medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Childhood listening and associated cognitive difficulties persist into adolescence

Katsuaki Kojima[1,2,3], Li Lin[3], Lauren Petley[3,4], Nathan Clevenger[3], Mark Bodik[3,5], David R. Moore[3,6,7]

1. Perinatal Institute, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039
2. Department of Pediatrics, University of Cincinnati College of Medicine, 3230 Eden Avenue, Cincinnati, OH 45267
3. Communication Sciences Research Center, Cincinnati Children's Hospital Medical Center, 240 Albert Sabin Way, Cincinnati, OH 45229
4. Department of Psychology, Clarkson University, 8 Clarkson Ave, Potsdam, NY 13699
5. Cornell University, Ithaca, NY 14850
6. Departments of Otolaryngology and Neuroscience, University of Cincinnati College of Medicine, Cincinnati, OH
7. Manchester Centre for Audiology and Deafness, University of Manchester, UK

Keywords: hearing loss, speech-in-noise, listening difficulty, cognition, pediatric, longitudinal, auditory processing disorder, NIH Cognition Toolbox

Corresponding author: Katsuaki.kojima@cchmc.org

Data sharing statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Research highlights

- Children aged 6-13 years with clinically normal audiograms are often reported by their caregivers to have listening difficulties
- In a four-year longitudinal evaluation, children with listening difficulties were found to have persistent challenges in competing speech segregation and generally poor cognitive performance
- Maternal education, spatial segregation of speech, and cognitive performance independently predicted the degree of the children's listening difficulties

Abstract (213 words /250 words)

Our objective in this study was to examine children with and without listening difficulty (LiD) longitudinally to ask whether LiD was developmentally transient or persistent. Like many other developmental problems, we hypothesized that LiD persists into adolescence. Behavioral and physiological data were collected from children initially aged 6-13 years at baseline and two and four years later. Of 169 enrolled participants, 147, 100, and 31 children completed required testing in study visits at baseline, 2 years, and 4 years later, respectively. All children had clinically normal audiograms at all visits. Listening skills scores from a caregiver report scale (ECLiPS), and auditory and cognitive skills were significantly poorer in the LiD group than in a typically developing (TD) group throughout the study period. In both groups, age-adjusted listening and auditory skills increased over time. Using the longitudinal data, a parsimonious prediction model for ECLIPS scores (pooled across groups) was created. The final model included maternal education, spatial listening skills, and cognitive performance, which explained 54.8% of the variance in the ECLiPS score. Children with LiD but normal audiograms have, relative to TD children, persistent listening difficulties, challenges in competing speech tasks, and poor cognitive performance through adolescence. The degree of LiD can be independently predicted by maternal education, cognitive performance, and spatial listening skills.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

Language development is critically dependent on sensitive hearing, as classically measured using pure tone audiometry (Tomblin et al. 2015; Carhart and Jerger 1959). However, active listening to sounds is also of great importance since attention must be focused to accurately comprehend the meaning of speech and other sound sequences (Fleta 2014; Bulut and Karasakaloglu 2017). Listening difficulty (LiD) has been defined as difficulty hearing and understanding speech and other suprathreshold sounds despite having normal pure-tone hearing sensitivity (Boothalingam et al. 2019; Pienkowski 2017; Farah et al. 2014; Parthasarathy et al. 2020; Petley et al. 2021). Childhood LiD can be assessed using a validated and standardized caregiver questionnaire, the ECLiPS (Barry et al. 2015). The prevalence of LiD in children is unknown. Previously, approximately 1–5% of children have been shown to consult audiology services without having a hearing loss (Hind et al. 2011; Brewer et al. 2016). This is within the range of young children who do have hearing loss (Schmucker et al. 2019). LiD has been closely associated with multiple other developmental problems, including attention deficit hyperactivity disorder (Sharma et al. 2014), reading disorders (Dawes and Bishop 2010), developmental language disorder (Ferguson et al. 2011), and autism spectrum disorder (Azouz et al. 2014).

To determine the contribution of hearing, listening, and non-auditory factors to LiD, we recently completed the longitudinal Sensitive Indicators of Childhood Listening Difficulties (SICLiD) project (Moore et al. 2020; Hunter et al. 2021; Petley et al. 2021). We characterized LiD in 6- to 13-year-old children with normal audiograms during the baseline, cross-sectional phase of the project (Wave 1) using various caregiver reports and auditory, physiological, and cognitive test measures (Petley et al. 2021). We found that these children had mostly normal ear (Hunter et al. 2021) and brainstem (Hunter et al., submitted) function but higher speech hearing signal/noise thresholds (LiSN-S; (Cameron and Dillon 2007)) and lower cognitive scores (NIH Cognition Toolbox; (Weintraub et al. 2013b)) than typically developing (TD) children (Petley et al. 2021).

Whether LiD is a transient childhood problem that resolves with age or persists into adulthood is unclear. Answering this question is essential for developing management strategies and understanding the mechanisms and consequences of LiD. Little is currently known about what may be effective treatments for LiD, although some that have been proposed show promise, including computer-based training (Loo et al. 2016; Weihing et al. 2015), classroom amplification (Rosenberg 2002; Johnston et al. 2009; Keith and Purdy 2014), and low-gain or remote microphone hearing aids (Keith and Purdy 2014; Schafer et al. 2014)). Persistent LiD may have a more significant impact on the individual, family, and society over a long time. More comprehensive assessment and more aggressive treatment strategies may therefore be warranted.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Young adults with a history of LiD in childhood reported continuing communication difficulty almost 10 years after a clinical diagnosis (Del Zoppo et al. 2015). Still, longitudinal quantitative data on listening, auditory, and cognitive abilities in children with earlier LiD is lacking. We hypothesized that LiD would be long-lasting because LiD was seen across ages (between 6 and 13 years old) in our previous cross-sectional data (Petley et al., 2021) and because of the general persistence through adolescence or into adulthood of other associated developmental problems of attention (Hechtman et al. 2016; Roy et al. 2017), reading (Shaywitz et al. 1999; Lohvansuu et al. 2021), and language (St Clair et al. 2011).

For both cross-sectional and longitudinal research, the reliability of measures is critical for accurate evaluation. One specific issue recently identified was the reliability of the NIH Cognition Toolbox (Taylor et al. 2020), a test suite inspired by a desire for greater standardization of neurobehavioral tests (Weintraub et al. 2013b). As a secondary aim, we assessed long-term test-retest reliabilities of the NIH Cognition Toolbox and other auditory measures (ECLiPS, LiSN-S) we previously examined. Finally, building on a model that explained 42% of the variance in our cross-sectional data (Petley et al. 2021), we aimed to develop an enhanced model using longitudinal data that better characterizes LiD.

MATERIALS AND METHODS

The Institutional Review Board of Cincinnati Children's Hospital (CCH) Research Foundation approved this study.

Participants

The LiD group's eligibility requirements included caregiver-reported listening difficulties (see Procedure), age between 6 and 13 years old upon enrollment, English native language, and the absence of any otologic, neurologic, psychiatric, or intellectual inability to complete testing procedures (Petley et al. 2021). Typically developing (TD) participants were eligible based on the same criteria. Additionally, they could not have any caregiver-reported listening difficulties, developmental delay, language, attention, or learning disorder.

One hundred sixty-six participants (74 with LiD, 92 TD) were enrolled; 19 withdrew or otherwise exited the study (6 with LiD, 13 TD). The remaining 147 participants (68 with LiD, 79 TD) completed the assessments at baseline (Wave 1). After the initial study visit, we contacted the participants after ~2 years (Wave 2) and 4 years (Wave 3) to collect longitudinal data. A total of 100 (52 with LiD, 48 TD) and 31 (16 with LiD, 15 TD) participated in Wave 2 and 3, respectively (Table 1). The sample size for Wave 3 was notably smaller than other waves, partly because of the COVID-19 pandemic. Due to attrition and updates to testing procedures, the sample size for each assessment was variable (see Table 1). All participants had clinically normal hearing bilaterally, as detailed below.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Procedure

The recruitment procedure and test details are described in our previous reports (Petley et al. 2021; Moore et al. 2018; Hunter et al. 2021). Briefly, participants with LiD were recruited from Cincinnati Children's Hospital and the broader community. Those who opted in completed (i) a consent form, (ii) the background questionnaire, and (iii) the ECLiPS (Barry and Moore 2021) to confirm eligibility for the LiD or TD group. According to institutional policy, children aged 11 and above were also assented using a child-friendly version of the consent document. All participants received financial compensation for their participation. During the study visit, each participant completed various behavioral and audiological testing.

Background Questionnaire

This questionnaire asked about the participant's medical history, parental education, and demographic information. Key variables were date of birth, language, race and ethnicity, maternal education (a proxy for socioeconomic position (Sherar et al. 2016)), history of ear and hearing problems (including pressure equalization, PE, tube placement), diagnosis of, or treatment for, learning problems (attention disorders, developmental delays, speech-language disorders), neurological (e.g., history of head trauma) or psychiatric conditions, school interventions, and birth history (prematurity, NICU stay). Demographic variables, including maternal education, age, gender, and race, obtained from this questionnaire, are summarized at each time point in Table 1 for the LiD and TD groups.

Caregiver Evaluation of Children's Listening

The ECLiPS (Barry and Moore 2021) was completed in each wave to profile participants' listening and communication abilities. It contains 38 simple statements (items) describing behaviors commonly observed in children. Caregivers were asked to rate their child by how much they agreed with each statement on a five-point scale ranging from strongly disagree to strongly agree. The ratings were averaged to derive scores, which were scaled by age, on five subscales (speech and auditory processing [SAP], environmental and auditory sensitivity [EAS], language/literacy/laterality [L/L/L], memory and attention [M&A], and pragmatic and social skills [PSS]), each containing 6 to 9 distinct items. These scales were further collected under Language, Listening, Social, and Total aggregate (composite) scores. All scales and composites were standardized for a population mean of 10 (SD = 3) based on British data (Barry et al. 2015). All of the subscales of the ECLiPS have high test-retest reliability, with intraclass correlations (ICCs) above 0.8 (Barry and Moore 2021). Construct validity has been demonstrated through convergence with other established tests that measure similar skills, including the Children's Auditory

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Processing Performance Scale (CHAPS; (Smoski et al. 1998)), the CCC-2 (Bishop 2006), and the Social Communication Questionnaire (Rutter et al. 2003).

Audiometry

All participants received a full pure-tone hearing threshold assessment at octave intervals from 0.25 to 16 kHz (plus 10 and 12 kHz), bilaterally, using Sennheiser HAD 300 headphones and a modified Hughson-Westlake procedure (see (Hunter et al. 2021)). Participants with elevated thresholds (>20 dB HL) in either ear and at any standard clinical frequency (0.25 - 8 kHz) were excluded and referred for further clinical care as appropriate.

Speech in Speech Hearing (LiSN-S)

The Listening in Spatialized Noise – Sentences (LiSN-S) test (Cameron and Dillon 2007; Brown et al. 2010) measures the ability to listen and repeat simple, spoken sentences in the presence of distracting sentences, developed using the same child-friendly criteria as the BKB sentences (Bench et al. 1979). The LiSN-S (US Edition (Brown et al. 2010)) was administered at each wave using a commercial CD played on a laptop, a task-specific soundcard, and Sennheiser HD 215 headphones. Participants were asked to repeat a series of target sentences (“T”), presented using virtual space generic head-related transfer functions (Humanski and Butler 1988), directly in front (0°: diotic) while ignoring two distracting speakers (“D1”, “D2”). There are four listening conditions in which the distractors change voice (different or same as target) and/or position (0° and 90°) (Humanski and Butler 1988). The LiSN-S is adaptive; the level of the target speaker decreases or increases in SNR relative to the distracting speech if the listener responds correctly or incorrectly during up to 30 sentences in each condition. The speech reception threshold (SRT) represents 50% correct SNR for the condition. The Low Cue condition is where distractors are in the same voice as the target and 0° relative to the listener. Distractors are in different voices and ±90° relative to the listener in the High Cue condition. The derived scores of the LiSN-S are the Talker Advantage and Spatial Advantage, so-called because each is the difference between SRTs from two conditions. This subtraction process should, to varying extent, separate auditory from cognitive influences (Petley et al. 2021). Spatial Advantage thus represents the participant’s ability to separate the spatial source of the speech using purely auditory cues, while Talker Advantage demonstrates the participant’s ability to distinguish individual talkers, which involves both acoustic and linguistic information (Perrachione et al. 2011; Quinto et al. 2020). Test-retest comparisons for the four listening conditions of the LiSN-S showed significant improvement from the first test to the second but no differences in the advantage scores (Cameron et al. 2011).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A spatial "Pattern Score" is a quantitative clinical measure of the benefit of adding virtual spatial cues to the information in the Low Cue condition of the LiSN-S (i.e., target and distracting stimuli presented diotically) (Cameron and Dillon 2011). The Pattern Score has been used as both a diagnostic score and a treatment outcome for a specific auditory disorder termed Spatial Processing Disorder (SPD) (Cameron et al. 2011, 2012; Cameron et al. 2014).

Cognition (National Institutes of Health, NIH Toolbox)

Each participant's cognitive skills were assessed using the NIH Toolbox for the Assessment of Neurological and Behavioral Function, Cognition Domain (Weintraub et al. 2013a; Weintraub et al. 2013b). According to Toolbox recommendations, participants completed testing online or via an iPad app in a private sound-attenuated booth or a quiet room at each wave. The Toolbox contains up to eight standardized cognitive instruments measuring different aspects of fluid or crystallized reasoning. The precise composition of the testing battery depends on user choice and participant age.

All participants in this study completed the Picture Vocabulary Test (PVT), Flanker Inhibitory Control and Attention Test (Flanker), Dimensional Change Card Sort test (DCCS), and Picture Sequence Memory test (PSMT). Each test produced a US age-corrected standardized score, and the scores of all four tests combine to calculate a single Early Childhood Composite. The PVT is an adaptive test in which the participants are presented with an audio recording of a word and select which of the four pictures most closely matches the word's meaning. In the Flanker, which tests inhibition/attention, the participant reports the direction of a central visual stimulus (left or right, fish or arrow) in a string of five similar, flanking stimuli that may be congruent (same direction as target) or incongruent (opposite direction). The DCCS tests cognitive flexibility (attention switching). Target and test card stimuli vary along two dimensions, shape and color. Participants are asked to match test cards to the target card according to a specified dimension that varies for each trial. Both the PVT and DCCS score accuracy and reaction time. PSMT assesses episodic memory by presenting an increasing number of illustrated objects and activities, each with a corresponding audio-recorded descriptive phrase. Picture sequences vary in length from 6 to 18 pictures depending on age, and participants are scored on the cumulative number of adjacent pairs remembered correctly over two learning trials. The Early Childhood Composite (ECC) obtained by combining these scores is a measure of overall cognitive skill for children 6 years and older.

Children at least 8 years of age qualify for additional subtests from the Toolbox, contributing to a Total Composite (TC) measure of general cognitive skill. These additional tests were administered to all children 8 years of age and older (i.e., all children for Waves 2 and 3). Fluid Composite measures are the list sorting working memory (LSWM), the pattern comparison processing speed (PCPS) tests, and the DCCS, Flanker, and PSMT. LSWM assesses working memory by asking participants to arrange objects

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

presented visually and auditorily (food and animals) in order of size. PCPS requires participants to respond as quickly as possible to whether two visually presented cards are the same or different. Participants completed the Reading Recognition (RR) test in addition to the PVT for the Crystallized Composite measure. The RR requires participants to read words and letters aloud.

The Toolbox tests demonstrate high reliability in pediatric samples with short testing intervals between 7 and 21 days, with intraclass correlations (ICCs) between 0.76 and 0.99 (Weintraub et al. 2013b). Recent evidence suggests that their reliability over longer test-retest intervals (1 to 2 years) is substantially lower, with ICCs between 0.24 and 0.85 (Taylor et al. 2020). The convergent validity of the Toolbox has been assessed against a range of published tests used in clinical practice that measure similar cognitive capacities (Weintraub et al. 2013b). These assessments yielded convergent validity correlations between 0.48 and 0.93, suggesting that the Toolbox instruments index the desired constructs.

Analysis

The primary analysis was divided into three sequential parts. Part 1 examined differences between TD and LiD groups with respect to demographics, auditory, and cognitive function using the ECLiPS, LiSN-S, and NIH Cognition Toolbox. Demographic information included age, sex, race, ethnicity, and maternal education for each group. We developed mixed-effect models (Verbeke 1997; Fitzmaurice et al. 2012) for each demographic characteristic to assess differences between groups (TD and LiD) in the longitudinal data. A mixed-effect model is advantageous in analyzing longitudinal data because it accounts for various correlation structures among the repeated measures and is flexible in handling missing data. With the assumption of the data missing at random (MAR), the mixed-effect model used all the available data with Restricted Maximum Likelihood (REML) estimation. We ran Little's test (Little 1988) to assess if missing data could be assumed as missing completely at random.

For each demographic characteristic, we created a mixed-effect model using group (TD vs. LiD), wave (Wave 1, 2, and 3), and interaction (Group x Wave) as fixed effects and participant as a random effect. Maternal education was collapsed into four groups: up to high school education (High school), some college education (College), Bachelor's degree (Bachelor), and post-graduate education (Post-grad). Race was collapsed into two groups: white and non-white. Ethnicity was collapsed into two groups: Hispanic or Latino and others. Demographic characteristics that are statistically different between groups ($p < .05$) may confound the statistical analysis that compared behavioral testing results between TD and LiD. Therefore, such demographic variables were included in the analyses to control potential confounding.

Next, we evaluated longitudinal changes in listening and cognitive skills of TD and LiD children using the ECLiPS sub-scores and total scaled score, LiSN-S Advantage, Cue and Pattern scores, and NIH

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Toolbox sub-tests and composite scores. For each test, we created a mixed-effect model using group (TD vs. LiD), wave (Wave 1 vs. 2), interaction (Group x Wave), and maternal education (potential confounder) as fixed effects and participant as a random effect. We used standardized scores adjusted for age on the test date for all analyses. To control inflation, p-values were adjusted for multiple comparisons using Benjamini-Hochberg (B-H) adjustment (Benjamini and Hochberg 1995) on eight ECLiPS sub-scores, four LiSN-S scores, and seven NIH-TB sub-scores. We used false discovery rate (FDR) < 5%. We did not include Wave 3 data in these mixed-effect models because of the limited sample size relative to other waves. Instead of whole group-level comparisons, we created separate mixed-effect models using data only of Wave 3 participants to evaluate longitudinal changes among those participants.

The goal of Part 2 of the analysis was to evaluate the test-retest reliability of the measures. We calculated Pearson correlation coefficients and intra-class correlation coefficients (ICC) of test results between Waves 1 and 2. We calculated the coefficients for all participants and, separately, for the TD and LiD groups.

In Part 3, we aimed to identify the functional domains that contributed to LiD across all participants in the longitudinal data. Our prior cross-sectional study created a model that predicted 42% of the variance in ECLiPS Total scaled scores at Wave 1 (Petley et al. 2021). This model included four predictors: the SCAN-3:C Composite Score, the LiSN-S Talker Advantage score, and the NIH PV and DCCS scores. Here, we tested if this model predicted ECLiPS Total scaled scores at Wave 2. Next, we developed a new mixed effect model using the longitudinal data to predict ECLiPS Total Scaled scores. Candidate predictors were maternal education, five LiSN-S scores (Cue, Advantage, and Pattern scores), seven NIH Toolbox subtests, and four NIH Toolbox composite scores. We created separate models for subtests and composite scores for NIH Toolbox testing to avoid including repetitive test results in a model. Owing to its smaller sample size, Wave 3 was omitted from this longitudinal model.

All candidate predictors for the mixed-effect model for Part 3 of the analysis were first examined against the ECLiPS Total scaled score via univariate analyses using a mixed-effect model. These models used participants as a random effect. Predictor variables and the interaction term (predictor x wave) were used as fixed effects. Only variables with $p < 0.1$ in the univariate analysis were included in the backward selection. Tukey-Kramer multiple adjustment was applied for pairwise comparisons among the levels of the significant variables. Beta weights for each predictor variable show strength of the effects on the ECLiPS Total Scaled score. Analysis used SAS version 9.4 (SAS Institute, Cary, N.C.). A two-sided significance level was set at 0.05.

RESULTS

Group differences in caregiver-reported listening skills were maintained at Wave 2

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

By design, all ECLiPS scores were lower in the LiD group than in the TD group at Wave 1 (Petley et al. 2021) (Figure 1A; Table 2). The LiD group continued to have lower ECLiPS sub-scores and total scores than the TD group at Wave 2 (group effect, $p < .0001$). The relationship between TD and LiD did not change across the two waves following B-H adjustment (wave effect for ECLiPS sub-scores, $p = 0.01$ - 0.67; Group x Wave, $p = 0.02 – 0.91$, B-H critical p value 0.0063).

LiD group had consistently lower cognitive scores

All NIH Cognition Toolbox sub-test scores and composite scores for the LiD group were significantly lower than for the TD group (group effect, $p < .0001$; Figure 1B; Table 3) across Waves 1 and 2. There was no significant Group x Wave interaction following B-H adjustment ($p = 0.01 – 0.84$, B-H critical p value 0.0071). Wave differences for the Flanker Test scores ($p = .008$) were trending but not significant after B-H adjustment.

LiD group had consistently lower skills in segregating sentences

All LiSN-S z-scores for the LiD group were lower than the TD group across waves (Figure 1C; Table 2; group effect, $p < .015$; Group x Wave, $p = 0.12 – 1.00$). Across groups, Spatial Advantage scores were significantly higher at Wave 2 than at Wave 1 ($p = .007$). Other LiSN-S z-scores did not differ between Waves 1 and 2. Spatial Pattern Scores did not differ significantly between groups or waves (group effect, $p = 0.41$; wave effect, $p = 0.11$; Group x Wave, $p = 0.86$). Thirteen children in Wave 1 (7 LiD, 6 TD, $p = 0.57$), but only three children in Wave 2 (LiD only, $p = 0.24$) had Pattern Scores meeting the diagnostic criteria for SPD.

Differences between the TD and LiD groups persisted through adolescence

We assessed longitudinal changes in listening skills, speech-in-speech skills, and cognitive function across the three waves based on only those 30 participants who completed all three waves (Figure 2). All ECLiPS sub-scores and total scores in Wave 3 remained higher for the TD than the LiD group (group effect, $p < .0001$), as in the main analysis of Waves 1 and 2 data. However, in contrast to the main analysis, ECLiPS SAP (wave effect $p = .004$; Group x Wave, $p = 0.18$) and composite Listening scores (wave effect $p = .003$; Group x Wave, $p = 0.93$) were significantly higher across both groups in Wave 2 than Wave 1. Similarly, ECLiPS Total scores in Wave 3 were significantly higher than Wave 1(wave effect, $p = 0.028$; Group x Wave, $p = 0.71$). All NIH Cognition Toolbox scores and two LiSN-S z scores (Low Cue and Talker Advantage) in the TD group were significantly higher than in the LiD group (group effect, $p < .003$). There was no significant main wave effect or Group x Wave interaction for either NIH Cognition Toolbox or LiSN-S scores. In Wave 2, missing values were random (Little's $p = 0.74$).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

However, for Wave 3, missing values were non-random, suggesting systemic reasons for participant attrition (Little's $p = 0.01$).

LiD group associated with lower maternal education

Of the demographic information assessed, only maternal education was significantly lower for the LiD group ($p < .0001$), with significant Group x Wave interaction (Wave 1 vs. Wave 3, $p < .0001$; Table 1). Adding maternal education to the mixed-effect models comparing listening, cognitive, and auditory skills between TD and LiD did not, however, change the relationships described above, affirming that maternal education was not confounding.

ECLiPS and NIH-TB scores demonstrated good reliability between Waves 1 and 2

Part 2 of the analysis showed ECLiPS total scaled scores had good reliability between Waves 1 and 2 (Figure 3; Table 4). Reliability did not differ between TD and LiD groups (TD: $r = 0.52$, LiD: $r = 0.52$, $p = 1.000$). All sub-test and composite NIH Toolbox scores were significantly correlated between Waves 1 and 2 (ICCs ranging from 0.48 to 0.82) and consistent between TD and LiD ($p = 0.07 – 0.96$). All LiSN-S scores were significantly correlated, with a moderate agreement between waves (ICCs ranging from 0.28 to 0.42).

Cognitive, sensory, and socioeconomic factors independently contributed to parent-reported listening skills

For Part 3 of the analysis, we first found the Wave 1 prediction model (Petley et al. 2021) accounted for 42% of the variance of ECLiPS Total Scaled scores at Wave 2. Our new model, using the longitudinal data of both Waves 1 and 2, explained 54.8% of the variance of the ECLiPS total scaled score ($\chi^2(2) = 48.21$, $p < 0.0001$). The new model included maternal education (beta 0.70, for Bachelor vs. High school maternal education groups, $p = .0038$), NIH total composite score (beta 0.49, $p < .0001$) and LiSN-S Spatial Advantage (beta 0.16, $p = .0023$). Consistent with the importance of maternal education, the ECLiPS mean Total Scaled scores were 3.1 points higher for children in the Bachelor maternal education group than for those in the High school group.

DISCUSSION

Impact of study

We found that children with LiD, classified using the ECLiPS caregiver questionnaire, had persistently impaired performance across multi-year testing on auditory and cognitive tasks relative to TD children. ECLiPS scores and NIH Toolbox testing showed strong test-retest reliability across two years.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Caregiver-reported listening skills were independently predicted by maternal education, LiSN-S Spatial Advantage scores, and NIH Toolbox total composite scores. To our knowledge, this study is the first to show an association between the degree of childhood LiD and maternal education, independent of auditory and cognitive skills, suggesting that socioeconomic factors may make essential contributions to this poorly understood problem. Our results carry clinical significance because the long-lasting nature of childhood LiD indicates an increased impact on affected individuals that may warrant earlier and more aggressive management.

Reliability

A prior longitudinal study of childhood LiD was based on a self-reported questionnaire with a follow-up response rate of approximately 13% (Del Zoppo et al. 2015). The current study used a large and heterogeneous sample of children and had a follow-up rate at Wave 2 of 68%. Additionally, loss to follow-up at Wave 2 was random, suggesting a low likelihood of sampling bias. Therefore, findings from Waves 1 and 2 likely represent the typical progression of LiD. Unfortunately, due to COVID, the follow-up rate of Wave 3 dropped significantly, and sampling bias was introduced. A separate longitudinal analysis of Wave 3 participants showed overall longer-term results that were consistent with the primary analysis of Waves 1 and 2.

All ECLiPS, LiSN-S, and NIH Toolbox scores showed a statistically significant correlation between Waves 1 and 2. These correlations remained significant in the sub-group analysis performed separately in the TD and LiD groups. The high reliability of the Toolbox scores found here, despite long testing intervals of 2 years, appears to allay other recent concerns about these scores (Taylor et al. 2020). This may be attributable to our consistent testing procedure, conducted by research staff at a single site, compared to testing involving multiple sites (Taylor et al. 2020). LiSN-S age-adjusted scores were also significantly correlated between Waves 1 and 2. However, the agreement between waves varied from poor to good, and correlation coefficients were overall lower than in a previous study (Cameron et al. 2011). The higher retest reliability for that previous study may be explained by its shorter testing intervals (2-4 months) and older (above 12 years old) participants (Cameron et al. 2011).

Persistent listening challenges

ECLiPS scores of TD children were consistent with a prior study (Barry et al. 2015), indicating that the significant difference between TD and LiD was due to persistent listening challenges in LiD rather than a supra-normal performance of the TD group. The protracted, poorer performance of the LiD group, also reported in other developmental learning disabilities, suggests that existing management strategies aren’t adequately addressing these children's challenges. In separate work (Hunter et al., in

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

preparation), we found more than half the children in the LiD group had been referred to CCH for specialist assessment and received interventions before and/or during the study period. Intervention efficacy needs to be evaluated separately in a future research study using quantitative outcome measures specific to each treatment. Secondary analysis of Wave 3 participants showed improved ECLiPS scores at Waves 2 and 3 compared to Wave 1 but without a Group x Wave interaction. This suggests that both groups' listening skills improved over time, relative to published norms (Barry and Moore 2021), which may be interpreted as accelerated development within this specific cohort.

Cognitive challenges

One of the striking findings of our study was a persistent cognitive challenge of the LiD group across all the domains tested by the NIH Toolbox. Our prior study showed an association between LiD and impaired cognition at Wave 1 (Petley et al. 2021). The findings of the current study were even more robust because we showed this association in the longitudinal data and across all the NIH Toolbox subscale and composite scores. Our prediction model identified NIH Total Composite score as the single most predictive cognitive measure over any combination of NIH Toolbox subscales, confirming the broad cognitive challenges associated with LiD. A recent review has suggested that an intervention strategy for treating LiD should be deficit-specific (Dillon and Cameron 2021). While attractive in principle, the results presented here suggest that such a strategy may need to employ an unrealistically large number of interventions. An alternate approach may be to identify and remediate LiD in infancy, as is currently under investigation for childhood language (Ferjan Ramirez et al. 2020) and autism spectrum disorders (Whitehouse et al. 2021).

Spatial hearing

In the main analysis of Wave 1 and 2 data, all LiSN-S age-adjusted scores were higher in the TD than in the LiD group. This shows that children with LiD have challenges in both the auditory and cognitive processing represented in the Cue and Advantage scores. In our cross-sectional study of Wave 1 data, TD children scored significantly higher only in the LiSN-S Low Cue and Talker Advantage scores (Petley et al., 2021). Interestingly, in the secondary analysis of Wave 3 participants in the current study, children with LiD scored lower in Low Cue and Talker Advantage only, similar to the prior study. Taken together, the data suggest that Low Cue and Talker Advantage may be more sensitive measures for differentiating TD and LiD than other LiSN-S scores.

Although the number of cases was small, the prevalence of SPD decreased at Wave 2, despite persistent listening, auditory, and cognitive challenges in the LiD group. Moreover, none of the participants received specific treatment for SPD (Cameron et al. 2012). These findings indicate that SPD

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

may be a developmentally transient condition. Furthermore, neither the Pattern score nor the frequency of SPD differed significantly between the TD and LiD groups, suggesting that SPD does not represent the listening problems identified by the ECLiPS (Petley et al. 2021). Future research clarifying the clinical significance and natural history of SPD would be beneficial. Among LiSN-S scores, only the Spatial Advantage age-adjusted scores increased significantly from Waves 1 to 2 in both groups. This could be due to developmental changes or inconsistent normalization and is unlikely to be training effects after a 2-year separation between tests.

Prediction model

Using a longitudinal design and a data-driven approach, we developed a model explaining the degree of listening difficulty with greater accuracy than in our prior study (Petley et al. 2021). In the new model, Talker Advantage provided less independent information about listening difficulty than Spatial Advantage. Similarly, maternal education independently contributed to the model that includes NIH Toolbox total composite and Spatial Advantage. Our final model seems plausible because it consists of auditory, cognitive, and social factors.

Socioeconomic factors in listening difficulties

To our knowledge, the current study is the first to show the independent contribution of socioeconomic factors to listening skills in children with LiD or APD. Prior studies showed maternal education is associated with early childhood development (Jeong et al. 2017), children's academic outcomes (Harding et al. 2015), language skills (Magnuson et al. 2009), attention problems (Hjern et al. 2010), and auditory brainstem responses (Skoe et al. 2013). It is, therefore, unsurprising that maternal education was associated with the degree of LiD, potentially through multiple pathways. However, the finding of our study is most noteworthy because the degree of LiD was associated with maternal education independent of cognitive and sensory functions. Maternal education and socioeconomic factors should be a part of the evaluation of children with LiD. A further study investigating the mechanisms connecting socioeconomic factors with LiD is warranted.

Limitations

There are several limitations to our study. Due to the small sample size of Wave 3, we ran a separate analysis focusing on longitudinal changes of Wave 3 participants to reduce sampling bias. Though this method reduced bias, the analysis remained underpowered. Comprehensive and conventional auditory processing testing, such as the SCAN-3, was lacking in Waves 2 and 3. Despite these limitations, our findings are significant because we showed long-lasting challenges of childhood LiD that may

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

influence clinical management strategy, and we identified an essential social factor independently contributing to LiD.

Conclusion

Children with LiD and clinically normal audiograms have persistent auditory, listening, and cognitive challenges through at least early adolescence. The degree of LiD can be independently predicted by maternal education, cognitive processing, and spatial listening skills.

ACKNOWLEDGEMENTS

This research was supported by grant 5R01DC014078 from the National Institute of Deafness and other Communication Disorders (DRM), 2UL1TR001425-05A1 by the National Center for Advancing Translational Sciences of the National Institutes of Health (KK), and by the Cincinnati Children's Research Foundation. DRM received support from NIHR Manchester Biomedical Research Centre.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. Participant numbers and demographic information

	Wave 1		Wave 2		Wave 3	
	TD	LiD	TD	LiD	TD	LiD
N	79	68	48	52	15	16
Age, year (SD)	9.3 (2.0)	9.7 (1.9)	12.1 (2.2)	11.9 (2.0)	13.4 (1.9)	14.5 (2.2)
Female, n (%)	33 (42)	24 (35)	24 (50)	16 (31)	6 (40)	6 (38)
Maternal education, n (%)						
High school	1 (1)	12 (18)	1 (2)	9 (18.4)	0 (0)	3 (19)
College	8 (10)	17 (25)	3 (6)	15 (30.6)	0 (0)	6 (38)
Bachelor	50 (63)	25 (37)	31 (65)	16 (32.7)	11 (73)	3 (19)
Post-grad	20 (25)	14 (21)	13 (27)	9 (18.4)	4 (27)	4 (25)
Race, non-white (%)	13 (16)	17 (25)	5 (10)	16 (31)	1 (7)	9 (56)
Ethnicity, Hispanic or Latino (%)	3 (4)	4 (6)	2 (4)	3 (6)	1 (7)	2 (12)

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2. Auditory and listening test scores

	Wave 1		Wave 2		Wave 3	
Testing measures*	TD	LiD	TD	LiD	TD	LiD
	n = 79	n = 67	n = 47	n = 49	n = 13	n = 15
ECLiPS scores						
Speech and auditory processing	10.4 (1.7)	3.0 (1.7)	10.5 (2.5)	4.2 (2.6)	11.4 (2.2)	6.0 (3.3)
Environmental and auditory sensitivity	10.8 (2.5)	4.8 (2.6)	9.8 (2.7)	4.9 (2.7)	10.6 (3.1)	7.9 (3.2)
Language/literacy/laterality	10.5 (2.8)	3.9 (2.4)	10.2 (2.6)	3.9 (2.4)	10.9 (2.2)	4.9 (2.8)
Memory and attention	10.7 (2.5)	4.2 (2.0)	10.8 (3.0)	4.8 (2.7)	12.2 (3.5)	6.5 (3.4)
Pragmatic and social skills	10.9 (2.4)	5.1 (1.8)	10.9 (2.4)	5.7 (2.2)	12.1 (3.1)	7.9 (3.4)
Language	10.6 (2.4)	3.6 (2.1)	10.6 (2.9)	3.8 (2.4)	11.8 (3.2)	5.3 (2.9)
Listening	10.7 (2.4)	3.4 (1.7)	10.8 (2.6)	4.2 (2.3)	12.1 (3.1)	6.3 (3.4)
Social	11.0 (2.4)	3.6 (2.1)	10.6 (2.9)	3.8 (2.4)	11.6 (3.4)	7.7 (3.6)
Total	10.8 (2.4)	3.0 (1.8)	10.5 (2.7)	3.5 (2.3)	11.8 (3.1)	5.9 (3.2)
LiSN-S scores	n = 76	n = 65	n = 43	n = 47	n = 15	n = 15
Low Cue, z score	-0.13 (1.00)	-0.76 (1.18)	0.16 (1.16)	-0.96 (1.10)	0.33 (1.32)	-0.87 (1.08)
High Cue, z score	0.28 (1.04)	-0.39 (1.32)	0.43 (0.81)	-0.31 (1.26)	0.68 (0.91)	0.25 (1.10)
Talker Advantage, z score	-0.05 (0.81)	-0.54 (0.99)	0.20 (0.95)	-0.34 (0.96)	0.55 (1.05)	-0.55 (1.01)
Spatial Advantage, z score	-0.23 (1.36)	-0.66 (1.69)	0.24 (1.20)	-0.29 (1.52)	0.60 (1.07)	-0.17 (0.88)
Pattern Score	7.43 (1.83)	7.25 (2.14)	7.79 (1.13)	7.64 (2.09)	8.28 (1.46)	8.76 (1.23)

*All values are mean (SD) unless otherwise noted.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3. Cognitive test scores

	Wave 1		Wave 2		Wave 3	
Testing measures*	TD	LiD	TD	LiD	TD	LiD
NIH Cognition Toolbox Composite Scores	n = 56	n = 49	n = 42	n = 45	n = 15	n = 15
Picture vocabulary	114 (14)	98 (13)	112 (14)	93 (11)	113 (15)	91(16)
Flanker test	101(15)	90 (14)	98 (12)	86 (14)	107 (15)	80 (12)
List sorting working memory†	110 (15)	94 (14)	110 (14)	92 (13)	117 (15)	89 (12)
Dimensional change card sort test	104 (15)	89 (13)	108 (19)	90 (15)	113 (21)	89 (17)
Pattern comparison processing speed test†	103 (22)	87 (20)	115 (22)	86 (22)	123 (19)	90 (22)
Picture sequence memory test	112 (19)	95 (20)	113 (18)	99 (13)	128 (17)	95 (14)
Oral reading recognition test†	107 (13)	90 (12)	110 (12)	92 (11)	106 (14)	85 (11)
Fluid cognition composite†	109 (18)	85 (17)	114 (17)	84 (17)	126 (14)	81 (18)
Crystalized cognition composite†	112 (14)	93 (12)	113 (13)	91 (10)	111 (15)	86 (14)
Total composite†	112 (15)	87 (14)	116 (15)	85 (13)	123 (15)	80 (15)
Early Childhood composite	112 (16)	89 (14)	112 (16)	87 (13)	122 (16)	82 (18)

*All values are mean (SD) unless otherwise noted.

†Data missing for 9 children in Wave 1 (TD: n = 51, LiD: n = 45).

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 4. Pearson correlation coefficients and estimated reliability coefficients between waves 1 and 2

	Pearson correlation coefficient	ICC (95%CI), Unadjusted	ICC (95%CI), Adjusted by Wave
ECLIPS total scaled score	0.85	0.86 (0.80, 0.90)	0.86 (0.80, 0.90)
LiSN-S, z score			
Low Cue	0.29	0.27 (0.13-0.49)	0.28 (0.13-0.49)
High Cue	0.35	0.32 (0.17-0.52)	0.33 (0.18-0.53)
Talker advantage	0.32	0.29 (0.14-0.51)	0.33 (0.17-0.54)
Spatial Advantage	0.45	0.39 (0.24-0.56)	0.42 (0.27-0.58)
NIH-toolbox age-corrected standard score			
Picture vocabulary	0.66	0.65 (0.51-0.77)	0.65 (0.52-0.77)
Flanker test	0.51	0.42 (0.26-0.60)	0.48 (0.32-0.64)
List sorting working memory	0.66	0.68 (0.53-0.79)	0.68 (0.55-0.80)
Dimensional change card sort test	0.48	0.53 (0.35-0.69)	0.52 (0.35-0.69)
Pattern comparison processing speed test	0.58	0.56 (0.40-0.71)	0.56 (0.40-0.70)
Picture sequence memory test	0.44	0.45 (0.27-0.64)	0.45 (0.27-0.64)
Oral reading recognition test	0.70	0.68 (0.54-0.79)	0.70 (0.57-0.80)
Fluid cognition composite	0.74	0.75 (0.64-0.84)	0.76 (0.64-0.84)
Crystallized cognition composite	0.80	0.79 (0.69-0.86)	0.79 (0.69-0.86)
Total composite	0.82	0.82 (0.73-0.88)	0.82 (0.74-0.88)
Early childhood composite	0.73	0.73 (0.61-0.82)	0.74 (0.63-0.83)

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

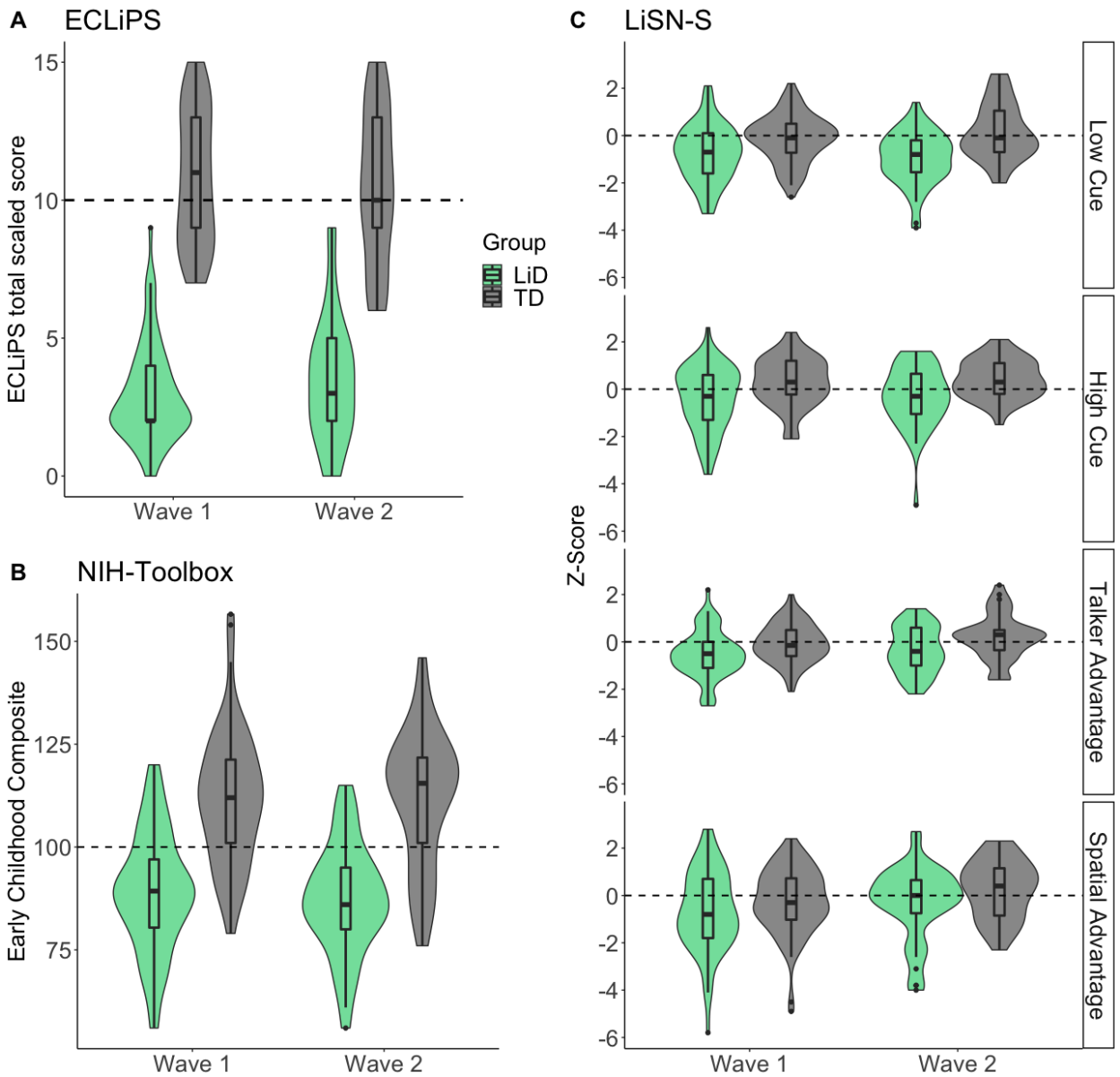

Figure 1. All test scores were consistently lower in the LiD group across Waves 1 and 2. A. Violin plots of ECLiPS total scaled scores show the probability density of the data. Violin plots are overlaid with boxplots indicating each group's median and interquartile range in Waves 1 and 2. The horizontal dashed line reflects the expected standard score (here, 10). Wave 3 data are not included in the figure due to the significantly smaller sample compared with Waves 1 and 2 (see Tables 1-3). **B.** Early Childhood Composite scores from the NIH Cognition Toolbox. **C.** LiSN-S Cue and Advantage scaled Z-Scores. LiD indicates listening difficulty; TD, typically developing.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

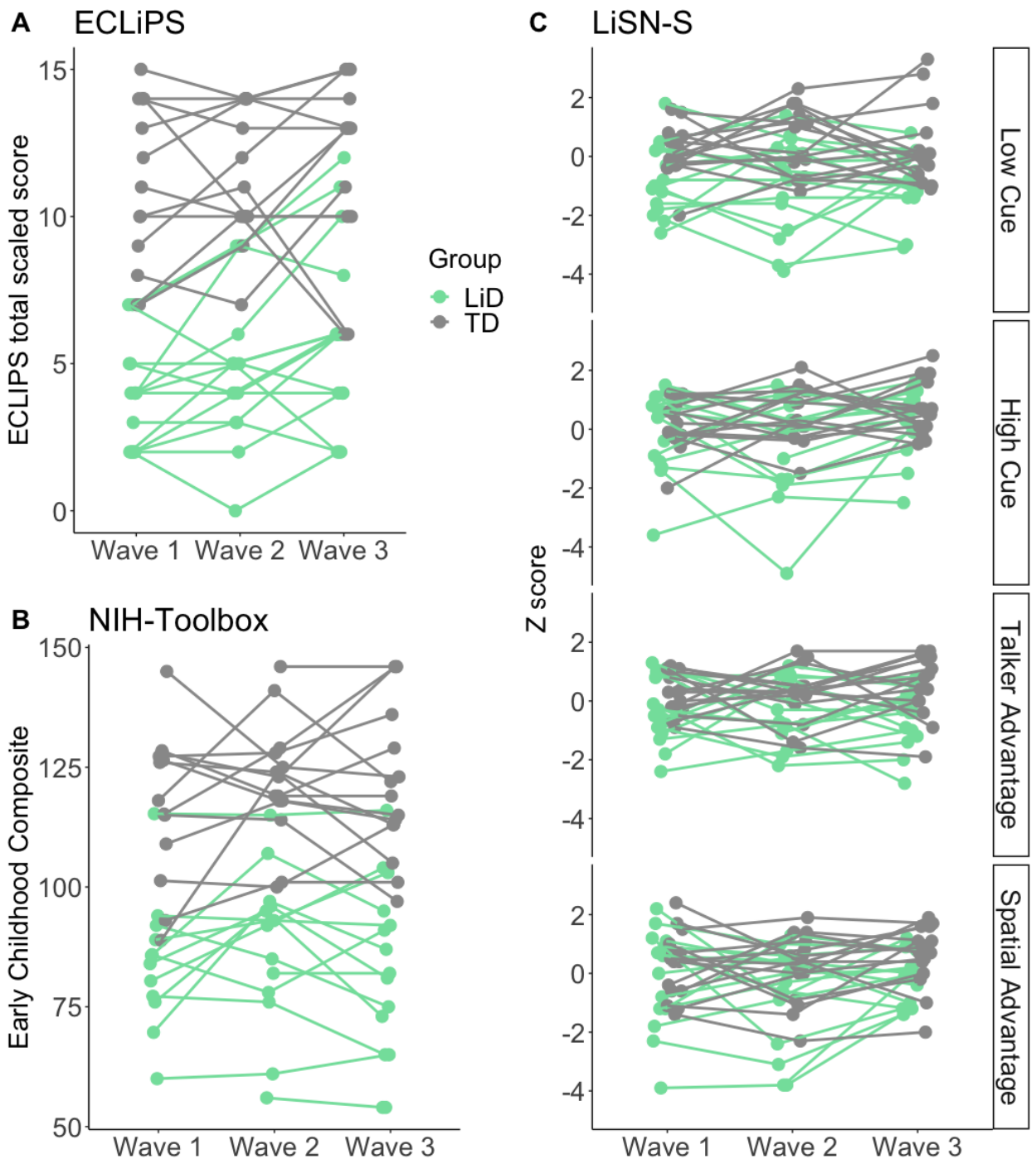

Figure 2. Individual data of Wave 3 participants showed improved listening skills over 3 waves. A. Line plots of ECLiPS total scaled scores across 4 years. **B.** Same as A for NIH-Toolbox Early Childhood Composite scaled scores. **C.** Same as A for LiSN-S Cue and Advantage scores. LiD indicates listening difficulty; TD, typically developing.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

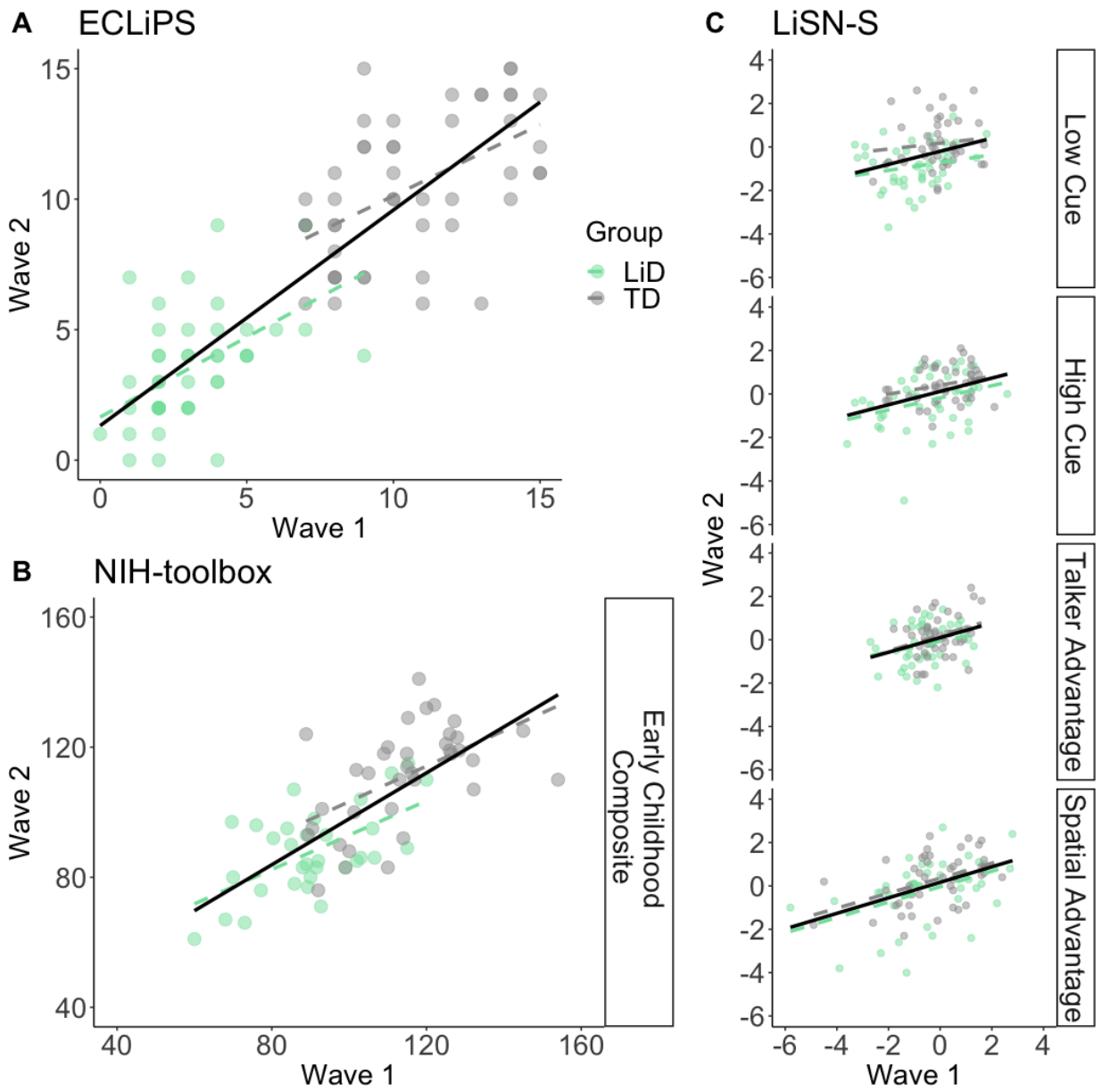

Figure 3. Listening skills and cognitive abilities showed moderate to excellent correlations between Waves 1 and 2. Auditory skills showed poor agreement between waves. **A.** Scatter plots of ECLiPS total scaled scores at Wave 1 and 2. Dashed lines show regression lines for each group. The black line indicates the regression line for the entire data combining groups. **B.** Same as A for Early Childhood Composite. **C.** Same as A for LiSN-S Cue and Advantage scores. LiD indicates listening difficulty; TD, typically developing.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

Azouz, Hanan Galal, Mona Khalil, Rania M Abdou, and Mohamed Sakr. 2014. 'The correlation between central auditory processing in autistic children and their language processing abilities', *International Journal of Pediatric Otorhinolaryngology*, 78: 2297-300.

Barry, J. G., and D. R. Moore. 2021. 'ECLIPS: Evaluation of children's listening and processing skills (2nd ed.)', *Cincinnati Children's Hospital*.

Barry, J. G., D. Tomlin, D. R. Moore, and H. Dillon. 2015. 'Use of Questionnaire-Based Measures in the Assessment of Listening Difficulties in School-Aged Children', *Ear Hear*, 36: e300-13.

Bench, J., A. Kowal, and J. Bamford. 1979. 'The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children', *Br J Audiol*, 13: 108-12.

Benjamini, Yoav, and Yosef Hochberg. 1995. 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal statistical society: series B (Methodological)*, 57: 289-300.

Bishop, Dorothy VM. 2006. *CCC-2: Children's Communication Checklist-2* (Pearson).

Boothalingam, S., D. W. Purcell, C. Allan, P. Allen, and E. Macpherson. 2019. 'Auditory Localization and Spatial Release From Masking in Children With Suspected Auditory Processing Disorder', *Ear Hear*, 40: 1187-96.

Brewer, C. C., C. K. Zalewski, K. A. King, O. Zobay, A. Riley, M. A. Ferguson, J. E. Bird, M. M. McCabe, L. J. Hood, D. Drayna, A. J. Griffith, R. J. Morell, T. B. Friedman, and D. R. Moore. 2016. 'Heritability of non-speech auditory processing skills', *Eur J Hum Genet*, 24: 1137-44.

Brown, D. K., S. Cameron, J. S. Martin, C. Watson, and H. Dillon. 2010. 'The North American Listening in Spatialized Noise-Sentences test (NA LiSN-S): normative data and test-retest reliability studies for adolescents and young adults', *J Am Acad Audiol*, 21: 629-41.

Bulut, Berker, and Nuri Karasakaloglu. 2017. 'Benefiting from Listening in Vocabulary Development', *Journal of Education and Training Studies*, 5: 99-109.

Cameron, S., and H. Dillon. 2007. 'Development of the Listening in Spatialized Noise-Sentences Test (LISN-S)', *Ear Hear*, 28: 196-211.

Cameron, S., and H. Dillon. 2011. 'Development and evaluation of the LiSN & learn auditory training software for deficit-specific remediation of binaural processing deficits in children: preliminary findings', *J Am Acad Audiol*, 22: 678-96.

Cameron, S., H. Dillon, H. Glyde, S. Kanthan, and A. Kania. 2014. 'Prevalence and remediation of spatial processing disorder (SPD) in Indigenous children in regional Australia', *Int J Audiol*, 53: 326-35.

Cameron, S., H. Glyde, and H. Dillon. 2011. 'Listening in Spatialized Noise-Sentences Test (LiSN-S): normative and retest reliability data for adolescents and adults up to 60 years of age', *J Am Acad Audiol*, 22: 697-709.

Cameron, S., H. Glyde, and H. Dillon. 2012. 'Efficacy of the LiSN & Learn auditory training software: randomized blinded controlled study', *Audiol Res*, 2: e15.

Carhart, Raymond, and James F Jerger. 1959. 'Preferred method for clinical determination of pure-tone thresholds', *Journal of speech and hearing disorders*, 24: 330-45.

Dawes, P., and D. V. Bishop. 2010. 'Psychometric profile of children with auditory processing disorder and children with dyslexia', *Arch Dis Child*, 95: 432-6.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Del Zoppo, C., L. Sanchez, and C. Lind. 2015. 'A long-term follow-up of children and adolescents referred for assessment of auditory processing disorder', *Int J Audiol*, 54: 368-75.

Dillon, H., and S. Cameron. 2021. 'Separating the Causes of Listening Difficulties in Children', *Ear Hear*, 42: 1097-108.

Farah, R., V. J. Schmithorst, R. W. Keith, and S. K. Holland. 2014. 'Altered white matter microstructure underlies listening difficulties in children suspected of auditory processing disorders: a DTI study', *Brain Behav*, 4: 531-43.

Ferguson, M. A., R. L. Hall, A. Riley, and D. R. Moore. 2011. 'Communication, listening, cognitive and speech perception skills in children with auditory processing disorder (APD) or Specific Language Impairment (SLI)', *J Speech Lang Hear Res*, 54: 211-27.

Ferjan Ramirez, N., S. R. Lytle, and P. K. Kuhl. 2020. 'Parent coaching increases conversational turns and advances infant language development', *Proc Natl Acad Sci U S A*, 117: 3484-91.

Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied longitudinal analysis* (John Wiley & Sons).

Fleta, Teresa. 2014. 'Active listening for second language learning in the early years.' in, *Early years second language education* (Routledge).

Harding, Jessica F, Pamela A Morris, and Diane Hughes. 2015. 'The relationship between maternal education and children's academic outcomes: A theoretical framework', *Journal of Marriage and Family*, 77: 60-76.

Hechtman, L., J. M. Swanson, M. H. Sibley, A. Stehli, E. B. Owens, J. T. Mitchell, L. E. Arnold, B. S. Molina, S. P. Hinshaw, P. S. Jensen, H. B. Abikoff, G. Perez Algorta, A. L. Howard, B. Hoza, J. Etcovitch, S. Houssais, K. D. Lakes, J. Q. Nichols, and M. T. A. Cooperative Group. 2016. 'Functional Adult Outcomes 16 Years After Childhood Diagnosis of Attention-Deficit/Hyperactivity Disorder: MTA Results', *J Am Acad Child Adolesc Psychiatry*, 55: 945-52 e2.

Hind, S. E., R. Haines-Bazrafshan, C. L. Benton, W. Brassington, B. Towle, and D. R. Moore. 2011. 'Prevalence of clinical referrals having hearing thresholds within normal limits', *Int J Audiol*, 50: 708-16.

Hjern, A., G. R. Weitoft, and F. Lindblad. 2010. 'Social adversity predicts ADHD-medication in school children--a national cohort study', *Acta Paediatr*, 99: 920-4.

Humanski, R. A., and R. A. Butler. 1988. 'The contribution of the near and far ear toward localization of sound in the sagittal plane', *J Acoust Soc Am*, 83: 2300-10.

Hunter, L. L., C. M. Blankenship, L. Lin, N. T. Sloat, A. Perdew, H. Stewart, and D. R. Moore. 2021. 'Peripheral Auditory Involvement in Childhood Listening Difficulty', *Ear Hear*, 42: 29-41.

Jeong, Joshua, Dana Charles McCoy, and Günther Fink. 2017. 'Pathways between paternal and maternal education, caregivers' support for learning, and early child development in 44 low-and middle-income countries', *Early Childhood Research Quarterly*, 41: 136-48.

Johnston, Kristin N, Andrew B John, Nicole V Kreisman, James W Hall III, Carl C Crandell, Kristin N Johnston, Andrew B John, Nicole V Kreisman, James W Hall III, and Carl C Crandell. 2009. 'Multiple benefits of personal FM system use by children with auditory processing disorder (APD)', *International Journal of Audiology*, 48: 371-83.

Keith, William J., and Suzanne C. Purdy. 2014. 'Assistive and Therapeutic Effects of Amplification for Auditory Processing Disorder', *Semin Hear*, 35: 027-38.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Little, Roderick JA. 1988. 'A test of missing completely at random for multivariate data with missing values', *Journal of the American statistical Association*, 83: 1198-202.

Lohvansuu, K., M. Torppa, T. Ahonen, K. Eklund, J. A. Hamalainen, P. H. T. Leppanen, and H. Lyytinen. 2021. 'Unveiling the Mysteries of Dyslexia-Lessons Learned from the Prospective Jyvaskyla Longitudinal Study of Dyslexia', *Brain Sci*, 11.

Loo, J. H., S. Rosen, and D. E. Bamiou. 2016. 'Auditory Training Effects on the Listening Skills of Children With Auditory Processing Disorder', *Ear Hear*, 37: 38-47.

Magnuson, Katherine A, Holly R Sexton, Pamela E Davis-Kean, and Aletha C Huston. 2009. 'Increases in maternal education and young children's language skills', *Merrill-Palmer Quarterly (1982-)*: 319-50.

Moore, D. R., K. Hugdahl, H. J. Stewart, J. Vannest, A. J. Perdew, N. T. Sloat, E. K. Cash, and L. L. Hunter. 2020. 'Listening Difficulties in Children: Behavior and Brain Activation Produced by Dichotic Listening of CV Syllables', *Front Psychol*, 11: 675.

Moore, D. R., S. L. Sieswerda, M. M. Grainger, A. Bowling, N. Smith, A. Perdew, S. Eichert, S. Alston, L. W. Hilbert, L. Summers, L. Lin, and L. L. Hunter. 2018. 'Referral and Diagnosis of Developmental Auditory Processing Disorder in a Large, United States Hospital-Based Audiology Service', *J Am Acad Audiol*, 29: 364-77.

Parthasarathy, A., K. E. Hancock, K. Bennett, V. DeGruttola, and D. B. Polley. 2020. 'Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing', *Elife*, 9.

Perrachione, T. K., S. N. Del Tufo, and J. D. Gabrieli. 2011. 'Human voice recognition depends on language ability', *Science*, 333: 595.

Petley, L., L. L. Hunter, L. Motlagh Zadeh, H. J. Stewart, N. T. Sloat, A. Perdew, L. Lin, and D. R. Moore. 2021. 'Listening Difficulties in Children With Normal Audiograms: Relation to Hearing and Cognition', *Ear Hear*.

Pienkowski, M. 2017. 'On the Etiology of Listening Difficulties in Noise Despite Clinically Normal Audiograms', *Ear Hear*, 38: 135-48.

Quinto, A., S. Abu El Adas, and S. V. Levi. 2020. 'Re-Examining the Effect of Top-Down Linguistic Information on Speaker-Voice Discrimination', *Cogn Sci*, 44: e12902.

Rosenberg, Gail Gegg. 2002. "Classroom acoustics and personal FM technology in management of auditory processing disorder." In *Seminars in hearing*, 309-18. Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New ….

Roy, A., L. Hechtman, L. E. Arnold, J. M. Swanson, B. S. G. Molina, M. H. Sibley, A. L. Howard, and M. T. A. Cooperative Group. 2017. 'Childhood Predictors of Adult Functional Outcomes in the Multimodal Treatment Study of Attention-Deficit/Hyperactivity Disorder (MTA)', *J Am Acad Child Adolesc Psychiatry*, 56: 687-95 e7.

Rutter, Michael, A Bailey, and Catherine Lord. 2003. 'SCQ', *The Social Communication Questionnaire. Torrance, CA: Western Psychological Services*.

Schafer, Erin C, Jordan Traber, Paige Layden, Aneeqa Amin, Katie Sanders, Danielle Bryant, and Nicole Baldus. 2014. "Use of wireless technology for children with auditory processing disorders, attention-deficit hyperactivity disorder, and language disorders." In *Seminars in Hearing*, 193-205. Thieme Medical Publishers.

Schmucker, C., P. Kapp, E. Motschall, J. Loehler, and J. J. Meerpohl. 2019. 'Prevalence of hearing loss and use of hearing aids among children and adolescents in Germany: a systematic review', *BMC public health*, 19: 1277.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Sharma, M., I. Dhamani, J. Leung, and S. Carlile. 2014. 'Attention, memory, and auditory processing in 10- to 15-year-old children with listening difficulties', *J Speech Lang Hear Res*, 57: 2308-21.

Shaywitz, Sally E., Jack M. Fletcher, John M. Holahan, Abigail E. Shneider, Karen E. Marchione, Karla K. Stuebing, David J. Francis, Kenneth R. Pugh, and Bennett A. Shaywitz. 1999. 'Persistence of Dyslexia: The Connecticut Longitudinal Study at Adolescence', *Pediatrics*, 104: 1351-59.

Sherar, L. B., T. P. Griffin, U. Ekelund, A. R. Cooper, D. W. Esliger, E. M. van Sluijs, L. Bo Andersen, G. Cardon, R. Davey, K. Froberg, P. C. Hallal, K. F. Janz, K. Kordas, S. Kriemler, R. R. Pate, J. J. Puder, L. B. Sardinha, A. F. Timperio, and A. S. Page. 2016. 'Association between maternal education and objectively measured physical activity and sedentary time in adolescents', *J Epidemiol Community Health*, 70: 541-8.

Skoe, E., J. Krizman, and N. Kraus. 2013. 'The impoverished brain: disparities in maternal education affect the neural response to sound', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33: 17221-31.

Smoski, W, Michael A Brunt, and J Curtis Tannahill. 1998. 'Children's auditory performance scale (CHAPS)', *Tampa, FL: Educational Audiology Association*.

St Clair, M. C., A. Pickles, K. Durkin, and G. Conti-Ramsden. 2011. 'A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI)', *J Commun Disord*, 44: 186-99.

Taylor, B. K., M. R. Frenzel, J. A. Eastman, A. I. Wiesman, Y. P. Wang, V. D. Calhoun, J. M. Stephen, and T. W. Wilson. 2020. 'Reliability of the NIH toolbox cognitive battery in children and adolescents: a 3-year longitudinal examination', *Psychol Med*: 1-10.

Tomblin, J. B., M. Harrison, S. E. Ambrose, E. A. Walker, J. J. Oleson, and M. P. Moeller. 2015. 'Language Outcomes in Young Children with Mild to Severe Hearing Loss', *Ear Hear*, 36 Suppl 1: 76S-91S.

Verbeke, Geert. 1997. 'Linear mixed models for longitudinal data.' in, *Linear mixed models in practice* (Springer).

Weihing, Jeffrey, Gail D Chermak, and Frank E Musiek. 2015. "Auditory training for central auditory processing disorder." In *Seminars in hearing*, 199-215. Thieme Medical Publishers.

Weintraub, S., P. J. Bauer, P. D. Zelazo, K. Wallner-Allen, S. S. Dikmen, R. K. Heaton, D. S. Tulsky, J. Slotkin, D. L. Blitz, N. E. Carlozzi, R. J. Havlik, J. L. Beaumont, D. Mungas, J. J. Manly, B. G. Borosh, C. J. Nowinski, and R. C. Gershon. 2013a. 'I. NIH Toolbox Cognition Battery (CB): introduction and pediatric data', *Monogr Soc Res Child Dev*, 78: 1-15.

Weintraub, S., S. S. Dikmen, R. K. Heaton, D. S. Tulsky, P. D. Zelazo, P. J. Bauer, N. E. Carlozzi, J. Slotkin, D. Blitz, K. Wallner-Allen, N. A. Fox, J. L. Beaumont, D. Mungas, C. J. Nowinski, J. Richler, J. A. Deocampo, J. E. Anderson, J. J. Manly, B. Borosh, R. Havlik, K. Conway, E. Edwards, L. Freund, J. W. King, C. Moy, E. Witt, and R. C. Gershon. 2013b. 'Cognition assessment using the NIH Toolbox', *Neurology*, 80: S54-64.

Whitehouse, A. J. O., K. J. Varcin, S. Pillar, W. Billingham, G. A. Alvares, J. Barbaro, C. A. Bent, D. Blenkley, M. Boutrus, A. Chee, L. Chetcuti, A. Clark, E. Davidson, S. Dimov, C. Dissanayake, J. Doyle, M. Grant, C. C. Green, M. Harrap, T. Iacono, L. Matys, M. Maybery, D. F. Pope, M. Renton, C. Rowbottam, N. Sadka, L. Segal, V. Slonims, J. Smith, C. Taylor, S. Wakeling, M. W. Wan, J. Wray, M. N. Cooper, J. Green, and K.

medRxiv preprint doi: https://doi.org/10.1101/2022.08.11.22278673; this version posted August 12, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Hudry. 2021. 'Effect of Preemptive Intervention on Developmental Outcomes Among Infants Showing Early Signs of Autism: A Randomized Clinical Trial of Outcomes to Diagnosis', *JAMA Pediatr*, 175: e213298.