Review Article

Clinical Characteristics Associated With Stuttering Persistence: A Meta-Analysis

Cara M. Singer,[a] Alison Hessling,[b] Ellen M. Kelly,[c] Lisa Singer,[d] and Robin M. Jones[c]

Purpose: The purpose of this meta-analytic study was to identify clinical characteristics, defined as child factors that can be assessed by a speech-language pathologist as part of a routine speech-language evaluation that may differentiate children who persist in stuttering from children who eventually recover from stuttering. Clinical characteristics explored included sex, age at onset, family history of stuttering, stuttering frequency and severity, speech-language skills, and temperament.
Method: Studies were identified through electronic databases, journals, and reference lists of relevant reports (e.g., research articles). Eligible studies followed young children who stutter (i.e., under 6 years old) for at least 24 months, assessed a potential clinical marker at study entry, and determined talker group classification (i.e., persistent or recovered) at study

completion. Sex and family history differences were estimated using risk ratios; all other differences were estimated using Hedges's *g*. Heterogeneity and methodological differences among studies were evaluated.
Results: Eleven studies (41 reports) met eligibility criteria. Persistent children were older at stuttering onset and exhibited higher frequencies of stuttering-like disfluencies, lower speech sound accuracy, and lower expressive and receptive language skills than recovered children. Males and children with a family history of stuttering were also more likely to persist.
Conclusions: Clinical characteristics were identified that are associated with increased risk for stuttering persistence. Future studies have the potential to translate these clinical characteristics into prognostic markers for stuttering persistence risk.

Approximately 5%–8% of preschool-aged children (e.g., children aged 2;0–5;11 [years;months]) exhibit a childhood onset fluency (or stuttering) disorder (e.g., Månsson, 2000; Reilly et al., 2009; for a review, see Yairi & Ambrose, 2013). Stuttering is a neurodevelopmental communication disorder characterized by stuttering-like disfluencies (e.g., sound–syllable repetitions, whole-word repetitions, and prolongations; Tumanova et al., 2014). Whereas diagnostic criteria for stuttering often vary across

studies (for a discussion, see Gordon & Luper, 1992), criteria typically include an elevated frequency of stuttering-like disfluencies (i.e., 3% frequency of stuttered disfluencies or above) and/or parent concern for stuttering. The onset of childhood stuttering typically occurs when a child is between 2 and 5 years of age (Månsson, 2000; Yairi & Ambrose, 2013). Within 5 years of stuttering onset, approximately 80% of children fall below the diagnostic threshold for stuttering, thereby exhibiting recovery (Yairi & Ambrose, 2013, 2005). Yairi and Ambrose (2005) reported that, of the children who recovered within 5 years of onset, 26% recovered within 18 months, 40% recovered within 2 years, and 80% recovered within 3 years.

Although it is just one of many factors to consider when making treatment decisions, the high rate of recovery in childhood stuttering has influenced decisions of *when* and *how* to treat young children who stutter (e.g., Bernstein Ratner, 2018; Nippold, 2018). Onslow and Packman (1999) referred to the interaction between recovery and treatment procedures as "*the* most pressing issue for modern clinicians

[a]Department of Communication Sciences and Disorders, Grand Valley State University, Grand Rapids, MI
[b]Department of Communication Sciences and Disorders, Baylor University, Waco, TX
[c]Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN
[d]Department of Radiation Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Boston, MA

Correspondence to Cara M. Singer: singerca@gvsu.edu

who have clinical contact with children who begin to stutter" (p. 114). It has long been recommended that speech-language pathologists (SLPs) consider both the length of time the child has been stuttering and the child's risk for persistence when making treatment decisions for young children (e.g., Kelman & Nicholas, 2008; Zebrowski, 1997). However, assessing a child's risk for persistence is dependent on the individual SLP's assessment of risk (i.e., prognostic) factors of stuttering chronicity within a clinical context.

To date, evidence for potential risk factors has stemmed from prospective cohort studies in which young children who stuttered were followed for a period of time and classified as either recovered or persistent (e.g., Ambrose et al., 2015; Spencer & Weber-Fox, 2014; Yairi & Ambrose, 2005). The multifactorial nature of developmental stuttering (for multifactorial perspectives on stuttering, see Conture & Walden, 2012; A. Smith & Weber, 2017) and phenomena that differentiate children who stutter from children who do not stutter (e.g., sex differences) have motivated the selection of characteristics explored in longitudinal studies. Overall, important contributors to stuttering include linguistic, speech-motor, temperament, neurological, and genetic factors. Accordingly, these contributors have been explored for their prognostic utility using a variety of approaches, including collection of case history information (e.g., Yairi & Ambrose, 2005), physiological measurement (e.g., Zengin-Bolatkale et al., 2018), brain morphometric characteristics (Garnett et al., 2018), and behavioral measures (e.g., Ambrose et al., 2015). However, the ability of practicing SLPs to assess and interpret these factors varies (i.e., clinical utility).

Clinical characteristics, defined as child factors that can be assessed by an SLP as part of routine clinical care, are of particular interest in this study because they have high clinical utility and are often included as part of a comprehensive speech-language evaluation for stuttering (for an example of an evaluation protocol for developmental stuttering, see Clark et al., 2017). Examples of clinical characteristics include features of the child such as sex, family history of stuttering, age at stuttering onset, demonstrated and described stuttering behaviors, speech-language skills, and temperament.

Findings from prospective cohort studies on these clinical characteristics have been widely inconsistent; in fact, there are both significant and nonsignificant findings between persistent and recovered children in all of these areas. For example, whereas males have historically been considered to be at greater risk to persist, many studies have found statistically nonsignificant differences related to risk between males and females (cf. Dworzynski et al., 2007; Kefalianos et al., 2017). Similarly, there have been inconsistent (and nonsignificant) findings of differences between persistent and recovered children from studies that reported on age at onset (Ambrose et al., 2015; cf. Dworzynski et al., 2007; Roehl, 2018), the presence of a positive family history of stuttering (Kefalianos et al., 2017; Rommel et al., 2001; cf. Yairi & Ambrose, 2005), stuttering behaviors (e.g., Ambrose et al., 2015; Garnett et al., 2018; cf. Yairi &

Ambrose, 2005), speech and language skills (e.g., Ryan, 2001; Singer et al., 2019; cf. Spencer & Weber-Fox, 2014), and temperament/emotion (e.g., Zengin-Bolatkale et al., 2018; cf. Ambrose et al., 2015). It should also be noted that some of these characteristics (e.g., temperament/emotion) are in their relative infancy of empirical study relative to persistence; therefore, there is relatively little data currently available.

Researchers often cite the small sample sizes of many prospective cohorts as contributing to nonsignificant findings (e.g., Singer et al., 2019; Spencer & Weber-Fox, 2014). Furthermore, inconsistent findings have often been attributed to other methodological differences across studies, such as the diagnostic criteria for stuttering and the age of children at study entry (Ambrose et al., 2015), but these differences between studies have not been empirically explored.

There have been previous attempts to synthesize the literature for clinical application (Clark et al., 2017; Guitar & Conture, 2006; Walsh et al., 2018) that have identified many of these clinical characteristics (e.g., sex, age at onset, family history of stuttering) as prognostic indicators, but no one has *quantitatively* synthesized the empirical data in order to advance our understanding of characteristics associated with stuttering persistence. Therefore, given the lack of a quantitative analysis and previously discussed inconsistencies in the literature, a systematic synthesis of the evidence is warranted to identify which of these variables might be indicative of increased risk for stuttering persistence. Furthermore, an evaluation of methodological characteristics across prospective studies is warranted to identify and select studies that meet rigorous standards and to explore differences between studies that might moderate findings. In the absence of a systematic review, the practitioner (e.g., SLP) is left to subjectively synthesize the knowledge base without sufficient empirical support for assessment and treatment decisions related to stuttering persistence.

To address these limitations, this study evaluated the relation of clinical characteristics to stuttering persistence using a meta-analytic approach. We used a meta-analysis to quantify statistical estimates of overall differences in clinical characteristics between persistent and recovered children as well as assess whether methodological characteristics moderate differences between studies. We focused on clinical characteristics that are identifiable when a child is early in stuttering development (i.e., younger than 6 years of age) and often included in initial evaluations of stuttering in order to assess their potential prognostic values. This study also assessed risk of bias within, and the heterogeneity between, the individual primary prognostic studies. We addressed two research questions:

1. What sources of bias are frequently found in prospective cohort studies of children who stutter that might impact study findings?

2. Are there clinical characteristics that differentiate children who persist in stuttering and children who eventually recover from stuttering in early childhood?

Method

This study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher et al., 2009) guidelines. The methods were registered online with PROSPERO (CRD42019118590).

Eligibility Criteria

Reports of prospective cohort studies were included in this study if they met the following criteria. First, studies must have been longitudinal in nature, following participants for at least 24 months to increase the likelihood that children were correctly identified as recovered. Second, participants must have been under 6 years of age at study entry (to target children near stuttering onset). Third, assessments must have been used to identify participants (a) as stuttering at study entry and (b) as either recovered or persistent at study completion. Fourth, at least one clinical characteristic must have been assessed at study entry when all participants were stuttering and reported separately for the persistent and recovered groups.

Inclusionary criteria for clinical characteristics were based on clinical utility. Specifically, we included data collected from caregiver reports (e.g., interviews, questionnaires, scales), normative assessments, children's speech samples, and child reports, all of which are commonly used in assessments of children who stutter (e.g., Clark et al., 2017). Data that require special equipment and are not commonly used by clinicians in the assessment of childhood stuttering, such as physiological, brain morphometry, and genetic analyses (except for family history and pedigree analyses), were excluded.

Eligible *studies* (i.e., investigations of the same cohort of participants) were identified by searching and evaluating associated *reports* (published articles, doctoral dissertations, books, etc.) that resulted from those studies. Individual reports were eligible for inclusion if they included quantitative results of a potential clinical characteristic of stuttering persistence that permitted the calculation of an effect size. Multiple reports from the same study could be included for different clinical characteristics. There were no publication year or language restrictions.

Search Strategy

Studies were identified by several means to minimize publication bias and maximize the likelihood of finding all relevant studies. PsycINFO, ProQuest Dissertations, and PubMed were searched in January 2020 using the following search terms: Ti,ab ((stammer* OR stutter*) AND (recover* OR persist* OR longitudinal) AND child*). The first author also searched websites of selected journals (i.e., *Journal of Fluency Disorders, Journal of Speech, Language, and Hearing Research, Journal of Communication Disorders, American Journal of Speech-Language Pathology, International Journal of Speech-Language Pathology, International Journal of Language & Communication Disorders*) and the American Speech-Language-Hearing Association using key terms. Authors were contacted if additional information was needed.

Study Selection

Research reports from these search methods were narrowed during a two-step double-screened review. The first author and a reliability coder reviewed the title and abstract of each report using Microsoft Excel. Any report that compared children who stutter that persisted (i.e., Persistent group) and children who stutter that recovered (i.e., Recovered group) was included in the next stage of the review process. In this second stage, the first and second authors read each report to identify whether it was eligible or to identify reason(s) for ineligibility. During both steps, disagreements were discussed until a consensus was reached. Lastly, the reference lists of included reports were searched to identify additional eligible reports.

Data Management

For all reports that met inclusionary criteria, the first and second authors extracted and entered all data in a Microsoft Excel spreadsheet. Data extracted from reports included author, year of publication, funding source, journal, language, document type, recruitment source, eligibility requirements, time-since-onset criteria at study entry, group classification criteria, length of follow-up, percent participants treated, criteria for length of recovery, sample sizes, mean ages, sex distributions of the persistent and recovered groups, reporting of attrition, and methods to control for confounding variables. The coders then recorded data for each clinical characteristic reported in each study. When data were reported for multiple visits, only data obtained closest to stuttering onset were extracted.[1] When multiple measures of a characteristic were reported (e.g., clinician-reported stuttering severity and parent-reported stuttering severity), all data were extracted.

Risk of Bias Assessment

To explore the first research question, risk of bias stemming from study characteristics was assessed with the Quality In Prognosis Studies tool (Hayden et al., 2013). For each study, the first and second authors used the Quality In Prognosis Studies to individually assign ratings of low, moderate, or high risk of bias to each of six domains: study participants, prognostic factor measurement, study attrition, outcome measurement, study confounding, and statistical analysis and reporting. As described in Hayden et al. (2013), the reviewers evaluated characteristics related to each domain (participant inclusionary requirements for the study participants domain, validity of assessment measure used for prognostic factor measurement, etc.) when

---

[1]The only exception to this rule was data extracted from Yairi and Ambrose (2005). Data collected less than 12 months post onset were extracted instead of data collected 0–6 months post onset because fewer children were observed at the earlier time point.

assigning the level of bias. Study confounding was assessed for explorations of speech-language skills and temperament, for which confounding factors are known (e.g., sex; Barbu et al., 2015; Chentsova-Dutton & Tsai, 2007), but not for studies that only assessed sex and/or family history. The first and second authors made a conservative a priori decision that a study's risk of bias for outcome measurement would be at least moderate if participants were followed for less than 3 years. For any disagreement, the coders discussed the domain until a consensus was reached. An a priori decision was made that studies rated as being at high risk for bias would be excluded from the meta-analysis.

Furthermore, characteristics related to the risk of bias were assessed for each report. When two reports with participant overlap reported similar data, data from the report with the lowest level of bias across the six domains were extracted for analysis. For example, there was considerable participant overlap in Roehl (2018) and Spencer and Weber-Fox (2014). Data from Roehl were extracted for receptive vocabulary, over data from Spencer and Weber-Fox, due to group classification being based on longer length of follow-up (i.e., lower risk of bias related to the outcome measurement).

Effect Size Synthesis

Data were analyzed using the *metafor* package in R. Effect sizes were calculated for each individual clinical characteristic (e.g., speech sound skills, receptive language) that was measured in at least two studies. Separate measures of an individual characteristic (e.g., clinician-reported stuttering severity and parent-reported stuttering severity) were analyzed independently whenever possible. Thus, multiple effect sizes were sometimes calculated from a single study.

To explore our second research question, overall effect sizes were synthesized using a random effects model. A random effects model assumes there is between- and within-study heterogeneity that results in multiple true population effect sizes. Effect sizes from individual studies were weighted using inverse variance weights that are sample-size dependent (i.e., studies with greater sample sizes receive more weight). Risk ratios were calculated to estimate differences related to sex and the presence/absence of a known family history of stuttering. Hedges's *g* (Hedges, 1982), standardized mean difference comparisons that correct for small sample sizes, were calculated to estimate differences related to continuous measures (e.g., speech-language scores).

Homogeneity tests were conducted using the Q, I^2, and τ^2 statistics. The Q-statistic reflects the amount of heterogeneity; I^2 and τ^2 reflect the *proportion* and the *amount* of true heterogeneity, respectively. I^2 values of 25%, 50%, and 75% are considered as low, moderate, and high proportions of heterogeneity (Higgins et al., 2003).

Publication Bias Assessment

Lastly, publication bias was evaluated to determine whether findings may be influenced by including only studies that reported larger-than-average effects, which are more likely to be published. When at least 10 studies were included in a meta-analysis of a given clinical characteristic, publication bias was investigated by creating funnel plots and conducting an Egger regression test for funnel plot asymmetry (Egger et al., 1997).

Results

Study Selection

Study (i.e., investigations of the same cohort of participants) selection results are detailed in Figure 1. From an initial pool of 405 reports, 78 reports passed the title and abstract review. The interrater reliability index, prior to reaching consensus, was .93 for this stage of the process. Exclusionary criteria were evaluated in the order listed in Figure 1. Reasons individual reports were excluded at the full-text review stage are provided in Appendix A. The full-text review resulted in 41 reports meeting eligibility criteria; interrater reliability index, prior to reaching consensus, was .94 for the full-text stage. After identifying these 41 reports, we found they derived from 11 unique studies. Descriptions of these 11 studies and their associated 41 reports may be found in Table 1. Studies are identified based on names designated by researchers within reports (e.g., Vanderbilt Developmental Stuttering Project, Twins Early Development Study) or, when a name was not available, based on location (e.g., Erasmus University Medical Centre Study, California – Long Beach Study).

Risk of Bias Assessment

Results pertaining to the risk of bias assessment for each study (i.e., Research Question 1) are provided in Figure 2. The most common sources of bias identified across studies included study participation that limited generalizability of findings based on inclusionary/exclusionary criteria; limited follow-up (i.e., less than 3 years), which could impact group classification; insufficient reports of attrition; and lack of reports on methods to control for the effect of known confounding variables. Four of the studies (36%) had eligibility criteria that negatively impacted how representative the sample was of the general population of children who stutter. Specifically, Arenas et al. (2017) only included participants with hearing loss, Kloth et al. (1999) only included children who had at least one parent with a history of stuttering, Dworzynski et al. (2007) only included twins, and the Vanderbilt Developmental Stuttering Project (Erdemir et al., 2018; Singer et al., 2019, 2020; Zengin-Bolatkale et al., 2018) only included children who scored above the 17th percentile on assessments of speech and language skills (excluding stuttering). Of the seven studies that reported data on attrition, none reported between-group analyses of differences between children who did and did not complete the study. Last, of the nine studies that explored the prognostic value of speech-language skills and/or temperament, six studies did not consistently report methods used to control for known confounding factors.

Figure 1. Flow chart depicting selection of studies.
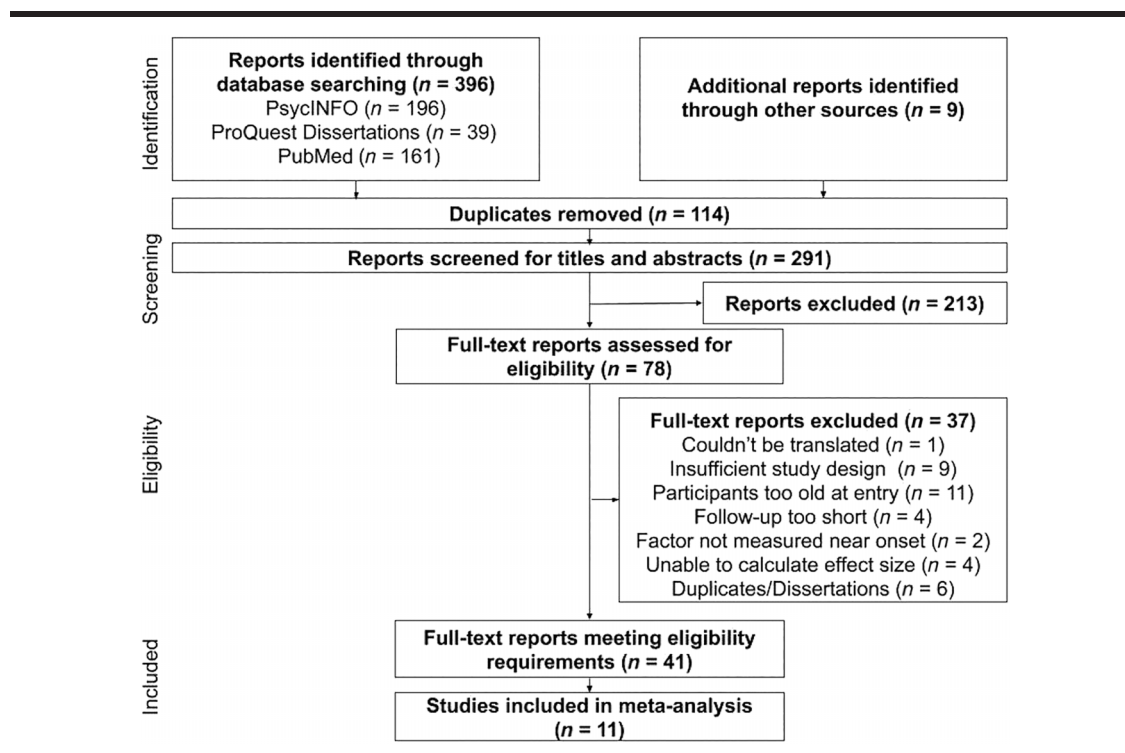
None of the studies were determined to exhibit high risk of bias, so all eligible data were included in subsequent analyses.

Effect Size Synthesis

Based on the nature, availability, and quality of the clinical characteristics included in the 11 studies, study outcomes were grouped into the following categories: sex, age at onset, family history of stuttering, stuttering behaviors, speech-language behaviors, and temperament. The results described below are divided and reported by category. Meta-analyses with only a few studies, but of theoretical and empirical interest to stuttering persistence, are provided for suggestive value (for a similar approach, see Ntourou et al., 2011).

Forest plots are included for each category. The location of the box along the *x*-axis indicates an individual study's effect size, the length of the lines connected to the box indicates the 95% confidence interval (CI), and the size of the box represents the individual study's weight in the analysis. The location of the center of the diamond along the *x*-axis indicates the estimated mean effect size of all the studies when combined, and the width of the diamond represents the 95% CI. If the diamond crosses the dashed line (i.e., the line of no effect), no effect is indicated. Heterogeneity statistics are reported within the forest plots. Detailed information for each included report can be found in Appendix B. Moderator and publication bias analyses are reported if there was a sufficient number of studies to conduct such analyses.

Sex

As shown in Figure 3, males are 1.48 times more likely to persist than females (risk ratio = 1.48, 95% CI [1.10, 2.00]). Low heterogeneity was detected across the 11 studies ($I^2 = 22.4\%$, $\tau^2 = 0.05$); therefore, follow-up metaregressions were not conducted.

Because 11 studies were included, a funnel plot (see Figure 4) was constructed to explore possible publication biases. Funnel plots that depict most of the studies have strong, positive effects (i.e., depicted on the right side of the funnel) indicate that findings are vulnerable to publication bias (i.e., the phenomenon that studies with significant effects are more likely to be published than studies will null effects). The relative symmetry of the funnel plot in Figure 4 suggests the findings were not significantly vulnerable to publication bias. Furthermore, there was no evidence of small study effects or publication bias based on an Egger's regression test ($z = -0.035$, $p = .972$).

Age at Onset

As seen in Figure 5, persistent children are reported to have a later onset than recovered children (Hedges's $g = 0.43$, 95% CI [0.16, 0.71]). A low level of heterogeneity was observed across the six studies ($I^2 = 0.0\%$, $\tau^2 = 0.00$).

Family History of Stuttering

Three types of family history of stuttering were analyzed separately: any stuttering (persistent and/or recovered), persistent stuttering, and recovered stuttering. The

Table 1. Summary of included studies with related reports listed in parentheses.

Included studies		Stuttering-related measures					Participants		
Study (reports)	Country	Entry age (yr.)	Entry TSO (mo.)	Criteria for group classification	Follow-up length	Attrition	Recruitment	Inclusionary/ exclusionary criteria	Percent receiving treatment
Outcomes of Children with Hearing Loss (Arenas et al., 2017)	USA	3	Freely vary	Parent report	At least 2 years or until 8 years old	R	Multisite	Hearing loss between 25 and 75 dB	34.8% of the recovered group and 40% of the persistent group
Twins Early Development Study (Dworzynski et al., 2007)	England	≤ 5	< 12	Parent report	Until 7 years old	R	Population-Based	Twins	NR
Erasmus University Medical Centre Study (Franken et al., 2018)	Netherlands	3–5	Freely vary	3% SS; SSI > 9; parent- and clinician-rated stuttering severity of at least 2 on an 8-point scale	Up to 9 years	R	Clinical	No known neurological, intellectual, or psychosomatic problems	91% of recovered group and 100% of persistent group
Illinois Stuttering Research Project (Ambrose et al., 1997; Hall et al., 1999; Mahurin-Smith & Ambrose, 2013; Paden et al., 2002; Paden & Yairi, 1996; Paden et al., 1999; Sawyer, 2005; Subramanian et al., 2003; Throneburg & Yairi, 2001; Watkins & Yairi, 1997; Watkins et al., 1999; Yairi & Ambrose, 1992, 1999, 2005; Yairi et al., 1996)	USA	Freely vary	< 12	3% SS; parent- and clinician-rated stuttering severity of at least 2 on an 8-point scale	4 years	R	Single-site	No known neurological disorders	0% of recovered group and 89% of persistent group (Yairi & Ambrose, 2005)

(table continues)

Table 1. *(Continued).*

| Included studies		Stuttering-related measures					Participants		
Study (reports)	**Country**	**Entry age (yr.)**	**Entry TSO (mo.)**	**Criteria for group classification**	**Follow-up length**	**Attrition**	**Recruitment**	**Inclusionary/ exclusionary criteria**	**Percent receiving treatment**
Early Language in Victoria Study (Kefalianos et al., 2017)	Australia	2	< 12	Stuttering severity rating of at least 2 on a 10-point scale by parents and clinician; 12 months for recovery	At least 3 years	R	Population-Based	NR	13.4% of recovered group and 16.7% of persistent group
Netherlands Study (Kloth et al., 1999)	Netherlands	2–5	< 12	Parent report	At least 4 years	R	Single-site	Positive parental history of stuttering	NR
Purdue Stuttering Project (Bostian, 2017; Gerwin et al., 2019; Hilger et al., 2016; Kreidler et al., 2017; Leech et al., 2017; Roehl, 2018; Spencer & Weber-Fox, 2014; Walsh et al., 2018)	USA		Freely vary	3% SS; clinician-rated stuttering of severity 2 or higher on an 8-point scale	2–5 years	NR	Multisite	No known neurological disorders or injury	Only treatment prior to study reported
Ulm Study (Brosch et al., 2001, 2002; Hage, 2001; Rommel et al., 2001)	Germany	3–6	Freely vary	NR	3–4.5 years	R	Clinical	NR	Treatment allowed; specifics not reported
California – Long Beach Study (Ryan, 2001)	USA	< 6	Freely vary	3.0 SW/min; trends of SW/min	Up to 10 years	NR	Clinical	No previous treatment	Treatment reported for persistent children; not recovered

Table 1. *(Continued).*

Included studies		Stuttering-related measures					Participants		
Study (reports)	Country	Entry age (yr.)	Entry TSO (mo.)	Criteria for group classification	Follow-up length	Attrition	Recruitment	Inclusionary/ exclusionary criteria	Percent receiving treatment
Subtypes and Risk Factors in Childhood Stuttering Study (Ambrose et al., 2015; Buhr, 2007; Hollister et al., 2017)	USA	Freely vary	< 12	3% SS; parent and clinician stuttering severity; recovery needed to be exhibited for 12 months based on parent report	Up to 5 years	NR	Multisite	No known neurological disorders	Only reported for small subset (e.g., Hollister et al., 2017)
Vanderbilt's Developmental Stuttering Project (Erdemir et al., 2018; Singer et al., 2019, 2020; Zengin-Bolatkale et al., 2018)	USA	3–5	Freely vary	3% SW and above 10 on SSI; recovery required absence of parent concern	At least 2 years	R	Single-site	No known neurological conditions and scores above 17th percentile on speech-language measures	23% of recovered group and 20% of persistent group

Note. TSO = time since onset; R = reported; NR = not reported; SS = stuttered syllables; SSI = Stuttering Severity Instrument (Riley, 1972); SW = stuttered words.

Figure 2. Risk of bias summary figure. The study confounding source of bias category was only assessed for studies that explored speech-language skills and temperament.
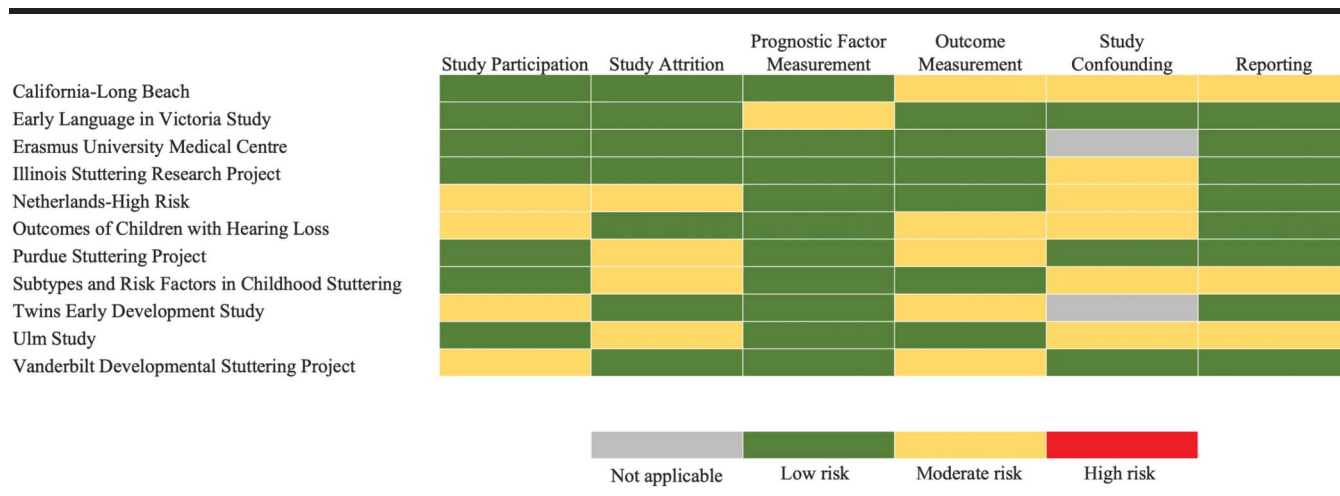
nature of data collection on family history of stuttering varied across the studies: Kefalianos et al. (2017) collected data on first-degree family members, Yairi and Ambrose (2005) and Walsh et al. (2018) collected data on first- and second-degree relatives, and Franken et al. (2018) collected data on any known family members. Rommel et al. (2001) did not report whether specific familial relations were targeted. First-degree relatives include parents and siblings; second-degree relatives include grandparents, aunts, uncles, and cousins.

For all analyses pertaining to family history, data reported on the proportion of the children in the persistent and recovered groups with and without family history were included. For example, data related to family history of any stuttering reported in Yairi and Ambrose (2005) were based on the proportion of *children* in each group with a family history of stuttering and were therefore included. In contrast, data from Yairi and Ambrose (2005) specific to family history of persistent and recovered stuttering pertained to the proportion of *relatives* that exhibited either persistent or recovered stuttering and were therefore excluded.

As seen in Figure 6, children with a family history of any stuttering are 1.89 times more likely to persist than children without a family history of stuttering (risk ratio = 1.89, 95% CI [1.27, 2.82]). Low heterogeneity was detected across this model ($I^2 = 0.0\%$, $\tau^2 = 0.00$). Significant differences in risk for persistence between children with and without a family history of persistent or recovered stuttering were not detected, but high heterogeneity was detected across both models (I^2 range: 45.0%–51.2%; τ^2 range: 1.06–1.10). Due to the small number of studies, metaregressions could not be conducted.

Figure 3. Forest plot comparing risk for stuttering persistence between males and females. As indicated by the diamond falling to the right of the line of no effect, males are at increased risk for persistence compared to females. CI = confidence interval; F = female; M = male; RE = random effects.
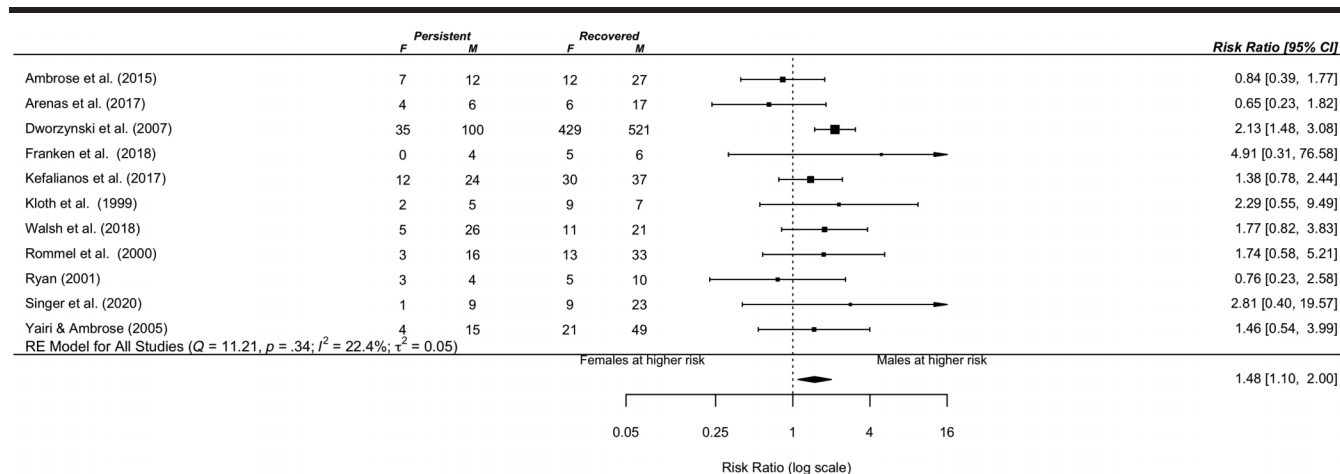
Stuttering Behaviors

Seven stuttering behaviors were analyzed separately: clinician-rated stuttering severity, parent-rated stuttering severity, and frequency of stuttering-like disfluencies, non–stuttering-like disfluencies, part-word repetition disfluencies, single-syllable whole-word repetition disfluencies, and dysrhythmic phonations (i.e., audible/inaudible prolongations). To assess a child's stuttering frequency, studies utilized varied methods to collect child speech samples, as described in Appendix C.

As seen in Figure 7, persistent children produced a higher frequency of stuttering-like disfluencies than recovered children (Hedges's g = 0.53, 95% CI [0.18, 0.87]). Low heterogeneity was detected across the five studies (I^2 = 27.2%, τ^2 = 0.04). Persistent children did not differ in stuttering severity based on clinician[2] or caregiver assessment, or the frequency of non–stuttering-like disfluencies, part-word repetition disfluencies, single-syllable whole-word repetition disfluencies, or dysrhythmic phonations when compared to recovered children. Low heterogeneity was detected across studies reporting data for stuttering severity and non–stuttering-like disfluencies (I^2 = 0.00, τ^2 = 0.00). Moderate-to-high heterogeneity was detected across studies reporting data for individual stuttering-like disfluency types (I^2 range: 63.5%–77.0%; τ^2 range: 0.15–0.31), but due to the small number of studies, moderator analyses were not conducted.

Speech-Language Behaviors

Five speech-language skills were quantified via norm-referenced measures: speech sound accuracy, receptive vocabulary, receptive language, expressive vocabulary, and expressive language. Three speech-language skills were quantified with language sample analysis: mean length of utterance (MLU), Index of Productive Syntax (IPSYN; Scarborough, 1990), and Developmental Sentence Scoring (Lee & Canter, 1971).

As seen in Figure 8, persistent children scored significantly lower on measures of speech sound accuracy (Hedges's g = −0.54, 95% CI [−0.88, −0.21]), receptive language (Hedges's g = −0.46, 95% CI [−0.77, −0.17]), and expressive language (Hedges's g = −0.43, 95% CI [−0.69, −0.16]) than recovered children. Low heterogeneity was detected across these models (I^2 range: 0.0%–10.4%, τ^2 range: 0.00–0.01). Persistent and recovered children did not differ significantly on measures of receptive vocabulary or expressive vocabulary. Moderate heterogeneity was detected across these models (I^2 range: 59.5%–61.9%, τ^2 range: 0.12–0.21), but due to the small number of studies, moderator analyses were not conducted.

Sensitivity analyses were conducted to explore whether findings change when data evaluated to be at greater risk for

bias related to speech-language behaviors are excluded. Data from Singer et al. (2020) were evaluated to have moderate risk of bias due to excluding children who scored below the 17th percentile on any one speech-language measure. Similarly, data from Kefalianos et al. (2017) were also evaluated to have moderate risk of bias due to the speech-language evaluation occurring when participants were 2 years of age, which might have been before some children began stuttering. Results from the sensitivity analyses indicated that significant findings did not change. Persisting children scored lower on measures of speech sound accuracy (Hedges's g = −0.62, 95% CI [−0.99, −0.24]), receptive language (Hedges's g = −0.43, 95% CI [−0.77, −0.08]), and expressive language (Hedges's g = −0.44, 95% CI [−0.73, −0.15]) than recovered children. Furthermore, persisting children did not differ significantly from recovered children on measures of receptive vocabulary (Hedges's g = −0.31, 95% CI [−0.82, 0.20]). When expressive vocabulary data from Singer et al. (2020) and Kefalianos et al. (2017) were excluded, there was an insufficient number of studies to conduct a meta-analysis.

Findings related to language samples are provided but should be interpreted with caution considering data were only available in two or three studies and language sample procedures varied. For example, the length of language samples analyzed varied across reports: Kloth et al. (1999) analyzed 10-min speech samples, Hollister et al. (2017) analyzed 100-utterance speech samples, Buhr (2007) analyzed 50-utterance speech samples, and Watkins et al. (1999) analyzed 250- to 300-utterance speech samples. Furthermore, Hollister et al. hand calculated IPSYN, whereas Kloth et al. used a computer program.

As seen in Figure 8, persistent and recovered children, on average, did not exhibit significantly different expressive language skills when quantified using IPSYN, MLU, or Developmental Sentence Scoring. Low heterogeneity was found across these models (I^2 = 0.00%, τ^2 = 0.00).

Temperament

Three temperament characteristics were explored: negative affectivity, surgency, and effortful control. These characteristics were reported in Ambrose et al. (2015) and Zengin-Bolatkale et al. (2018) using the Children's Behavior Questionnaire (Rothbart et al., 2001). Temperament data from Kefalianos et al. (2017), measured using the approach/withdrawal scale of the Short Temperament Scale for Children (Sanson et al., 1994), were determined to be too dissimilar from the other data to be included in the analysis. Results are interpreted with caution because of the small number of currently available studies for analysis.

As seen in Figure 9, persistent and recovered children did not significantly differ in negative affectivity, surgency, and effortful control. High heterogeneity was detected across the studies for negative affectivity and surgency (I^2 range: 95.2%–98.5%, τ^2 range: 2.98–16.67), but moderator analyses could not be conducted due to the small number of studies. Low heterogeneity was detected across the studies for effortful control (I^2 = 0.00, τ^2 = 0.00).

---

[2]Two studies (Franken et al., 2018; Roehl, 2018) reported data on two measures of clinician-rated stuttering severity. Due to the small number of studies included in the analysis, robust variance estimation procedures (Tipton, 2015) could not be used and only data on one measure per study could be included (i.e., the first measure reported).

Figure 4. Contour funnel plot of studies reporting risk related to sex. Asymmetry related to the distribution of the data points may be due to publication bias, differences in study methods, and the direction of the relation.
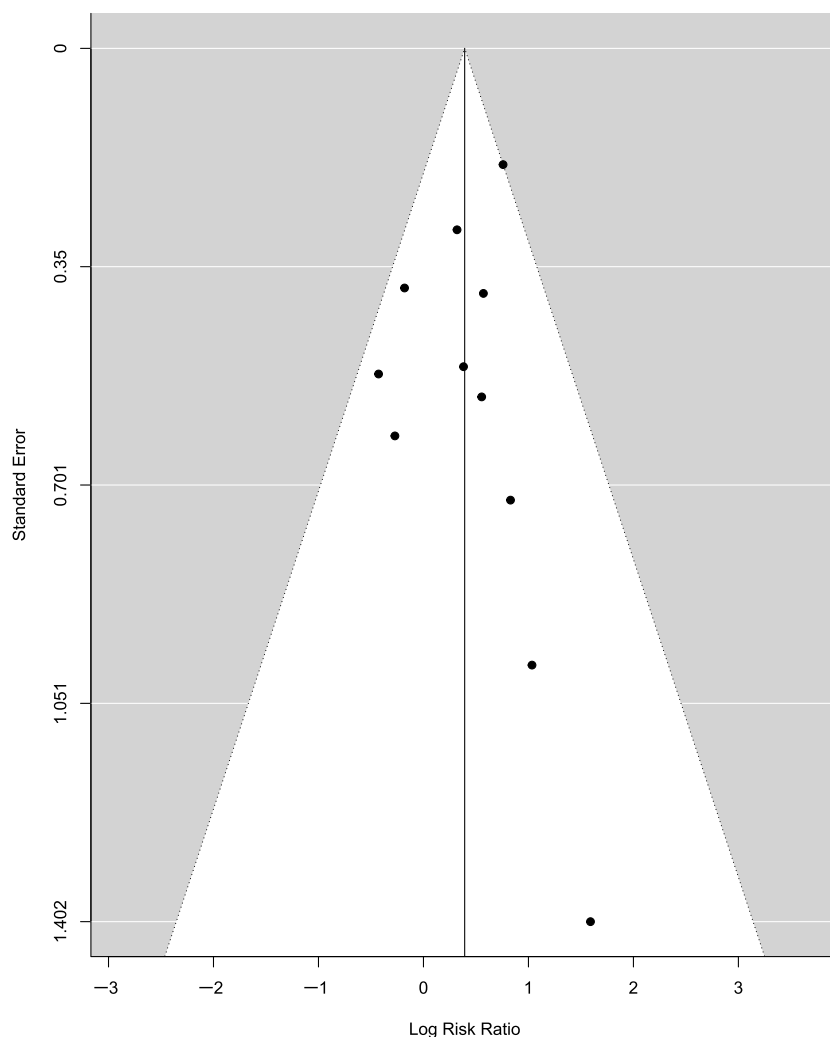
Summary of Descriptive Statistics for Clinical Characteristics

Table 2 presents the descriptive statistics for the clinical characteristics of the persistent and recovered groups. Only data that could be combined (e.g., standard scores could not be combined with raw scores) were used to calculate these descriptive statistics, and therefore, values in the table represent a subset of the data included in the meta-analyses.
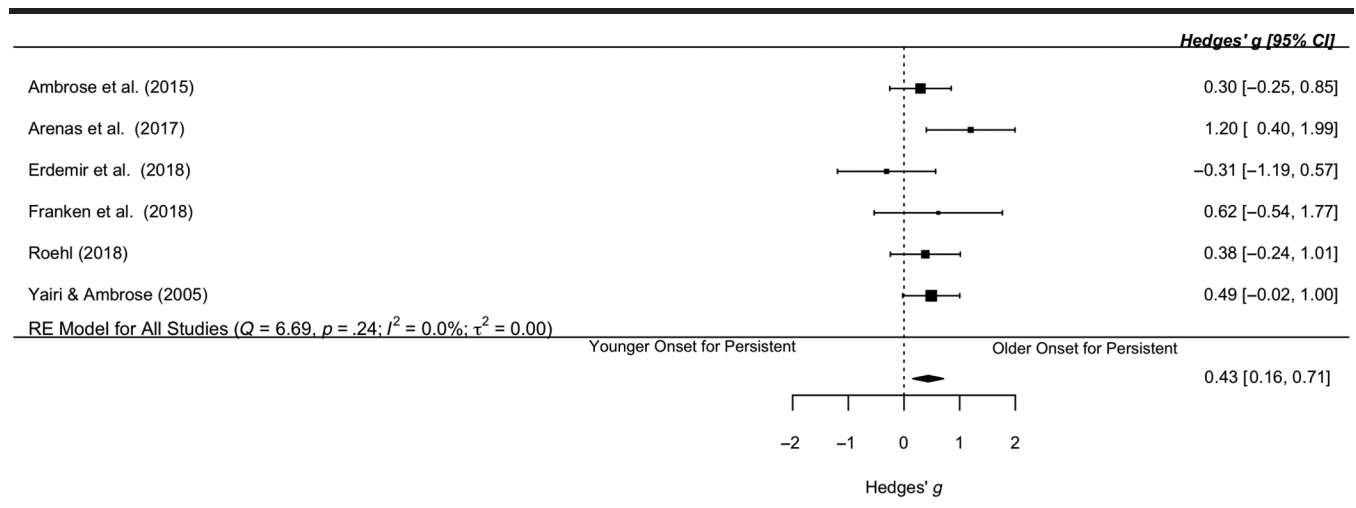
Discussion

As the first quantitative synthesis of prospective cohort studies to identify clinical characteristics in early development (i.e., 2- to 6-year-old age range) that differentiate children who recover from stuttering versus those who persist, this study represents the highest level of evidence currently available for clinical characteristics related to stuttering

persistence. Given the small number of studies for each individual model and the multiple sources of potential bias (e.g., inclusionary/exclusionary criteria, limited follow-up) within studies, findings must be interpreted accordingly. Children who persist, compared to children who eventually recover, (a) were more likely to be male; (b) begin stuttering at a later age; (c) have known family histories of stuttering (persistent and/or recovered); (d) produce higher stuttering frequencies; and (e) perform lower on measures of speech sound accuracy, expressive language, and receptive language. The modest effect sizes of these between-group differences and risk ratios suggest that persistent children, as a group, exhibit vulnerabilities and/or characteristics that might yield heightened risk for the development of chronic stuttering.

It is possible that additional differences will be detected when more data become available. For example, findings related to negative and positive emotionality, individual stuttering-like disfluency types, expressive vocabulary, and

Figure 5. Forest plot comparing mean age at stuttering onset for persistent and recovered children. As indicated by the diamond falling to the right of the line of no effect, persistent children were older at stuttering onset than recovered children. CI = confidence interval; RE = random effects.

	Hedges' g [95% CI]
Ambrose et al. (2015)	0.30 [−0.25, 0.85]
Arenas et al. (2017)	1.20 [0.40, 1.99]
Erdemir et al. (2018)	−0.31 [−1.19, 0.57]
Franken et al. (2018)	0.62 [−0.54, 1.77]
Roehl (2018)	0.38 [−0.24, 1.01]
Yairi & Ambrose (2005)	0.49 [−0.02, 1.00]
RE Model for All Studies (Q = 6.69, p = .24; I^2 = 0.0%; τ^2 = 0.00)	0.43 [0.16, 0.71]

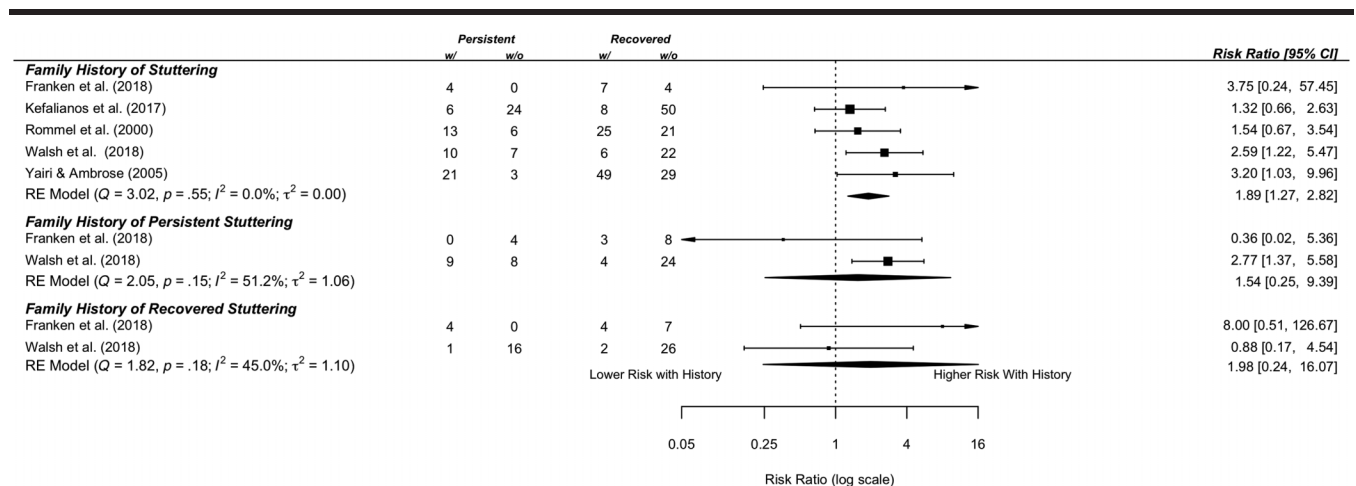Younger Onset for Persistent — Older Onset for Persistent

−2 −1 0 1 2

Hedges' g

family history of persistent stuttering were based on only two or three studies—with one study reporting a significant difference in each case. Furthermore, given the high heterogeneity within these models, the evidence to support the association between these characteristics and stuttering persistence may presently be characterized as *insufficient*.

Furthermore, data that were available, but could not be included in this study, are also important to our understanding of potential clinical characteristics related to stuttering persistence. Data that were not included (a) used metrics that could not be combined with data from included studies or (b) were only reported in one study. For example, Yairi and Ambrose (2005), and other reports associated with the Illinois Developmental Stuttering Project (e.g.,

Ambrose et al., 1997), while providing additional support for the association between family history of persistent stuttering and stuttering persistence, used an alternative metric to the one consistent with the studies included in our analyses. Specifically, they found that children who persist have a higher proportion of relatives with a family history of persisting stuttering than children who recover. Additionally, data from Yairi and Ambrose (2005) and Spencer and Weber-Fox (2014) provide evidence for novel clinical factors. Yairi and Ambrose (2005) found that persistent children exhibit relatively stable stuttering-like disfluencies, whereas children who recover exhibit sharper declines within the first 18 months of onset. Spencer and Weber-Fox found that children who persist performed lower on nonword repetition

Figure 6. Forest plot comparing risk for stuttering persistence between children with and without family histories of stuttering. As indicated by the diamond falling to the right of the line of no effect, children with a family history of stuttering are at increased risk for stuttering persistence compared to children without a family history of stuttering. No risk differences were found based on family history of persistent and recovered stuttering. CI = confidence interval; RE = random effects.

	Persistent w/	Persistent w/o	Recovered w/	Recovered w/o	Risk Ratio [95% CI]
Family History of Stuttering					
Franken et al. (2018)	4	0	7	4	3.75 [0.24, 57.45]
Kefalianos et al. (2017)	6	24	8	50	1.32 [0.66, 2.63]
Rommel et al. (2000)	13	6	25	21	1.54 [0.67, 3.54]
Walsh et al. (2018)	10	7	6	22	2.59 [1.22, 5.47]
Yairi & Ambrose (2005)	21	3	49	29	3.20 [1.03, 9.96]
RE Model (Q = 3.02, p = .55; I^2 = 0.0%; τ^2 = 0.00)					1.89 [1.27, 2.82]
Family History of Persistent Stuttering					
Franken et al. (2018)	0	4	3	8	0.36 [0.02, 5.36]
Walsh et al. (2018)	9	8	4	24	2.77 [1.37, 5.58]
RE Model (Q = 2.05, p = .15; I^2 = 51.2%; τ^2 = 1.06)					1.54 [0.25, 9.39]
Family History of Recovered Stuttering					
Franken et al. (2018)	4	0	4	7	8.00 [0.51, 126.67]
Walsh et al. (2018)	1	16	2	26	0.88 [0.17, 4.54]
RE Model (Q = 1.82, p = .18; I^2 = 45.0%; τ^2 = 1.10)					1.98 [0.24, 16.07]

Lower Risk with History — Higher Risk With History

0.05 0.25 1 4 16

Risk Ratio (log scale)

Figure 7. Forest plot comparing mean stuttering behaviors for persistent and recovered children. As evidenced by the positioning of the diamond relative to the line of no effect, persistent children exhibited a higher frequency of stuttering-like disfluencies than recovered children. All other differences were nonsignificant. CI = confidence interval; RE = random effects; SLD = stuttering-like disfluencies; SSI = Stuttering Severity Instrument.
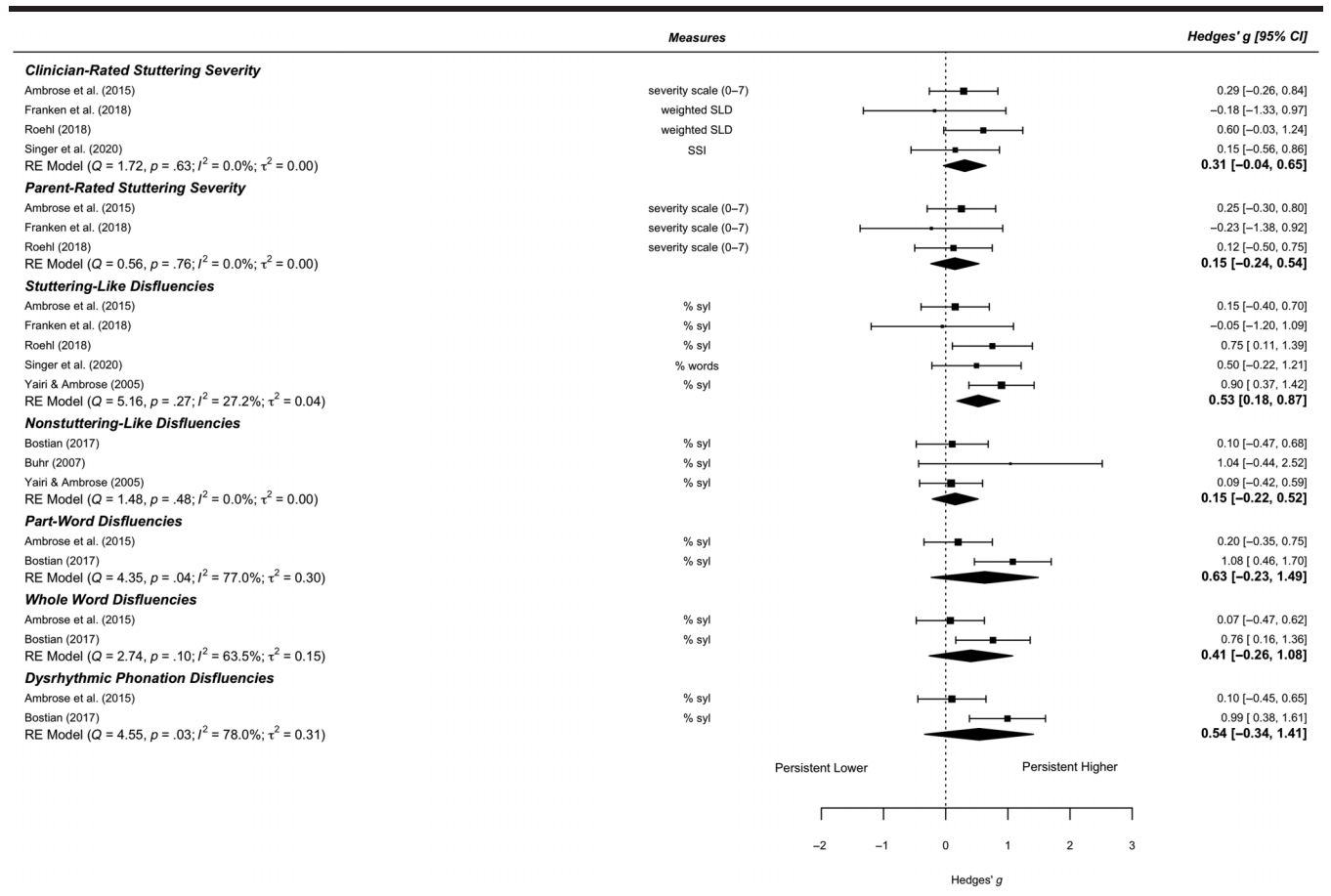
tasks compared to children who eventually recover. While singular, these findings from prospective cohort studies can be integrated with the present meta-analytic findings to help clinicians identify empirically supported clinical characteristics related to stuttering persistence.

Overall, findings of this systematic review and meta-analysis confirm, but also extend, previous clinically oriented syntheses developed for pediatricians (Guitar & Conture, 2006) and SLPs (Clark et al., 2017; Walsh et al., 2018) to identify characteristics associated with stuttering persistence. As a result of this study, characteristics (e.g., age at onset, male sex) reported in previous syntheses are now supported with clear, unequivocal quantitative *empirical evidence* of their relations to stuttering persistence. Furthermore, this study permitted identification of specific aspects of language or speech production that are significantly associated with stuttering chronicity (e.g., performance on standardized measures of expressive and receptive language) and others that are not (e.g., MLU). Present findings also highlight the association between stuttering frequency and stuttering chronicity, which has received little prior attention (i.e.,

individual studies rarely found significant differences between recovered and persistent groups). Lastly, given the quantitative nature of this study, descriptive statistics of the clinical characteristics for the persistent and recovered children are reported that may be used for clinical consideration and future empirical comparisons.

Theoretical Connections

Present findings relate to the "multifactorial dynamic pathways" model (A. Smith & Weber, 2017), which implicates multiple domains (genetics, speech, language, speech motor, temperament, etc.) in the development of childhood stuttering. Specifically, the present results may be interpreted as evidence that there are genetic (e.g., male sex, family history) and speech-language (e.g., lower speech sound accuracy and/or language) vulnerabilities that contribute to an individual's risk for persistent stuttering. The nonsignificant between-group findings related to temperament/emotional contributions may be explained by the rigor of the present analyses and the limited sample sizes of the included studies.

Figure 8. Forest plot comparing mean speech-language skills for persistent and recovered children. As indicated by the positioning of the diamonds relative to the line of no effect, persistent children exhibit lower speech sound accuracy, receptive language, and expressive language than recovered children. All other differences were nonsignificant. APP-R = Assessment of Phonological Processes–Revised (Hodson, 1986); BBTOP-CI = Bankson–Bernthal Test of Phonology—Consonant Inventory (Bankson & Bernthal, 1990); CDI = MacArthur–Bates Communicative Development Inventories (Fenson et al., 1993); CI = confidence interval; DSS = Developmental Sentence Scoring (Lee & Canter, 1971); EVT = Expressive Vocabulary Test (Williams, 1997); GFTA = Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 2000); IPSYN = Index of Productive Syntax (Scarborough, 1990); MLU = mean length of utterance; PLS = Preschool Language Scales (Zimmerman et al., 1979); PPVT = Peabody Picture Vocabulary Test (Dunn & Dunn, 1997); Reynell = Reynell Language Development Scale (Reynell, 1983); SPELT = Structured Photographic Expressive Language Test (Dawson et al., 2003); TACL = Test for Auditory Comprehension of Language (Carrow-Woolfolk, 1999); TELD = Test of Early Language Development (Hresko et al., 1999).
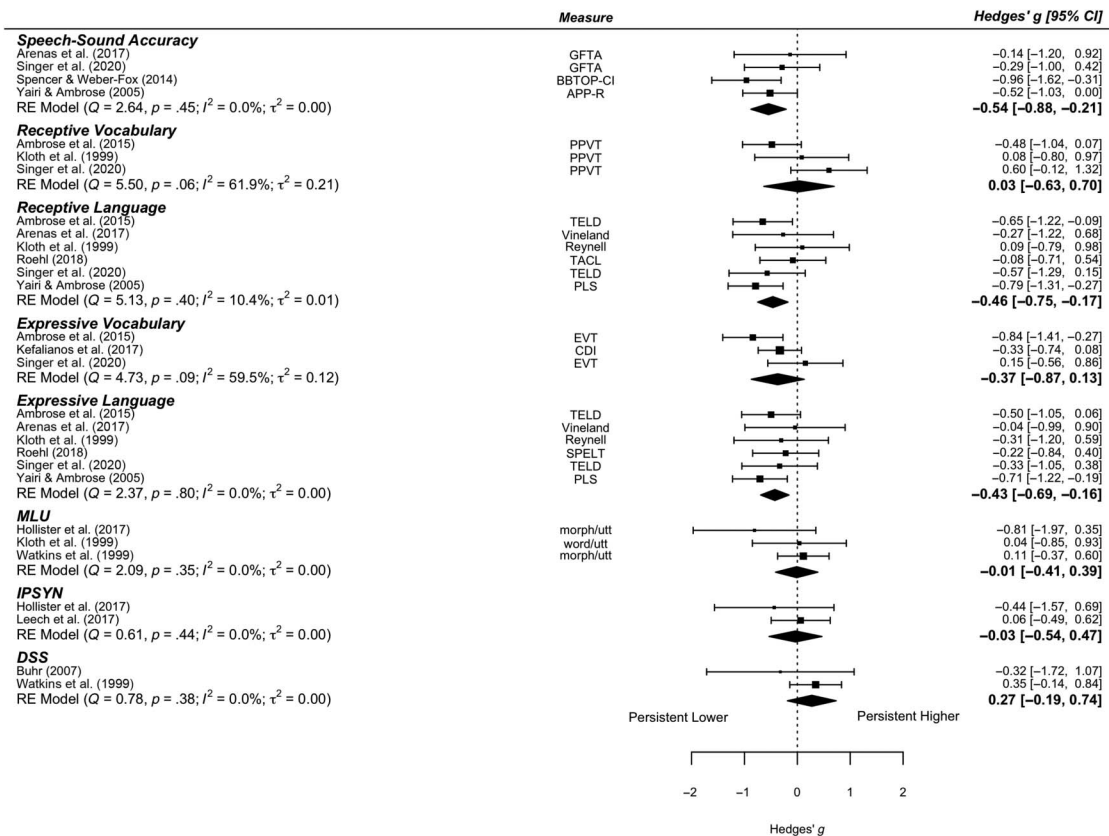
Figure 9. Forest plot comparing temperament for persistent and recovered children. As indicated by all diamonds crossing the line of no effect, differences are nonsignificant. CI = confidence interval; RE = random effects.
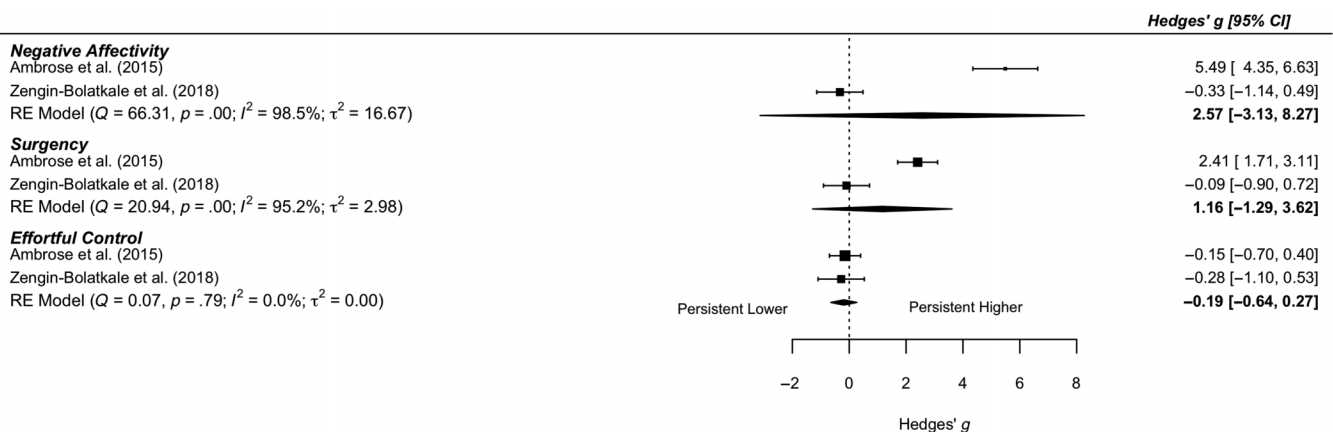
Table 2. Summary descriptive statistics of clinical characteristics for the persistent and recovered groups.

Clinical characteristics	Persistent group		Recovered group	
	M (SD)	N	M (SD)	N
Sex (% male)	74.4	297	57.7	1,301
Age at onset (months)	39.6 (13.1)	78	34.4 (8.2)	180
Family history of stuttering				
Any stuttering (% positive)	69.2	78	46.8	237
Persisting stuttering (% positive)	42.8	21	18	39
Recovered stuttering (% positive)	23.8	21	15.4	39
Speech-language behaviors				
Speech sound[†,‡] (standard score)	90.0 (17.0)	34	100.5 (16.1)	64
Receptive language[†,‡] (standard score)	111.8 (15.9)	64	123.1 (17.8)	167
Expressive language[†,‡] (standard score)	105.5 (15.4)	64	115.9 (18.8)	167
Receptive vocabulary[‡] (standard score)	107.6 (17.7)	29	109.7 (11.6)	71
Expressive vocabulary[§] (standard score)	106.9 (14.0)	29	114.1 (11.8)	71
DSS	6.6 (1.5)	26	6.1 (1.5)	66
IPSYN	78.1 (10.6)	36	79.5 (9.2)	27
MLU[‡] (morphemes/utterance)	4.0 (0.9)	30	4.0 (1.0)	67
Stuttering behaviors				
Stuttering frequency[ǀ] (% syllable)	8.7 (7.7)	62	6.1 (4.8)	157
Nonstuttering frequency (% syllable)	4.8 (2.0)	42	4.8 (2.3)	103
Part-word repetitions (% syllable)	3.6 (3.4)	38	2.6 (3.0)	68
Whole-word repetitions (% syllable)	3.0 (3.0)	38	2.5 (1.8)	68
Dysrhythmic phonations (% syllable)	2.4 (2.4)	38	1.6 (1.8)	68
Weighted SLD	11.9 (7.1)	20	9.0 (9.1)	37
Parent Rating Scale (0–7)	4.0 (7.5)	39	3.5 (1.1)	76
Temperament				
Effortful control	5.0 (0.5)	27	5.1 (0.4)	61
Negative affectivity	4.2 (0.3)	27	3.8 (0.3)	61
Surgency	5.1 (0.4)	27	4.7 (0.5)	61

Note. Weighted means and standard deviations reported. DSS = Developmental Sentence Scoring (Lee & Canter, 1971); IPSYN = Index of Productive Syntax (Scarborough, 1990); MLU = mean length of utterance; weighted SLD = weighed stuttering-like disfluencies (Ambrose & Yairi, 1999).

[†]Denotes Arenas et al. (2017) raw data excluded. [‡]Denotes Kloth et al. (1999) nonstandardized data excluded. [§]Denotes Kefalianos et al. (2017) raw data excluded. [ǀ]Denotes Singer et al. (2020) data based on percent words excluded.

Therefore, additional empirical studies will be necessary to establish a more comprehensive understanding of the contribution of these variables to stuttering chronicity and related theoretical models (for an overview of the potential role of these variables in stuttering, see Conture & Walden, 2012; A. Smith & Weber, 2017).

Lack of sufficient data is also true for other elements of multifactorial models such as the importance of environment, a factor that has minimally been explored in relation to stuttering chronicity (e.g., mother's communicative style; Kloth et al., 1999). Ultimately, the continued empirical assessment of factors associated with risk for stuttering persistence will contribute to the development, expansion, and clarification of related theoretical models, which, in turn, will likely lead to translational advancements. The present findings, while limited, at times, by the number of available studies, can be used to inform current clinical practice and future research studies.

Clinical Implications

Findings support and augment previous recommendations that SLPs conduct comprehensive speech-language evaluations when working with young children who stutter and their families (e.g., Clark et al., 2017; Walsh et al., 2018). First, given that the number of stuttering-like disfluencies was found to differentiate children who eventually recover from those who persist, whereas stuttering severity did not, a disfluency analysis may be an important component within a comprehensive evaluation and may provide more prognostic value than a severity rating alone (for guidelines and a tutorial for disfluency analysis, see Yairi & Ambrose, 2005, and Yaruss, 1998, respectively). Second, this study reports descriptive statistics for persistent and recovered children, providing reference points for the mean values and variability within each clinical characteristic. For example, the mean standard scores for speech-language tests are reported for children who persisted and children who eventually recovered. Whereas the utility of these data are limited at present due to the overlap in scores/characteristics between children who eventually recover and children who persist, they may provide some guidance that is useful to clinicians as they evaluate whether a child is presenting with characteristics associated with higher risk for persistence until empirically supported cutoff scores are available. Third, by reporting effect sizes for each clinical characteristic, present findings might be used to identify a hierarchy of evaluation components to aid SLPs in the identification of factors

that may be related to risk for stuttering persistence (as well as concomitant challenges that may need to be considered and/or addressed when planning treatment for these children). Factors with greater effect sizes might be considered to have stronger prognostic value (larger magnitude of effect based on available empirical evidence) than factors with smaller effect sizes and therefore may be more important to include in an evaluation. Table 3 summarizes the characteristics that might increase a child's risk for stuttering persistence, from those with the strongest to those with the weakest effect sizes, along with related evaluation procedures based on current levels of evidence. By way of clinical example, an SLP might advise parents/caregivers that their child (who stutters) is more at risk for persistence when there is a known family history of stuttering, the child is male, and she or he started to stutter after about age 4 years. In addition, a higher rate of stuttering-like disfluencies would confer greater concern for stuttering, whether for the short or long term, with elevated frequencies more characteristic of persistent stuttering. Lastly, below average scores for speech sound, receptive, and/or expressive language skills may also be associated with increased risk for persistence, and the SLP should consider these when addressing stuttering, speech and/or language, based on the individual child's profile, needs, and impact of each concern on overall communication.

Multifaceted, comprehensive assessments that include pertinent characteristics with insufficient evidence, at present and/or those not included in this study, are still warranted and provide valuable information. For example, we suggest that evaluating the impact of stuttering on the child's communication (e.g., willingness to talk, frustration with speaking) and the family's concerns (e.g., not knowing how to help their child who is struggling to communicate) should be considered. Although minimal research has been conducted on whether these factors are related to stuttering persistence, which led to our inability to include them in our study, these factors, particularly the child's negative attitudes toward talking, have long been theorized to be related to (continued) stuttering (e.g., approach–avoidance hypothesis;

Sheehan, 1952; demands and capacities model; Starkweather, 1982) and important to consider when making treatment decisions (e.g., Kelman & Nicholas, 2008). Furthermore, data on clinical characteristics currently described as *insufficient* should not be interpreted as errant or unimportant—but underexplored or preliminary at present—and may provide useful information when considering treatment approaches. Last, it is important to emphasize the importance of individual differences between children. Present findings are based on between-group findings, which are not representative of every child or of individual children. For these reasons, it is important that clinicians implement evidenced-based practice, which will include the consideration of not only the present findings (i.e., empirical evidence) but also their own clinical expertise, and individual client's and families' values and characteristics (e.g., Dollaghan, 2007).

Implications for Future Research

Findings related to the sources of bias that were frequently found in the included studies (the first research question) elucidate important study design and reporting decisions that researchers pursing prospective cohort studies on young children who stutter should consider in future research. The most common sources of bias were restrictive inclusionary/exclusionary criteria, limited follow-up (i.e., less than 3 years), insufficient reports of attrition, and lack of reports on methods to control for the effect of known confounding variables. These sources of bias should be considered carefully in future study design and reporting decisions for prospective cohort studies of young children who stutter. Reducing potential sources of bias within and across prospective cohort studies helps increase both the validity and consistency of resultant findings.

Whereas the present findings identify clinical characteristics that are related to stuttering persistence, additional research is needed to understand how SLPs can use these characteristics to assess a specific child's risk for persistence. For example, cumulative risk (i.e., the more risk factors

Table 3. Clinical characteristics related to stuttering chronicity (with established and insufficient evidence) organized by descending effect size.

Clinical characteristic	Evaluation procedure
Established evidence	
Known family history of (any) stuttering	Caregiver report
Male sex	Caregiver report
Older age at onset	Caregiver report
Lower speech sound skills	Speech Sound Test
Higher rate of stuttering-like disfluencies	Disfluency sample
Lower receptive language	Global Receptive Language Test
Lower expressive language	Global Expressive Language Test
Insufficient evidence	
Greater negative/positive reactivity	Caregiver questionnaire
Known family history of persistent stuttering	Caregiver report
Higher rate of individual stuttering-like disfluency types	Disfluency sample
Expressive vocabulary	Expressive vocabulary measure

Note. Established evidence characterized by statistically significant summary results. Insufficient evidence characterized by high heterogeneity across meta-analytic models, with at least one primary study reporting significant findings.

exhibited, the greater the risk for persistence) has been proposed as important (e.g., Zebrowski, 1997), but has not yet been empirically validated. Also, additional research is needed to determine whether there are clinical characteristics that are more valuable for predicting persistence than others and, thus, might be weighted more heavily and/or be more critical to assess. Also, at present, it is unclear how individual clinical characteristics relate to, influence, and/or are influenced by other characteristics and dynamic processes of child development that occur over time (e.g., epigenetic influences).

In addition to focusing on how to best assess risk of stuttering persistence, there are other directions for future research. First, group-level explorations, such as those provided in this study, are limited in their ability to account for individual differences (i.e., heterogeneity) within groups. The reporting of data at the level of the individual and/or subgroups would allow for the important exploration of whether certain profiles (e.g., a constellation of factors) might put a given child at greater risk for persistence than others. To explore profiles, large samples are necessary. Just as sharing and assimilating data across studies were essential to this study, we believe working toward the establishment of large, multisite databases is essential to the continued advancement of our understanding of stuttering development, assessment, and treatment. These databases would help to combat the low sample size issues that have plagued nearly all prospective cohort studies.

Furthermore, research on these clinical characteristics might benefit from the development and evaluation of a standard protocol for assessing young children who are stuttering. Developing a *flexible* standard protocol (e.g., using varied standardized measures of receptive and/or expressive language, and/or speech sound production/articulation skills) may facilitate the pooling of data and aligning of research and clinical practices. Based on present findings, we suggest this protocol should include collection of speech samples for disfluency analysis, administration of norm-referenced tests for speech sound accuracy and expressive and receptive language, caregiver interviews incorporating questions about age at onset and family history of stuttering, and a caregiver report measure of temperament. By adopting similar assessment methods while still collecting additional data of interest (physiological measures of reactivity, brain imaging, etc.), clinicians and researchers could contribute valuable data central to furthering our understanding of stuttering persistence.

Caveats

Present findings were limited by our data inclusion and analysis decisions when conducting the present meta-analysis. For example, more data would have been available had we allowed studies with older children (i.e., above the age of 6 years) at study entry. In contrast, had we adopted more restricted criteria for group classification (e.g., required direct evaluation of stuttering), less but more homogenous data would have been available. Another implication of our efforts to focus only on studies that included children under

the age of 6 years is that some of the participants may change classification after the study concludes (e.g., a child may subsequently recover or stuttering may reemerge; see Jones et al., 2008, for an example of a study that identified relapse in young children). This is an unavoidable reality of developmental stuttering that underscores the need for long-term follow-up of large cohorts to better understand the factors that predict outcomes.

Additionally, inherent to any meta-analysis, we acknowledge that, despite our extensive search strategy, we may not have identified some relevant reports. Last, due to the small number of studies available for some analyses, it is possible that false negatives were obtained for some variables. Overall, however, we believe that the strengths of this study outweigh the limitations, allowing us to evaluate clinical characteristics related to stuttering persistence based on the present literature base and to identify potential areas worthy of future investigation.

Conclusions

This study used meta-analytic methods to identify clinical characteristics that differentiate children who later persist versus those who recover from stuttering in early childhood. Male sex, later age at stuttering onset, a family history of stuttering, a higher rate of stuttering-like disfluencies, lower speech sound accuracy, and lower receptive and expressive language skills were found to be related to stuttering persistence. Insufficient evidence was available to support a relation between stuttering persistence and lower expressive vocabulary, greater negative and positive reactivity, and higher rates of individual stuttering-like disfluency types. Findings represent the highest level of empirical evidence for clinical risk factors associated with stuttering persistence to date and further our understanding of the nature of persistent stuttering. The inclusion of these clinical characteristics in evaluations of stuttering in young children will provide valuable prognostic data that will aid SLPs in making treatment decisions. Future studies may consider investigating whether the presence of multiple risk factors yields increased risk for persistence (i.e., cumulative risk) as well as whether there are risk factors or groupings of risk factors that confer greater risk than others.

Author Contributions

Cara Michelle Singer: Conceptualization (Lead), Data curation (Lead), Formal analysis (Lead), Investigation (Lead), Methodology (Lead), Visualization (Lead), Writing - Original Draft (Lead). **Alison Hessling:** Data curation (Supporting), Investigation (Supporting), Validation (Supporting), Writing - Review & Editing (Supporting). **Ellen M. Kelly:** Conceptualization (Supporting), Methodology (Supporting), Writing - Review & Editing (Supporting). **Lisa Singer:** Data curation (Supporting), Validation (Supporting), Writing - Review & Editing (Supporting). **Robin M. Jones:** Conceptualization (Supporting), Funding acquisition (Lead), Investigation (Supporting), Methodology (Supporting),

Acknowledgments

References

References marked with an asterisk indicate study data that were included in the meta-analysis.

Ambrose, N. G. (1996). The genetic basis of persistence and recovery in stuttering [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign, United States.

Ambrose, N. G., Cox, N. J., & Yairi, E. (1997). The genetic basis of persistence and recovery in stuttering. *Journal of Speech, Language, and Hearing Research, 40*(3), 567–580. https://doi.org/10.1044/jslhr.4003.567

Ambrose, N. G., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language, and Hearing Research, 42*(4), 895–909. https://doi.org/10.1044/jslhr.4204.895

*Ambrose, N. G., Yairi, E., Loucks, T. M., Seery, C. H., & Throneburg, R. (2015). Relation of motor, linguistic and temperament factors in epidemiologic subtypes of persistent and recovered stuttering: Initial findings. *Journal of Fluency Disorders, 45,* 12–26. https://doi.org/10.1016/j.jfludis.2015.05.004

Andrews, G., & Harris, M. (1964). *The syndrome of stuttering.* Spastics Society Medical Education.

*Arenas, R. M., Walker, E. A., & Oleson, J. J. (2017). Developmental stuttering in children who are hard of hearing. *Language, Speech, and Hearing Services in Schools, 48*(4), 234–248. https://doi.org/10.1044/2017_LSHSS-17-0028

Bankson, N. W., & Bernthal, J. E. (1990). *Bankson–Bernthal Test of Phonology.* Pro-Ed.

Barbu, S., Nardy, A., Chevrot, J.-P., Guellaï, B., Glas, L., Juhel, J., & Lemasson, A. (2015). Sex differences in language across early childhood: Family socioeconomic status does not impact boys and girls equally. *Frontiers in Psychology, 6,* 1874. https://doi.org/10.3389/fpsyg.2015.01874

Bernstein Ratner, N. (2018). Selecting treatments and monitoring outcomes: The circle of evidence-based practice and client-centered care in treating a preschool child who stutters. *Language, Speech, and Hearing Services in Schools, 49*(1), 13–22. https://doi.org/10.1044/2017_LSHSS-17-0015

*Bostian, A. (2017). *Disfluency characteristics of children aged 4–5 years associated with persistence and recovery of stuttering* [Unpublished master's thesis]. Purdue University.

Brosch, S., Hage, A., Kalehne, P., & Johannsen, H. S. (1999). Stuttering children and the probability of remission—The role of cerebral dominance and speech production. *International Journal of Pediatric Otorhinolaryngology, 47*(1), 71–76. https://doi.org/10.1016/S0165-5876(98)00178-5

Brosch, S., Häge, A., & Johannsen, H. S. (2002). Prognostic indicators for stuttering: The value of computer-based speech analysis. *Brain and Language, 82*(1), 75–86. https://doi.org/10.1016/S0093-934X(02)00005-6

Brosch, S., Höge, A., & Johannsen, H. S. (2001). Value of acoustic speech analysis for prognostic assessment of stuttering in children: Partial results of a prospective longitudinal study. *HNO, 49,* 289–297.

*Buhr, A. (2007). *The roles of syntactic and prosodic components of grammar in early childhood stuttering* [Unpublished doctoral dissertation]. The University of Iowa.

Carrow-Woolfolk, E. (1999). *Test for Auditory Comprehension of Language–Third Edition.* AGS.

Chang, S. E. (2005). Regional differences in brain structure underlying childhood stuttering persistence and recovery: An MRI study [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign, United States.

Chentsova-Dutton, Y. E., & Tsai, J. L. (2007). Gender differences in emotional response among European Americans and Hmong Americans. *Cognition and Emotion, 21*(1), 162–181. https://doi.org/10.1080/02699930600911333

Chow, H. M., & Chang, S.-E. (2017). White matter developmental trajectories associated with persistence and recovery of childhood stuttering. *Human Brain Mapping, 38*(7), 3345–3359. https://doi.org/10.1002/hbm.23590

Clare, G. (2008). Conversational turn-taking between persistent and recovered stutterers and fluent interlocutors: A conversation and statistical analysis [Unpublished master's thesis]. University College London, England.

Clark, C. E., Tumanova, V., & Choi, D. (2017). Evidence-based multifactorial assessment of preschool-aged children who stutter. *Perspectives of the ASHA Special Interest Groups, 2*(4), 4–27. https://doi.org/10.1044/persp2.SIG4.4

Conture, E. G., & Walden, T. A. (2012). Dual diathesis-stressor model of stuttering. *Theoretical Issues of Fluency Disorders, 41,* 94–127.

Davis, S., Shisca, D., & Howell, P. (2007). Anxiety in speakers who persist and recover from stuttering. *Journal of Communication Disorders, 40*(5), 398–417. https://doi.org/10.1016/j.jcomdis.2006.10.003

Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *Structured Photographic Expressive Language Test–Third Edition.* Janelle Publications.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders.* Brookes.

Dunn, L. M., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test–III.* AGS.

*Dworzynski, K., Remington, A., Rijsdijk, F., Howell, P., & Plomin, R. (2007). Genetic etiology in cases of recovered and persistent stuttering in an unselected, longitudinal sample of young twins. *American Journal of Speech-Language Pathology, 16*(2), 169–178. https://doi.org/10.1044/1058-0360(2007/021)

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315,* 629–634. https://doi.org/10.1136/bmj.315.67109.29

*Erdemir, A., Walden, T. A., Jefferson, C. M., Choi, D., & Jones, R. M. (2018). The effect of emotion on articulation rate in persistence and recovery of childhood stuttering. *Journal of Fluency Disorders, 56,* 1–17. https://doi.org/10.1016/j.jfludis.2017.11.003

Feld, H. (2006). A study into the predictors for stuttering: Investigating the EXPLAN Model and language-based theories [Unpublished master's thesis]. University College London, England.

Fenson, L., Dale, P. S., & Reznick, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual.* Singular.

Forster, D. C. (1997). Speech-motor control and interhemispheric relations in recovered and persistent stuttering [Unpublished dissertation/thesis]. Carleton University, Ottawa, Canada.

Forster, D. C., & Webster, W. G. (2001). Speech-motor control and interhemispheric relations in recovered and persistent stuttering. *Developmental Neuropsychology, 19*(2), 125–145. https://doi.org/10.1207/S15326942DN1902_1

*Franken, M.-C. J. P., Koenraads, S. P. C., Holtmaat, C. E. M., & van der Schroeff, M. P. (2018). Recovery from stuttering in preschool-age children: 9-Year outcomes in a clinical population. *Journal of Fluency Disorders, 58*, 35–46. https://doi.org/10.1016/j.jfludis.2018.09.003

Garnett, E. O., Chow, H. M., & Chang, S.-E. (2019). Neuroanatomical correlates of childhood stuttering: MRI indices of white and gray matter development that differentiate persistence versus recovery. *Journal of Speech, Language, and Hearing Research, 62*(8S), 2986–2998. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0356

Garnett, E. O., Chow, H. M., Nieto-Castañón, A., Tourville, J. A., Guenther, F. H., & Chang, S.-E. (2018). Anomalous morphology in left hemisphere motor and premotor cortex of children who stutter. *Brain, 141*(9), 2670–2684. https://doi.org/10.1093/brain/awy199

Gerwin, K., Brosseau-Lapré, F., Brown, B., Christ, S., & Weber, C. (2019). Rhyme production strategies distinguish stuttering recovery and persistence. *Journal of Speech, Language, and Hearing Research, 62*(9), 3302–3319. https://doi.org/10.1044/2019_JSLHR-S-18-0294

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition*. AGS.

Gordon, P. A., & Luper, H. L. (1992). The early identification of beginning stuttering I: Protocols. *American Journal of Speech-Language Pathology, 1*(3), 43–53. https://doi.org/10.1044/1058-0360.0103.43

Guitar, B., & Conture, E. G. (2006). *The child who stutters: To the pediatrician* (3rd ed.). Stuttering Foundation of America.

Hage, A. (2001). Cognitive and linguistic abilities in young children: Are they able to predict the further development of stuttering. *Sprache Stimme Gehör, 25*(1), 20–24. https://doi.org/10.1055/s-2001-11978

Häge, A., Rommel, D., Schulze, H., & Johannsen, H. S. (1994). Stuttering in childhood: Etiology and follow-up. Initial results of a 5-year longitudinal study. *Folia Phoniatrica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP), 46*(6), 298–304.

Hall, K. D., Amir, O., & Yairi, E. (1999). A longitudinal investigation of speaking rate in preschool children who stutter. *Journal of Speech, Language, and Hearing Research, 42*(6), 1367–1377. https://doi.org/10.1044/jslhr.4206.1367

Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine, 158*(4), 280–289. https://doi.org/10.7326/0003-4819-158-4-201302190-00009

Häge, A., Rommel, D., Schulze, H., & Johannsen, H. S. (1994). Stuttering in childhood: Etiology and follow-up. Initial results of a 5-year longitudinal study. *Folia Phoniatrica et Logopaedica, 46*(6), 298–304. https://doi.org/10.1159/000266332

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*(2), 490–499. https://doi.org/10.1037/0033-2909.92.2.490

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*, 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hilger, A. I. (2015). *New evidence that bimanual motor timing performance is not a significantt factor in developmental stuttering* [Unpublished master's thesis]. Purdue University.

Hilger, A. I., Zelaznik, H., & Smith, A. (2016). Evidence that bimanual motor timing performance is not a significant factor in developmental stuttering. *Journal of Speech, Language, and Hearing Research, 59*(4), 674–685. https://doi.org/10.1044/2016_JSLHR-S-15-0172

Hodson, B. W. (1986). *The Assessment of Phonological Processes–Revised*. Interstate.

*Hollister, J., Van Horne, A. O., & Zebrowski, P. (2017). The relationship between grammatical development and disfluencies in preschool children who stutter and those who recover. *American Journal of Speech-Language Pathology, 26*(1), 44–56. https://doi.org/10.1044/2016_AJSLP-15-0022

Howell, P., Bailey, E., & Kothari, N. (2010). Changes in the pattern of stuttering over development for children who recover or persist. *Clinical Linguistics & Phonetics, 24*(7), 556–575. https://doi.org/10.3109/02699200903581034

Howell, P., & Davis, S. (2011). Predicting persistence of and recovery from stuttering by the teenage years based on information gathered at age 8 years. *Journal of Developmental & Behavioral Pediatrics, 32*(3), 196–205. https://doi.org/10.1097/DBP.0b013e31820fd4a9

Howell, P., Davis, S., & Williams, R. (2008). Late childhood stuttering. *Journal of Speech, Language, and Hearing Research, 51*(3), 669–687. https://doi.org/10.1044/1092-4388(2008/048)

Howell, P., Davis, S., & Williams, S. M. (2006). Auditory abilities of speakers who persisted, or recovered, from stuttering. *Journal of Fluency Disorders, 31*(4), 257–270. https://doi.org/10.1016/j.jfludis.2006.07.001

Hresko, W. P., Reid, D. K., & Hammill, D. D. (1999). *Test of Early Language Development–Third Edition (TELD-3)*. Pro-Ed.

Jameson, A. M. (1955). Stammering in children—Some factors in the prognosis. *Speech, 19*, 60–67.

Johnson, W. (1959). *The onset of stuttering: Research findings and implications*. University of Minnesota Press.

Jones, M., Onslow, M., Packman, A., O'Brian, S., Hearne, A., Williams, S., Ormond, T., & Schwarz, I. (2008). Extended follow-up of a randomized controlled trial of the Lidcombe Program of Early Stuttering Intervention. *International Journal of Language & Communication Disorders, 43*(6), 649–661. https://doi.org/10.1080/13682820801895599

*Kefalianos, E., Onslow, M., Packman, A., Vogel, A., Pezic, A., Mensah, F., Conway, L., Bavin, E., Block, S., & Reilly, S. (2017). The history of stuttering by 7 years of age: Follow-up of a prospective community cohort. *Journal of Speech, Language, and Hearing Research, 60*(10), 2828–2839. https://doi.org/10.1044/2017_JSLHR-S-16-0205

Kelly, P., & Perry, I. J. (2011). Audit of outcomes from preschool stammering intervention in a defined catchment population in the west of Ireland. *Irish Medical Journal, 104*(5), 138–139.

Kelman, E., & Nicholas, A. (2008). *Practical intervention for early childhood stammering: Palin PCI Approach*. Speechmark Publishing.

*Kloth, S. A. M., Kraaimaat, F. W., Janssen, P., & Brutten, G. J. (1999). Persistence and remission of incipient stuttering among high-risk children. *Journal of Fluency Disorders, 24*(4), 253–265. https://doi.org/10.1016/S0094-730X(99)00016-9

Kreidler, K. (2015). Neural systems for language processing in early childhood may help predict eventual stuttering persistence and recovery [Unpublished master's thesis]. Purdue University, West Lafayette, IN, United States.

Kreidler, K., Hampton Wray, A., Usler, E., & Weber, C. (2017). Neural indices of semantic processing in early childhood

distinguish eventual stuttering persistence and recovery. *Journal of Speech, Language, and Hearing Research, 60*(11), 3118–3134. https://doi.org/10.1044/2017_JSLHR-S-17-0081

Lee, L. L., & Canter, S. M. (1971). Developmental Sentence Scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*(3), 315–340. https://doi.org/10.1044/jshd.3603.315

*Leech, K., Bernstein Ratner, N., Brown, B., & Weber, C. (2017). Preliminary evidence that growth in productive language differentiates childhood stuttering persistence and recovery. *Journal of Speech, Language, and Hearing Research, 60*(11), 3097–3109. https://doi.org/10.1044/2017_JSLHR-S-16-0371

Mahurin-Smith, J., & Ambrose, N. G. (2013). Breastfeeding may protect against persistent stuttering. *Journal of Communication Disorders, 46*(4), 351–360. https://doi.org/10.1016/j.jcomdis.2013.06.001

Mansson, H. (2005). Stammens kompleksitet og diversitet [Stuttering's complexity and variability]. *Dansk Audiologopædi, 41*, 13–33.

Månsson, H. (2000). Childhood stuttering: Incidence and development. *Journal of Fluency Disorders, 25*(1), 47–57. https://doi.org/10.1016/S0094-730X(99)00023-6

Mohan, R., & Weber, C. (2015). Neural systems mediating processing of sound units of language distinguish recovery versus persistence in stuttering. *Journal of Neurodevelopmental Disorders, 7*(1), 28. https://doi.org/10.1186/s11689-015-9124-7

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

Nippold, M. A. (2018). Stuttering in preschool children: Direct versus indirect treatment. *Language, Speech, and Hearing Services in Schools, 49*(1), 4–12. https://doi.org/10.1044/2017_LSHSS-17-0066

Ntourou, K., Conture, E. G., & Lipsey, M. W. (2011). Language abilities of children who stutter: A meta-analytical review. *American Journal of Speech-Language Pathology, 20*(3), 163–179. https://doi.org/10.1044/1058-0360(2011/09-0102)

Onslow, M., & Packman, A. (1999). The Lidcombe Programme and natural recovery: Potential choices of initial management strategies for early stuttering. *Advances in Speech Language Pathology, 1*(2), 113–121. https://doi.org/10.3109/14417049909167165

Paden, E. P., Ambrose, N. G., & Yairi, E. (2002). Phonological progress during the first 2 years of stuttering. *Journal of Speech, Language, and Hearing Research, 45*(2), 256–267. https://doi.org/10.1044/1092-4388(2002/020)

Paden, E. P., & Yairi, E. (1996). Phonological characteristics of children whose stuttering persisted or recovered. *Journal of Speech and Hearing Research, 39*(5), 981–990. https://doi.org/10.1044/jshr.3905.981

Paden, E. P., Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering II: Initial status of phonological abilities. *Journal of Speech, Language, and Hearing Research, 42*(5), 1113–1124. https://doi.org/10.1044/jslhr.4205.1113

Panelli, C. A., McFarlane, S. C., & Shipley, K. G. (1978). Implications of evaluating and intervening with incipient stutterers. *Journal of Fluency Disorders, 3*(1), 41–50. https://doi.org/10.1016/0094-730X(78)90005-0

Reilly, S., Onslow, M., Packman, A., Cini, E., Conway, L., Ukoumunne, O. C., Bavin, E. L., Prior, M., Eadie, P., Block, S., & Wake, M. (2013). Natural history of stuttering to 4 years of age: A prospective community-based study. *Pediatrics, 132*(3), 460–467. https://doi.org/10.1542/peds.2012-3067

Reilly, S., Onslow, M., Packman, A., Wake, M., Bavin, E. L., Prior, M., Eadie, P., Cini, E., Bolzonello, C., & Ukoumunne, O. C. (2009). Predicting stuttering onset by the age of 3 years: A prospective, community cohort study. *Pediatrics, 123*(1), 270–277. https://doi.org/10.1542/peds.2007-3219

Reynell, J. (1983). *Reynell Developmental Language Scale Manual–Revised Edition*. Nfer-Nelson Publishing.

Riley, G. D. (1972). A Stuttering Severity Instrument for Children and Adults. *Journal of Speech and Hearing Disorders, 37*(3), 314–322.

Robertson, C. (2006). Factors influencing persistence and recovery in stuttering: The role of intelligence, self-esteem and bullying [Unpublished master's thesis]. University College London, England.

*Roehl, L. (2018). *Neural systems supporting nonlinguistic auditory processing in young children who persist and recover from stuttering* [Master's thesis]. Michigan State University.

*Rommel, D., Hage, A., Kalehne, P., & Johannsen, H. S. (2001). Development, maintenance, and recovery of childhood stuttering: Prospective longitudinal data three years after first contact. In K. L. Baker, L. Rustin, & F. Cook (Eds.), *Proceedings of the Fifth Oxford Dysfluency Conference*. Kevin L. Baker.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*(5), 1394–1408. https://doi.org/10.1111/1467-8624.00355

*Ryan, B. P. (2001). A longitudinal study of articulation, language, rate, and fluency of 22 preschool children who stutter. *Journal of Fluency Disorders, 26*(2), 107–127. https://doi.org/10.1016/S0094-730X(01)00095-X

Saez, A. (2002). An investigation of stuttering recovery and articulation status [Unpublished master's thesis]. Southern Connecticut State University, New Haven, CT, United States.

Sanson, A. V., Smart, D. F., Prior, M., Oberklaid, F., & Pedlow, R. (1994). The structure of temperament from three to seven years: Age, sex, and demographic influences. *Merrill-Palmer Quarterly, 40*, 233–252.

Sawyer, J. (2005). *Characteristics of disfluency clusters in stuttering and normally fluent preschool children: A longitudinal study* [Doctoral dissertation]. University of Illinois.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*(1), 1–22. https://doi.org/10.1017/S0142716400008262

Sheehan, J. (1952). Theory and treatment of stuttering as an approach—Avoidance conflict. *The Journal of Psychology, 36*(1), 27–49. https://doi.org/10.1080/00223980.1953.9712875

Shimada, M., Toyomura, A., Fujii, T., & Minami, T. (2018). Children who stutter at 3 years of age: A community-based study. *Journal of Fluency Disorders, 56*, 45–54. https://doi.org/10.1016/j.jfludis.2018.02.002

Singer, C. M., Walden, T. A., & Jones, R. M. (2019). Differences in the relation between temperament and vocabulary based on children's stuttering trajectories. *Journal of Communication Disorders, 78*, 57–68. https://doi.org/10.1016/j.jcomdis.2019.01.004

*Singer, C. M., Walden, T. A., & Jones, R. M. (2020). Attention, speech-language dissociations and stuttering chronicity. *American Journal of Speech-Language Pathology, 29*(1), 157–167. https://doi.org/10.1044/2019_AJSLP-19-00039

Smith, A., & Weber, C. (2017). How stuttering develops: The multifactorial dynamic pathways theory. *Journal of Speech, Language,*

and *Hearing Research, 60*(9), 2483–2505. https://doi.org/10.1044/2017_JSLHR-S-16-0343

Smith, K. (2007). *Factors influencing persistence and recovery in stuttering: The role of ethnicity, self-esteem, and bullying* [Unpublished master's thesis]. University College London, England.

Smith, K., Iverach, L., O'Brian, S., Mensah, F., Kefalianos, E., Hearne, A., & Reilly, S. (2017). Anxiety in 11-year-old children who stutter: Findings from a prospective longitudinal community sample. *Journal of Speech, Language, and Hearing Research, 60*(5), 1211–1222. https://doi.org/10.1044/2016_JSLHR-S-16-0035

Spencer, C. (2013). *Preschool language and phonological proficiencies in predicting stuttering recovery or persistence* [Unpublished master's thesis]. Purdue University, West Lafayette, IN, United States.

*Spencer, C., & Weber-Fox, C. (2014). Preschool speech articulation and nonword repetition abilities may help predict eventual recovery or persistence of stuttering. *Journal of Fluency Disorders, 41,* 32–46. https://doi.org/10.1016/j.jfludis.2014.06.001

Starkweather, G. W. (1982). Stuttering in children: An overview. *Journal of Childhood Communication Disorders, 6*(1), 5–14. https://doi.org/10.1177/152574018200600102

Subramanian, A., Yairi, E., & Amir, O. (2003). Second formant transitions in fluent speech of persistent and recovered preschool children who stutter. *Journal of Communication Disorders, 36,* 59–75. https://doi.org/10.1016/S0021-9924(02)00135-1

Throneburg, R. N. (1997). *Characteristics of disfluencies of preschool children who stutter: A longitudinal study regarding persistence and recovery* [Unpublished doctoral dissertation]. University of Illinois Urbana-Champaign, United States.

Throneburg, R. N., & Yairi, E. (2001). Durational, proportionate, and absolute frequency characteristics of disfluencies: A longitudinal study regarding persistence and recovery. *Journal of Speech, Language, and Hearing Research, 44*(1), 38–51. https://doi.org/10.1044/1092-4388(2001/004)

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20,* 375–393. https://doi.org/10.1037/met0000011

Tumanova, V., Conture, E. G., Lambert, E. W., & Walden, T. A. (2014). Speech disfluencies of preschool-age children who do and do not stutter. *Journal of Communication Disorders, 49,* 25–41. https://doi.org/10.1016/j.jcomdis.2014.01.003

Usler, E., Smith, A., & Weber, C. (2017). A lag in speech motor coordination during sentence production is associated with stuttering persistence in young children. *Journal of Speech, Language, and Hearing Research, 60*(1), 51–61. https://doi.org/10.1044/2016_JSLHR-S-15-0367

Usler, E., & Weber-Fox, C. (2015). Neurodevelopment for syntactic processing distinguishes childhood stuttering recovery versus persistence. *Journal of Neurodevelopmental Disorders, 7*(1), 4. https://doi.org/10.1186/1866-1955-7-4

*Walsh, B., Usler, E., Bostian, A., Mohan, R., Gerwin, K. L., Brown, B., Weber, C., & Smith, A. (2018). What are predictors for persistence in childhood stuttering? *Seminars in Speech and Language, 39*(4), 299–312. https://doi.org/10.1055/s-0038-1667159

Watkins, R. V., & Yairi, E. (1997). Language production abilities of children whose stuttering persisted or recovered. *Journal of Speech, Language, and Hearing Research, 40*(2), 385–399. https://doi.org/10.1044/jslhr.4002.385

*Watkins, R. V., Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering III: Initial status of expressive language abilities. *Journal of Speech, Language, and Hearing Research, 42*(5), 1125–1135. https://doi.org/10.1044/jslhr.4205.1125

Williams, K. (1997). *Expressive Vocabulary Test–Second Edition (EVT-2)*. AGS.

Yairi, E., & Ambrose, N. (1992). Onset of stuttering in preschool children. *Journal of Speech and Hearing Research, 35*(4), 782–788. https://doi.org/10.1044/jshr.3504.782

Yairi, E., & Ambrose, N. (1999). Early childhood stuttering I: Persistence and recovery rates. *Journal of Speech, Language, and Hearing Research, 42*(1), 1097–1112. https://doi.org/10.1044/jslhr.4205.1097

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders, 38*(2), 66–87. https://doi.org/10.1016/j.jfludis.2012.11.002

*Yairi, E., & Ambrose, N. G. (2005). *Early childhood stuttering for clinicians by clinicians*. Pro-Ed.

Yairi, E., Ambrose, N. G., Paden, E. P., & Throneburg, R. N. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering. *Journal of Communication Disorders, 29*(1), 51–77. https://doi.org/10.1016/0021-9924(95)00051-8

Yaruss, S. J. (1998). Real-time analysis of speech fluency: Procedures and reliability training. *American Journal of Speech-Language Pathology, 7*(2), 25–37. https://doi.org/10.1044/1058-0360.0702.25

Zebrowski, P. M. (1997). Assisting young children who stutter and their families: Defining the role of the speech-language pathologist. *American Journal of Speech-Language Pathology, 6*(2), 19–28. https://doi.org/10.1044/1058-0360.0602.19

*Zengin-Bolatkale, H., Conture, E. G., Walden, T. A., & Jones, R. M. (2018). Sympathetic arousal as a marker of chronicity in childhood stuttering. *Developmental Neuropsychology, 43*(2), 135–151. https://doi.org/10.1080/87565641.2018.1432621

Zimmerman, I., Steiner, V., & Pond, R. (1979). *Preschool Language Scale*. The Psychological Corporation.

Appendix A

Reasons Reports Excluded at Full-Text Stage

Report	Reason for exclusion
Ambrose (1996)	Duplicate/dissertation
Andrews & Harris (1964)	Unable to calculate effect size
Brosch et al. (1999)	Follow-up too short
Chang (2005)	Participants too old at study entry
Chow & Chang (2017)	Participants too old at study entry
Clare (2008)	Participants too old at study entry
Davis et al. (2007)	Participants too old at study entry
Forster (1997)	Insufficient study design
Forster (1997)	Duplicate/dissertation
Forster & Webster (2001)	Insufficient study design
Garnett et al. (2019)	Participants too old at study entry
Garnett et al. (2018)	Participants too old at study entry
Häge et al. (1994)	Follow-up too short
Feld (2006)	Insufficient study design
Hilger (2015)	Duplicate/dissertation
Howell & Davis (2011)	Participants too old at study entry
Howell et al. (2010)	Insufficient study design
Howell et al. (2006)	Insufficient study design
Howell et al. (2008)	Participants too old at study entry
Jameson (1955)	Participants too old at study entry
Johnson (1959)	Participants too old at study entry
Kelly & Perry (2011)	Unable to calculate effect size
Kreidler (2015)	Duplicate/dissertation
Månsson (2000)	Unable to calculate effect size
Mansson (2005)	Couldn't be translated
Mohan & Weber (2015)	Insufficient study design
Panelli et al. (1978)	Unable to calculate effect size
Reilly et al. (2013)	Follow-up too short
Robertson (2006)	Insufficient study design
Saez (2002)	Insufficient study design
Shimada et al. (2018)	Follow-up too short
K. Smith (2007)	Insufficient study design
K. Smith et al. (2017)	Insufficient study design
Spencer (2013)	Duplicate/dissertation
Throneburg (1997)	Duplicate/dissertation
Usler & Weber-Fox (2015)	Factor not measured near onset
Usler et al. (2017)	Factor not measured near onset

Report Characteristics by Study

Report	Type	Additional eligibility criteria[a]	Persisting	Recovered	Follow-up length	Confounding	Clinical characteristic reported
California - Long Beach							
Ryan (2001)	PR		7 (4 M) 50 mo. (11)	15 (10 M)	2–10 years	NR	**S,** [SS], [RV], [GL], [SF]
Early Language in Victoria Study							
Kefalianos et al. (2017)	PR		36 (24 M) 48 mo.	67 (37 M) 48 mo.	At least 3 years	Adjusted values reported	S, **EV, FH,** T
Erasmus University Medical Centre Study							
Franken et al. (2018)	PR		4 (4 M) 52.25 mo. (7.0)	11 (6 M) 43.36 mo. (8.6)	At least 9 years	N/A	**S, A, FH, FH-P, FH-R, PSS, CSS, SF**
Illinois Stuttering Research Project							
Watkins et al. (1999)	PR		22 (18 M) 43 mo.	62 (40 M) 39 mo.	At least 4 years	Data stratified by age	S, **MLU**, NDW,* NTW,* **DSS**
Yairi & Ambrose (2005)	BK	Within 12 mo. of onset	19 (15 M) 43.16 mo. (9.52)	70 (49 M) 36.87 mo. (7.66)	At least 4 years	Multiple methods reported	S, SS,[b] EL, EV, FH, FH-P, [PSS], [CSS], SF
Netherlands – High Risk							
Kloth et al. (1999)	PR		7 (5 M) 40 mo. (6.45)	16 (7 M) 36 mo. (9.8)	At least 4 years	NR	**S, EL, RL, MLU**
Outcomes of Children with Hearing Loss							
Arenas et al. (2017)	PR		7	11	At least 2 years or until 8 years old	NR	**S, SS, EL, RL**
Purdue Stuttering Project							
Bostian (2017)	UP	4–5 years old at entry	19 (14 M) 58.05 mo. (1.71)	29 (22 M) 53.58 mo. (1.04)		N/A	S, SF, **OD, PW, SW, DP,** MR,* CSS
Leech et al. (2017)	PR		28 (15 M) 62 mo.	22 (11 M) 57 mo.	2 years	Age and SES not statistically different	S, A, SS, RL, EL, NRT,* **IPSYN**
Roehl (2018)	UP		16 (12 M) 57.72 mo. (1.7)	26 (17 M) 54.84 mo. (1.3)	2–5 years	NR	S, **A, EL, RL, SF, CSS, PSS**
Spencer & Weber-Fox (2014)	PR		19 (15 M) 57.11 mo. (6.71)	21 (14 M) 53.33 mo. (5.36)	12–48 months	Age, nonverbal reasoning, SES matching	S, **SS,** EL, RL, NRT*
Walsh et al. (2018)[c]	PR		31 (26 M)	32 (21 M)			**S, FH,[d] FH-P**,[d] FH-R[d]
Subtypes and Risk Factors							
Ambrose et al. (2015)	PR		19 (12 M) 43.63 mo. (11.96)	39 (27 M) 38.18 mo. (7.57)	up to 5 years	NR	**S, A, EL, RL, EV, EL,** [SS], [MLU], **CSS, PSS, SF, PW, SW, DP,** RU,* **T**
Buhr (2007)	UP	Language skills within normal limits	4 49 mo.	4 50 mo.	2 years	NR	GL,* DSS, MLU, SF, **OD**

(table continues)

Report Characteristics by Study

Report	Type	Additional eligibility criteria[a]	Persisting	Recovered	Follow-up length	Confounding	Clinical characteristic reported
Hollister et al. (2017)	PR	Excluded children > 43 months or if speech sample < 100 utterances	8 37.1 mo. (4.6)	5 35.2 mo. (5.8)	2 years	Age not statistically different	EL, RL, EV, RV, SF, CSS, **IPSYN, MLU**
Twins Early Development Study Dworzynski et al. (2007)	PR		135 (100 M)	950 (521 M)		N/A	**S**
The Ulm Study Rommel et al. (2001)	UP		19 (16 M) 62.4 mo.	46 (33 M) 58.8 mo.	3 years	NR	**S**, [A], **FH**, [RV], [L], [MLU], [NDW],
Vanderbilt Developmental Stuttering Project Erdemir et al. (2018)	PR		10 (9 M) 46.9 mo. (4.5)	10 (9 M) 46.1 mo. (6.9)	2–2.5 years	Sex, age, and language matched	S, **A,** SF, CSS
Singer et al. (2020)	PR	Under 5;0 years at entry	10 (9 M) 46.6 mo. (4.5)	32 (23 M) 45.0 mo. (6.8)	At least 2 years	Age and SES not statistically different	**S, CSS, SF, SS, EL, RL, EV, RL**
Zengin-Bolatkale et al. (2018)	PR		9 (8 M) 45.11 mo. (6.23)	23 (17 M) 47.11 mo. (6.84)	16–32 months	Age not statistically different	S, CSS, SF, SS, EL, RL, EV, RV, CRS,* **T**

Note. Bold indicates data were extracted for this study; brackets indicate data could not be used to estimate an effect size; asterisk indicates data not reported in another study. PR = published report; M = male; NR = not reported; S = sex; SS = speech sound accuracy; RV = receptive vocabulary; GL = global language; SF = stuttered-like disfluency frequency; mo. = months; EV = expressive vocabulary; FH = family history of stuttering; T = temperament; N/A = not available; A = age at onset; FH-P = family history of persistent stuttering; FH-R = family history of recovered stuttering; PSS = parent-rated stuttering severity; CSS = clinician-rated stuttering severity; MLU = mean length of utterance; NDW = number of different words; NTW = number of total words; DSS = Developmental Sentence Scoring (Lee & Canter, 1971); BK = book; EL = expressive language; RL = receptive language; UP = unpublished; OD = other disfluencies; PW = part-word repetition frequency; SW= single-word repetition frequency; DP = dysrhythmic phonation frequency; MR = max repetition units; NRT = nonword repetition task; SES = socioeconomic status; IPSYN = Index of Productive Syntax (Scarborough, 1990); RU = repetition units (mean); L = language; CRS = child response to stuttering.

[a]Criteria adopted in addition to the study-wide criteria. [b]Reported for 19 persistent children and 65 recovered children. [c]Additional data for calculated effect sizes collected through personal communication via e-mail with the first author (Walsh, personal communication, April 24, 2019). [d]Reported for 17 persistent children and 28 recovered children.

Appendix C

Speech Sample Characteristics by Study

	Speech sample characteristics			
Study	Duration	Word/syllable count	Communication partner	Activity examples
Erasmus Medical Center Study	30 min × 2	1,000 syllables	Caregiver	NR
Illinois Stuttering Research Project	NR	1,000 words	Primary caregiver	Play-Doh
Netherlands Study	30 min (10 min coded)	NR	Mother	NR
Purdue Stuttering Project	12 min × 2	750–1,000 words	Caretaker; SLP	Clay
Subtypes and Risk Factors in Childhood Stuttering	20 min × 2	1,000+ syllables	Primary caregiver; SLP	Play-Doh
Vanderbilt Developmental Stuttering Project	NR	300 words	Clinician	Farm set

Note. NR = not reported; SLP = speech-language pathologist.