Research Article

Using Propensity Score Matching to Address Clinical Questions: The Impact of Remote Microphone Systems on Language Outcomes in Children Who Are Hard of Hearing

Maura Curran,[a] Elizabeth A. Walker,[b] Patricia Roush,[c] and Meredith Spratford[d]

Purpose: Children who are hard of hearing (CHH) have restricted access to acoustic and linguistic information. Increased audibility provided by hearing aids influences language outcomes, but the benefits of hearing aids are often limited by acoustic factors and distance. Remote microphone (RM) systems further increase auditory access by reducing the negative consequences of these factors. The purpose of this article was to identify factors that influence likelihood of RM system receipt and to investigate the effects of RM systems in home settings on later language outcomes. We used propensity score matching to compare language outcomes between children with and without access to personal RM systems in home settings. This article provides a description of how and why to perform propensity score–matching analyses with clinical populations.
Method: Participants were 132 CHH. Through parent report, we identified children who received RM systems for home use by 4 years of age. Logistic regression was used to

determine factors that predict likelihood of RM system receipt in home settings. Propensity score matching was conducted on a subgroup of 104 participants. Performance on language measures at age 5 years was compared across propensity-matched children who did and did not receive RMs for personal use.
Results: Likelihood of RM receipt was associated with degree of hearing loss, maternal education, and location (recruitment site). Comparisons between matched pairs of children with and without RM systems in early childhood indicated significantly better discourse skills for children whose families owned RM systems, but no significant differences for vocabulary or morphosyntax.
Conclusion: Results provide preliminary evidence that the provision of personal RM systems for preschool-age CHH enhances higher-level language skills. The propensity score–matching technique enabled us to use an observational, longitudinal data set to examine a question of clinical interest.

C hildren who are hard of hearing (CHH) acquire spoken language while experiencing limited access to auditory–linguistic input. Hearing aids (HAs) are expected to provide increased access to acoustic–phonetic

components of spoken language. Thus, for individuals with mild-to-severe hearing loss, HAs are a common form of audiologic intervention. Unfortunately, HAs are limited in their capacity to increase speech audibility in complex listening situations, such as in noisy and reverberant environments. Increased distance from the speaker may also limit auditory access. An alternative type of hearing assistive technology is remote microphone (RM) systems, which increase the intensity of a signal relative to the background noise, due to the proximity of the microphone to the speaker's mouth.[1] RM systems increase the signal-to-noise ratio, reducing the

[a]Department of Communication Sciences and Disorders, University of Delaware, Newark
[b]Department of Communication Sciences and Disorders, University of Iowa, Iowa City
[c]Department of Otolaryngology–Head and Neck Surgery, University of North Carolina at Chapel Hill
[d]Audibility, Perception, and Cognition Laboratory, Boys Town National Research Hospital, Omaha, NE

Correspondence to Elizabeth A. Walker: elizabeth-walker@uiowa.edu

---

[1]Additional forms of wireless connectivity include infrared and Bluetooth technology (Atcherson, Childress, & Kennett, 2016). For the purposes of this article, we will focus on the use of RM systems that utilize frequency modulation or digital modulation to transmit the signal to a listener.

Disclosure: The authors have declared that no competing interests existed at the time of publication.

negative effects of noisy, reverberant room acoustics or distance between speakers and listeners (Anderson & Goldstein, 2004; Benoit, 1989; Crandell, 1993; Davies, Yellon, & Purdy, 2001; Flexer, 1997; Moeller, Donaghy, Beauchaine, Lewis, & Stelmachowicz, 1996; Schafer & Thibodeau, 2006). Favorable signal-to-noise ratios are critical for CHH to understand speech in complex listening situations (Finitzo-Hieber & Tillman, 1978; Nábělek & Pickett, 1974), and RM use in elementary and secondary schools improves speech perception for CHH in those settings (Hawkins, 1984; Larsen & Blair, 2008). However, there is a lack of evidence regarding the effects of home RM system access on language outcomes.

The goal of the current work was to compare language outcomes of CHH with and without access to personal RM systems in home settings. To analyze the outcomes of children in treatment and no-treatment conditions, we use a statistical approach known as propensity score matching. This approach is a useful technique for reducing selection bias in nonrandomized observational studies, as it attempts to mimic randomization by controlling for factors that predict the likelihood of group membership in an intervention (C. J. Morgan, 2017; P. L. Morgan, Frisco, Farkas, & Hibel, 2010). Thus, a secondary goal is to provide a tutorial on how to use propensity score matching to conduct rigorous and relatively unbiased comparisons between individuals who receive treatment versus those who do not. This statistical procedure can help inform researchers and clinicians about the effects of a specific intervention on functional outcomes.

Propensity Score Matching to Support Evidence-Based Practice

Clinically, there is a dearth of outcomes-based evidence to support families and audiologists when deciding whether to get an RM system in noneducational settings. Decisions regarding treatment options should be motivated based on the findings of high-quality research designs (Wake & Carew, 2016), such as randomized controlled trials (RCTs). RCTs are considered the gold standard approach for examining the effects of specific treatments on outcomes (Austin, 2011). Participants are assigned to random treatment groups, which theoretically eliminates potential confounds associated with baseline characteristics of the participants. As a result, researchers can compare the results in treated and untreated groups without the concern that there are systematic differences between the groups prior to the onset of intervention. The challenge with conducting RCTs in some clinical settings is that it is difficult and possibly unethical to withhold treatment from children who may potentially benefit from the intervention.

Propensity score matching is an alternative approach to RCTs (Austin, 2011; Shadish, Clark, & Steiner, 2008). Propensity scores can be used in research designs when it is not feasible to randomly assign participants to different interventions (e.g., examining how foster care impacts outcomes; Berzin, 2010) or withhold an intervention (e.g., looking at

the effects of breastfeeding on cognitive outcomes; Jiang, Foster, & Gibson-Davis, 2011). In propensity score matching, the first step is to determine the probability of receiving a treatment (in the current context, RM receipt) based on specific participant characteristics. Propensity scores then match members of the control group (children who did not receive RM systems) and members of the treatment group (children who did receive RM systems) on the background variables that are associated with the likelihood that an individual will receive an RM system and the variables that are associated with outcomes. Thus, the children who received personal RM systems for use in home settings can be compared on the outcome measures of interest (i.e., language scores) to a group of children who did not receive RM systems but are similar in their likelihood of RM receipt. In effect, the propensity score–matching approach mimics an RCT analysis by directly comparing outcomes for a treatment group and a control group that do not differ on relevant pretreatment factors. Propensity score matching has rarely been used to evaluate treatment outcomes in the communication sciences and disorders literature. The need for longitudinal data from a large number of children may be one reason propensity score matching has not been widely used in the field, despite the potential for this method to provide insight into the effect of treatment in typical clinical practice settings.

Lack of Evidence for Benefits of Personal RMs in Preschool-Age CHH

In recent years, service providers have recommended personal RM systems for infants, toddlers, and preschoolers in the home setting to overcome some of the limitations of HAs. Gabbard (2004) recommended that families use RM systems in the home once children begin walking or crawling. Increased distance between the child and caregiver decreases access to the speech signal, which could lead to fewer opportunities to be exposed to and engage in conversational interactions. However, there is limited information regarding the benefits of personal RM use in the home for young children. Moeller et al. (1996) examined the efficacy of RM system use in home settings in 10 preschool-age children with mild-to-severe hearing loss (2–6 years of age). They were divided into two groups of children: (a) an RM group, which was composed of children who were instructed to use a personal RM system in lieu of their personal bilateral HAs as often as possible ($n = 6$), and (b) an HA group, which is composed of children who wore bilateral HAs without RM systems ($n = 4$). Examiners assessed grammatical skills using conversational language samples at a baseline test visit, followed by assessments at 6-month intervals over a 2-year time span. Results indicated that there were no significant differences between the RM group and the HA group on grammatical measures obtained from the language samples. Given the small number of participants, the authors noted it was difficult to achieve between-groups differences that were statistically

significant. At an individual level, Moeller et al. observed that two of the six RM users made better-than-expected gains in complex grammar usage over time (gains of 49 and 51 months of change in a 24-month period). These gains were more substantial than any of the gains shown by children in the HA group, and Moeller and her colleagues concluded that RM systems may have contributed to steep growth in grammatical learning for these particular subjects.

One limitation of the Moeller et al. (1996) study was that they only examined morphosyntax (e.g., grammatical complexity, mean sentence length). Amplification may have differential effects on various aspects of language, such as form (morphosyntax), content (vocabulary), or use (discourse). Our research team (Outcomes of Children With Hearing Loss [OCHL] study) has shown that content skills in preschool-age CHH may be less vulnerable to delay compared with form and use, which rely more heavily on a clear signal to access subtle acoustic characteristics in language input and conversational interactions (Moeller & Tomblin, 2015; Moeller, Tomblin, & OCHL Collaboration, 2015). We have published evidence to support this "inconsistent access hypothesis" with regard to the effects of HAs on language acquisition (Tomblin, Harrison, et al., 2015; Walker, Holte, et al., 2015). To the best of our knowledge, no researchers have compared the potential relation between RM systems in combination with HAs and different language constructs. This article addresses this gap in the literature.

The article by Moeller and colleagues is the only peer-reviewed publication, to date, to conduct between-group comparisons of language outcomes of children who used RM systems in nonacademic settings to children who only wore HAs. It is difficult to generalize these findings from the 1990s to the current generation of CHH, as hearing assistive technology is vastly different now compared with 20 years ago. The population of children is different as well because of earlier age of identification and more consistent family-centered intervention practices. Children with mild-to-severe hearing loss who were born in the last two decades are far more likely to be identified at birth and fit with HAs during infancy compared with children born in the 1990s or earlier (Dalzell et al., 2000; Harrison, Roush, & Wallace, 2003; Holte et al., 2012; Spivak, Sokol, Auerbach, & Gershkovich, 2009). Contemporary practices for CHH are firmly grounded in the assumption that providing children with early, consistent auditory access through amplification will have positive long-term implications for academic, language, and psychosocial achievement (Moeller & Tomblin, 2015). However, there is still a need to conduct research to support aspects of service delivery for this population, including early provision of RM systems in home environments.

Prospective, longitudinal research is needed to determine the effectiveness of RMs for CHH. Mulla and McCracken (2014) published one such study, in which they prospectively evaluated the benefits of RM systems in seven preschool-age CHH over a period of up to 8 months. Five of the seven children established consistent daily RM use (defined by the authors as average use of 2 hr or more per day) based on parent report measures. The researchers assessed trends in language skills via a parent report developmental survey. Three out of the seven children showed substantial increases in their standard scores on the parent report measure and a decreased gap over time between their developmental age scores and their chronological ages. Mulla and McCracken concluded that the use of the RM system led to improved global language skills in those three children. However, they acknowledged that improvements could have been due to developmental maturation and not only the result of RM use. The lack of a comparison group of children who did not have access to RM systems makes it impossible to determine if the children's gains were due to using the RM system, developmental maturation, effectiveness of early intervention, or a combination of all three factors. The use of parent report measures and developmental age scores as the outcome measure are also problematic. Parents may have had positive expectations regarding the benefit of the RM system, which could have influenced parent report outcomes. Furthermore, developmental age scores are not psychometrically valid as outcome measures for a number of reasons. Unlike standard scores from norm-referenced standardized tests, developmental age scores do not account for variance around the mean, cannot be represented on an equal interval scale, and are less sensitive to differences in ability levels (Hishinuma & Tadaki, 1997; Sullivan, Winter, Sass, & Svenkerud, 2014).

In summary, there are numerous gaps in our knowledge regarding whether children benefit from receipt of RM systems in the home. Such information could inform public health and educational policy, given that RM systems are costly and often covered through state or local public health or educational funding (Walker, Curran, Spratford, & Roush, in press). Evidence-based research is needed to guide the allocation of such resources.

This article addresses the following research questions:

1. What factors predict receipt of a personal RM system for home use?

2. Do CHH who have access to personal RM systems for home use have better language outcomes (i.e., vocabulary, morphosyntax, and discourse) than CHH without personal RM systems?

Because the OCHL project consisted of observational data with CHH, we were unable to randomly assign children to treatment versus no-treatment conditions. Therefore, propensity score matching was used to document whether differences in language outcomes at age 5 years existed between groups of children who did or did not receive RM systems for home use by 4 years of age. This article will describe the advantages and disadvantages of using propensity score matching to answer clinical research questions when an RCT is not an option.

Method

Participants

Participants were a cohort of 132 children who completed language assessments at the 5-year-old test visit as part of the OCHL study. Of these, 55 of 132 (41.7%) received an RM for home use for a minimum of 1 year prior to data collection. Children were recruited from three primary sites: University of Iowa, Boys Town National Research Hospital, and University of North Carolina at Chapel Hill (see Tomblin, Walker, et al., 2015, for additional information about recruitment). Children who participated in the OCHL study met the following criteria: (a) a permanent bilateral hearing loss, with better-ear pure-tone average (BEPTA) in the mild-to-severe hearing loss range; (b) at least one primary caregiver who spoke English in the home; (c) use of a spoken language communication approach; and (d) no significant delays in cognitive, visual, or motor domains. Children had to be within 1.5 SDs of the norm-referenced mean on at least one of the two nonverbal subtests, Block Design and Matrix Reasoning, of the Wechsler Preschool and Primary Scale of Intelligence–Third Edition (Wechsler, 2002) to qualify for participation. Vision had to be within normal or corrected-to-normal limits. Children who had motor impairments that precluded completing visual reinforcement audiometry, conditioned play audiometry, or pointing to items on standardized tests were excluded.

The mean BEPTA was 48.44 dB HL (SD = 13.01). Ninety-six children did not pass their newborn hearing screen. The mean age at confirmation of hearing loss was 5.36 months (median = 2.75, SD = 8.67), and the mean age at HA fitting was 8.88 months (median = 5, SD = 10.64) for children who were referred based on the newborn hearing screen. Thirty-six children were identified with hearing loss after the newborn period due to passing the newborn hearing screen, not being screened, or lack of follow-up following the newborn hearing screen referral. Of those 36 later-identified participants, the mean age at confirmation of hearing loss was 25.92 months (median = 27.5, SD = 13.49), and the mean age at HA fitting was 27.03 months (median = 28, SD = 13.89).

Data Collection

All study procedures were approved by the institutional review boards of each institution. As previously described in other articles (e.g., Tomblin, Walker, et al., 2015), children participated in two visits a year for children under the age of 2 years and one visit a year for children older than 2 years. At the initial visit, parents completed an intake interview with an examiner who documented demographic characteristics (e.g., maternal education level, sex, age at HA fitting, whether the child owned an RM system, and, if so, age at RM receipt). Starting in the second year of data collection for the study, parents completed a questionnaire regarding RM systems. In this questionnaire, parents reported whether their child had a personal RM system for home settings (home settings included any situation that

did not involve day care or preschool). If so, parents estimated the average amount of time the child wore the personal RM system per day outside the day care or preschool setting. A separate questionnaire asked about RM receipt in preschool settings. The RM questionnaire can be found online (http://ochlstudy.org/assessment-tools.html).

Audiologic Assessment

A certified audiologist with pediatric experience completed all hearing assessments. A test assistant participated in assessments as needed. The audiologist obtained air-conduction and bone-conduction thresholds at 500, 1000, 2000, and 4000 Hz at a minimum, using visual reinforcement audiometry, conditioned play audiometry, or conventional audiometry depending on the age of the child. All attempts were made to obtain ear-specific thresholds utilizing insert earphones, supra-aural headphones, or the child's own earmolds paired with insert earphones. Audiologists obtained sound-field thresholds if the child would not tolerate the testing with earphones or headphones. The four-frequency BEPTA was calculated for subsequent analyses.

Language Assessments

Trained examiners completed all language assessments with children within 3 months of the children's fifth birthdays, whenever possible. In two cases, children were tested 4 months after their fifth birthday. Children wore HAs during the language assessments but did not use RM systems (if they were available) because of the proximity between the examiner and the child and optimal acoustic environments for testing. HAs were checked by audiologists prior to all speech-language assessments (Tomblin, Walker, et al., 2015), and participants had full access to visual cues during testing. We included results from measures selected to tap different skills within the broader domain of language: vocabulary, morphosyntax, and discourse.

Vocabulary. The Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007) is a norm-referenced, standardized measure of receptive vocabulary. Standard scores were derived from the raw score total based on the chronological age of the child.

Morphosyntax. The Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2004) Word Structure is a subtest that assesses expressive morphological development using picture stimuli. Scaled scores were derived from the raw score total based on the chronological age of the child.

Discourse. The Preschool Language Assessment Instrument–Second Edition (PLAI-2; Blank, Rose, & Berlin, 2003) is a norm-referenced, standardized assessment of children's discourse abilities. Assessment items are categorized into four levels of abstraction, increasing in difficulty. The first level, Matching, contains language that is highly contextualized (e.g., the examiner points to a picture of a cup and asks, "What is this called?"). The second level, Selective Analysis, asks the child to integrate multiple concepts (e.g., "What shape is the bowl?"). The third level,

Reordering, assesses the child's ability to respond to questions or commands that are subtle or linguistically constrained (e.g., "Tell me what we put in the bowl before we added the egg."). The fourth level, Reasoning, represents the highest level of abstraction. Children must predict outcomes or provide justifications for responses, similar to the decontextualized language that may be used in a classroom setting (e.g., "What will happen to the cookies when we put them in the oven?"). Receptive and expressive scaled scores were utilized in the analysis.

Statistical Analysis

Likelihood of RM Receipt

To address the question of what factors were associated with likelihood of RM receipt in home settings, we utilized logistic regression models. In other words, given that we knew which children in our data set had received RMs for home use, we were able to identify individual differences that were associated with RM receipt. The dichotomous dependent variable was RM receipt in home settings (i.e., owning an RM vs. not owning an RM). The independent variables were sex, testing site (University of Iowa, Boys Town National Research Hospital, and University of North Carolina at Chapel Hill), maternal education level, BEPTA, race/ethnicity (i.e., White, Black, Hispanic, Latino or other Spanish origin, American Indian or Alaskan native, Asian, or Native Hawaiian or Pacific Islander), and family history of hearing loss. Family history was operationalized in two ways: parent report of hearing loss in a parent or older sibling and parent report of hearing loss in an older sibling only. The two variations on the family history variable were tested in separate models. Because of the large number of maternal education categories collected, arbitrary ordinal levels were introduced in the data analysis. These levels consisted of high school, some college, college, and graduate school/postgraduate. Dummy variables represented categorical variables in the statistical models.

Propensity Score Matching

The results from the logistic regressions predicting home RM receipt, as described above, were then used to analyze the effect of home RM receipt on child language outcomes via propensity score matching. First, the logistic regression model was implemented in Statistical Analysis System (SAS) software. Each participant was assigned a single score reflecting the participant's probability of receiving an RM. This score consisted of a weighted composite, taking into account all variables that affected RM receipt, based on the logistic regression analysis. Variables included in the weighted propensity score calculation were BEPTA, maternal education, and testing site. Any participant with missing data for any of the predictors included in the final model was omitted from the model. A propensity-matching SAS Macro (Fraeman, 2010) was then implemented to identify pairs of participants who did and did not receive RM systems for personal use. We then examined the distribution of baseline predictors included in the final model in both groups to ensure that our groups were well matched prior to RM receipt. The RM and non-RM groups were similar in the distribution of all baseline characteristics used for matching, including BEPTA, maternal education, and testing site. Thus, we matched pairs according to the overall likelihood of RM receipt and ensured that groups were matched according to the full distribution of baseline predictors. After determining that groups exhibited similar baseline profiles, paired t tests were carried out to compare language performance at the age of 5 years on receptive vocabulary (PPVT-4; Dunn & Dunn, 2007), expressive morphosyntax (CELF-4 Word Structure; Semel et al., 2004), and receptive and expressive discourse (PLAI-2; Blank et al., 2003).

With respect to selection of predictors to include in the propensity-matching procedure, it should be noted that the inclusion of a wide array of predictors is intended to mimic the effects of random assignment in an RCT: Children in both groups should be similar enough that treatment status is not confounded with any other baseline difference between treated and untreated children. That is, groups must present with similar characteristics prior to intervention. For example, a model that resulted in groups differing in BEPTA would have been discarded, even if participant groups were matched on overall propensity score, because we would not be able to assume that any between-groups difference in outcomes arose from RM receipt rather than BEPTA differences. Thus, we ensured that models resulted in RM and non-RM groups that were similar at baseline on all characteristics of interest. As is recommended in the propensity score literature (e.g., Austin, 2011), we did not require that predictors of RM receipt reach traditional levels of significance, but rather retained predictors that resulted in more closely matched pairs of participants, even when $p > .05$ in the logistic regression. For example, if omission of one predictor resulted in groups that differed at baseline, we retained that predictor in the model. In this way, we could strengthen the inference that postintervention differences in language skill were not caused by preintervention differences across our two groups of participants.

In addition to the need to include predictors that fall short of traditional levels of significance, propensity score matching requires that all predictors of treatment receipt relate to participant status prior to any participant receiving the treatment of interest. Inclusion of variables that may present after initiation of treatment for a portion of the participant group may result in use of predictors that were affected by the treatment itself or that are confounded with outcomes measures, rather than predictors that lead to receipt or nonreceipt of an intervention. For example, use of an RM system at school was not included as a predictor because participants may have been more likely to use an RM system in the school setting due to prior possession of an RM system for home use or may have received an RM system for home use following initiation of use in the school setting. In order to align with best practices for using propensity score techniques, we did not include any variables that may have occurred following

RM receipt by any child in the sample, including family involvement measures, provision of intervention services, child language performance prior to the age of 5 years, and whether children also had an RM system in the preschool setting. We acknowledge that these variables, particularly use of an RM in the preschool setting, may have affected the outcomes for children who owned RM systems in home settings. Forty-three percent of preschool-age children who had a school RM system also had a home RM system, but we were not able to control for this potential confound when using propensity score–matching techniques. We chose to focus on predictors evident at birth and outcomes at 5 years of age largely to maximize our sample size: Limiting the sample to children who received an RM after a certain age or to outcomes prior to 5 years would both reduce the number of participants we were able to include in the analysis.

Exclusion of North Carolina Residents

Participants who were residents of North Carolina exhibited a 100% probability of RM receipt in this sample. Thus, inclusion of North Carolina residents in the logistic analysis and propensity score matching led to two problems. First, the model did not reflect the effect of predictors other than state of residence for participants outside North Carolina. That is, inclusion of this subgroup leads "location" to be weighted extremely heavily in the final model, which, in turn, obscures the predictive power of other variables for children who resided outside North Carolina. Removal of the North Carolina residents reveals that, although significant, location of testing does not predict RM receipt as strongly as would be expected on the basis of the full sample.

The second problem lies in the requirements of propensity score matching. All treated–untreated pairs must have similar overall propensity scores. All children tested in North Carolina had scores above 0.80, with most over 0.90. The only untreated children with scores in this range were participants tested at North Carolina but residing in other states. This meant that the North Carolina residents were mostly unmatched. In addition, inclusion of this subgroup led to difficulties in matching children from other states. That is, the fact that location was weighted so heavily made it challenging to identify groups matched on factors such as BEPTA, because these predictors did not alter the overall score that strongly.

The extremely high rate of RM receipt for this subgroup of North Carolina residents likely reflects a combination of the local practice of recommendation of RM receipt from the audiology clinic at the University of North Carolina Hospital in Chapel Hill, where the majority of participants received clinical services, and RM funding patterns in the state. Because of these concerns, we excluded North Carolina residents from both analyses: the logistic regression to predict likelihood of RM receipt and the propensity score matching. Participants who were tested at the North Carolina site but who resided in other states did not exhibit a 100% probability of RM receipt and, thus, remained in the analysis.

Following exclusion of North Carolina residents, 104 participants remained in the sample.

Outcome Measures

With respect to outcome measures, not all participants completed all outcome measures, with missing data being greatest for the CELF-4 (Semel et al., 2004). To obtain the maximum possible number of pairs that could be compared on each outcome measure, we restricted candidate pair members to include only those participants who had completed the target measure. Because comparisons are conducted across pairs, any pair in which one participant exhibited missing data would be dropped entirely from the comparison. This means that a participant who did have complete data for an outcome measure may not be included in the final comparison because the matched participant did not complete the target measure. All participants with complete data for predictors of RM receipt were used to calculate probability of treatment through use of the logistic regression procedures, regardless of which outcome measures each participant had completed. However, when we implemented the SAS Macro to identify matched pairs of participants, we restricted the procedure such that the program could only select participants who were not missing CELF-4 data. That is, when evaluating CELF-4 results, no participant who had CELF-4 data could be paired with any participant who was missing CELF-4 results. Thus, each pair consisted of two participants with complete CELF-4 data. This was repeated for the PPVT-4 (Dunn & Dunn, 2007) and the PLAI-2 (Blank et al., 2003), as different participants had missing data for the three outcomes.

Identification of Matched Pairs

With respect to the identification of matched pairs, a perfect match between treated and untreated participants would present as a sequence of pairs with a caliper distance of 0.00. The caliper distance is the absolute value of the difference between the propensity score of the untreated participant and the treated participant in a matched pair. Because perfect matching is not possible, we must select a maximum difference that is allowable. The matching program will then select matches from within the target range. Differences from 0.08 to 0.25 have been used within propensity-matching schemes in order to best identify participants similar enough for comparison while accounting for the natural variability within a sample. In our sample, we used a maximum caliper difference of 0.10 in propensity scores between members of a matched pair. This distance allowed us to identify a maximum number of pairs for analysis while ensuring that we were not overly restricting variability in baseline measures and that our groups were similar on all baseline measures. Thus, we permitted a child in the RM group with a propensity score of 0.71 to be matched with a child in the non-RM group with a score of anywhere from 0.61 to 0.81. A child in the RM group with a propensity score of 0.71 could not be matched with a child in the non-RM group with a score of 0.60. Participants

without a candidate match in an acceptable caliper distance were dropped from the analysis, as is typical in propensity score–matching techniques. This procedure does not require participants to be matched on each variable of interest but rather focuses on overall probability of RM receipt, given the child's profile.

Results
Factors That Predict Personal RM Receipt for Home Settings

In order to match participants, we first completed our logistic regression to determine factors that predicted likelihood of RM receipt. Results from this stage of the analysis feed into the later matching procedures. Table 1 displays the final model of likelihood of RM receipt in home settings for the 104 participants included in the final analysis. BEPTA was a significant predictor of RM receipt, with higher probability of RM receipt for children with greater hearing loss (odds ratio [OR] = 1.06, p = .004). Primary testing site also emerged as a predictor; children tested by Boys Town National Research Hospital were significantly less likely to receive an RM system than children tested by the University of North Carolina at Chapel Hill (OR = 0.09, p = .03). This analysis included children who were tested at the University of North Carolina at Chapel Hill but were not North Carolina residents (n = 6). The same pattern was noted for the University of Iowa compared with the University of North Carolina at Chapel Hill, but this difference was not significant (OR = 0.13, p = .18). Neither sex, race, nor family history of hearing loss emerged as significant predictors of RM receipt. Maternal education level was not significantly associated with RM receipt when included in the full model (OR = 0.62, p = .69), although it did predict receipt in a univariate analysis (p = .02). As shown in Table 1, mothers with a college education presented with lower probability of RM receipt than children whose mothers had a high school education. We are unsure why the estimates presented in this direction. It may have been due to availability of funds to purchase an RM (e.g., Medicaid) or an interaction between education level and other factors

such as testing site or BEPTA. However, no interaction between maternal education and other predictors emerged as significant in our testing. It is possible that a larger sample would allow for further exploration of the effect of maternal education on RM receipt.

Effect of RM Receipt on Language Outcomes

As stated in the Method section, predictors that do not reach significance are included in propensity score models if there is an indication that omitting such predictors results in a confound between baseline characteristics and treatment assignment. This was the case with maternal education; omitting this predictor resulted in mismatched treated and untreated groups. More specifically, omission of maternal education resulted in a higher rate of participants with maternal education at the high school level in the group that did not receive RM systems than in the group that did receive RM systems. Thus, omission of maternal education resulted in a confound where groups were dissimilar on both treatment status and maternal education at baseline, even though maternal education was not significant in the logistic regression itself. Of note, omission of sex, race, and family history from the propensity score–matching model did not result in groups that were unbalanced on any of these variables of interest. As a result, maternal education was retained in the model predicting RM receipt, whereas sex, race, and family history were omitted.

The propensity score–matching procedures for participants with PPVT-4 (Dunn & Dunn, 2007) data resulted in identification of 27 matched pairs of treated (children with a personal RM for home use) and untreated (children without a personal RM) participants, for a total of 54 participants. Procedures restricted to participants with CELF-4 Word Structure (Semel et al., 2004) data resulted in identification of 13 matched pairs of treated and untreated participants, for a total of 26 participants. Procedures restricted to participants with PLAI-2 (Blank et al., 2003) data resulted in identification of 14 pairs of treated and untreated participants, for a total of 28 participants. Participant descriptions for each comparison (PPVT-4, CELF-4, PLAI-2) may be seen in Tables 2, 3, and 4. Figure 1 shows

Table 1. Model of likelihood of personal remote microphone receipt in home settings.

Fixed effects	Estimate	SE	χ^2	Pr(> |z|)	OR
Intercept	−3.19	1.02	9.85	.002	
Test site (Boys Town)[a]	−0.91	0.41	4.96	.03	0.09
Test site (Iowa)[a]	−0.57	0.43	1.77	.18	0.13
Better ear PTA	0.06	0.02	8.31	.004	1.06
Maternal education (high school)[b]	0.14	0.49	0.09	.77	0.65
Maternal education (college)[b]	−0.19	0.48	0.16	.69	0.62
Maternal education (graduate school)[b]	0.05	0.40	0.01	.90	0.79

Note. Pr = probability; Boys Town = Boys Town National Research Hospital; Iowa = University of Iowa; PTA = pure-tone average; OR = odds ratio.
[a]Reference group: University of North Carolina at Chapel Hill. [b]Reference group: some college.

Table 2. Matched PPVT-4 pairs included in propensity score–matching analysis.

Group	Total *n*	Mothers with high school education	Testing site	Mean better ear PTA (*SD*)	Mean propensity score (*SD*)	Mean PPVT-4 standard score (*SD*)
Home RM group	27	4	IA: 10 BT: 15 NC: 2	50.12 (11.23)	0.34 (0.16)	101.46 (12.11)
No-RM group	27	4	IA: 9 BT: 17 NC: 1	52.88 (13.52)	0.36 (0.16)	94.44 (15.88)

Note. PPVT-4 = Peabody Picture Vocabulary Test (normative mean = 100, *SD* = 15); PTA = pure-tone average; RM = remote microphone; IA = University of Iowa; BT = Boys Town National Research Hospital; NC = University of North Carolina at Chapel Hill.

individual data for the matched pairs on each of the outcome measures. Note that, as may be expected, there is some overlap across outcome measures in terms of the specific RM–non-RM participant pairs included. For example, the CELF-4 and PLAI-2 analyses contained four identical treated–untreated pairs. However, the majority of pairs differ across outcome measures. This is to be expected as missing data differed for each outcome measure, and we restricted matching to those participants with data for each measure of interest.

Vocabulary

Twenty-seven pairs of participants completed the PPVT-4 (Dunn & Dunn, 2007; see Table 2). No significant difference between groups emerged on the paired *t* test for PPVT-4 standard scores ($t = 0.55$, $p = .59$, $d = .22$). Of note, performance fell well within the typical range for this measure, with mean standard scores of 101.46 (*SD* = 12.11) and 94.44 (*SD* = 15.88) for pairs of participants who did and did not receive RM systems, respectively. Results did not support a statistically significant effect of home RM system receipt on receptive vocabulary.

Morphosyntax

Thirteen pairs of participants completed the CELF-4 Word Structure subtest (Semel et al., 2004; see Table 3). No significant difference between groups emerged on the paired *t* test for Word Structure scaled scores ($t = 0.55$,

$p = .61$, $d = −.21$). Results did not support a statistically significant effect of home RM system receipt on expressive morphology.

Discourse

Fourteen pairs of participants completed the PLAI-2 (Blank et al., 2003; see Table 4 and Figure 2 for PLAI-2 matched pair description and outcomes). The results of the *t* test for PLAI-2 Receptive Discourse scaled scores were significant ($t = 2.32$, $p = .03$, $d = .95$). Children who received a personal RM system had significantly higher scores on Receptive Discourse ability, with a large effect size. The *d* of .95 indicates that participants who received a personal RM outperformed participants who did not receive a personal RM by close to 1 *SD* on the Receptive Discourse measure. There was no significant difference in performance between groups on PLAI-2 Expressive Discourse scaled scores ($t = 1.76$, $p = .07$), although the difference trended in the same direction as the receptive scaled scores and the effect size was medium ($d = .57$).

Discussion

The primary objective of this study was to explore the association between personal RM receipt in home settings and later language outcomes. Findings related to the impact of RM systems in the home on language outcomes were mixed, with a significant positive effect of RM receipt

Table 3. Matched CELF-4 WS pairs included in propensity score–matching analysis.

Group	Total *n*	Mothers with high school education	Testing site	Mean better ear PTA (*SD*)	Mean propensity score (*SD*)	Mean CELF-4 WS scaled score (*SD*)
Home RM group	13	3	IA: 7 BT: 6 NC: 0	55.96 (11.24)	0.36 (0.15)	8.62 (4.41)
No-RM group	13	1	IA: 6 BT: 7 NC: 0	56.38 (11.97)	0.38 (0.15)	9.54 (4.29)

Note. CELF-4 WS = Clinical Evaluation of Language Fundamentals–Fourth Edition Word Structure subtest (normative mean = 10, *SD* = 3); PTA = pure-tone average; RM = remote microphone; IA = University of Iowa; BT = Boys Town National Research Hospital; NC = University of North Carolina at Chapel Hill.

Table 4. Matched PLAI-2 pairs included in propensity score–matching analysis.

Group	Total n	Mothers with high school education	Testing site	Mean better ear PTA (SD)	Mean propensity score (SD)	Mean PLAI-2 Exp scaled score (SD)	Mean PLAI-2 Rec scaled score (SD)
Home RM group	14	3	IA: 6 BT: 7 NC: 1	52.96 (7.10)	0.35 (0.15)	10.71 (4.12)	10.50* (2.50)
No-RM group	14	1	IA: 6 BT: 7 NC: 1	55.46 (12.14)	0.39 (0.12)	8.21 (3.58)	8.21* (2.72)

Note. PLAI-2 Exp = Preschool Language Assessment Instrument–Second Edition Expressive Language (normative mean = 10, SD = 3); PLAI-2 Rec = Preschool Language Assessment Instrument–Second Edition Receptive Language (normative mean = 10, SD = 3); PTA = pure-tone average; RM = remote microphone; IA = University of Iowa; BT = Boys Town National Research Hospital; NC = University of North Carolina at Chapel Hill.

*p = .03.

on discourse measures and nonsignificant findings on measures of vocabulary and morphosyntax.

Factors That Predict Personal RM Receipt for Home Settings

We currently lack evidence regarding the effects of personal RM receipt for preschool-age CHH on language outcomes and do not yet know whether specific factors are associated with more or less benefit from personal RMs. Thus, results from this analysis reflect the current state of clinical practice in the absence of clear evidence to guide decision making regarding RM receipt.

The results from the logistic regression indicate that children with milder hearing loss receive personal RM systems less frequently than children with greater degrees of

Figure 1. Individual data for propensity score–matched pairs for the Preschool Language Assessment Instrument–Second Edition (PLAI-2) Receptive Discourse (top left), PLAI-2 Expressive Discourse (top right), Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4) Word Structure (bottom left), and Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; bottom right). RM = remote microphone.
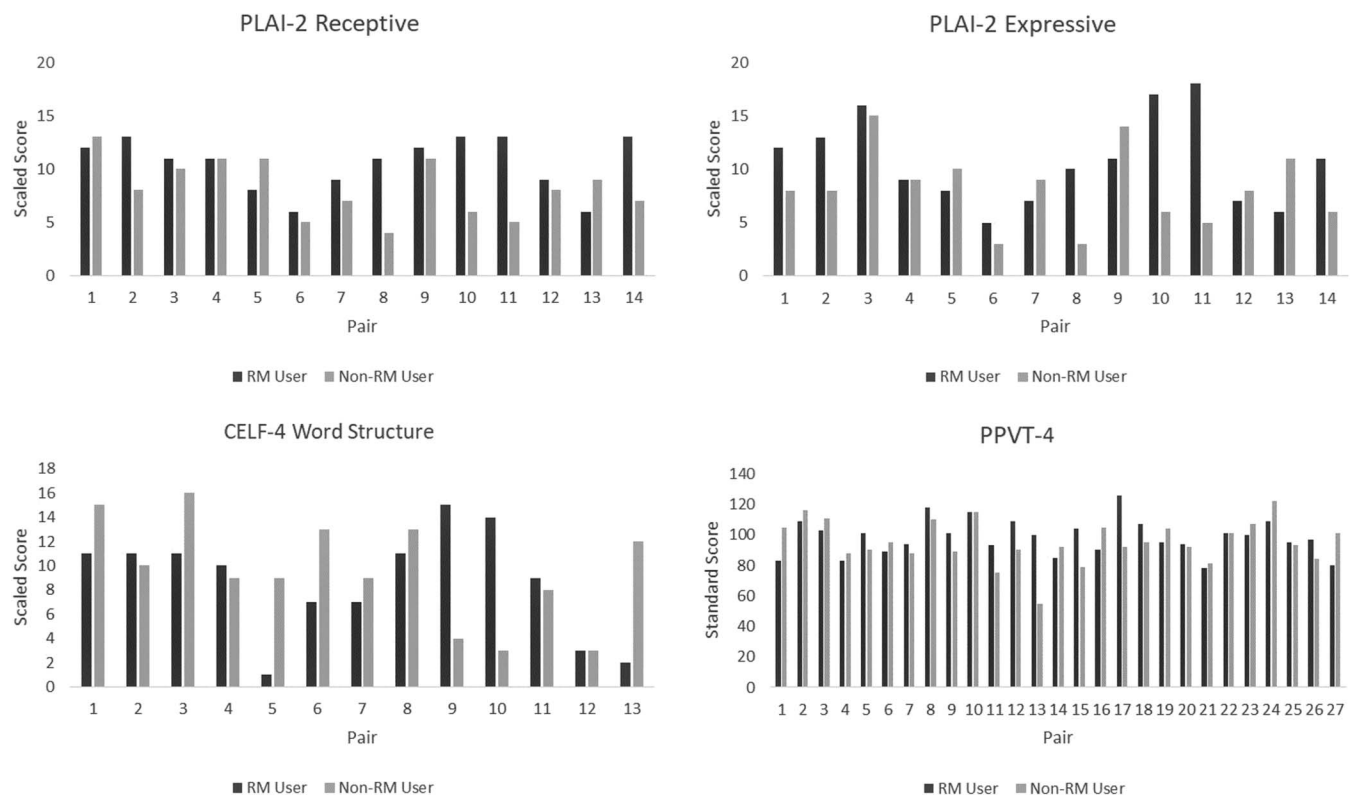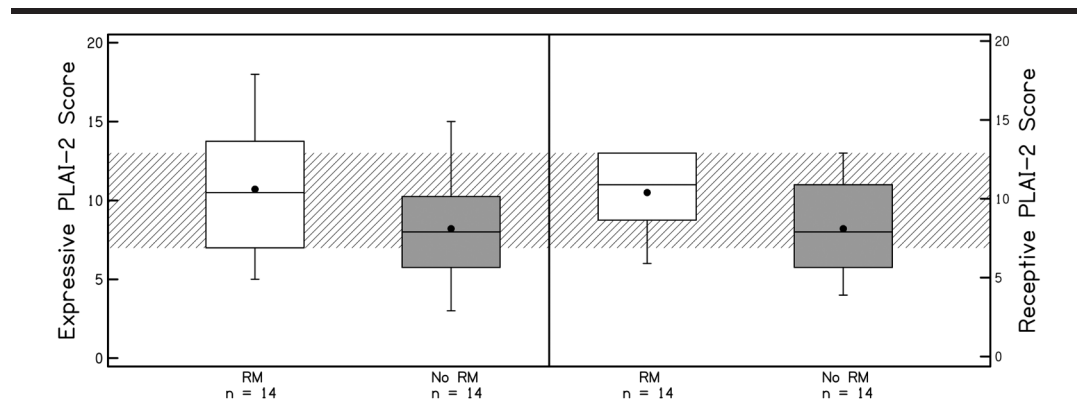
Figure 2. Performance of matched pairs on Preschool Language Assessment Instrument–Second Edition (PLAI-2) Expressive and Receptive Discourse scores. RM = remote microphone.

hearing loss. We did not specifically investigate why degree of hearing loss is an influential factor in RM receipt. We speculate that, because of the child's poor hearing thresholds, parents and professionals may understand that children with moderately severe hearing loss benefit from the increased intelligibility and audibility of speech provided with use of RM systems compared with their HAs. Children with milder hearing loss, in contrast, may not be considered to have sufficient lack of auditory access to warrant the need or expense of RM systems, especially when the benefit from HAs may not be clear to parents or professionals (Walker et al., 2017). However, children with mild–moderate degrees of hearing loss also remain at a disadvantage compared with peers with typical hearing when listening in background noise or to speech originating from a distance. During sensitive periods of development, children with mild–moderate hearing loss may be able to benefit from the use of RM technology in multiple areas such as language, social skills, or behavioral management. Additional research is needed to understand the advantages and disadvantages of using RM systems in children with milder degrees of hearing loss. These results provide evidence that a variety of factors lead children to be more or less likely to receive RM systems for personal use and may provide guidance to researchers, audiologists, and individuals training audiologists to make clinical decisions about RM provision.

New technology and equipment may support utilization of RM systems for populations that may have been overlooked for candidacy previously. It should be noted that the data in the current study were collected between 2009 and 2013 and that the RM technology worn by these participants may already be considered outdated. For example, microphone/transmitters can now be packaged into one piece instead of two, eliminating the connecting wire and making the technology even smaller and more convenient to use. In addition, when the questionnaires for this study were administered, it was assumed that the children were using analog frequency modulation (FM) radio transmission to deliver the signal from the transmitter to the receiver. We did not query parents regarding the type of

RM transmission being used. In recent years, digital modulation (DM) technology has been introduced as an alternative to FM systems. DM systems offer concrete advantages in speech recognition over FM technology. The audio bandwidth in DM systems is wider than in FM systems, thus providing more access to high-frequency information (Wolfe, Morais, Schafer, Agrawal, & Koch, 2015), although the HA may still impose bandwidth restrictions as the transducer (Kimlinger, McCreery, & Lewis, 2015). Digital signal processing also provides better speech recognition in moderate-to-high noise levels compared with FM systems because of the adaptive gain capacities that are available in DM (Thibodeau, 2014; Wolfe et al., 2013). These improvements may make RM technology even more beneficial to use in home and community settings, allowing greater utilization with children who have mild-to-severe hearing loss. Unfortunately, the current results do not capture the use of FM versus DM technology in the home and community setting; future research is needed to investigate the frequency of use and utility of DM systems in the home and community.

Effect of Home RM Receipt on Language Outcomes

A strength of this study is that propensity score–matching techniques enabled us to examine the effect of RM systems on child language outcomes at the age of 5 years using a rigorous, unbiased statistical technique. This represents an expansion upon past research, as we were able to examine larger sample sizes, narrower age ranges, and multiple aspects of language compared with previous studies (Moeller et al., 1996). The use of propensity score matching allowed us to leverage the rich OCHL data set for analysis and to compare outcomes of children who had received an RM system with children who did not have an RM system at home, after controlling for factors that predict likelihood of RM receipt. Previous research indicated potential advantages in parent report measures but lacked the inclusion of a no-RM control group (Mulla & McCracken, 2014). In the current data, we found no evidence

of an effect of personal RM receipt on receptive vocabulary. In fact, both children who did and did not receive personal RM systems exhibited performance well within typical limits on the PPVT-4 (Dunn & Dunn, 2007). These results are consistent with previous findings in the OCHL study, which indicates that vocabulary is a relative strength for CHH (Tomblin, Harrison, et al., 2015). There was also no evidence of an effect of RM on production of morphosyntactic markers on the CELF-4 Word Structure (Semel et al., 2004). Based on the finding that morphological development appears to be a vulnerable domain of language for CHH (Moeller et al., 2015), we might predict that enhanced auditory access via RM systems would support expressive morphology, but that was not observed in the current findings.

In contrast to the lack of significant differences for vocabulary and grammar, we saw a large effect of RM receipt for Receptive Discourse and a medium-sized effect for Expressive Discourse. This may indicate that RM systems support development of higher-level language skills. Recent research by Benitez-Barrera, Angley, and Tharpe (2018) may provide some explanation for how RM systems support language discourse skills. Benitez-Barrera et al. examined the impact of RM systems in the home on caregiver talk to young children with hearing loss. They found that caregivers tend to talk more frequently at a distance when using an RM system compared with using HAs alone with preschool-age children with moderate-to-profound hearing loss. Greater access to distant speech may enhance discourse skills because it will increase the frequency of exposure to communicative contexts. Additional investigation in a prospective study would be required to strongly state that there is a causal relationship between RM systems and language discourse.

Given our need to exclude predictors of RM receipt, which occurred after children in the sample began receiving RM systems, the current findings should be regarded as tentative. These findings support inclusion of higher-level discourse measures in future investigations of the effect of RM systems on child outcomes in order to more completely test the relationship between RM systems and language outcomes. If discourse skills truly are an area particularly affected by RM systems, future studies that focus solely on vocabulary or morphosyntax may not capture the effect of RM systems for young CHH.

Limitations and Future Directions

As we have described in previous articles (Tomblin, Walker, et al., 2015), the OCHL cohort consists of an economically advantaged cohort of children compared with the U.S. population as a whole. In addition, children from specific states (i.e., North Carolina) showed a higher likelihood of receiving RM systems compared with children from other states. As a result, the results of the current study may not be generalized to all children with hearing loss, which would include children from more diverse socioeconomic backgrounds and from states that did not participate in the multicenter study.

Another limitation of the current study is that we could not examine whether higher amounts of RM use led to better language outcomes. Because RM systems may be covered by public funds (Walker et al., in press), consideration of the effect of RM receipt is highly relevant for public policy concerns. However, examination of RM use may be a stronger test of the effect of RMs on outcomes. Unfortunately, we were unable to confirm that all children who owned an RM system actually used it. Seventeen parents of the 29 children who were included in the RM receipt group completed the RM interview at least once by the time their child was 4 years old. We were able to document that 16 of 17 parents used the RM system in specific contexts and one of 17 reported never using the RM system at the time of the RM interview. Unfortunately, we did not begin documenting RM usage until a couple of years after the longitudinal study had begun, in part because we did not initially anticipate that so many children would have personal RM systems. Because of time constraints during testing visits, 12 of the children never received the RM interview even after it was included in the test battery. Because of uncertainties of how often the RM system was being utilized by individual children and the fact that this is a preliminary report on the effects of RM systems, we chose not to examine how the amount of RM use influenced outcomes. Instead, we compared children who received an RM system for home use with children who did not have an RM system. We acknowledge that it is possible that the null findings and low effect sizes for the PPVT-4 (Dunn & Dunn, 2007) and CELF-4 (Semel et al., 2004) comparisons between the RM and non-RM groups may have been due to limited usage in the RM group. Future studies should carefully track average hours of daily use to establish how often the RM system is being used (Benitez-Barrera et al., 2018) and whether amount of use influences language outcomes, as it does with HAs (Tomblin, Harrison, et al., 2015).

The current study was also restricted by the experimental design of the larger longitudinal study. As stated in the Introduction, a common problem in trying to investigate the treatment effects of RM systems on language outcomes is the difficulty in conducting RCTs. Because the OCHL project consisted of observational data of CHH, we did not randomly assign children to treatment versus no-treatment conditions. The advantage of the propensity score–matching analysis was that it allowed us to control for baseline differences in the treatment versus no-treatment groups. Unfortunately, we were also constrained by the requirements for calculating the propensity scores. First, any participants with missing data for the predictors had to be omitted from the analysis. Second, we excluded residents of North Carolina from the analysis because of their high probability of receiving an RM system, which made it unfeasible to find children in the no-treatment group who matched on the other predictors. Third, best practices in propensity score matching requires that the predictors in the analysis occur prior to initiation of the treatment. Thus, we could neither control for nor examine the roles

of a number of possibly influential factors, including whether the children received and/or used an RM system in preschool, quality and quantity of early intervention, or level of family involvement in intervention. It is possible that children who had an RM system in preschool had improved access to speech in a specialized learning environment that may have influenced language outcomes at 5 years. Research on the impact of preschool RM systems, intervention services, and family involvement is greatly needed.

Conclusions

The current results supporting the provision of personal RM systems for preschool-age CHH were mixed. There is a need to further examine practices and attitudes surrounding fitting and using RM in the home setting from an intervention perspective. Many families of young children, especially those with mild hearing loss, may be underserved in terms of having access to personal RM systems. Additional high-quality research is needed to determine how RM systems used at home and preschool benefit auditory–linguistic access and spoken language outcomes of young CHH.

Propensity score–matching techniques provided a valuable avenue to test the relationship between personal RM receipt and language outcomes in CHH. This approach enabled us to leverage an observational, longitudinal data set to examine a question of clinical interest. However, propensity score matching has its own inherent limitations. Our ability to select a measure of intervention, predictors of intervention receipt, and outcomes of interest were all restricted by the requirements to limit our question to variables present prior to receipt of intervention by any single child and our need to maximize sample size. Overall, this type of analysis provides key evidence regarding the types of outcomes that may be useful to include in future prospective work.

Acknowledgments

References

Anderson, K. L., & Goldstein, H. (2004). Speech perception benefits of FM and infrared devices to children with hearing aids in a typical classroom. *Language, Speech, and Hearing Services in Schools, 35*(2), 169–184.

Atcherson, S. R., Childress, T., & Warren Kennett, S. (2016). Assistive hearing and access technologies and auditory–verbal therapy. In W. Estabrooks, K. MacIver-Lux, & E. Rhoades (Eds.), *Auditory–verbal therapy for young children with hearing loss and their families and the practitioners who guide them* (pp. 201–218). San Diego, CA: Plural.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424.

Benitez-Barrera, C., Angley, G., & Tharpe, A. M. (2018). Remote microphone system use at home: Impact on caregiver talk. *Journal of Speech, Language, and Hearing Research, 61,* 399–409.

Benoit, R. (1989). Home use of FM amplification systems during the early childhood years. *Hearing Instruments, 40*(3), 8–12.

Berzin, S. C. (2010). Understanding foster youth outcomes: Is propensity scoring better than traditional methods? *Research on Social Work Practice, 20*(1), 100–111.

Blank, M., Rose, S. A., & Berlin, L. J. (2003). *Preschool Language Assessment Instrument–Second Edition (PLAI-2)*. Austin, TX: Pro-Ed.

Crandell, C. C. (1993). Speech recognition in noise by children with minimal degrees of sensorineural hearing loss. *Ear and Hearing, 14*(3), 210–216.

Dalzell, L., Orlando, M., MacDonald, M., Berg, A., Bradley, M., Cacace, A., . . . Greenberg, E. (2000). The New York State universal newborn hearing screening demonstration project: Ages of hearing loss identification, hearing aid fitting, and enrollment in early intervention. *Ear and Hearing, 21*(2), 118–130.

Davies, M. G., Yellon, L., & Purdy, S. C. (2001). Speech-in-noise perception of children using cochlear implants and FM systems. *Australian and New Zealand Journal of Audiology, 23*(1), 52–62.

Dunn, D., & Dunn, L. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. Minneapolis, MN: NCS Pearson.

Finitzo-Hieber, T., & Tillman, T. W. (1978). Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *Journal of Speech and Hearing Research, 21*(3), 440–458.

Flexer, C. (1997). Individual and sound-field FM systems: Rationale, description, and use. *Volta Review, 99*(3), 133–162.

Fraeman, K. (2010). *An introduction to implementing propensity score matching with SAS®*. Bethesda, MD: United BioSource.

Gabbard, S. (2004). The use of RM technology for infants and young children. In R. C. Seewald & J. M. Bamford (Eds.), *A sound foundation through early amplification: Proceedings of an international conference* (pp. 155–161). Stafa, Switzerland: Phonak AG.

Harrison, M., Roush, J., & Wallace, J. (2003). Trends in age of identification and intervention in infants with hearing loss. *Ear and Hearing, 24*(1), 89–95.

Hawkins, D. B. (1984). Comparisons of speech recognition in noise by mildly-to-moderately hearing-impaired children using hearing aids and FM systems. *Journal of Speech and Hearing Disorders, 49*(4), 409–418.

Hishinuma, E. S., & Tadaki, S. (1997). The problem with grade and age equivalents: WIAT as a case in point. *Journal of Psychoeducational Assessment, 15*(3), 214–225.

Holte, L., Walker, E., Oleson, J., Spratford, M., Moeller, M. P., Roush, P., . . . Tomblin, J. B. (2012). Factors influencing follow-up to newborn hearing screening for infants who are hard of hearing. *American Journal of Audiology, 21*(2), 163–174.

Jiang, M., Foster, E. M., & Gibson-Davis, C. M. (2011). Breastfeeding and the child cognitive outcomes: A propensity score

matching approach. *Maternal and Child Health Journal, 15*(8), 1296–1307.

Kimlinger, C., McCreery, R., & Lewis, D. (2015). High-frequency audibility: The effects of audiometric configuration, stimulus type, and device. *Journal of the American Academy of Audiology, 26*(2), 128–137.

Larsen, J. B., & Blair, J. C. (2008). The effect of classroom amplification on the signal-to-noise ratio in classrooms while class is in session. *Language, Speech, and Hearing Services in Schools, 39*(4), 451–460.

Moeller, M. P., Donaghy, K. F., Beauchaine, K. L., Lewis, D. E., & Stelmachowicz, P. G. (1996). Longitudinal study of FM system use in nonacademic settings: Effects on language development. *Ear and Hearing, 17*(1), 28–41.

Moeller, M. P., & Tomblin, J. B. (2015). An introduction to the outcomes of children with hearing loss study. *Ear and Hearing, 36,* 4S–13S.

Moeller, M. P., Tomblin, J. B., & OCHL Collaboration. (2015). Epilogue: Conclusions and implications for research and practice. *Ear and Hearing, 36,* 92S–98S.

Morgan, C. J. (2017). Reducing bias using propensity score matching. *Journal of Nuclear Cardiology, 25,* 404–406.

Morgan, P. L., Frisco, M. L., Farkas, G., & Hibel, J. (2010). A propensity score matching analysis of the effects of special education services. *The Journal of Special Education, 43*(4), 236–254.

Mulla, I., & McCracken, W. (2014). Frequency modulation for preschoolers with hearing loss. *Seminars in Hearing, 35,* 206–216.

Nábělek, A. K., & Pickett, J. (1974). Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids. *The Journal of the Acoustical Society of America, 56*(2), 628–639.

Schafer, E. C., & Thibodeau, L. M. (2006). Speech recognition in noise in children with cochlear implants while listening in bilateral, bimodal, and FM system arrangements. *American Journal of Audiology, 15*(2), 114–126.

Semel, E., Wiig, E. H., & Secord, A. W. (2004). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)*. San Antonio, TX: Psychological Corporation.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103*(484), 1334–1344.

Spivak, L., Sokol, H., Auerbach, C., & Gershkovich, S. (2009). Newborn hearing screening follow-up: Factors affecting hearing aid fitting by 6 months of age. *American Journal of Audiology, 18*(1), 24–33.

Sullivan, J. R., Winter, S. M., Sass, D. A., & Svenkerud, N. (2014). Assessing growth in young children: A comparison of raw, age-equivalent, and standard scores using the Peabody Picture Vocabulary Test. *Journal of Research in Childhood Education, 28*(2), 277–291.

Thibodeau, L. (2014). Comparison of speech recognition with adaptive digital and FM remote microphone hearing assistance technology by listeners who use hearing aids. *American Journal of Audiology, 23*(2), 201–210.

Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., & Moeller, M. P. (2015). Language outcomes in young children with mild to severe hearing loss. *Ear and Hearing, 36,* 76S–91S.

Tomblin, J. B., Walker, E. A., McCreery, R. W., Arenas, R. M., Harrison, M., & Moeller, M. P. (2015). Outcomes of children with hearing loss: Data collection and methods. *Ear and Hearing, 36,* 14S–23S.

Wake, M., & Carew, P. (2016). Science, not philosophy, will help deaf and hard-of-hearing children reach their potential. *Pediatrics, 137*(1), 1–2.

Walker, E. A., Curran, M., Spratford, M., & Roush, P. (in press). Remote microphone systems for preschool-age children who are hard of hearing. *International Journal of Audiology*.

Walker, E. A., Holte, L., McCreery, R. W., Spratford, M., Page, T., & Moeller, M. P. (2015). The influence of hearing aid use on outcomes of children with mild hearing loss. *Journal of Speech, Language, and Hearing Research, 58*(5), 1611–1625.

Walker, E. A., Spratford, M., Ambrose, S. E., Holte, L., & Oleson, J. (2017). Service delivery to children with mild hearing loss: Current practice patterns and parent perceptions. *American Journal of Audiology, 26*(1), 38–52.

Wechsler, D. (2002). *The Wechsler Preschool and Primary Scale of Intelligence–Third Edition (WPPSI-III)*. San Antonio, TX: Psychological Corporation.

Wolfe, J., Morais, M., Schafer, E., Agrawal, S., & Koch, D. (2015). Evaluation of speech recognition of cochlear implant recipients using adaptive, digital remote microphone technology and a speech enhancement sound processing algorithm. *Journal of the American Academy of Audiology, 26*(5), 502–508.

Wolfe, J., Morais, M., Schafer, E., Mills, E., Mülder, H. E., Goldbeck, F., . . . Peters, B. R. (2013). Evaluation of speech recognition of cochlear implant recipients using a personal digital adaptive radio frequency system. *Journal of the American Academy of Audiology, 24*(8), 714–724.