/2008)

Note. ATP = adenosine triphosphate; BI = behavioral inhibition; CE = central executive; RTV = reaction time variability; WM = working memory.

al., 2006; Leth-Steensen, Elbaz, & Douglas, 2000; Russell et al., 2006; Schmiedek, Oberauer, Wilhelm, Süb, & Wittmann, 2007; Tamm et al., 2012). A large body of research indicates that individuals with ADHD exhibit increased RT variability across a wide range of tasks, including tasks measuring reaction time on motor speed, choice decision, vigilance, behavioral inhibition, cognitive interference, working memory, visual saccade, and visual discrimination (e.g., Alderson, Rapport, & Kofler, 2007; Buzy et al., 2009; Klein et al., 2006; Willcutt et al., 2008). In addition, RT variability has been proposed as an underlying trait (Hicks, Mayo, & Clayton, 1989; Russell et al., 2006) or potential endophenotype of the disorder (Castellanos et al., 2005; Sonuga-Barke & Castellanos, 2007).

The heightened interest in RT variability as a potential core and etiologically important feature of ADHD is reflected by the growing number of narrative summary and meta-analytic reviews on the topic within the past 8 years. These reviews focus primarily on examining two issues—the extent to which children and adults with ADHD differ from typically developing control groups in RT variability, and whether the magnitude of between-group differences is greater for this measure relative to traditional test metrics such as mean reaction time. An initial narrative review of 42 child and adult studies indicated that ADHD and control groups differed significantly in intraindividual variability in the vast majority of published studies, and that these differences were of greater magnitude relative to conventional metrics such as mean reaction time, stop signal reaction time, and errors (Klein et al., 2006). These findings, however, were based on comparing the significance levels reported between and among studies as opposed to quantifying the different outcome measures in comparable metrics (effect sizes) while controlling for study power (cf. Howard, Maxwell, & Fleming, 2000).

An initial meta-analytic review of 39 RT variability studies of children with ADHD updated the Klein et al. (2006) review, and indicated moderate-to-large magnitude ($d = 0.71$) ADHD-related intraindividual variability relative to typically developing children (Willcutt et al., 2008). Four additional meta-analytic reviews examined RT variability on specific tasks or with specific subgroups. Hervey, Epstein, and Curry (2004) examined seven adult studies reporting RT variability during continuous performance tasks, and three additional meta-analyses focused exclusively on the stop signal task: Alderson et al. (2007) reviewed 12 child studies; Lijffijt, Kenemans, Verbaten, and van Engeland (2005) examined 29 child and adult studies; and Lipszyc and Schachar (2010) evaluated 38 child and adult studies reporting RT variability during stop-signal tasks. Collectively, these four meta-analytic reviews reported moderate RT variability effect sizes of 0.61, 0.73, 0.65, and 0.71, respectively; however, their interpretations and conclusions may be premature due to the limited variety of tasks and/or age groups and resultant small percentage of available studies included in the reviews (i.e., between 3% and 12% of the 319 studies included in the current meta-analytic review). In addition, none of the studies examined differences among ADHD subtypes or corrected for measurement unreliability (Hunter & Schmidt, 2004), and only Lipszyc and Schachar (2010) corrected for publication bias.

The current meta-analytic review includes 319 studies of RT variability in children, adolescents, and adults with ADHD relative to typically developing (TD) and clinical control groups across a wide range of laboratory tasks (Tables S1–S8), and corrects for measurement unreliability and publication bias (Hunter & Schmidt, 2004; Lipsey & Wilson, 2001) to address the primary limitations of previous meta-analytic reviews. A series of fundamental questions concerning the measurement and specificity of RT variability, and the degree to which it is modifiable by pharmacological or motivational interventions are also addressed in the review.

1.1. Specificity of RT variability

Three fundamental issues regarding specificity are addressed in our review to determine the extent to which RT variability (a) is unique to one or more of the three ADHD subtypes, (b) differs between individuals with ADHD relative to typically developing groups, and (c) represents a performance pattern characteristic of other clinical disorders rather than being unique to ADHD. The specificity of RT variability for the three ADHD subtypes (Inattentive, Hyperactive/Impulsive, and Combined subtypes) is addressed initially to determine whether its occurrence is limited to one or more of the three ADHD subtypes. The results of these analyses also have potential implications for interpreting studies that collapse data across ADHD subtypes when comparing them with typically developing and clinical control groups (e.g., between-group effect size differences may be deflated if RT variability is limited to a single ADHD subtype).

Studies comparing ADHD to typically developing and other clinical disorder groups are analyzed subsequently to address issues related to the potential specificity of RT variability (Zakzanis, 2001). The specificity of RT variability to ADHD is key, given the myriad disorders that feature clinically impairing attention problems and/or impulsive behavior (cf. Youngstrom, Arnold, & Frazier, 2010). An initial, selective meta-analytic review indicated that RT variability was not specific to ADHD, but rather a common feature of many different childhood disorders (Willcutt et al., 2008). This review, however, included only a small subset of currently available studies (12%), and did not directly compare ADHD to clinical control groups. Recent studies comparing children with ADHD to children with other clinical disorders are equivocal, with studies reporting decreased (Geurts et al., 2008), similar (Oosterlaan & Sergeant, 1995), or increased (O'Brien et al., 1992) RT variability in children with ADHD. Substantive differences exist across studies (e.g., comparison groups, diagnostic methods), however, and these differences must be investigated systematically via meta-analysis to determine the specificity of increased RT variability in ADHD.

1.2. Estimating RT variability

Early studies relied primarily on global measures of RT variability such as standard deviation (SD) and standard error (SE), and these metrics remain the most common indices of RT variability despite their well-documented conceptual and statistical limitations (Castellanos et al., 2005; Geurts et al., 2008; Leth-Steensen et al., 2000; Schmiedek et al., 2007). For example, global measures of RT dispersion such as SD correlate highly with measures of overall mean reaction time (MRT; $r = .92$; Wagenmakers & Brown, 2007). In ADHD studies, correlations between MRT and SD of RT (SDRT) range from .70 to .90 (Epstein et al., 2003; Spencer et al., 2009), and medication-related improvements in these metrics are also highly correlated ($r = .80$; Spencer et al., 2009). These correlations imply multicollinearity in measurement (Tabachnick & Fidell, 2007), and question the extent to which the MRT and SD variables are measures of the same underlying construct (Klein et al., 2006).

Several authors have calculated dispersion around mode RT, or computed the coefficient of variability (CV) to address this issue. CV is a metric that reflects global variability after accounting for overall reaction time (CV = SDRT/MRT). The resultant metric is uncorrelated with mean reaction time, and has been argued to be a more appropriate index of RT variability (Bellgrove, Hawi, Kirley, Gill, & Robertson, 2005). The CV metric (like SD and SE) rests on the assumption that reaction times are normally distributed, however, which does not appear to be the case (Castellanos, Sonuga-Barke, Milham, & Tannock, 2006; Schmiedek et al., 2007). Specifically, reaction time distributions tend to be positively skewed, especially for individuals with ADHD, as a result of a subset of abnormally slow responses (Castellanos et al., 2006; Schmiedek et al., 2007). In addition, critics of this approach argue that it removes variance attributable to SDRT from SDRT (Klein et al., 2006). To address this issue, some researchers have used ex-Gaussian or spectral power-based/signal processing methods, which attempt to dissect global variability based on disparate assumptions.

Ex-Gaussian methods are preferable when RT variability results from a randomly occurring subset of abnormally slow responses. This method separates each child's RT distribution into exponential ("Ex-") and normal (Gaussian) components. Within the normal component, this approach provides estimates of mu (μ) and sigma (σ). Mu reflects the mean reaction time of the normal distribution, and sigma reflects the variability of this normal component. If a child's RTs are normally distributed, then mu and sigma will equal MRT and SDRT, respectively, and tau (τ) will equal zero. Tau reflects the exponential component of the RT distribution, and reflects the subset of extremely slow responses that otherwise have a strong influence on MRT and SDRT calculation. Tau is similar conceptually to a distribution's skewness, but is considered a more reliable metric (Schmiedek et al., 2007).

In contrast, signal processing methods are preferred when RT variability results from nonrandom, periodic fluctuations whose rhythmicity will be reflected in increased power to specific spectral bands. Conceptually, signal processing methods examine the temporal pattern of a RT series to determine if the abnormally long RTs identified by other methods occur in a predictable, temporal sequence (Castellanos et al., 2005; Geurts et al., 2008). Recent studies, however, have questioned the veracity of conclusions based on power in specific frequency bands. For example, task characteristics (e.g., interstimulus intervals, stimulus predictability) and the choice of task appear to influence the peak frequency band that will be obtained (Johnson et al., 2007). In addition, Geurts et al. (2008) demonstrated that small changes in data preparation, method of spectral estimation, and subset of trials analyzed can significantly impact the peak frequency band obtained.

In summary, although many studies have utilized multiple RT variability indices, the relative sensitivity of these correlated but

theoretically dissimilar methods for detecting ADHD-related variability remains unknown. The current meta-analysis addresses this issue by examining the extent to which the metric used to estimate RT variability moderates the magnitude of between-group effect sizes. This analysis will allow us to determine whether specific metrics are better able to differentiate individuals with ADHD from other groups, or whether the additional complexity of ex-Gaussian and spectral-based analyses fails to result in larger effect sizes (Spencer et al., 2009).

1.3. Intraindividual variability: random or periodic phenomenon?

Researchers disagree regarding whether ADHD-related RT variability occurs randomly, or coincides temporally with periodic fluctuations in underlying physiological or neuronal processes (Castellanos et al., 2005; Geurts et al., 2008). The answer to this question has potentially important implications for ADHD treatment and etiological models. For example, discovering that RT variability results from predictable fluctuations in performance over time and is unique to ADHD could inform the development of novel interventions and important targets for outcome assessment. Periodic performance fluctuations would have direct implications for etiological models of ADHD to the extent that these fluctuations could be linked with coinstantaneous physiological and neuronal processes. Although this link remains hypothetical in humans, preliminary evidence of neuronal and behavioral synchronization in rats makes this a compelling possibility (Castellanos et al., 2005).

The current meta-analysis synthesizes extant studies employing spectral power-based analyses of RT data and compares findings for specific power spectrums relative to global spectral power (power across all measured bands) across studies. Convergent findings of increased RT variability in specific spectral bands would provide compelling support for etiological models such as the default mode network model (Sonuga-Barke & Castellanos, 2007). In contrast, equivocal results across studies and peak frequency band effect sizes approximating effect size magnitude of other RT variability indices would contradict hypotheses regarding predictable, periodic fluctuations in performance (Geurts et al., 2008).

1.4. Intraindividual variability in ADHD: ubiquitous phenomenon or dependent on task, context, and/or state?

If intraindividual variability is a ubiquitous characteristic of ADHD (Klein et al., 2006; Russell et al., 2006) it should be observable across tasks, contexts, and states. Initial reviews suggest that children with ADHD demonstrate increased intraindividual variability across a wide range of tasks (Lipszyc & Schachar, 2010; Tamm et al., 2012; Willcutt et al., 2008). In addition, a factor analysis of laboratory tasks indicated that RT variability appears to be best characterized as a single factor in children with ADHD, suggesting that it could reflect a stable characteristic despite considerable differences in task demands (Klein et al., 2006). Studies manipulating task characteristics such as working memory demands and inter-stimulus intervals (Buzy et al., 2009; Epstein et al., 2006) report conflicting results and limit conclusions that RT variability represents a stable neurocognitive deficit in ADHD. For example, Klein et al. (2006) reported that RT variability across a wide variety of tasks loaded on a single factor for children with ADHD, whereas Tillman, Thorell, Brocki, and Bohlin (2008) reported a nonsignificant relation between RT variability metrics on two commonly used laboratory tasks ($r = .03$, *ns*). These conflicting findings suggest that environmental factors (e.g., task demands) may significantly moderate the magnitude, if not the appearance, of increased intraindividual variability in ADHD. The current meta-analysis examines whether intraindividual variability is a ubiquitous feature of ADHD by analyzing between-study heterogeneity and potential moderators to examine the extent to which task factors and internal physiologic and/or cognitive states (e.g., treatment, motivation/incentive effects) may impact the association between ADHD and RT variability.

1.5. Is intraindividual variability causally related to ADHD symptoms, or an outcome of other proposed core deficits?

RT variability and ADHD symptoms. If RT variability reflects a core neuropsychological deficit in ADHD, it should predict ADHD behavioral symptoms and functional impairments. In this regard, the results to date are equivocal. For example, divergent results indicate that RT variability is related to both inattention and hyperactivity/impulsivity (Epstein et al., 2003), inattention only (Wåhlstedt, Thorell, & Bohlin, 2009), hyperactivity only (Buzy et al., 2009), or the interaction between inattention and hyperactivity (Clarke et al., 2007). In addition, RT variability is frequently attributed to periodic lapses in attention (e.g., Hervey et al., 2006); however, Schmiedek et al. (2007) investigated this claim and concluded that attentional lapse models could not account for RT variability. Similarly, Epstein et al. (2010) found that temporal performance-based indices of inattention (omission errors) and behavioral inhibition (commission errors, successful inhibitions) could not account for ADHD-related impairments in RT variability (Cohen's *d* changed minimally, from .78 to .70). The present meta-analytic review addresses this issue by examining heterogeneity among ADHD subtypes and comparing obtained effect sizes to estimate the extent to which RT variability is related to inattentive, hyperactive/impulsive, or both symptom clusters.

1.5.1. RT variability and other proposed core neuropsychological impairments

Studies examining the relation among neurocognitive impairments have direct implications for etiological models of ADHD. Extant models make specific, divergent predictions regarding the role of RT variability in ADHD (Table 1). The evidence to date, however, is incomplete and equivocal. For example, Schmiedek et al. (2007) found that RT variability was strongly related to working memory and higher-order reasoning ($r = -.71$ to $-.72$), as well as processing speed ($r = -.58$). There is contradictory evidence, however, regarding whether RT variability is affected by manipulating working memory demands in children with ADHD (Buzy et al., 2009; Klein et al., 2006). For example, Finke et al. (2011) concluded that their findings supported working memory model (Rapport et al., 2001) predictions regarding the role of working memory deficits, rather than perceptual processing or energetic factors, as etiological factors underlying increased RT variability in ADHD. In contrast, Verté, Geurts, Roeyers, Oosterlaan, and Sergeant (2006) indicated that working memory, response inhibition, and RT variability are distinct but related cognitive domains. In another study, RT variability was influenced by a host of cognitive processes, including visual processing, working memory, response selection and preparation, and response execution/motor response, whereas RT variability did not predict ADHD-related impairments in processing speed (Jacobson et al., 2011). In contrast, motivational/volitional control factors have been found to be related (Andreou et al., 2007; Kuntsi, Wood, van der Meere, & Asherson, 2009) and unrelated (Aase, Meyer, & Sagvolden, 2006; Epstein et al., 2011) to RT variability, and findings regarding behavioral inhibition are similarly equivocal (Buzy et al., 2009; Epstein et al., 2010; Lipszyc & Schachar, 2010). The current meta-analysis addresses this issue by examining heterogeneity in RT variability effect sizes, and testing the extent to which any between-study heterogeneity is attributable to several model-implied moderators (e.g., behavioral inhibition task demands, effects of incentives).

1.6. Does treatment (e.g., medication) decrease or normalize RT variability in ADHD?

The clinical model of psychopathology hypothesizes that interventions aimed at improving suspected underlying neurological substrate(s) and core psychological/cognitive features of ADHD should produce the greatest level and breadth of therapeutic change (National Advisory Mental Health Council's Workgroup, 2010;

Rapport et al., 2001). Interventions aimed at peripheral behaviors, on the other hand, should show limited generalization upward to core features, and minimally affect other peripheral symptoms. For this reason, a key test of RT variability's role as an ADHD core deficit will be the extent to which pharmacological and psychosocial treatments known to improve ADHD behavioral symptoms also decrease RT variability. Studies of medication effects on RT variability in ADHD are equivocal to date, with studies reporting significant (Heiser et al., 2004; Teicher, Lowen, Polcari, Foley, & McGreenery, 2004) and nonsignificant (Aggarwal & Lillystone, 2000; van der Meere, Gunning, & Stemerdink, 1999) treatment-related changes. The current meta-analysis addresses this issue by examining the magnitude of treatment-related changes in RT variability relative to baseline and placebo conditions, and the extent to which treatment type is a significant moderator of this relationship.

1.7. Are the RTs of children with ADHD slower and more variable, or just more variable?

The performance of individuals with ADHD across a wide range of laboratory tasks has long been characterized as *slower* and more variable. Several previous meta-analytic reviews have examined mean reaction time (MRT), and consistently reported small-to-moderate effect sizes ranging from 0.29 to 0.66 (Alderson et al., 2007; Frazier, Demaree, & Youngstrom, 2004; Lijffijt et al., 2005; Lipszyc & Schachar, 2010; Oosterlaan, Logan, & Sergeant, 1998). However, MRT is influenced heavily by RT variability (Schmiedek et al., 2007; Wagenmakers & Brown, 2007), and several studies suggest that the MRT of children with ADHD is not slower than their peers after accounting for RT variability using ex-Gaussian methods (e.g., Buzy et al., 2009; Epstein et al., 2003). These findings could be related to study power, however, and a meta-analytic approach is needed to determine the magnitude of MRT differences after accounting for RT variability. Specifically, the current meta-analysis can test the extent to which children with ADHD demonstrate slower MRT after accounting for RT variability by examining studies using ex-Gaussian estimation (i.e., *mu*, which reflects MRT after accounting for variability in both the normal and exponential components of the RT distribution) and/or reporting MRT after accounting for RT variability (e.g., ANCOVA). Obtained effect sizes can be compared to studies examining the opposite relationship—RT variability after controlling for MRT (i.e., tau and CV metrics)—to draw conclusions regarding the presence of ADHD-related impairments in one or both of these neurocognitive indices (processing speed, variability).

1.8. The current meta-analysis

In summary, the current meta-analysis is a comprehensive review of 319 studies reporting on reaction time variability in children, adolescents, and adults with ADHD relative to (a) typically developing groups, (b) clinical control groups, and (c) themselves (i.e., subtype comparisons, treatment and motivation effects). Through meta-analytic synthesis, analysis, adequately powered moderator investigation, and best case analysis, the current review seeks to inform current debate regarding the veracity of RT variability as an ADHD core deficit, with implications for the evaluation of etiological models and treatment interventions for children and adults with ADHD.

2. Method

2.1. Literature searches

A three-phase literature search was conducted using Medline, PubMed, PsycInfo, PsycArticles, PsycBooks, ERIC, Dissertation Abstracts International, and Social Science Citation Index. Search terms included permutations of the ADHD diagnostic label (ADHD, ADD, attention deficit, attention problems, inattent*, hyperact*, hyperkinesis, minimal brain dysfunction/damage, MBD), variability, reaction time (RT), variability metrics (SDRT, coefficient of variation, CV, sigma, tau, RT of SE, Slow-*, frequency, signal processing), and tasks frequently used to derive RT variability data (TOVA, Conners' CPT, stop signal, reaction time, motor speed, SRT, CRT, *n*-back, CPT, Flanker, Stroop, go/no-go, vigilance, inhibition, attention, KITAP, Attention Network Test, ANT). An asterisk following a root word instructs search engines to look for any derivative of the word that is followed by the asterisk (e.g., hyperactive, hyperactivity). No search delimiters were selected to avoid missing studies due to database misclassification. To further expand the initial study base, the options “apply related words” and “also search within the full text of the articles” were selected across all databases. Searches were conducted with and without an ADHD search term included. Searches were conducted independently by all co-authors and repeated until no new studies were located. After the initial searches, studies cited by articles reporting RT variability in ADHD were examined (Phase II backward search), and a forward search (Phase III) was conducted using the Social Science Citation Index to locate studies citing those that reported RT variability in ADHD. Listserv requests for unpublished data were sent to APA Divisions 12 (Clinical), 40 (Neuropsychology), and 53 (Clinical Child), which were considered the most likely divisions for members conducting neurocognitive research with individuals with ADHD. In addition, emails were sent to authors of studies published within the last 5 years that investigated RT variability but did not report sufficient data for effect size calculation ($N = 13$ of 19 who responded with data). Finally, a data request was displayed during two, 2011 ISRCAP poster sessions. These procedures generated 3,404 peer-reviewed studies, dissertations, and unpublished manuscripts written since 1962. All search processes were completed and study recruitment was closed on August 1, 2011.

2.2. Inclusion and exclusion criteria

Inclusion and exclusion criteria are described below, with the number of studies omitted for each criterion in parentheses; Several studies failed to meet multiple inclusion criteria; the counts below reflect the first failed criteria identified. The following served as inclusion criteria for the review: (a) English language (29) studies of (b) children, adolescents, and/or adults with a primary diagnosis of ADHD or related labels (e.g., hyperactive, attention problems) completing one or more laboratory tasks from which RT variability data is collected during a non-medication condition (baseline or placebo) (2,761); (c) inclusion of a typically developing control group, inclusion of one or more clinical control groups, comparison between ADHD participants' performance across pre- and post-treatment conditions, comparison among ADHD subtypes, and/or comparison of ADHD participants' performance with and without external motivators (149); (d) RT variability data reported, or statistics reported from which effect size can be estimated (31); and (e) estimated intelligence >80 (0). Exclusion criteria included: (a) gross neurological, sensory, or motor impairment, history of a seizure disorder, or psychosis (12); and (b) repeat data (e.g., study published in journal and as book chapter; follow-up longitudinal study) (69). Studies reporting variability for only time estimation (15), on-task attention (3), or other non-reaction time indices (16) were excluded.

For studies reporting repeat data with the same task(s) and metric(s) (e.g., SDRT), the newest study with the largest sample size was included. For repeat studies reporting different RT variability metrics for the same tasks across studies, preference was given to studies reporting ex-Gaussian or spectral power-based indices relative to SD/SE metrics, because of the relative paucity of studies reporting the former and arguments for the benefits of ex-Gaussian over Gaussian metrics (e.g., Leth-Steensen et al., 2000). In all cases, decisions were made prior to effect size calculation to minimize experimenter bias.

A total of 319 studies published (or conducted, for unpublished findings) between 1972 and 2011 met study criteria and were included in one or more sets of analyses (128 of the 319 studies contributed data to two or more analytic Tiers). These 319 studies (294 published

studies, 12 dissertations, and 13 unpublished data provided by authors) provided 943 total effect sizes. Tier I examined 35 studies ($k = 41$ independent subsamples) comparing ADHD subtypes (82 effect sizes). In Tier II, 270 studies were included in analyses of ADHD relative to typically developing groups (572 effect sizes). Ten of these 270 studies reported data for multiple, independent subsamples (defined as ADHD and control samples with non-overlapping participants), resulting in a total Tier II study size of $k = 283$. Tier III examined 71 studies comparing ADHD with one or more clinical control groups (163 effect sizes). In Tier IV, 48 studies ($k = 52$ independent subsamples) reported comparisons of ADHD groups on versus off medication or pre/post treatment (126 effect sizes).

2.3. Coding of moderators

All potential moderator variables were coded according to the characteristics reported in Tables S1 to S8 (Supplementary online). Continuous variables were used whenever possible to facilitate regression-based approaches that allow simultaneous examination of multiple potential moderators (Hedges & Pigott, 2004). Task duration and number of trials were obtained from published task manuals when available. A publication year of 2011 was assigned for studies in press and unpublished data provided by authors. Categorical variables were coded ordinally, where higher values are associated with an addition to the variable in question (e.g., adding matched controls, diagnostic tools). Two age-related variables were created: a categorical variable describing studies as child (age 12 and below), adolescent (ages 13 to 18), or adult (ages 18 +) based on reported age means and ranges[1]; and an effect size calculated as the magnitude of age differences for between-group comparisons (negative values indicate that the ADHD group was younger). The latter variable was created due to evidence of developmental changes in RT variability (Eckert & Eichorn, 1977; Williams, Hultsch, Strauss, Hunter, & Tannock, 2005), coupled with concerns raised during data collection that the non-significant group differences in age reported in many studies may be attributable to low power.

Diagnostic Method was coded as an index of study quality based on the recommendations for gold standard diagnosis of ADHD used to code study quality in previous meta-analytic reviews (Alderson et al., 2007; Kofler et al., 2008; Lipszyc & Schachar, 2010). Studies were classified into ordinal groups, wherein higher values reflect more rigorous diagnostic procedures: 0 = referral or previous diagnosis only; 1 = single informant questionnaire and/or interview; 2 = multiple informant questionnaires and/or interviews; 3 = multiple informant report based on standardized and normed questionnaires and gold standard semi-structured/structured clinical interview. An additional ordinal variable was coded based on the number of demographic characteristics upon which each study matched their ADHD and comparison group. Behavioral inhibition demands for each task were classified as High (e.g., stop signal, go/no-go) or Low (e.g., CPT, simple/choice RT tasks) according to established criteria (Alderson et al., 2007; Lijffijt et al., 2005).

Moderator analyses were conducted using a tiered approach, wherein basic demographic and categorical variables (e.g., age group) were analyzed first using the mixed effects maximum likelihood Analog to ANOVA approach recommended by Lipsey and Wilson (2001). Additional continuous and dichotomous moderators were examined using random effects regression for meta-analysis if significant between-study heterogeneity remained at the overall study or subgroup level after accounting for categorical demographic variables.

2.4. Computation of effect sizes

Means, *SD*s, and sample sizes for each group were used to compute Hedges' *g* effect sizes using Comprehensive Meta-Analysis (v2.2). When these data were unavailable, effect sizes were estimated using reported test statistics. For between-group comparisons, these statistics included each group's sample size and *t* or *p* values, each group's means and the comparison *p* value, or reported effect sizes converted to Hedges' *g*. For within-subject comparisons (i.e., Tier IV medication and incentive effects), a pre-post correlation of .5 was assumed when these data were not reported as recommended (Smith, Glass, & Miller, 1980).[2] Hedges' *g* effect sizes are Cohen's *d* effect sizes corrected for study sample size due to the upward bias in effect size magnitude of small *N* studies. Hedges' *g* effect sizes are in standard deviation units, such that an effect size of 1.0 indicates that two groups differ by one standard deviation (Zakzanis, 2001). An effect size of 0.2 is interpreted as small (detectable only through statistics), 0.5 is medium (detectable to a careful observer), and 0.8 is large (obvious to any observer; Cohen, 1988). Overall effect sizes were computed under a random effects model (cf. Hunter & Schmidt, 2004) in which each study is weighted by its inverse variance weight ($1/SE^2$).

Meta-analysis macros for SPSS using random/mixed effects were used for all moderator analyses as recommended (Lipsey & Wilson, 2001). Random effects models with inverse variance weighting were used for effect size calculation and all moderator analyses to correct for study-level sampling error (Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). Artifact correction was conducted at the overall rather than individual study level using published internal consistency (coefficient of equivalence) and test–retest reliability (coefficient of stability) coefficients. We were unable to correct effect sizes at the individual study level because the majority of studies used experimental tasks without established psychometric properties (i.e., internal consistency and test–retest data were available for 19.5% and 23.4% of the 943 effect sizes, respectively). Reliability data not reported at the individual study level were collected primarily from published test manuals (e.g., Conners & MHS Staff, 2000; Conners & MHS Staff, 2001; Leark, Greenberg, Kindschi, Dupuy, & Hughes, 2007) and psychometric studies (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956; Saville et al., 2011). Both coefficients of equivalence (internal consistency) and stability (test–retest) were used to correct the bare-bones overall effect sizes (Hunter & Schmidt, 2004). As demonstrated by Hunter and Schmidt (2004), this method is superior to not correcting for error when reliability information is available for some but not most studies. Across sources, the average internal consistency for RT variability metrics was .84. Mean test–retest reliability was .76 based on 1-week re-administration in all cases except the Conners CPT (3-month). No corrections to the group assignment variable were conducted given our goal of assessing diagnostic rigor as a potential moderator following previous ADHD meta-analyses (Alderson et al., 2007; Kofler et al., 2008; Lipszyc & Schachar, 2010).

2.4.1. Multiple effect sizes

Most studies reported data sufficient to calculate multiple effect sizes (Tables S1 to S8). The most common reasons included reporting RT variability data across multiple tasks, reporting multiple RT variability metrics, or both. Separate effect sizes were calculated for each task and metric to be comprehensive and allow studies to be included in as many analysis subsets as possible. To meet the independence assumption, only one effect size was used for each study in any given analysis (Lipsey & Wilson, 2001). This effect size reflected the average of all relevant effect sizes for that particular analysis (e.g., for ADHD-TD group comparisons, only effect sizes from tasks completed during non-medication conditions were used).

[1] All moderator results were unchanged when a continuous age variable based on reported age means was used in lieu of this categorical age variable.

[2] The assumed pre-post correlation of .5 is a convention recommended by Smith et al. (1980, appendix 7, page 214) that did not change the results significantly if varied within a wide range as large as .1 to .9 (i.e., intervention effects remained within three hundredths of a decimal place from the effect sizes shown in Table 2).

2.4.2. Publication bias: the file drawer problem

Ten studies did not provide data sufficient to calculate effect size, but reported no significant between-group differences. These studies were retained in the analysis and assigned an effect size of 0.00 because omitting them would artificially inflate overall effect size estimates due to publication bias (Rosenthal, 1995). In addition, 33 studies reported insufficient data for effect size calculation, but either contained detailed Figures from which this data could be estimated (20), or their authors responded to email queries and provided data (13). Four tests of publication bias were used for each analysis subtest (Fail-safe *N*, Begg & Mazumdar's (1994) rank correlation test, Egger's test of the intercept, and Duval & Tweedie's trim-and-fill procedure; Lipsey & Wilson, 2001). These results are provided in Appendix B. For analyses where significant publication bias was detected, overall effect sizes were corrected using the methods recommended by Duval and Tweedie (2000).

3. Results

3.1. Overview

Five datasets (Tiers) were created to address the primary questions raised in the Introduction. In Tier I, ADHD subtypes are compared to inform inclusion criteria for subsequent analyses. Tier II examines questions related to comparisons of individuals with ADHD to typically developing individuals, whereas Tier III addresses specificity questions involving comparison of ADHD groups to groups with other forms of psychopathology (clinical control groups). Tier IV addresses treatment and motivation/incentive effects on the magnitude of ADHD RT variability estimates. Finally, Tier V investigates the relation between response variability and overall (mean) response time to critically evaluate the characterization of individuals with ADHD as *slower* and more variable. Within each Tier, we initially report overall ('moderator-independent') effect sizes, followed by heterogeneity tests to determine whether moderator analyses are warranted. We then analyze potential categorical and continuous moderators, respectively, followed by 'best case' analysis (Lipsey & Wilson, 2001). This 'best case' section provides final effect sizes after accounting for methodological differences across studies. Studies included in each Tier are listed in Tables S1–S8 (supplementary online), and stem-and-leaf histograms of obtained effect sizes for each Tier are reported in Tables 3–6. All analyses are based on random effects models; effect sizes are corrected for artifact and publication bias (Duval & Tweedie, 2000; Hunter & Schmidt, 2004). Effect sizes for each analytic tier are summarized in Table 2.

3.2. Tier I: ADHD subtype comparisons

Potential differences among ADHD subtypes were analyzed first to examine the extent to which increased RT variability is attributable to one or both primary symptom clusters (inattention, hyperactivity/impulsivity), and determine which subtypes should be included in subsequent analytic tiers. A total of 41 studies (Table 3; Tables S7–S8) were included in the analyses comparing ADHD subtypes (Total *N*: ADHD-C = 2,810; ADHD-I = 2,245; ADHD-H = 304).

Table 2
Reaction time (RT) variability: analysis summary.

	ADHD subtypes (Tier I) $k = 41$	ADHD vs. typically developing (Tier II) $k = 283$	ADHD vs. clinical control (Tier III) $k = 71$	Treatment effects (Tier IV) $k = 52$
Hedges' *g* effect size corrected for:				
Sampling error only	0.19 (0.05 to 0.32)	0.71 (0.66 to 0.76)	0.24 (0.11 to 0.36)	−0.51 (−0.61 to −0.41)
Sampling error and:				
Publication bias	0.01, *ns* (−0.13 to 0.15)	0.57 (0.51 to 0.62)	0.11, *ns* (−0.01 to 0.24)	−0.56 (−0.67 to −0.46)
Measurement unreliability	0.24 (0.07 to 0.41)	0.89 (0.83 to 0.95)	0.30 (0.14 to 0.46)	−0.64 (−0.76 to −0.51)
Publication bias and measurement unreliability	*0.01, ns (−0.16 to 0.18)*	*0.71 (0.65 to 0.78)*	*0.14, ns (−0.02 to 0.29)*	*−0.70 (−0.83 to −0.57)*
Moderated Hedges' *g* effect sizes				
Age group				
Child	–	0.76 (0.68 to 0.84)	0.25 (0.09 to 0.41)	–
Adolescent	–	–	0.08, *ns* (−0.24 to 0.39)	–
Adult	–	0.46 (0.31 to 0.61)	−0.06, *ns* (−0.46 to 0.34)	–
ADHD subtype comparison				
ADHD-C vs. ADHD-I	0.35 (0.13 to 0.57)	–	–	–
ADHD-C vs. ADHD-H	0.24, *ns* (−0.23 to 0.70)	–	–	–
ADHD-I vs. ADHD-H	−0.37, *ns* (−1.00 to 0.23)	–	–	
Treatment type				
MPH/stimulants	–	–	–	−0.74 (−0.87 to −0.61)
All others	–	–	–	−0.21, *ns* (−0.47 to 0.05)

Note: Hedges' *g* effect sizes (95% confidence intervals in parentheses) are Cohen's *d* effect sizes corrected for sample size due to the upward bias of small *N* studies. Effect sizes are considered significantly different from 0.0 (statistically significant at $p < .05$) if their 95% confidence interval does not include 0.0. Moderator subgroup effect sizes are corrected for sampling error, measurement unreliability, and publication bias. After accounting for the moderators listed for each analysis Tier, no significant heterogeneity was detected either overall or within each moderator subgroup, indicating that studies did not differ in the overall magnitude of their results by more than expected based on study-level sampling error. Positive values in Tier I indicate increased RT variability for ADHD-C groups relative to ADHD-I and ADHD-H groups. Positive values in Tiers II and III indicate increased RT variability for ADHD groups relative to typically developing (Tier II) and clinical control (Tier III) groups. Negative values in Tier IV reflect decreases in RT variability associated with treatment.
ADHD-C = ADHD-Combined Subtype; ADHD-H = ADHD-Hyperactive/Impulsive Subtype; ADHD-I = ADHD-Inattentive Subtype; MPH = methylphenidate; *ns* = non-significant (95% confidence interval includes 0.0; $p > .05$).

Table 3
ADHD subtype comparisons: stem and leaf histogram of 41 Hedges' g effect sizes.

Stem	Leaf
1.3	**5**
1.2	.
1.1	.
1.0	.
0.9	4, **7**
0.8	.
0.7	0, 1, 5
0.6	2, 2, 6
0.5	0, 1, 1, 8
0.4	0, **8**, **9**
0.3	1
0.2	2, 3, 5, 5, 6, **8**, **1**
0.1	5, 8
0.0	0, 0, 0, 1, **2**
−0.0	1, 5, 6, 7, **2**
−0.1	1, 9
−0.2	2
−0.3	5
−0.4	0

Note: Stem and leaf plots provide a histogram of effect sizes across studies. The "stem" reflects the ones and tenths digits, and the "leaf" reflects the hundredths digit for each obtained effect size. Each leaf indicates one unique effect size. For example, the 0.9 stem has two leaves: 4, and 7, indicating that two of the included studies obtained effect sizes of 0.94 and 0.97, respectively. **Bold font** represents ADHD-C/ADHD-H comparisons, regular font represents ADHD-C/ADHD-I comparisons. Positive values indicate increased RT variability for ADHD-C groups relative to ADHD-I or ADHD-H groups.

3.2.1. Moderator-independent RT variability differences

As shown in Table 2, individuals with ADHD-C were not significantly more variable relative to the ADHD-I and ADHD-H subtypes ($g = 0.01$, 95% CI $= -0.16$ to 0.18). Significant heterogeneity existed across studies, however, supporting examination of potential moderators of between-study differences in obtained effect sizes ($Q = 85.25$, $df = 40$, $p < .0001$).

3.2.2. Categorical moderators of ADHD Subtype between-study differences

Most studies examined children (34 of 41 studies) and reported only SD or SE (34 of 41 studies), precluding examination of age group or metric as potential moderators. All additional potential moderators were ordinal or continuous, allowing us to examine them simultaneously via mixed effects regression for meta-analysis (Lipsey & Wilson, 2001).

3.2.3. Continuous moderators of ADHD subtype between-study differences

A mixed effects regression was conducted using the following variables defined above: ADHD Diagnostic Method, Percent Female, Number of Trials, Behavioral Inhibition Demands (Low/High), Age Mean, and Subtype Comparison (ADHD-C vs. ADHD-I or ADHD-H). These variables were not significantly intercorrelated after correcting for multiple comparisons (all $p > .05$). Between-group age effect size was not entered as a potential moderator because most studies reported demographic information for the overall ADHD sample rather than by subtype. Results indicated that the model explained a moderate degree of between-study variance ($R^2 = .33$, $Q_R = 28.11$, $df = 5$, $p = .015$), such that no residual between-study variance remained after accounting for the model ($Q_E = 28.11$, $df = 25$, $p = .30$). Only Subtype Comparison significantly predicted between-study effect size magnitude differences, such that studies tended to have larger effects sizes when comparing ADHD-C to ADHD-I groups, relative to studies comparing ADHD-C to ADHD-H groups ($\beta = -.48$, $p = .003$). All other variables were nonsignificant at $p \geq .22$.

3.2.4. Best case estimation

To determine the expected effect sizes after considering comparison group, separate estimates were calculated for ADHD-C/ADHD-I, ADHD-C/ADHD-H, and ADHD-I/ADHD-H comparisons. Examination of studies comparing ADHD-C to ADHD-I ($k = 36$) and ADHD-H ($k = 8$) subtypes revealed that ADHD-C and ADHD-H groups did not differ significantly (95% CI included 0.0), whereas ADHD-C groups demonstrated small but significantly increased magnitude RT variability relative to ADHD-I groups ($g = 0.35$, 95% CI $= 0.13$ to 0.57; 76% overlap). ADHD-H groups demonstrated a nonsignificant trend toward increased RT variability relative to ADHD-I groups across the 7 studies reporting these data ($g = -0.37$, 95% CI $= -1.00$ to 0.23). These findings suggest that RT variability may be related somewhat more strongly to hyperactivity/impulsivity than inattention, but must be considered preliminary given the small number of studies reporting data on the ADHD-H subtype ($N = 304$ ADHD-H participants). Given this pattern of results, all three subtypes were included in subsequent analyses; a continuous moderator (% ADHD-C) was coded to account for the minimal influence of subtype in analyses where significant heterogeneity was detected.

3.3. Tier II: ADHD vs. typically developing group comparisons

A total of 283 studies reporting data on 9,780 individuals with ADHD and 12,024 typically developing control participants were included in analyses comparing ADHD and typically developing groups (Table 4; Tables S1–S2).

Table 4
ADHD vs. typically developing individuals: Stem and leaf histogram of 283 Hedges' g effect sizes.

Stem	Leaf
2.6	**1**
2.5	.
2.4	.
2.3	.
2.2	7
2.1	**6**
2.0	**1**
1.9	**1**
1.8	*5*
1.7	*3*, **1**, **1**
1.6	.
1.5	6, **0**, **2**, **5**, **9**, **9**
1.4	**4**, **6**, **8**, **9**
1.3	*5*, *6*, **1**, **1**, **2**, **2**, **2**, **2**, **3**, **7**, **8**, **8**
1.2	*6*, **2**, **4**
1.1	4, 5, *4*, **0**, **0**, **0**, **0**, **1**, **2**, **5**, **5**, **6**, **6**, **7**, **7**, **8**
1.0	*2*, **1**, **1**, **1**, **1**, **3**, **3**, **5**, **6**, **7**, **7**, **8**
0.9	0, 2, 5, 6, *2*, *3*, *3*, **0**, **0**, **0**, **1**, **1**, **1**, **3**, **4**, **5**, **6**, **7**, **7**, **7**, **7**, **8**, **8**, **8**, **9**, **9**, **9**, **9**, **9**
0.8	0, 1, 2, 2, 9, *3*, **0**, **0**, **0**, **0**, **1**, **1**, **2**, **2**, **2**, **3**, **4**, **4**, **5**, **5**, **6**, **6**, **6**, **6**, **7**, **7**, **7**, **8**, **8**, **9**
0.7	0, 0, 1, 7, 7, 7, *0*, *1*, *2*, *2*, *3*, *5*, *6*, *9*, **1**, **2**, **3**, **3**, **5**, **6**, **6**, **7**, **7**, **7**, **8**, **8**, **8**, **8**, **9**, **9**, **9**
0.6	0, 3, 4, 5, 6, 6, 6, **0**, **0**, **0**, **1**, **1**, **2**, **2**, **3**, **3**, **4**, **4**, **4**, **4**, **5**, **5**, **5**, **5**, **6**, **6**, **6**, **6**, **7**, **7**, **8**, **8**, **8**, **8**, **8**, **8**, **9**, **9**
0.5	2, 3, 4, 5, 6, 6, 8, 9, **0**, **0**, **1**, **1**, **2**, **4**, **4**, **6**, **8**, **8**, **8**, **8**
0.4	2, 2, 3, 5, **0**, **1**, **3**, **4**, **5**, **7**, **8**, **8**, **9**
0.3	4, 4, 4, 4, 7, 8, **1**, **1**, **2**, **3**, **3**, **3**, **4**, **5**, **7**, **8**, **9**
0.2	5, 7, **0**, **1**, **2**, **3**, **3**, **4**, **6**, **7**, **8**, **9**, **9**
0.1	0, 0, 1, 5, 6, **3**, **3**, **6**, **6**, **8**, **8**
0.0	**0**, **0**, **0**, **0**, **0**, **9**, **9**
−0.0	*5*, **1**, **9**
−0.1	8
−0.2	4, **1**, **7**, **9**
−0.3	.
−0.4	*0*, **5**
−0.5	.
−0.6	5

Note: Stem and leaf plots provide a histogram of effect sizes across studies. The "stem" reflects the ones and tenths digits, and the leaf reflects the hundredths digit for each obtained effect size. Each leaf indicates one unique effect size. For example, the 1.2 stem has three leaves: 6, 2, and 4, indicating that three of the included studies obtained effect sizes of 1.26, 1.22, and 1.24, respectively. **Bold font** represents child studies, *italicized* font represents adolescents, and regular font represents adults. Positive values indicate increased RT variability for ADHD groups relative to typically developing groups.

3.3.1. Moderator-independent RT variability differences

As shown in Table 2, individuals with ADHD exhibited moderate-to-large increases in RT variability relative to TD groups ($g = 0.71$, 95% CI = 0.65 to 0.78; 57% population overlap). The overall test of homogeneity was significant, suggesting that there is more variance among effect sizes than would be expected based on study-level error alone, and supports the analysis of potential moderators ($Q = 741.73$, $p < .0001$).

3.3.2. Categorical moderators of ADHD-TD between-study differences

Based on the tiered approach described above, Variability Metric and Age Group (child, adolescent, adult) were examined initially. Variability Metric (SD/SE, CV, tau, sigma, spectral power-based metrics) was examined first to inform inclusion criteria for subsequent analyses. Because many studies reported multiple RT variability metrics (e.g., all studies reporting tau also reported sigma and CV or SD/SE), we elected to compute effect sizes separately for each subgroup and compare the obtained effect sizes using confidence interval analyses (Cumming & Finch, 2005). This method was selected for practical reasons as a compromise between meeting the independence assumption (Rosenthal, 1995) and including as many studies as possible in moderator analyses. Bias-corrected results indicated no significant differences (all $p > .05$) among most metrics.[3] Effect sizes for studies reporting sigma ($g = 0.39$, 95% CI = 0.15 to 0.63, $k = 7$) were smaller relative to tau ($p < .01$), SD/SE, and CV (all $p < .05$). Based on this pattern of results, all relevant studies are included in subsequent analyses.

Results of the mixed effects (maximum likelihood estimation) Analog to ANOVA (Lipsey & Wilson, 2001) revealed that Age Group exerted a significant impact on obtained effect sizes, with significantly larger effect sizes for children and adolescents relative to adults (Child = Adolescent > Adult studies; $p < .01$). Based on this finding, the child and adolescent groups were combined and a dichotomous variable (child/adolescent, adult) was used. The Analog to ANOVA test for homogeneity for this dichotomous moderator indicated significant between-group differences ($Q_B = 11.42$, $df = 1$, $p = .0007$), with child/adolescent studies ($g = 0.76$, 95% CI = 0.68 to 0.84; 55% population overlap) associated with larger effect sizes relative to adult studies ($g = 0.46$, 95% CI = 0.31 to 0.61; 70% population overlap). After accounting for Age Group, the residual variance was nonsignificant ($Q_W = 280.02$, $df = 281$, $p = .51$). In addition, within-group residual variance was nonsignificant for both child/adolescent ($Q = 238.23$, $df = 233$, $p = .39$) and adult ($Q = 41.80$, $df = 48$, $p = .72$) groups, indicating that within each age group, effect sizes did not differ more than expected based on study-level sampling error. These findings indicate that the Age Group moderator fully accounted for the heterogeneity in the effect size distribution, and that additional moderator analyses are not warranted.[4] That is, effect sizes across studies were homogeneous despite considerable between-study differences in diagnostic methods, inhibitory demands, task characteristics, and variability metric.

3.3.3. Continuous moderators of ADHD-TD between-study differences

No continuous moderators were examined given the homogeneity of effect sizes described above (i.e., the Age Group moderator fully accounted for between-study heterogeneity, indicating that additional moderator analyses are not warranted). However, age group was correlated with additional potential moderators to examine the extent to which the obtained age effect was attributable to potential multicollinearity among moderators (e.g., child and adult studies may vary systematically on one or more additional moderators, leading to the appearance of an age effect that is attributable instead to a secondary variable). Age Group coded dichotomously (child/adolescent, adult) was not correlated significantly with most planned moderators including Demographic Matching, Diagnostic Moniker, and Task Duration (all $p > .40$). Age Group was non-significantly related to the percentage of ADHD-Combined (relative to ADHD-I and -H) at $p = .06$, but demonstrated small magnitude relationships with Percent Female ($r = .33$, $p < .0005$), Diagnostic Method ($r = .12$, $p = .04$), Number of Trials ($r = .14$, $p = .02$), Age Effect Size ($r = .14$, $p < .02$), and Publication Year ($r = .13$, $p = .02$). After correcting for multiple comparisons, only the small magnitude relationship between Age Group and Percent Female remained significant (adult studies were associated with a higher proportion of females). Thus, the most parsimonious conclusion is that the Age Group moderator effect is attributable to between-study age differences rather than secondary moderator effects.

3.3.4. Best case estimation

Given that no significant variance remained within the overall sample or either subsample (child/adolescent, adult) after accounting for Age Group, the best case estimates are equal to the artifact- and bias-corrected mixed effects effect sizes for each age group. Specifically, despite considerable between-study differences in methodology, diagnostic methods, inhibitory demands, task characteristics, and measurement of variability, ADHD and typically developing children consistently differed by 0.76 standard deviations ($g = 0.76$; 95% CI = .68 to .84), corresponding to a medium-to-large effect size and approximately 55% population overlap (Zakzanis, 2001). That is, only 45% of children with ADHD perform outside the typically developing range on RT variability tasks.[5] Adults with ADHD demonstrated smaller magnitude RT variability differences (child/adolescent vs. adult: $p < .01$) relative to controls across studies, differing from typically developing groups by 0.46 standard deviations ($g = 0.46$; 95% CI = .31 to .61; 70% population overlap).

3.4. Tier III: ADHD vs. Clinical Control Groups

A total of 71 studies comparing 6,486 individuals with ADHD and 10,176 individuals with other clinical disorders were included in the ADHD-Clinical Control analyses (Table 5; Tables S3–S4).

3.4.1. Moderator-independent RT variability differences

As shown in Table 2, individuals with ADHD did not differ significantly from clinical control groups with regards to RT variability (i.e., 95% CI includes 0.0). All analyses were repeated and results were nearly identical after removing 7 studies reporting comparisons with physical health disorders (4), subclinical ADHD (1), and malingerers (2). Significant heterogeneity existed across studies, supporting the examination of potential moderators of between-study differences ($Q = 271.46$, $df = 70$, $p < .0001$).

[3] Effect sizes were: Spectral power-based metrics ($g = 0.63$, 95% CI = 0.35 to 0.90, $k = 9$), SD/SE ($g = 0.70$, 95% CI = 0.62 to 0.77, $k = 253$), CV ($g = 0.78$, 95% CI = 0.63 to 0.93, $k = 35$), and tau ($g = 0.99$, 95% CI = 0.64 to 1.34, $k = 8$). When only studies reporting Slow-4 (.027 to .073 Hz) were examined, a similar effect size ($g = 0.59$, 95% CI = 0.27 to 0.91, $k = 6$) was found relative to the overall effect size for spectral power-based metrics (Table S9). Similar magnitude effect sizes were found for Slow-3 (.073 to .17 Hz; $k = 4$) and Slow-4 ($k = 6$), which were generally larger than effect sizes for Slow-5 (.01 to .027 Hz; $k = 2$), but these findings must be considered preliminary given the small number of studies reporting these metrics.

[4] Interpretation of results for all analytic Tiers was unchanged when meta-regression was run using a continuous age variable instead of the categorical age variable. For Tier II, when entered into meta-regression simultaneously with age effect size, percent female, percent ADHD-C, number of trials, task duration, matching, diagnostic moniker, and diagnostic method, only age emerged as a significant moderator ($p = .026$; all other $p > .050$). This model explained significant between-study heterogeneity (Q_R [9] = 23.75, $p = .005$, age $\beta = -0.15$ indicating that effect sizes decrease with increasing age), such that no residual variance remained (Q_E [264] = 270.16, $p = .38$).

[5] This value corresponds to 100% negative predictive power (NPP) but only 45% positive predictive power (PPP) if the diagnostic cut-off score is set at the edge of the Typically Developing range, indicating that RT variability is not likely to be useful diagnostically. Changing the cut-off score can increase PPP at the cost of decreased NPP (i.e., more true positives but also more false positives; Zakzanis, 2001).

Table 5
ADHD vs. clinical control groups: stem and leaf histogram of 71 Hedges' g effect sizes.

Stem	Leaf
1.1	**3**
1.0	.
0.9	**5**
0.8	1, **2**, **2**, **3**, **4**, **5**
0.7	0, **3**, **5**
0.6	0, **1**, **2**, **7**
0.5	*3*, *3*, **3**, **3**, **4**, **6**, **9**
0.4	1, 2, 6, **0**, **0**, **3**, **4**, **4**, **5**
0.3	8, *6*, **2**, **3**, **4**, **5**, **6**, **7**, **9**
0.2	5, 8, *0*, *2*, *4*, **0**
0.1	*0*, **4**, **8**
0.0	6, 7, **0**, **0**, **0**, **0**
−0.0	**2**, **3**
−0.1	**1**, **2**, **3**
−0.2	1, **0**
−0.3	**1**, **3**
−0.4	.
−0.5	.
−0.6	3, *4*, **0**
−0.7	**5**, **6**
−0.8	**1**
−0.9	.
−1.0	.
−1.1	3

Note: **Bold font** represents child studies, *italicized* font represents adolescents, and regular font represents adults. Positive values indicate increased RT variability for ADHD groups relative to clinical control groups.

3.4.2. Categorical moderators of ADHD-clinical control between-study differences

Variability metric was not analyzed as a potential moderator because SD and/or SE were used almost exclusively in studies comparing ADHD and clinical control groups (63 of 71 studies). Age group (child, adolescent, adult) was examined initially using mixed effects (maximum likelihood estimation) Analog to ANOVA (Lipsey & Wilson, 2001). Results are shown in Table 2 and revealed significant ADHD-Clinical Control group differences for studies including children ($g = 0.25$, 95% CI = 0.09 to 0.41, 82% population overlap; $k = 50$) but not for studies involving adolescents or adults (both 95% CIs include 0.0; 100% population overlap).

No significant heterogeneity was observed across studies after accounting for Age Group ($Q_w = 66.78$, $df = 68$, $p = .52$). Likewise, no significant heterogeneity was observed within age categories: child ($Q_w = 41.30$, $df = 49$ $p = .77$), adolescent ($Q_w = 4.81$, $df = 6$, $p = .57$), and adult ($Q_w = 20.67$, $df = 13$, $p = .08$). These findings provide strong support for the conclusion that Age Group was sufficient to account for between-study heterogeneity in the effect size distribution, and that additional moderator analyses (or further examination of specific subgroups within the Clinical Control category) are not warranted.

3.4.3. Continuous moderators of ADHD-clinical control between-study differences

No continuous moderators were examined given the homogeneity of effect sizes described above. Only Percent Female ($r = .29$, $p = .01$) was correlated significantly with Age Group after correcting for multiple comparisons (adult studies included a higher proportion of females).

3.4.4. Best case estimation

Given that no significant variance remained after accounting for Age Group, either within the overall sample or within the child, adolescent, or adult subsamples, best case estimates are equal to the artifact- and bias-corrected effect sizes for each subgroup. Specifically, despite considerable between-study methodological differences, ADHD and clinical control children demonstrated small magnitude between-group differences ($g = 0.25$, 95% CI = 0.09 to 0.41), reflecting approximately 82% population overlap. In contrast, adults and adolescents with ADHD were indistinguishable from clinical control groups (both 95% CI include 0.0; Table 2).

3.5. Tier IV: treatment and incentive effects

A total of 52 studies (Table 6; Tables S5–S6) were included in the analyses of treatment effects on RT variability for individuals with ADHD (Total $N = 1{,}779$).

3.5.1. Moderator-independent RT variability differences

Across studies, extant treatments resulted in large RT variability decreases for individuals with ADHD ($g = -0.70$, 95% CI = −0.83 to −0.57). Significant heterogeneity existed across studies, supporting the examination of potential moderators of between-study differences ($Q = 185.71$, $df = 51$, $p < .0001$).

3.5.2. Categorical moderators of treatment effects on RT variability

We were unable to examine Variability Metric and Age Group as potential moderators because 46 of 52 studies reported only SD and/or SE, and 45 of 52 were child studies. To examine Treatment Type, studies were categorized as methylphenidate (MPH)/stimulant treatment ($k = 41$) or Other ($k = 11$). The Other category contained 11 studies: 4 studies using multiple medications or different medication classes across participants (e.g., "usual dose"), 2 studies examining the impact of biofeedback, and 1 study each examining atomoxetine, modafinil, caffeine, cognitive training, and risperidone. Analog to ANOVA results indicated that Treatment Type explained significant between-study differences ($Q_B = 12.62$, $df = 1$, $p < .0001$), such that no significant between-study residual differences remained after accounting for Treatment Type ($Q_W = 51.03$, $df = 50$, $p = .43$). As shown in Table 2, MPH/stimulants ($g = -0.74$, 95% CI = −0.87 to −0.61; 55% population overlap) were associated with large magnitude decreases in RT variability. In contrast, nonstimulant treatments did not change RT variability significantly in individuals with ADHD (95% CI included 0.0).

Within-group residual variance was nonsignificant for both MPH/stimulant ($Q = 39.40$, $df = 40$, $p = .50$) and other ($Q = 11.62$, $df = 10$, $p = .31$) groups, indicating that within each treatment group, effect sizes did not differ more than expected based on

Table 6
ADHD treatment effects: stem and leaf histogram of 52 Hedges' g effect sizes.

Stem	Leaf
0.3	3
0.2	.
0.1	6
0.0	9, **1**
−0.0	0, 4, **0**
−0.1	5, 6, **3**
−0.2	2, 5, **4**, **5**, **6**
−0.3	**4**, **5**, **9**, **9**
−0.4	2, **0**, **1**, **2**, **2**, **5**, **8**, **8**, **9**
−0.5	2, **0**, **4**, **7**, **8**
−0.6	**1**, **5**, **8**
−0.7	**0**, **4**, **5**, **6**, **8**
−0.8	**3**
−0.9	**2**, **4**, **8**
−1.0	0, **2**, **5**, **9**
−1.1	.
−1.2	.
−1.3	.
−1.4	**3**, **3**
−1.5	**2**

Note: **Bold font** represents methylphenidate (MPH) or other psychostimulant medication, normal font represents other, non-stimulant treatments. Negative values indicate decreased RT variability associated with treatment.

study-level sampling error. These findings indicate that Treatment Type was sufficient to account for between-study heterogeneity in the effect size distribution, and that additional moderator analyses (or further examination of studies within the "Other" category) are not warranted.

3.5.3. Continuous moderators of treatment effects on RT variability

No continuous moderators were examined given the homogeneity of effect sizes described above. Treatment type was not related significantly to any other potential moderators after correcting for multiple comparisons (all $p > .32$).

3.5.4. Best case estimation

Given that no significant variance remained within the overall sample or either subsample after accounting for treatment type, the best case estimates are equal to the artifact- and bias-corrected effect sizes for each treatment category. Specifically, stimulant medication exerted a robust, large magnitude effect on RT variability for individuals with ADHD ($g = -0.74$, 95% CI $= -0.87$ to -0.61), indicating that at post-treatment approximately 45% of individuals scored outside of the pretreatment range. In contrast, all other treatments tested to date were ineffective as evidenced by the nonsignificant effect size (i.e., 95% CI included 0.0; Table 2) and lack of significant heterogeneity among non-stimulant treatment studies.

3.5.5. Clinical significance

To further address the clinical significance of stimulant medication-related decreases in RT variability, 12 of the 52 treatment studies were located that reported both stimulant medication effects and a typically developing comparison group. Comparison of medicated ADHD children with typically developing children revealed an overall nonsignificant effect size ($g = 0.12$, 95% CI $= -0.09$ to 0.33), with no detectable RT variability differences in 9 of 12 studies (all 95% CIs include 0.0). Two studies reported continued moderately increased RT variability in the medicated ADHD group ($g = 0.61$ and 0.44; Hermens et al., 2005; Tucha, Prell et al., 2006), and one study reported that medicated individuals with ADHD were less variable than the typically developing comparison group ($g = -0.62$; Greenberg, 1987). Collectively, these studies provide additional insight into the large magnitude impact of stimulant medication on RT variability and suggest that RT variability is significantly improved or normalized in most medicated individuals with ADHD. Caution is warranted when interpreting these results, however, given that most medication studies did not include the typically developing comparison group needed to make judgments regarding clinically meaningful change (Jacobson & Truax, 1991).

3.5.6. Impact of incentives (motivation)

An additional set of analyses was conducted on 8 studies reporting data on the impact of incentives on RT variability in children with ADHD (Total $N = 335$).[6] Five studies compared ADHD performance without and with external incentives (Douglas & Parry, 1983; Epstein et al., 2011; Kuntsi et al., 2009; Shanahan, Pennington, & Willcutt, 2008; Uebel et al., 2010), one study compared trials within a single task in which children could or could not earn money (Scheres, Milham, Knutson, & Castellanos, 2007), and two studies reported comparisons with typically developing children without and with incentives from which a change in between-group effect magnitude could be computed (Andreou et al., 2007; Scheres, Oosterlaan, & Sergeant, 2001). Negative effect sizes indicate decreased RT variability associated with external incentives/motivators.

After correcting for measurement unreliability, an overall small effect was obtained ($g = -0.38$, 95% CI $= -0.63$ to -0.14, 73% overlap). Significant between-study heterogeneity was detected ($Q = 26.61$, $p < .0005$). Qualitative inspection revealed that 7 of the 8 studies reported nonsignificant to small magnitude effects (range $= -.42$ to .03) and one study reported large magnitude effects ($g = -0.90$; Kuntsi et al., 2009). With this study removed, an overall small effect size was obtained ($g = -0.28$, 95% CI $= -0.43$ to -0.13; 82% overlap) and between-study heterogeneity was nonsignificant at $p = .21$, suggesting that between-study heterogeneity was attributable to this outlier. No moderators were examined due to this lack of heterogeneity and power concerns associated with the small *k*.

Collectively, these results suggest that external motivators are associated with small magnitude RT variability decreases that would generally be undetectable to a careful observer (i.e., detectable only with statistics; Cohen, 1988), in contrast to the large magnitude between-group differences obtained in Tier I and large magnitude effects associated with stimulant medication (Tier IV).

3.6. Tier V: Mean reaction time after accounting for variability.

A final set of analyses was conducted to test the oft-reported conclusion that individuals with ADHD are *slower* and more variable relative to their peers (Table S10). Specifically, we examined ADHD-TD comparison studies reporting mean reaction time (MRT) after accounting for RT variability to determine the extent to which these individuals exhibit overall slowed reaction times, or whether the finding of small to moderate magnitude MRT effect sizes in previous meta-analytic reviews ($g = 0.29$ to 0.66; Alderson et al., 2007; Frazier et al., 2004; Lijffijt et al., 2005; Lipszyc & Schachar, 2010; Oosterlaan et al., 1998) is likely attributable to a subset of abnormally slow responses (i.e., attributable to RT variability). Eight studies were located that reported data for MRT after accounting for variability (Total *N*: ADHD = 447; TD = 457); results should therefore be interpreted with caution. Seven of the studies used ex-Gaussian approaches, wherein mu (μ) reflects MRT after accounting for RT variability. The remaining study reported a *t*-test comparison of residual MRT scores after covarying RT variability.

The overall bias-corrected effect size was nonsignificant ($g = -0.19$, 95% CI $= -0.46$ to 0.08), with the small magnitude negative effect size reflecting a trend toward individuals with ADHD demonstrating *faster* RTs relative to typically developing controls after accounting for their subset of abnormally slow responses. The *Q* test for heterogeneity was nonsignificant ($Q = 12.05$) at $p = .06$. Qualitative inspection of the data revealed a consistent trend, with ADHD individuals demonstrating somewhat faster RTs after accounting for RT variability in seven of the eight studies (Table S10). In addition, the only study with a significant between-group effect size reported *faster* RTs for ADHD relative to TD children. Re-running the analysis eliminating the only study to report slower (albeit nonsignificant) RTs in individuals with ADHD (Leth-Steensen et al., 2000) resulted in a significant overall effect size ($g = -0.22$, 95% CI $= -0.42$ to -0.02) that reflects the trend across studies that children with ADHD are moderately faster than TD children on RT tasks after accounting for between-group variability differences.

This pattern of results contrasts the large magnitude between-group effect sizes for RT variability after accounting for MRT found in Tier I (i.e., CV: $g = 0.78$; tau: $g = 0.99$), suggesting directionality of effects. Taken together, these findings suggest that ADHD-related RT variability may underlie previous findings of slower motor processing speed (MRT) in ADHD, rather than both neuropsychological constructs being impaired or attributed to a common (third variable) explanation. This conclusion must be considered tentative, however, given that it is based on eight studies reporting on 904 total participants.

[6] Four additional studies (Aase & Sagvolden, 2006; Aase et al., 2006; Frank, Santamaria, O'Reilly, & Willcutt, 2007; Oosterlaan & Sergeant, 1998) reported ADHD-TD between-group differences during reinforcement tasks, but did not report comparison of ADHD and/or TD participants with vs. without external motivators/reinforcement. Hedges' *g* effect sizes in these studies ranged from 0.16 to 1.68 (0.34, 0.16, 0.78, and 1.68, respectively).

4. Discussion

Individuals with ADHD are described frequently as ubiquitously *slower and more variable* than their unaffected peers, and ADHD-related reaction time (RT) variability is considered by many to reflect a unique, stable, and etiologically important characteristic of the disorder. The present review critically evaluated these claims through meta-analytic synthesis and analysis of the 319 published and unpublished studies of RT variability in children, adolescents, and adults with ADHD relative to typically developing groups, clinical control groups, and themselves (i.e., ADHD subtype comparisons, treatment and motivation effects). Overall, results revealed that children/adolescents (Hedges' $g = 0.76$) and adults ($g = 0.46$) with ADHD demonstrated robust, medium-to-large magnitude increases in intraindividual RT variability relative to typically developing individuals, even after accounting for sampling error, measurement unreliability, publication bias, and ADHD subtype. Individuals with ADHD continued to demonstrate large magnitude increased RT variability after accounting for motor processing speed (MRT), whereas slower MRT was no longer detectable after accounting for RT variability. This pattern of results suggests directionality with regards to the RT variability/MRT relationship, and contradicts ADHD models positing slowed processing speed as a core deficit in ADHD (Russell et al., 2006). That is, individuals with ADHD in the current review tended to be *more variable but not slower* than their typically developing peers after controlling for their increased performance variability. Conclusions regarding intact processing speed in ADHD must be considered tentative, however, given that these analyses were based on only 904 total participants ($k = 8$). Consistent with extant models (Castellanos et al., 2005; Russell et al., 2006), ADHD-related variability appears to be attributable primarily to a subset of abnormally slow responses (tau), rather than ubiquitous variability across all trials of a given task (sigma).

Comparison with previous meta-analytic reviews of neurocognitive differences between ADHD and typically developing individuals suggests that RT variability effect sizes (ES) are similar to or larger than effect sizes reported for most identified impairments, including behavioral inhibition ($ES = 0.00$ to 0.64; Alderson et al., 2007; Lijffijt et al., 2005; Lipszyc & Schachar, 2010; Oosterlaan et al., 1998), objectively-measured inattention (0.64 to 1.34) and impulsivity (0.51 to 0.98; Huang-Pollock, Karalunas, Tam, & Moore, 2012; Losier, McGrath, & Klein, 1996; Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005), interference effects (0.24 to 0.56; Frazier et al., 2004; Lansbergen, Kenemans, & van Engeland, 2007; van Mourik, Oosterlaan, & Sergeant, 2005), Full Scale IQ (0.61; Frazier et al., 2004), verbal short-term memory (0.47 to 0.72; Frazier et al., 2004; Martinussen, Hayden, Hogg-Johnson, & Tannock, 2005; Willcutt et al., 2005), and visuospatial short-term memory (0.63 to 0.85; Martinussen et al., 2005; Willcutt et al., 2005). In contrast, obtained RT variability effect sizes were smaller than those reported for ADHD-related impairments in classroom attention (1.4; Kofler et al., 2008) and visuospatial storage/rehearsal (1.06; Martinussen et al., 2005), and considerably smaller than central executive working memory effect sizes (2.01 to 2.05; Kasper, Alderson, & Hudec, 2012).

4.1. RT variability: specific to ADHD or marker for general psychopathology?

After accounting for age effects across studies comparing ADHD and typically developing groups, no significant between-study heterogeneity was detectable, indicating that the overall finding of medium-to-large magnitude ADHD-related RT variability was consistent across studies despite considerable differences in methodology, diagnostic methods, inhibitory demands, energetic factors, volitional/motivational influences, task characteristics, and variability metric. In addition, similar results were found across ADHD subtypes, such that ADHD-H and ADHD-C groups were not detectably different, and ADHD-I groups were somewhat less variable relative to ADHD-C and ADHD-H groups. In other words, increased ADHD-related variability is a highly reliable finding that appears to be independent of symptom presentation, task, context, and state variables. In addition, increased RT variability is relatively common in individuals with ADHD (45% of children and 30% of adults with ADHD score outside the typically developing range).

In contrast, we found that RT variability demonstrates minimal-to-no specificity for differentiating ADHD groups from groups of children and adults with other clinical disorders (only 18% of children and 0% of adolescents and adults with ADHD score outside the clinical control range). Given the large number of studies investigating variability in ADHD relative to typically developing ($k = 283$) and clinical control groups ($k = 71$), the logical conclusion is that RT variability is not specific to ADHD; instead, RT variability may be conceptualized more parsimoniously as a marker for general psychopathology or shared risk factor rather than a diagnostic marker for ADHD (Geurts et al., 2008; Willcutt et al., 2008). This conclusion is consistent with previous reports regarding RT variability's lack of specificity among clinical disorders (Geurts et al., 2008; Willcutt et al., 2008), but runs contrary to hypotheses regarding the diagnostic utility of RT variability in ADHD (Leth-Steensen et al., 2000).

In general, similar moderators of RT variability were found in the current study and that of Lipszyc and Schachar (2010), despite the current study including over eight times as many ADHD studies. Specifically, both reviews identified age group as a significant moderator, such that studies of adults relative to children were associated with smaller RT variability effect sizes. Lipszyc and Schachar (2010), however, found that demographic characteristics were no longer significant after accounting for between-study differences in diagnostic rigor, whereas the current study found that diagnostic rigor either did not predict between-study differences (Tier III) or was not needed given the lack of between-study heterogeneity (Tiers I, II, and IV). The overall finding that effect magnitude was independent of methodological quality also contradicts two additional meta-analyses that found diagnostic rigor to be associated with overall smaller effect sizes when comparing ADHD and typically developing groups on observed attention (Kofler et al., 2008; $k = 23$) and behavioral inhibition (Alderson et al., 2007; $k = 25$). These reviews, however, concluded that diagnostic rigor's impact was likely attributable to the unintended inclusion of non-ADHD children (with other forms of psychopathology) in the ADHD group of studies using less-than-gold-standard diagnostic procedures that fail to adequately provide for differential diagnosis among the myriad disorders featuring clinically impairing levels of inattention and/or impulsivity. If this explanation is correct, then it is not surprising that diagnostic rigor was a non-factor in the current review given the overall finding that RT variability is not specific to ADHD but rather conceptualized more parsimoniously as a general marker or shared risk factor among a broad range of psychopathology.

4.2. Treatment effects

Large magnitude decreases in RT variability were associated with stimulant medication for ADHD groups ($g = -0.74$, 95% CI $= -0.87$ to -0.61), whereas no significant changes were found for non-stimulant treatments across the 52 studies reporting these data. In addition, stimulant medication normalized RT variability at the group level in 9 of the 12 studies comparing medicated ADHD individuals with their typically developing peers. These findings are consistent with previous meta-analyses demonstrating the superiority of stimulants relative to non-stimulants for decreasing parent- and teacher-reported ADHD behavioral symptoms ($d = -1.53$ to -1.83; Faraone, Biederman, Spencer, & Aleardi, 2006; van der Oord, Prins, Oosterlaan, & Emmelkamp, 2008), oppositional behavior ($d = -0.61$ to -1.08), and social problems ($d = -0.62$ to -1.06; van der Oord et al., 2008), as well as previous meta-analytic findings of large-magnitude decreases

in objectively-measured inattention ($d = -1.59$) and impulsivity ($d = -0.80$; Losier et al., 1996). In contrast, stimulant medication appears to exert a non-significant impact on academic functioning across studies (95% CI includes 0.0; van der Oord et al., 2008), reflecting the finding that fewer than half of children with ADHD demonstrate significant stimulant-related improvements in their academic performance (Rapport, Denney, DuPaul, & Gardner, 1994). Clearly, novel interventions are needed for children with ADHD.

4.3. Implications for ADHD etiological models

Across studies, RT variability is attributed most commonly to periodic *lapses of attention* that are either random or periodic and result in a subset of abnormally slow responses that skew RT distributions (Table 1; Tamm et al., 2012). This explanation is consistent with evidence that the abnormally slow RTs that skew RT distributions and contribute to higher tau scores are often preceded or followed by omission errors (Epstein et al., 2010), as well as previous findings of large magnitude visual attention deficits in ADHD (Kofler et al., 2008). It is inconsistent, however, with experimental and meta-analytic conclusions that at least some attentional processes may be intact in ADHD (Huang-Pollock & Nigg, 2003; Huang-Pollock, Nigg, & Carr, 2005; van der Meere & Sergeant, 1987) or explained by deficits in the central executive component of working memory (Burgess et al., 2010; Kofler, Rapport, Bolden, Sarver, & Raiker, 2010). Attentional lapse hypotheses are inconsistent also with the current finding that RT variability was associated somewhat more strongly with hyperactive relative to inattentive symptoms. In addition, Schmiedek et al. (2007) investigated the relation between attention and RT variability using ex-Gaussian and EZ-Diffusion modeling and concluded that attentional lapse models could not account for RT variability. Similarly, Epstein et al. (2010) found that temporal performance-based indices of inattention (omission errors) and behavior inhibition (both commission errors and successful inhibitions) could not account for ADHD-related impairments in RT variability (Cohen's *d* changed minimally, from .78 to .70). Thus, it appears likely that additional explanations are needed to account for the reliable finding that individuals with ADHD and other clinical disorders are *consistently inconsistent* in their performance on neurocognitive tasks.

The *default mode network* hypothesis (Table 1) was not supported by the current results. Based on the recommended cut-off of .10 Hz for default mode oscillations (Sonuga-Barke & Castellanos, 2007), the current results suggest that potential periodicity in ADHD response time distributions is not attributable specifically to default mode network intrusions. For example, similar, small-to-moderate magnitude effect sizes were obtained across the Slow-4 (.027 to .072 Hz) and Slow-3 (.073 to .17) bands. In addition, effect sizes for these spectral bands did not exceed the overall effect sizes across the entire spectral band, which contradicts hypotheses regarding predictable fluctuations in performance and supports previous recommendations against interpreting specific frequency bands (Geurts et al., 2008). However, only a small number of studies have reported spectral power-based comparisons ($k = 9$), indicating that the results must be considered preliminary. In addition, problems with spectral power-based analyses have been reported, such that the peak frequency obtained appears to be influenced less by participant performance and more by task characteristics, small changes in data preparation, method of spectral estimation, and subset of trials analyzed (Geurts et al., 2008; Johnson et al., 2007). For example, omission errors are handled typically by imputing the average RT of trials immediately before and after each missing data point. Given that omission errors comprise a sizable percentage of trials for individuals with ADHD (i.e., they fail to respond to valid targets during 9% to 40% of trials, relative to approximately half those values for non-ADHD children; Losier et al., 1996), the validity of the subsequent analyses may be considered tenuous. Further methodological studies are needed to determine the best method for examining potential periodicity in performance for individuals with ADHD and other clinical disorders.

Working memory and behavioral inhibition deficits are cited also as potential explanations for increased RT variability in ADHD and other disorders (Table 1). *Behavioral inhibition* deficits are unlikely to account for RT variability, however, given the overall finding of no significant between-study heterogeneity despite the inclusion of a variety of tasks that vary systematically with regards to their inhibitory demands. For example, included studies used a variety of tasks readily classified as possessing low (simple and choice RT tasks) and high (stop signal, change, and go/no-go tasks) inhibitory requirements, as well as by tasks reflecting both action cancellation (e.g., stop signal) and restraint (e.g., go/no-go) inhibitory subtypes (Table S2). In addition, behavioral inhibition did not moderate between-study heterogeneity among studies comparing ADHD subtypes ($p = .81$). Recent meta-analytic reviews conclude that behavioral inhibition processes are likely intact in ADHD (Alderson et al., 2007; Lijffijt et al., 2005), and that increased stop-signal reaction time (SSRT) in ADHD is likely attributable to ADHD-related RT variability rather than impaired inhibitory processes (Lijffijt et al., 2005). In other words, ADHD-related poor performance on behavioral inhibition tasks may reflect an outcome rather than a cause of increased RT variability (Russell et al., 2006).

Similar arguments may be made against *working memory* as an explanation for increased RT variability in ADHD and other clinical disorders; however, the working memory demands required by the wide variety of tasks included in the present review are less clearly delineated relative to these tasks' inhibitory demands. For example, emerging evidence indicates that working memory is related highly to reaction time performance and variability on a wide variety of neurocognitive tasks, including simple and choice reaction time tasks (Schmiedek et al., 2007) as well as more complex tasks including continuous performance, visual match-to-sample (Raiker, Rapport, Kofler, & Sarver, 2012), fluid reasoning, *n*-back updating (Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009), short-term memory (Engle, Tuholski, Laughlin, & Conway, 1999; Swanson & Kim, 2007) and behavioral inhibition paradigms (Alderson, Rapport, Hudec, Sarver, & Kofler, 2010; Garon, Bryson, & Smith, 2008). Thus, we were unable to rule out the hypothesis that shared working memory demands may account for the homogeneous effect sizes across studies.

Neuroenergetic factors also may account for increased RT variability in ADHD and/or other forms of psychopathology, as predicted by *cognitive neuroenergetic* and *subcortical deficit* models (Table 1). These models appear inconsistent, however, with the finding that individuals with ADHD do not exhibit longer responses latencies (i.e., slower mean RT) after accounting for their increased variability, as well as previous reports of attentional variability indicating that individuals with ADHD do not show performance decrements over time (Rapport, Kofler, Alderson, Timko, & DuPaul, 2009; Sergeant & Scholten, 1983, 1985). In addition, included studies differed considerably in event rate, which is manipulated frequently to examine the impact of arousal and activation as direct tests of the cognitive energetic model (Raymaekers, Antrop, van der Meere, Wiersema, & Roeyers, 2007). Systematic, experimental research is needed to examine the extent to which alternative core deficits can account for and are explained by RT variability in ADHD and other clinical disorders, whether RT variability is indeed a robust construct as evidenced in the present review, or whether additional explanatory models are needed to account for these and other ADHD-related impairments (Table 1).

4.4. Limitations

The unique contribution of the current study was the synthesis and analysis of a large body of literature investigating a potential core deficit of ADHD and comparing children, adolescents, and adults with ADHD to typically developing groups, clinical control groups, and themselves (i.e., subtype comparisons, treatment and motivation effects). Several

caveats require consideration when interpreting the present results despite these and other methodological refinements (e.g., artifact- and publication bias-corrected random effects models, best case estimation). Significant publication bias was detected for most analyses despite concerted efforts to obtain unpublished data, suggesting that a sizeable number of failure-to-replicate studies have been conducted but not published. Although we were able to statistically correct for this trend when computing overall effect sizes for each analysis, missing studies by necessity were not included in moderator analyses, where they may have impacted the significance and magnitude of examined moderators. In addition, the lack of between-study heterogeneity prevented us from directly examining the impact of several planned moderators hypothesized to reflect underlying mechanisms responsible for RT variability. For example, we were unable to investigate the potential impact of tasks emphasizing speed relative to accuracy, and were only able to examine the impact of behavioral inhibition demands in one of the five tiers. Thus, our conclusion that these factors do not impact RT variability in ADHD is based on the finding that the included studies differed systematically on these variables but not in obtained effect sizes, rather than based on a direct test (i.e., there was no heterogeneity for these constructs to explain).

We were unable to examine the relation between medication-related improvements in RT variability and medication-related improvements in other ADHD symptoms; studies examining this hypothesized mediation are needed to determine the extent to which RT variability underlies ADHD behavioral and functional impairments (or vice versa). In addition, we were only able to examine the relation between ADHD symptoms and RT variability indirectly through comparison of ADHD subtypes. Although examination of 36 studies comparing ADHD-C with ADHD-I subtypes suggested that RT variability is somewhat more related to hyperactivity/impulsivity than inattention, only 8 studies with 304 total ADHD-H participants have examined RT variability in this subtype. Similarly, a limited number of studies were available to address other questions. These include the potential periodicity of RT variability in ADHD, the impact of motivation, the extent to which ex-Gaussian statistics are associated with increased magnitude between-group differences, and the extent to which ADHD-related deficits in mean RT are attributable instead to RT variability. Although each analysis was based on several hundred individuals, these results must be considered tentative and require large-scale replication.

4.5. Summary and clinical implications

Collectively, the present meta-analysis of 319 studies revealed that children and adolescents with ADHD demonstrate large magnitude impairments in RT variability relative to their typically developing peers. Contrary to contemporary characterizations of individuals with ADHD as *slower and more variable*, the current meta-analytic findings indicate that ADHD individuals may be better characterized as *more variable but not slower* after accounting for their increased response variability. In addition, this increased variability appears to be attributable primarily to a subset of abnormally slow responses (tau), rather than ubiquitous variability across all trials of a given task (sigma). This increased RT variability decreases somewhat but remains robust in adulthood, with adults with ADHD continuing to demonstrate moderate magnitude increased RT variability. These large and moderate effect sizes reflect the finding that 47% of children/adolescents and 32% of adults with ADHD score outside the typically developing range, indicating that RT variability metrics are unlikely to be useful diagnostically (i.e., to ensure than no typically developing individuals are falsely classified as ADHD, 53% of ADHD children and 68% of ADHD adults would be misclassified as typically developing). In addition, individuals with ADHD are essentially indistinguishable from clinical control groups (i.e., 78% and 100% population overlap between ADHD and clinical control groups of children and adolescents/adults), indicating that RT variability lacks specificity and thus is not a viable diagnostic marker of ADHD. Clinically, these findings add to the disappointing but consistent finding that no laboratory or clinic-based measure/combination of measures has sufficient predictive power to diagnose ADHD (Rapport et al., 2001; Sonuga-Barke, Bitsakou, & Thompson, 2010).

The search for core deficits in ADHD parallels our search for the underlying causes of this potentially debilitating disorder of brain, behavior, and development, and holds promise for identifying novel intervention targets and expanding our understanding of the underlying mechanisms and processes responsible for the hallmark behavioral symptoms and functional impairments associated with ADHD. In the case of RT variability, the current review indicates large magnitude between-group differences that decrease somewhat but remain detectable in adulthood for individuals with ADHD as well as a broad range of clinical disorders. Thus, RT variability reflects a stable feature of ADHD and other clinical disorders that is robust to systematic differences across studies, including inhibitory processes, diagnostic methods, task characteristics, volitional/motivational influences, and metric used to index variability. The robustness of overall effect sizes despite considerable methodological and contextual differences across studies supports strongly a call for additional research identifying the mechanisms and processes responsible for, and attributable to, increased RT variability in ADHD and other clinical disorders. Importantly, different mechanisms and processes may be related to increased RT variability across clinical groups and different forms of psychopathology (i.e., *equifinality*). Future research examining mediators of the relationship between diagnostic status and increased RT variability are needed to address this issue. In addition, large-scale studies are needed that sample a wide variety of clinical and community participants, define groups based on RT variability, and examine commonalities among individuals with high RT variability (e.g., NIMH RDOC criteria: Insel et al., 2010). These investigations will be critical for addressing the extent to which RT variability is associated with particular neurological, neurocognitive, behavioral, and functional predictors and outcomes for individuals with a broad range of psychopathology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cpr.2013.06.001.

References[7]

Aase, H., Meyer, A., & Sagvolden, T. (2006). Moment-to-moment dynamics of ADHD behavior in South American children. *Behavioral and Brain Functions*, *2*, 11.

Aggarwal, A., & Lillystone, D. (2000). A follow-up pilot study of objective measures in children with attention deficit hyperactivity disorder. *Journal of Pediatrics and Child Health*, *36*, 134–138.

Alderson, R. M., Rapport, M. D., Hudec, K. L., Sarver, D. E., & Kofler, M. J. (2010). Competing core processes in attention-deficit/hyperactivity disorder (ADHD): Do working memory deficiencies underlie behavioral inhibition deficits? *Journal of Abnormal Child Psychology*, *38*, 497–507.

Alderson, R. M., Rapport, M. D., & Kofler, M. J. (2007). Attention-deficit/hyperactivity disorder and behavioral inhibition: A meta-analytic review of the stop-signal paradigm. *Journal of Abnormal Child Psychology*, *35*, 745–758.

Andreou, P., Neale, B. M., Chen, W., Christiansen, H., Gabriels, I., Heise, A., et al. (2007). Reaction time performance in ADHD: Improvement under fast-incentive condition and familial effects. *Psychological Medicine*, *37*, 1703–1715.

American Psychiatric Association. (2000). *American psychiatric association: Diagnostic and statistical manual of mental disorders* (4th ed.) Washington, DC.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*, 65–94.

Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2006). Young adult outcome of hyperactive children: Adaptive functioning in major life activities. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 192–202.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088–1101.

[7] Studies included in the meta-analysis but not cited in the text are listed in the Appendix and Tables S1-S8

Bellgrove, M. A., Hawi, Z., Kirley, A., Gill, M., & Robertson, I. H. (2005). Dissecting the attention deficit hyperactivity disorder (ADHD) phenotype: Sustained attention, response variability and spatial attentional asymmetries in relation to dopamine transporter (DAT1) genotype. *Neuropsychologia*, *43*, 1847–1857.

Bellgrove, M. A., Hester, R., & Garavan, H. (2004). The functional neuroanatomy of response variability: Evidence from a response inhibition task. *Neuropsychologia*, *42*, 1910–1916.

Biederman, J., & Spencer, T. (1999). Attention-deficit/hyperactivity disorder (ADHD) as a noradrenergic disorder. *Biological Psychiatry*, *46*, 1234–1242.

Burgess, G. C., Depue, B. E., Ruzic, L., Willcutt, E. G., Du, Y. P., & Banich, M. T. (2010). Attentional control activation relates to working memory in ADHD. *Biological Psychiatry*, *67*, 632–640.

Buzy, W. M., Medoff, D. R., & Schweitzer, J. B. (2009). Intra-individual variability among children with ADHD on a working memory task: An ex-Gaussian approach. *Child Neuropsychology*, *15*, 441–459.

Castellanos, F. X., Sonuga-Barke, E. J. S., Milham, M. P., & Tannock, R. (2006). Characterizing cognition in ADHD: Beyond executive dysfunction. *Trends in Cognitive Sciences*, *10*, 117–123.

Castellanos, F. X., Sonuga-Barke, E. J. S., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of ADHD-related intra-individual variability. *Biological Psychiatry*, *57*, 1416–1423.

Castellanos, F. X., & Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: The search for endophenotypes. *Nature Reviews Neuroscience*, *3*, 617–628.

Clarke, S. D., Kohn, M. R., Hermens, D. F., Rabbinge, M., Clark, C. R., Gordon, E., et al. (2007). Distinguishing symptom profiles in adolescent ADHD using an objective cognitive test battery. *International Journal of Adolescent Mental Health*, *19*, 355–367.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, N. J., & Douglas, V. J. (1972). Characteristics of the orienting response in hyperactive and normal children. *Psychophysiology*, *9*, 238–245.

Comprehensive Meta-Analysis (Version 2.2) [Computer software]. Englewood, NJ: Biostat.

Conners, C. K., & MHS Staff (Eds.). (2000). *Conners' Continuous Performance Test II: Computer Program for Windows Technical Guide and Software Manual*. North Tonawanda, NY: Multi-Health Systems.

Conners, C. K., & MHS Staff (Eds.). (2001). *Conners' Kiddie Continuous Performance Test Version 5: Computer Program for Windows Technical Guide and Software Manual*. North Tonawanda, NY: Multi-Health Systems.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*, 170–180.

Douglas, V. I., & Parry, P. A. (1983). Effects of reward on delayed reaction time task performance of hyperactive children. *Journal of Abnormal Child Psychology*, *11*, 313–326.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.

Eckert, H. M., & Eichorn, D. H. (1977). Developmental variability in reaction time. *Child Development*, *48*, 452–458.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid IQ: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*, 309–331.

Epstein, J. N., Brinkman, W. B., Froehlich, T., Langberg, J. M., Narad, M. E., Antonini, T. N., et al. (2011). Effects of stimulant medication, incentives, and event rate on reaction time variability in children with ADHD. *Neuropsychopharmacology*, *36*, 1060–1072.

Epstein, J. N., Conners, C. K., Hervey, A. S., Tonev, S. T., Arnold, L. E., Abikoff, H. B., et al. (2006). Assessing medication effects in the MTA study using neuropsychological outcomes. *Journal of Child Psychology and Psychiatry*, *47*, 446–456.

Epstein, J. N., Erkanli, A., Conners, C. K., Klaric, J., Costello, J. E., & Angold, A. (2003). Relations between continuous performance test performance measures and ADHD behaviors. *Journal of Abnormal Child Psychology*, *31*, 543–554.

Epstein, J. N., Hwang, M. E., Antonini, T., Langberg, J. M., Altaye, M., & Arnold, E. (2010). Examining predictors of reaction times in children with ADHD and normal controls. *Journal of the International Neuropsychological Society*, *16*, 138–147.

Faraone, S. V., Biederman, J., Spencer, T. J., & Aleardi, M. (2006). Comparing the efficacy of medications for ADHD using meta-analysis. *Medscape General Medicine*, *8*, 4.

Finke, K., Schwarzkopf, W., Müller, U., Frodl, T., Müller, H. J., Schneider, W. X., et al. (2011). Disentangling the adult attention-deficit hyperactivity disorder endophenotype: Parametric measurement of attention. *Journal of Abnormal Psychology*, *120*, 890–901.

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, *18*, 543–555.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31–60.

Geurts, H. M., Grasman, R. P. P. P., Verté, S., Oosterlaan, J., Roeyers, H., van Kammen, S. M., et al. (2008). Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome. *Neuropsychologia*, *46*, 3030–3041.

Halperin, J. J., & Schulz, K. P. (2006). Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder. *Psychological Bulletin*, *132*, 560–581.

Halperin, J. M., Trampush, J. W., Miller, C. J., Marks, D. J., & Newcorn, J. H. (2008). Neuropsychological outcome in adolescents/young adults with childhood ADHD: Profiles of persisters, remitters and controls. *Journal of Child Psychology and Psychiatry*, *49*, 958–966.

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, *9*, 426–445.

Heiser, P., Frey, J., Smidt, J., Sommerlad, C., Wehmeier, P. M., Hebebrand, J., et al. (2004). Objective measurement of hyperactivity, impulsivity, and inattention in children with hyperkinetic disorders before and after treatment with methylphenidate. *European Child & Adolescent Psychiatry*, *13*, 100–104.

Hervey, A. S., Epstein, J. N., & Curry, J. F. (2004). Neuropsychology of adults with attention-deficit/hyperactivity disorder: A meta-analytic review. *Neuropsychology*, *18*, 485–503.

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Arnold, L. E., Conners, C. K., et al. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology*, *12*, 125–140.

Hicks, R. E., Mayo, J. P., Jr., & Clayton, C. J. (1989). Differential psychopharmacology of methylphenidate and the neuropsychology of childhood hyperactivity. *International Journal of Neuroscience*, *45*, 7–32.

Howard, G. S., Jr., Maxwell, S. E., & Fleming, K. J. (2000). The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods*, *5*, 315–332.

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., & Moore, A. N. (2012). Evaluating vigilance deficits in ADHD: A meta-analysis of CPT performance. *Journal of Abnormal Psychology*, *121*, 360–371.

Huang-Pollock, C. L., & Nigg, J. T. (2003). Searching for the attention deficit in attention deficit hyperactivity disorder: The case of visuospatial orienting. *Clinical Psychology Review*, *23*, 801–830.

Huang-Pollock, C. L., Nigg, J. T., & Carr, T. H. (2005). Deficient attention is hard to find: Applying the perceptual load model of selective attention to attention deficit hyperactivity disorder subtypes. *Journal of Child Psychology and Psychiatry*, *46*, 1211–1218.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria: Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, *167*, 748–751.

Jacobson, L. A., Ryan, M., Martin, R. B., Ewen, J., Mostofsky, S. H., & Mahone, E. M. (2011). Working memory influences processing speed and reading fluency in ADHD. *Child Neuropsychology*, *17*, 209–224.

Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.

Johnson, K. A., Kelly, S. P., Bellgrove, M. A., Barry, E., Cox, M., Gill, M., et al. (2007). Response variability in attention deficit hyperactivity disorder: Evidence for neuropsychological heterogeneity. *Neuropsychologia*, *45*, 630–638.

Kasper, L. J., Alderson, R. M., & Hudec, K. L. (2012). Moderators of working memory deficits in children with ADHD: A meta-analytic review. *Clinical Psychology Review* (online early).

Klein, C., Wendling, K., Huettner, P., Ruder, H., & Peper, M. (2006). Intra-subject variability in attention-deficit hyperactivity disorder. *Biological Psychiatry*, *60*, 1088–1097.

Kofler, M. J., Rapport, M. D., & Alderson, R. M. (2008). Quantifying ADHD classroom inattentiveness, its moderators, and variability: A meta-analytic review. *Journal of Child Psychology and Psychiatry*, *49*, 59–69.

Kofler, M. J., Rapport, M. D., Bolden, J., Sarver, D. E., & Raiker, J. S. (2010). ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behavior. *Journal of Abnormal Child Psychology*, *38*, 149–161.

Kuntsi, J., Wood, A. C., van der Meere, J., & Asherson, P. (2009). Why cognitive performance in ADHD may not reveal true potential: Findings from a large population-based sample. *Journal of the International Neuropsychological Society*, *15*, 570–579.

Lansbergen, M. M., Kenemans, J. L., & van Engeland, H. (2007). Stroop interference and attention-deficit/hyperactivity disorder: A review and meta-analysis. *Neuropsychology*, *21*, 251–262.

Leark, R. A., Greenberg, L. M., Kindschi, C. L., Dupuy, T. R., & Hughes, S. J. (2007). *TOVA Professional Manual: Test of Variables of Attention Continuous Performance Test*. Los Alamitos, CA: TOVA Company.

Leth-Steensen, C., Elbaz, Z. K., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica*, *104*, 167–190.

Lijffijt, M., Kenemans, L., Verbaten, M. N., & van Engeland, H. (2005). A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: Deficient inhibitory motor control? *Journal of Abnormal Psychology*, *114*, 216–222.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. California: Sage.

Lipszyc, J., & Schachar, R. (2010). Inhibitory control and psychopathology: A meta-analysis using the stop signal task. *Journal of the International Neuropsychological Society*, *16*, 1064–1076.

Losier, B. J., McGrath, P., & Klein, R. M. (1996). Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: A meta-analytic review. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *37*, 971–987.

MacDonald, S. W., Nyberg, L., & Backman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*, *29*, 474–480.

Mannuzza, S., Klein, R. G., Bessler, A., Malloy, P., & LaPadula, M. (1993). Adult outcome of hyperactive boys. *Archives of General Psychiatry*, *50*, 565–576.

Martinussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 377–384.

National Advisory Mental Health Council's Workgroup (2010). *From discovery to cure: Accelerating the development of new and personalized interventions for mental illnesses*. Rockville, MD: NAMHC.

O'Brien, J. D., Halperin, J. M., Newcorn, J. H., Sharma, V., Wolf, L., & Morganstein, A. (1992). Psychometric differentiation of conduct disorder and attention deficit disorder with hyperactivity. *Developmental and Behavioral Pediatrics*, *13*, 274–277.

Oosterlaan, J., Logan, G. D., & Sergeant, J. A. (1998). Response inhibition in AD/HD, CD, comorbid AD/HD + CD, anxious, and control children: A meta-analysis of studies with the stop task. *Journal of Child Psychology and Psychiatry*, *39*, 411–425.

Oosterlaan, J., & Sergeant, J. A. (1995). Response choice and inhibition in ADHD, anxious and aggressive children: The relationship between S–R compatibility and stop signal task. In J. A. Sergeant (Ed.), *Eunethydis: European approaches to hyperkinetic disorder* (pp. 225–240). Amsterdam: University of Amsterdam.

Oosterlaan, J., & Sergeant, J. A. (1998). Response inhibition and response re-engagement in attention-deficit/hyperactivity disorder, disruptive, anxious and normal children. *Behavioural Brain Research*, *94*, 33–43.

Pelham, W. E., Foster, E. M., & Robb, J. A. (2007). The economic impact of attention-deficit/hyperactivity disorder in children and adolescents. *Journal of Pediatric Psychology*, *32*, 711–727.

Raiker, J. S., Rapport, M. D., Kofler, M. J., & Sarver, D. E. (2012). Objectively-measured impulsivity and attention-deficit/hyperactivity disorder (ADHD): Testing competing predictions from the working memory and behavioral inhibition models of ADHD. *Journal of Abnormal Child Psychology*, *40*, 699–713.

Rapport, M. D., Alderson, R. M., Kofler, M. J., Sarver, D. E., Bolden, J., & Sims, V. (2008). Working memory deficits in boys with attention-deficit/hyperactivity disorder (ADHD): The contribution of central executive and subsystem processes. *Journal of Abnormal Child Psychology*, *36*, 825–837.

Rapport, M. D., Chung, K., Shore, G., & Isaacs, P. (2001). A conceptual model of child psychopathology: Implications for understanding attention deficit hyperactivity disorder and treatment efficacy. *Journal of Clinical Child Psychology*, *30*, 48–58.

Rapport, M. D., Denney, C. B., DuPaul, G. J., & Gardner, M. J. (1994). Attention deficit disorder and methylphenidate: Normalization rates, clinical effectiveness, and response prediction in 76 children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 882–893.

Rapport, M. D., Kofler, M. J., Alderson, R. M., Timko, T., Jr., & DuPaul, G. J. (2009). Variability of attention processes in ADHD: Observations from the classroom. *Journal of Attention Disorders*, *12*, 563–573.

Raymaekers, R., Antrop, I., van der Meere, J. J., Wiersema, J. R., & Roeyers, H. (2007). HFA and ADHD: A direct comparison on state regulation and response inhibition. *Journal of Clinical and Experimental Neuropsychology*, *29*, 418–427.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*, 183–192.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., Jr., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*, 343–350.

Russell, V. A., Oades, R. D., Tannock, R., Killeen, P. R., Auerbach, J. G., & Sagvolden, T. (2006). Response variability in ADHD: A neuronal and glial energetics hypothesis. *Behavior & Brain Functions*, *2*, 30–45.

Sagvolden, T., Aase, H., Johansen, E. B., & Russell, V. A. (2005). A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behavioral and Brain Sciences*, *28*, 397–468.

Saville, C. W. N., Pawling, R., Trullinger, M., Daley, D., Intriligator, J., & Klein, C. (2011). On the stability of instability: Optimizing the reliability of intra-subject variability of reaction times. *Personality and Individual Differences*, *51*, 148–153.

Scheres, A., Milham, M. P., Knutson, B., & Castellanos, F. X. (2007). Ventral striatal hyporesponsiveness during reward anticipation in attention-deficit/hyperactivity disorder. *Biological Psychology*, *61*, 720–724.

Scheres, A., Oosterlaan, J., & Sergeant, J. A. (2001). Response inhibition in children with DSM-IV subtypes of AD/HD and related disruptive disorders: The role of reward. *Child Neuropsychology*, *7*, 172–189.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1089–1096.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süb, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology. General*, *136*, 414–429.

Sergeant, J. A. (2005). Modeling attention-deficit/hyperactivity disorder: A critical appraisal of the cognitive-energetic model. *Biological Psychiatry*, *57*, 1248–1255.

Sergeant, J. A., & Scholten, C. A. (1983). A stages-of-information approach to hyperactivity. *Journal of Child Psychology and Psychiatry*, *24*, 49–60.

Sergeant, J. A., & Scholten, C. A. (1985). On data limitations in hyperactivity. *Journal of Child Psychology and Psychiatry*, *26*, 111–124.

Shanahan, M. A., Pennington, B. F., & Willcutt, E. W. (2008). Do motivational incentives reduce the inhibition deficit in ADHD? *Developmental Neuropsychology*, *33*, 137–159.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.

Sonuga-Barke, E. J. S., Bitsakou, P., & Thompson, M. (2010). Beyond the dual pathway model: Evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 345–355.

Sonuga-Barke, E. J. S., & Castellanos, F. X. (2007). Spontaneous attentional fluctuations in impaired states and pathological conditions: A neurobiological hypothesis. *Neuroscience and Biobehavioral Reviews*, *31*, 977–986.

Spencer, S. V., Hawk, L. W., Jr., Richards, J. B., Shiels, K., Pelham, W. E., Jr., & Waxmonsky, J. G. (2009). Stimulant treatment reduces lapses in attention among children with ADHD: The effects of methylphenidate on intra-individual response time distributions. *Journal of Abnormal Child Psychology*, *37*, 805–816.

Swanson, H. L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, *35*, 151–168.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham, MA: Allyn & Bacon.

Tamm, L., Narad, M. E., Antonini, T. N., O'Brien, K. M., Hawk, L. W., Jr., & Epstein, J. N. (2012). Reaction time variability in ADHD: A review. *Neurotherapeutics*, *9*, 500–508.

Teicher, M. H., Lowen, S. B., Polcari, A., Foley, M., & McGreenery, C. E. (2004). Novel strategy for the analysis of CPT data provides new insight into the effects of methylphenidate on attentional states in children with ADHD. *Journal of Child and Adolescent Psychopharmacology*, *14*, 219–232.

Tillman, C. M., Thorell, L. B., Brocki, K. C., & Bohlin, G. (2008). Motor response inhibition and execution in the stop-signal task: Development and relation to ADHD behaviors. *Child Neuropsychology*, *14*, 42–59.

U.S. Census Bureau (2010). *Annual estimates of the population by sex and 5-year age groups for the United States: 2010 Census Summary File 1 (NC-EST2010)* (Retrieved June 2, 2012, from http://www.census.gov).

van der Meere, J., Gunning, B., & Stemerdink, N. (1999). The effect of MPH and clonidine on response inhibition and state regulation in children with ADHD. *Journal of Child Psychology and Psychiatry*, *2*, 291–298.

van der Meere, J., & Sergeant, J. (1987). A divided attention experiment with pervasively hyperactive children. *Journal of Abnormal Child Psychology*, *15*, 379–392.

van der Oord, S., Prins, P. J. M., Oosterlaan, J., & Emmelkamp, P. M. G. (2008). Efficacy of methylphenidate, psychosocial treatments and their combination in school-aged children with ADHD: A meta-analysis. *Clinical Psychology Review*, *28*, 783–800.

van Mourik, R., Oosterlaan, J., & Sergeant, J. A. (2005). The Stroop revisited: A meta-analysis of interference control in AD/HD. *Journal of Child Psychology and Psychiatry*, *46*, 150–165.

Verté, S., Geurts, H. M., Roeyers, H., Oosterlaan, J., & Sergeant, J. A. (2006). The relationship of working memory, inhibition, and response variability in child psychopathology. *Journal of Neuroscience Methods*, *151*, 5–14.

Wagenmakers, E. -J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*, 830–841.

Wåhlstedt, C., Thorell, L. B., & Bohlin, G. (2009). Heterogeneity in ADHD: Neuropsychological pathways, comorbidity and symptom domain. *Journal of Abnormal Child Psychology*, *37*, 551–564.

Willcutt, E. G., Pennington, B. F., Olson, R. K., Chhabildas, N. A., & Hulslander, J. (2005). Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: In search of the common deficit. *Developmental Neuropsychology*, *27*, 35–78.

Willcutt, E. G., Sonuga-Barke, E. J. S., Nigg, J. T., & Sergeant, J. A. (2008). Recent developments in neuropsychological models of childhood psychiatric disorders. *Advances in Biological Psychiatry*, *24*, 195–226.

Williams, B. R., Hultsch, D. F., Strauss, E. H., Hunter, M. A., & Tannock, R. (2005). Inconsistency in reaction time across the lifespan. *Neuropsychology*, *19*, 88–96.

Youngstrom, E. A., Arnold, L. E., & Frazier, T. W. (2010). Bipolar and ADHD comorbidity: Both artifact and outgrowth of shared mechanisms. *Clinical Psychology: Science and Practice*, *17*, 350–359.

Zakzanis, K. K. (2001). Statistics to tell the truth, the whole truth, and nothing but the truth: Formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. *Archives of Clinical Neuropsychology*, *16*, 653–667.