Purpose: This study investigated long-term speech intelligibility outcomes in 63 prelingually deaf children, adolescents, and young adults who received cochlear implants (CIs) before age 7 ($M = 2;11$ years;months, range = 0;8–6;3) and used their implants for at least 7 years ($M = 12;1$, range = 7;0–22;5).

Method: Speech intelligibility was assessed using playback methods with naïve, normal-hearing listeners.

Results: Mean intelligibility scores were lower than scores obtained from an age- and nonverbal IQ-matched, normal-hearing control sample, although the majority of CI users scored within the range of the control sample. Our sample allowed us to investigate the contribution of several demographic and cognitive factors to speech intelligibility. CI users who used their implant for longer periods of time exhibited poorer speech intelligibility scores. Crucially, results from a hierarchical regression model suggested that this difference was due to more conservative candidacy criteria in CI users with more years of use. No other demographic variables accounted for significant variance in speech intelligibility scores beyond age of implantation and amount of spoken language experience (assessed by communication mode and family income measures).

Conclusion: Many factors that have been found to contribute to individual differences in language outcomes in normal-hearing children also contribute to long-term CI users’ ability to produce intelligible speech.

Mounting evidence suggests that cochlear implants (CIs) are an effective medical intervention to facilitate spoken language development in severely to profoundly deaf children (Geers, Brenner, & Tobey, 2011; Niparko et al., 2010). A large number of early-implanted children exhibit speech perception and language comprehension skills within 1 standard deviation (SD) of the normative mean when tested under quiet conditions in the laboratory or clinic (Geers & Sedey, 2011). In addition to receptive language abilities, early-implanted children develop significant expressive language skills and are often able to produce highly intelligible speech when assessed using playback methods with naïve, normal-hearing (NH) listeners (Allen, Nikolopoulos, & O’Donoghue, 1998; Beadle et al., 2005; Chin, Bergeson, & Phan, 2012; Habib, Waltzman, Tajudeen, & Svirsky, 2010; Miyamoto, Kirk, Robbins, Todd, & Riley, 1996; Osberger, Maso, & Sam, 1993; Peng, Spencer, & Tomblin, 2004; Tobey, Geers, Brenner, Altuna, & Gabbert, 2003; Tobey, Geers, Sundararajan, & Lane, 2011). These expressive spoken language measures are considered important benchmarks of speech and language development in CI users because intelligible speech production requires a child to have the prerequisite speech perception abilities to learn and understand speech, the linguistic knowledge to plan and execute spoken utterances, and the motor control abilities to articulate meaningful sentences. Thus, the ability to produce intelligible speech is an important milestone in spoken language development in this clinical population.

Early investigations of speech intelligibility in CI users typically assessed changes in speech intelligibility within the first few years of CI use. CI users’ speech intelligibility rapidly improves after implantation, and many deaf children are able to produce speech that is intelligible to groups of NH, naïve listeners within 4 or 5 years after implantation (Allen et al., 1998). The gains observed in speech intelligibility continue to improve well after 5 years of CI use (Beadle et al., 2005; Miyamoto et al., 1996). These earlier research studies also identified several factors that predict future success in producing intelligible speech. Both earlier age at CI implantation and earlier onset of deafness are associated with better speech intelligibility outcomes (Osberger et al., 1993). It is important to note, however, that these studies reported considerable variability in the speech intelligibility of CI users and suggested that conventional demographic factors—
specifically, years of CI use and age of CI implantation—contribute to this variability.

As the children who have been implanted with multi-channel CIs approach and enter adulthood, it is now possible to investigate long-term speech and language outcomes in these individuals (Pisoni, Cleary, Geers, & Tobey, 1999; Geers et al., 2011; Ruffin, Kronenberger, Colson, Henning, & Pisoni, 2013; Svirsly, Robbins, Kirk, Pisoni, & Miyamoto, 2000). These long-term studies allow us to better understand speech intelligibility outcomes more than 10 years after CI implantation. Tobey and colleagues assessed speech intelligibility using the McGarr sentence materials (McGarr, 1981) longitudinally in a large group of CI users. These CI users were tested at age 8 to 9 years and then again at age 15 to 18 years. Tobey et al. (2003, 2011) reported 63.5% key word intelligibility at 8 to 9 years of age (average of 5.5 years of CI use) and about 85% key word intelligibility at 15 to 18 years of age (average of 13.3 years of CI use), a pattern that was consistent with previous work suggesting that speech intelligibility continues to improve over time with CI use (Tobey et al., 2003, 2011). Tobey et al.’s results also demonstrated that long-term speech intelligibility outcomes are generally quite good: CI users are capable of producing speech that is highly intelligible to naive NH listeners. However, similar to earlier studies, Tobey et al. also reported considerable variability in the speech intelligibility scores: Although the mean speech intelligibility was 85%, the standard deviation was 20%. Tobey et al. (2011) identified a number of demographic, family, and child factors that contributed to this variability, including gender, family size, and communication mode. CI users from smaller families who relied more strongly on auditory-oral modes of communication produced more intelligible speech than did users of total communication. The longitudinal nature of Tobey et al.’s (2011) data set also allowed for an investigation of speech intelligibility measures of a single group of CI users over time and documented the gains with additional years of CI experience.

Despite the benefits of this prior research, there are a number of important gaps in our knowledge of long-term speech intelligibility outcomes in prelingually deaf children implanted with CIs. First, studies of this population are relatively scarce. With the exception of the sample studied by Geers and colleagues (Tobey et al., 2011), the samples in other studies have been a combination of CI users with different ages of onset of deafness (prelingual and postlingual) or early and late ages of implantation. Therefore, a need exists for replication of speech intelligibility findings with a different sample of CI users. Second, there has been no research on prelingually deaf CI users beyond 15 years postimplantation. This knowledge, which is only now becoming available as the first large pediatric cohorts reach 15 and more years of CI use, is crucial for understanding development into early and middle adulthood. Third, the design of Tobey et al.’s (2011) study precluded the investigation of other demographic factors that may contribute to variability in speech intelligibility outcomes. Specifically, there was relatively little variability in the duration of CI use and age of implantation between participants, which may have attenuated the contribution of these factors as possible contributors to outcome variability. In addition, the restricted range in the year of CI implantation did not allow for an investigation of possible cohort effects—differences based on the year of CI implantation. These were deliberate design choices; Tobey et al.’s more homogenous sample allowed the researchers to better understand longitudinal changes in speech intelligibility but precluded the investigation of other demographic factors that may affect speech intelligibility outcomes.

Our recent work investigating long-term speech and language outcomes of CI users suggests that language outcomes and predictors of those language outcomes vary by cohort (Ruffin et al., 2013). Thus, year of implantation may be an important source of variability in understanding long-term outcomes and the applicability of findings from long-term outcome studies of individuals implanted 20 years ago to CI users implanted today. These potential sources of variability were investigated in the present study, which reports long-term speech intelligibility outcomes in a more diverse sample of CI users, many of whom used their CIs for 15 years or more.

To address these gaps in our understanding, we report the results of long-term speech intelligibility outcomes in 63 prelingually, profoundly deaf CI users. These CI users were all long-term users spanning a wide age range, who received their implants between 7 and 21 years prior to assessment. We used the same speech intelligibility assessment methods as Tobey et al. (2011) used but with a different CI sample that allowed us to investigate the effects of several additional variables on speech intelligibility outcomes. Using the same materials and methodology also allowed for an independent replication of the data reported in Tobey et al. (2011), which is important in order to document the validity of the speech intelligibility measure and the generalizability of any given sample. Our sample of long-term CI users also allowed for additional analyses, including an investigation of cohort effects for CI users who received their implants at different times. Our recent findings on the speech perception abilities in the same sample of CI users (Ruffin et al., 2013) has shown that changes over time in CI device characteristics and implantation criteria can interact with years of CI use, producing reliable differences in samples of CI users who have used their CIs for the same number of years. This pattern suggests that in addition to the years of CI use, the year of implantation—which in our sample is highly related to the duration of CI use—may have important effects on measures of speech intelligibility, and our sample allows for investigation of these possible effects.

Method

Participants

Recruitment. All CI users tested in this study were recruited through multiple venues, including the patient populations receiving clinical services at a large hospital-based
CI clinic and CI users who had participated in previous studies in our research center. The study was also advertised to local professionals and schools who had contact with CI users. NH participants were recruited from the community through the use of flyers posted in the same institution and local geographic areas from which the CI sample was recruited. E-mail and Internet sites affiliated with our CI clinic and university were also used for recruitment of NH participants.

**CI sample.** The sample of CI users included 63 children, adolescents, and young adults who met the following five inclusion criteria: (a) onset of severe-to-profound hearing loss (>70-dB hearing loss in the better-hearing ear) prior to age 3 years; (b) cochlear implantation prior to age 7 years; (c) at least 7 years of CI use at the time of testing; (d) consistent use of a currently available, state-of-the-art multichannel CI system; and (e) living in a home where English is the primary spoken language. Potential participants were excluded if their medical chart or parental report indicated any comorbid developmental or neurocognitive delays or disabilities other than hearing loss or if their nonverbal IQ score was greater than 2 SDs below the normative mean. This sample allowed us to investigate potential age cohort effects on speech intelligibility outcomes. Ruffin et al. (2013) found systematic differences in speech perception outcomes in CI users who received their implants in different years, reflecting changes over time in device characteristics and CI candidacy criteria. Table 1 provides demographic and device information about the three cohorts of CI users studied here. Age Cohort 1 consisted of individuals who used their implant for between 7 and 9 years, Age Cohort 2 consisted of individuals who used their implant for between 10 and 14 years, and Age Cohort 3 consisted of individuals who used their implant for more than 15 years. In our sample, all testing occurred within a 22-month period, so duration of CI use is strongly related to chronological year of CI implantation.

**NH control sample.** The NH control sample included 63 children, adolescents, and young adults who met the following inclusion criteria: (a) ages 7 to 25 years; (b) nonverbal IQ score within 2 SDs of the normative mean; (c) passed a basic audiometric hearing screening assessment (each ear was tested individually with headphones at frequencies of 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz at 20 dB); (d) reported no significant developmental or cognitive delays; and (e) mean, standard deviation, and range of age and nonverbal IQ were similar to those of the sample of CI users.

**Materials**

**Sentence Intelligibility Task.** Sentence intelligibility was assessed using the McGarr Sentence Intelligibility Test (McGarr, 1981). This methodology has previously been used to measure spoken language intelligibility in CI users (Dawson et al., 1995; Geers, 2002; Osberger et al. 1993; Pisoni & Geers, 2000; Tobey et al., 2003, 2011; Tobey & Hasenstab, 1991). The test materials consist of meaningful English sentences that are three, five, or seven syllables in length. In each of the sentences, one word is designated as the “key word.” These key words are either of high or low predictability, based on the predictability of the key word within each sentence. There were six sentences containing high-predictability key words and six containing low-predictability key words at each sentence length for a total of 36 sentences (McGarr, 1981).

**Other language and cognitive measures.** In addition to results from the sentence intelligibility task reported here, the CI users performed a battery of tasks designed to assess language and cognitive skills. We were interested in how other language and cognitive skills were related to our measure of speech intelligibility. Of particular interest to the present study are the Hearing in Noise Test Sentences for Children (HINT-C; Nilsson, Soli, & Gelnert, 1996); Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4; Dunn & Dunn, 2007); and the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003), Forward and Backward Digit Spans, Non-Verbal IQ, and McGarr Sentence Speaking Durations. Additional methodological information regarding the HINT-C, PPVT-4, and CELF-4 can be found in Ruffin et al. (2013); methodological information regarding the Forward and Backward Digit Span tasks and Non-Verbal IQ can be found in Kronenberger, Pisoni, Henning, and Colson (2013); and methodological information regarding McGarr Sentence Speaking Duration can be found in AuBuchon, Pisoni, and Kronenberger (in press).

**Procedure**

Prior to testing, both groups of participants had fully consented (with assent by children as appropriate) to the protocol approved by the university’s institutional review board. All testing was completed at a hospital-based clinic. All CI users were tested by licensed speech-language pathologists; NH participants were tested either by the same speech-language pathologists or by an experienced psychometric technician.

All participants reported age at time of testing, sex, race/ethnicity, and family income, which was assessed on a 1 (< $5,500) to 10 (≥ $95,000) scale (intermediate values of 3, 5, and 7 correspond to annual income values of $15,000–$24,999, $35,000–$49,999, and $65,000–$79,999, respectively). The CI sample also reported age at onset of deafness, age at CI implantation, and years of implant use. Additional variables recorded for the CI sample included communication mode, which was coded on a 1 (mostly sign) to 6 (auditory–oral) rating scale (see Geers & Brenner, 2003), and preimplant residual hearing (mean unaided pure-tone average in the better-hearing ear for the frequencies 500 Hz, 1000 Hz, and 2000 Hz at 20 dBHL).

For each of the 36 McGarr test sentences, the examiner said the target sentence aloud while showing participants a card with the printed sentence. The participant was then prompted to repeat the sentence back to the examiner. Digital audio recordings were made of each speaker’s vocal responses for later playback and acoustic measurement.

**Intelligibility transcription procedure.** Sentence intelligibility was assessed using orthographic transcriptions
Table 1. Sample description.

<table>
<thead>
<tr>
<th>Variable</th>
<th>All</th>
<th>7–9 Years</th>
<th>10–14 Years</th>
<th>15+ Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>63</td>
<td>19</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>Years of CI use</td>
<td>12.1 (3.9)</td>
<td>8.1 (0.8)</td>
<td>11.9 (1.4)</td>
<td>18.1 (2.3)</td>
</tr>
<tr>
<td>Best preimplant PTA</td>
<td>107.6 (11.0)</td>
<td>103.3 (13.3)</td>
<td>108.1 (10.2)</td>
<td>112.3 (7.2)</td>
</tr>
<tr>
<td>Onset of deafness (months)</td>
<td>2.8 (7.7)</td>
<td>0.0 (0.0)</td>
<td>2.8 (8.4)</td>
<td>6.6 (9.8)</td>
</tr>
<tr>
<td>Age at implantation (months)</td>
<td>35.8 (19.8)</td>
<td>27.9 (14.2)</td>
<td>35.2 (20.8)</td>
<td>47.7 (19.4)</td>
</tr>
<tr>
<td>Deafness duration (months)</td>
<td>33.0 (18.6)</td>
<td>27.9 (14.2)</td>
<td>32.4 (17.5)</td>
<td>41.1 (24.0)</td>
</tr>
<tr>
<td>Age at testing (years)</td>
<td>15.1 (4.9)</td>
<td>10.4 (1.7)</td>
<td>14.9 (2.6)</td>
<td>22.0 (3.6)</td>
</tr>
<tr>
<td>Communication modea</td>
<td>4.7 (0.8)</td>
<td>4.5 (0.9)</td>
<td>4.8 (0.7)</td>
<td>4.7 (1.1)</td>
</tr>
<tr>
<td>Incomeb</td>
<td>7.2 (2.5)</td>
<td>7.6 (2.6)</td>
<td>7.4 (2.2)</td>
<td>5.9 (3.1)</td>
</tr>
<tr>
<td>Gender (n; female/male)</td>
<td>28/35</td>
<td>6/13</td>
<td>18/12</td>
<td>4/10</td>
</tr>
<tr>
<td>Bilateral CI/bimodal user (n)</td>
<td>21</td>
<td>11</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Meningitis (n)</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Implant model/processing strategyc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC Clarion—MPS</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ABC Clarion HP—MPS</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ABC Clarion—HiRes</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>CC Nucleus 22—SPEAK</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>CC Nucleus 24—ACE</td>
<td>31</td>
<td>11</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>CC Nucleus 24—SPEAK</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CC Nucleus System 5—ACE</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ME Combi 40++—CIS</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ME Concert—FSP</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ME Sonata—CIS</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. Unless otherwise indicated, values are means accompanied by standard deviations in parentheses. PTA = unaided preimplant pure-tone average (in dB HL); ABC = Advanced Bionics Corporation; MPS = Multiple Pulsatile Stimulation; CC = Cochlear Corporation; SPEAK = Spectral Peak; ACE = Advanced Combination Encoder; ME = Med-El Corporation; CIS = Continuous Interleaved Sampling; FSP = Fine Structure Processing. 

aCommunication mode is coded on a scale from auditory—oral = 6 to mostly sign = 1 (Geers, 2002). 
bIncome was coded on a 1 (< $5,500/year) to 10 (≥ $95,000/year) scale. 
cRefers to the first implanted model or first used strategy. All bilateral CI users were implanted sequentially.

obtained from a large group of naïve undergraduate listeners who were all native speakers of American English. One hundred seventy-one undergraduate students at Indiana University participated in this study for partial course credit in an introductory psychology class or for payment of $10. All participants had NH as assessed by a hearing screening consisting of pure tones presented at 25 dB HL at 500 to 4000 Hz in the right and left ear. Participants also passed an orthographic transcription prescreening task to ensure task compliance and equipment function. None of the listeners reported any prior experience with deaf speakers or individuals who used a CI.

The undergraduate listeners were seated at a computer screen and presented with audio recordings of the McGarr sentences over high-quality headphones. They were instructed to orthographically transcribe what they believed the speaker said. Three undergraduate listeners transcribed each of the 36 utterances from a single CI speaker. One undergraduate listener transcribed each of the 36 utterances from a single NH speaker. Thus, we obtained three separate transcriptions of each CI speaker’s utterance and a single transcription for each NH speaker’s utterance. Sentence intelligibility was assessed in two ways: key word intelligibility and total sentence intelligibility.

**Key word intelligibility** refers to the proportion of key words that were correctly transcribed. The 36 sentences were originally constructed such that one word was designated as the key word and intelligibility was assessed based on the transcription of each key word in each sentence. Each of the 36 sentences was coded as either *intelligible* or *not intelligible* (“1” or “0”) on the basis of whether or not the undergraduate listener correctly transcribed the key word. For each sentence spoken by a CI user, the mean of the three judges was used to create a composite score representing a given speaker’s intelligibility for that sentence. Key word sentence intelligibility was calculated by taking the mean sentence intelligibility across all high- or low-predictability key words averaged over all three sentence lengths.

**Total sentence intelligibility** refers to the proportion of words in each test sentence that a listener correctly transcribed. Again, for each sentence spoken by a CI user, the mean of the three judges was used to create a composite score of the speaker’s intelligibility for that sentence. Total sentence intelligibility was calculated by calculating the mean of all sentences for each speaker. Finer-grained analyses were performed by calculating the mean of all sentences by predictability and sentence length, to yield a mean intelligibility for each sentence length of either high or low predictability.

### Results

The results are presented below in three sections. First, we present the overall speech intelligibility scores for these long-term CI users. We compared the speech intelligibility...
scores of these CI users with those of NH controls, as well as the speech intelligibility scores across the three age cohorts of CI users. Next, we present the demographic factors that are associated with speech intelligibility outcomes. To better understand the sources of variability in speech intelligibility, we built a hierarchical regression model to identify factors that account for unique variability in long-term speech intelligibility scores. Finally, because individuals with low speech intelligibility scores also tend to struggle with other long-term outcome measures, we discuss several other outcome measures that (a) correlate with speech intelligibility and (b) were found to be unrelated to speech intelligibility.

**Speech Intelligibility Scores**

Overall, the average speech intelligibility of the 63 CI users was very high. Figure 1 shows a rank-order plot of the speech intelligibility scores, as assessed by total sentence intelligibility, of the CI users and NH controls ordered from least to most intelligible. Examination of the figure shows that most CI users produced highly intelligible speech, and the variability observed in this sample is largely driven by the bottom third of the distribution. The rank-order plot for speech intelligibility using key word percentage correct was almost identical to the pattern shown in Figure 1, so it will not be presented here. Despite this overall high intelligibility, CI users’ average speech intelligibility was lower than the speech intelligibility scores obtained from the NH controls, as measured both by key word intelligibility scores, $t(124) = 5.40, p < .001$, and total sentence intelligibility scores, $t(124) = 5.27, p < .001$. However, the top third of the CI users did not exhibit lower intelligibility than the NH controls, $t(82) = 0.62, p < .6$, suggesting that the group differences were largely driven by the greater variability among CI users, especially in the lower-performing CI users.

Finer-grained analyses revealed differences in intelligibility between high- and low-predictability sentences and sentences of different lengths. A two-way, repeated-measures analysis of variance showed that both key word predictability, $F(1, 62) = 14.85, p < .01$, and sentence length, $F(2, 62) = 4.05, p < .05$, affected key word intelligibility. Sentences that contained high-predictability key words were more intelligible than sentences that contained low-predictability key words. Follow-up analyses showed that the main effect of sentence length reflects higher key word intelligibility only for five-syllable sentences rather than seven-syllable sentences, $t(62) = 2.62, p < .05$. We had no a priori reason to predict this difference, and a significant difference only between five- and seven-syllable sentences is not consistent with any post hoc hypothesis that we might impose on the data, so all subsequent analyses will present data averaged over the three McGarr sentence lengths.

A two-way, repeated-measures analysis of variance also showed that total sentence intelligibility was affected by both key word predictability, $F(1, 62) = 9.88, p < .01$, and sentence length, $F(2, 62) = 9.68, p < .001$. A significant interaction was also found between these two factors, $F(1, 62) = 9.00, p < .001$. Again, sentences with high-predictability key words were more intelligible than sentences with low-predictability key words. Follow-up analyses of the main effect of sentence length and sentence length by key word interaction showed that for high-predictability sentences, seven-syllable sentences were more intelligible than either three-syllable, $t(62) = 4.07, p < .001$, or five-syllable, $t(62) = 4.85, p < .001$, sentences. However, low-predictability, five-syllable sentences were less intelligible than either three-syllable, $t(62) = 3.44, p < .01$, or seven-syllable, $t(62) = 2.42, p < .05$, sentences. As with the sentence length effects reported for the measures of key word intelligibility, we had no a priori hypotheses that predicted this pattern of data, nor do we have any post hoc hypotheses that would predict this pattern. We suspect that this pattern of data may reflect more subtle differences between the McGarr test items (perhaps word frequencies, lexical density, or transitional probabilities between the words of the sentences or phonological properties of the sentences), or this may be a purely spurious result. Further, we do not have any indication as to whether these sentence length effects originate with speakers or with the undergraduate judges. For these reasons, all subsequent analyses averaged over the three sentence lengths and comparison data are presented for only high- and low-predictability sentences.

Tables 2 and 3 show the means, standard deviations, and ranges of the speech intelligibility scores for both the CI users and NH controls, for both key word intelligibility (see Table 2) and total intelligibility (see Table 3) of high- and low-predictability sentences. Again, these descriptive statistics show little difference in the maximum scores attained by CI users and NH controls and suggest that group differences reflect variability among CI users, as evidenced by lower minimum scores and higher standard deviations. Based on key word intelligibility, both CI users and NH controls produced more intelligible high-predictability sentences than low-predictability sentences: CI users, $t(62) = 3.85, p < .001$; NH controls: $t(62) = 2.21, p < .05$. A similar pattern was found for total sentence intelligibility, although only CI users produced more intelligible speech on sentences containing high-predictability key words because of ceiling effects among

---

**Figure 1.** Rank order plot of cochlear implant users’ (filled bars) and normal-hearing controls’ (open bars) total McGarr sentence intelligibility scores.
the NH controls: CI users, $t(62) = 3.14, p < .01$; NH controls, $t(62) = 1.42, p > .1$.

These speech intelligibility findings replicate earlier studies showing effects of key word predictability in McGarr sentences with CI users (Tobey et al., 2003, 2011). An important caveat to the interpretation of these context effects is that, as with the sentence length effects, we cannot identify the precise locus of these effects. Context effects could reflect a facilitatory effect of high-predictable context on the production of these sentences by the CI users, or these effects could reflect a facilitatory effect on the transcription of these sentences. These possible loci are not mutually exclusive, and both may play a role in the observed context effects. However, the present study cannot distinguish between these two interpretations.

In all subsequent analyses, we used total sentence intelligibility as the measure of speech intelligibility, rather than key word intelligibility used in previous studies. We used only one score for all of the analyses reported below because (a) key word intelligibility and total sentence intelligibility are highly correlated ($r = .98$) and (b) to the extent that these different scoring procedures reflect different constructs, total sentence intelligibility encompasses intelligibility of all words of the sentence, not just the key words, and this measure may be a more stable measure of speakers’ speech intelligibility.

### Speech Intelligibility by Age Cohort

Given the wide range of age of implantation among the CI users in this sample, we were also able to investigate speech intelligibility across the age cohorts of CI users. CI users were divided into age cohorts based on their duration of CI use, which in this sample was strongly correlated with the chronological year in which these CI users were implanted. As previously reported in Ruffin et al. (2013), several systematic differences were observed in individuals who have been implanted at different times. This analysis is useful and informative because of differences over the past 25 years in CI device characteristics, as well as CI implantation criteria and guidelines for age of implantation. Thus, findings on long-term outcomes of an individual implanted in the late 1980s and early 1990s may not be comparable to outcomes of a deaf child who is implanted now. Here, we report long-term speech intelligibility outcomes for the same three age cohorts that Ruffin et al. (2013) identified in their earlier study of speech perception and language outcome measures. This analysis allows us to better understand the parallels between speech perception and speech production that exist across different age cohorts of CI users and the demographic, device, and implantation candidacy criteria factors that might contribute to variability in both of these outcome measures.

Table 4 summarizes descriptive measures of speech intelligibility for the three age cohorts of CI users. The distribution of speech intelligibility scores across the three age cohorts as well as NH controls is displayed graphically in Figure 2. Despite similarities in the overall means and standard deviations, the underlying distributions of the three age cohorts of CI users varied. The shape of the distribution of the speech intelligibility scores also varied, showing a greater degree of skew for the two younger age cohorts. Thus, for the younger CI users, the distributions contain fewer lower-performing outliers, whereas the oldest cohort was more normally distributed. The greater skew for younger cohorts is evident in the percentage of CI users who were less than 80% intelligible. As shown in Table 4, a greater percentage of CI users in the older cohorts were less than 80% intelligible, again suggesting that despite similar mean speech intelligibility scores, younger cohorts are better characterized by overall high speech intelligibility scores, with a handful of less intelligible CI users exerting greater leverage on the distribution.

Similar to the results reported in Ruffin et al. (2013), we also found a larger number of lower-performing CI users in the oldest cohort. Ruffin et al. (2013) reported poorer speech perception skills in CI users who had been using their implants for more than 15 years. We replicate this finding here with measures of speech intelligibility. Specifically, the oldest cohort of CI users produced less intelligible speech overall than either of the two younger cohorts despite longer durations of CI use. This pattern suggests important parallels in outcomes of speech perception and production abilities: CI users who are more likely to struggle with speech perception are also more likely to have poorer speech production skills (see Pisoni, Cleary, Geers, & Tobey, 1999).

Figure 3 illustrates the relation between speech intelligibility and speech perception, along with the cohort effects. This figure shows that speech perception is related to speech
intelligibility (solid line); individuals with lower speech intelligibility scores also have poorer speech perception scores. Figure 3 also illustrates the observed cohort effects. CI users with poorer speech intelligibility tend to be older (dashed line) and have used their CI longer (dotted line), as illustrated by the bumps on the left side of the graph. Despite having used their implants longer, older CI users are more likely to exhibit poorer speech intelligibility. This complementary relationship between speech perception and speech production is consistent with the long-term language outcome data reported in Ruffin et al. (2013). We discuss the possible underlying factors of the observed cohort effects in the following section.

Sources of Variability in Speech Intelligibility

One goal of the present research is to better understand the sources of variability in speech intelligibility scores. Given the long-term outcome nature of this study, CI users’ demographics and several other measures were used to predict speech intelligibility. Attributes that contribute to poorer speech intelligibility may shed light on the underlying factors that contribute to the variability in speech intelligibility observed in the CI users in this sample.

Demographics. A great deal of research on speech and language outcomes of CI users has focused on explaining the enormous variability that is present in this clinical population. Although a large percentage of the variability in speech and language outcomes still remains unexplained, several demographic factors are known to be associated with speech and language outcomes in both CI and NH populations (e.g., Geers et al., 2011; Geers & Sedey, 2011; Pisoni et al., 1999). Our analyses replicated many of those findings. To better understand the unique contributions of demographic variables that may account for variability in the speech intelligibility scores, we used a multiple regression model that included all variables entered in a stepwise manner that are hypothesized to contribute to language outcomes.

Previous work suggests that age of CI implantation is an important predictor of speech and language outcomes (Archbold et al., 2008; Fryauf-Bertschy, Tyler, Kelsay, Gantz, & Woodworth, 1997; Habib et al., 2010; Niparko et al., 2010; Nikolopoulos, O’Donoghue, & Archbold, 1999; Table 4. Descriptive statistics of speech intelligibility (total sentence intelligibility) by age cohort.

<table>
<thead>
<tr>
<th>Cohort no.</th>
<th>n</th>
<th>Years of CI Use</th>
<th>M</th>
<th>SD</th>
<th>Skew</th>
<th>Percentage &lt; 80% Intelligible</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>7–9</td>
<td>89.0</td>
<td>9.2</td>
<td>−3.1</td>
<td>5.3</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>10–14</td>
<td>88.3</td>
<td>11.4</td>
<td>−2.5</td>
<td>16.7</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>15+</td>
<td>87.2</td>
<td>12.1</td>
<td>−1.6</td>
<td>21.4</td>
</tr>
</tbody>
</table>

Figure 2. Distributions of overall speech intelligibility percentages by age cohort. Open circles refer to an individual cochlear implant user’s score. Black bars refer to group means. Age Cohort 1 consisted of individuals who had used their implant for 7 to 9 years, Age Cohort 2 consisted of individuals who had used their implant for 10 to 14 years. Age Cohort 3 consisted of individuals who had used their implant for more than 15 years.

Figure 3. Chronological age, years of cochlear implant (CI) use, age of CI implantation, and speech perception composite by speech intelligibility cumulative sample. The speech perception composite score represents the average of normalized scores of the Lexical Neighborhood Task (LNT; Kirk, Pisoni, & Osberger, 1995), Auditory-Visual Neighborhood Sentence Test (AVLNST; Holt, Kirk, Pisoni, Burckhartzmeyer, & Lin, 2005), and Hearing in Noise Test Sentences for Children (HINT-C; Nilsson et al., 1996) tasks.
Osberger et al., 1993; Ruffin et al., 2013; Peng et al., 2004; Sharma, Dorman, & Spahr, 2002; Tye-Murray, Spencer, & Woodworth, 1995), so we first included this predictor into our regression model (see Table 5). Not surprisingly, age of implantation is a significant predictor of long-term speech intelligibility scores. This variable accounts for about 10% of the variability in the speech intelligibility scores.

The results of the age cohort analysis previously reported by Ruffin et al. (2013) found that chronological age and duration of CI use are associated with poorer speech and language outcomes. Consistent with their findings, we found negative correlations between speech intelligibility and both chronological age and duration of CI use (chronological age: \( r = -0.29, p < .05 \), shown in Figure 4; duration of implant use: \( r = -0.24, p = 0.057 \)). However, neither of these variables predicted any additional variability in speech intelligibility after age of implantation was included in the regression model. In fact, we found no further effect of age cohorts once age of implantation was added into the model. The negative correlations observed between chronological age and duration of CI use and speech intelligibility can be accounted for by the older average age of CI implantation in older individuals in the present sample, not by any additional unique variance associated with chronological age at test or duration of CI use, or even device characteristics. We do not claim that device characteristics have no impact on speech and language outcomes; in fact, there is evidence that they do (Lenarz, Joseph, Sönmez, Büchner, & Lenarz, 2011; Peng et al., 2004). We merely failed to find any evidence that device characteristics affected speech intelligibility in the present data set. Perhaps with a larger data set we may have found an effect of device characteristics on speech intelligibility.

Next, we added additional factors to the regression model that reflect the amount of spoken language input that a CI user experiences. The amount of speech input that a child receives has been shown to be a strong predictor of language outcomes in both hearing-impaired (DesJardin, & Eisenberg, 2007; Niparko et al., 2010) and NH, typically developing children (Hart & Risley, 1995; Huttenlocher, Haight, Bryk,

Table 5. Hierarchical regression model predicting speech intelligibility with communication mode, family income, and age of implantation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( p )</th>
<th>( \eta )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at implantation</td>
<td>-0.300</td>
<td>-2.265</td>
<td>&lt; .05</td>
<td>-0.300</td>
<td>.090</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at implantation</td>
<td>-0.241</td>
<td>-2.373</td>
<td>&lt; .05</td>
<td>-0.318</td>
<td></td>
</tr>
<tr>
<td>Communication mode</td>
<td>0.547</td>
<td>5.387</td>
<td>&lt; .001</td>
<td>0.606</td>
<td></td>
</tr>
<tr>
<td>Family income</td>
<td>0.286</td>
<td>2.819</td>
<td>&lt; .01</td>
<td>0.370</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** When chronological age or duration of cochlear implant use is added to Model 2, it fails to account for a significant amount of variance (chronological age: \( \beta = -0.147, t = -0.95, p > .3 \); duration of use: \( \beta = -0.116, t = -0.949, p > .3 \); age cohorts: \( \beta = 0.057, t = 0.43, p > .6 \)).

Two measures of speech input available for the CI users in this sample were communication mode and family income. **Communication mode** refers to the degree to which a CI user relies on signed versus auditory–oral communication (Geers et al., 2011; Geers & Moog, 1989). CI users who are exposed to more auditory–oral spoken language in their language-learning environment will likely have heard more spoken language and engaged in more expressive language skills with their caretakers. Indeed, previous research shows that a greater reliance on auditory–oral communication is associated with better language outcomes (Geers, Brenner, & Davidson, 2003; Geers & Sedey, 2011; Miyamoto et al., 1994; Seltzer, & Lyons, 1991; Weisleder & Fernald, 2014). Replicating this effect for speech intelligibility would add additional support for the important role of speech and language input as a predictor of language outcomes and document similarities in the foundational learning processes underlying language development in both CI users and NH populations.

Two measures of speech input available for the CI users in this sample were communication mode and family income. **Communication mode** refers to the degree to which a CI user relies on signed versus auditory–oral communication (Geers et al., 2011; Geers & Moog, 1989). CI users who are exposed to more auditory–oral spoken language in their language-learning environment will likely have heard more spoken language and engaged in more expressive language skills with their caretakers. Indeed, previous research shows that a greater reliance on auditory–oral communication is associated with better language outcomes (Geers, Brenner, & Davidson, 2003; Geers & Sedey, 2011; Miyamoto et al., 1994; Seltzer, & Lyons, 1991; Weisleder & Fernald, 2014). Replicating this effect for speech intelligibility would add additional support for the important role of speech and language input as a predictor of language outcomes and document similarities in the foundational learning processes underlying language development in both CI users and NH populations.

When included after age of implantation in the hierarchical regression model, both communication mode and family income accounted for additional unique and significant variance in speech intelligibility scores. Greater reliance on auditory–oral communication and higher family income was associated with higher speech intelligibility scores in this long-term sample of CI users. In fact, when these two factors indexing language input are put into a regression
model together, they jointly accounted for a little more than 40% of the variability in speech intelligibility (see Table 5). This finding is consistent with earlier research on both CI users and NH children, suggesting that the amount of speech input that a child receives is an important predictor of language outcomes. Thus, one of the most important contributors to language outcomes in CI users is exactly the same contributor to language outcomes in NH children: amount of language input that they receive in their language-learning environment (Hart & Risley, 1995; Huttenlocher et al., 1991; Weisleder & Fernald, 2014).

None of the remaining demographic factors predicted variability in the speech intelligibility scores. We failed to find any effects of gender in our data set, as was reported in Tobey et al. (2011). Furthermore, age of onset of deafness, preimplant residual hearing, and bilateral/bimodal use of CIs did not predict any additional variability in the speech intelligibility scores beyond age of implantation and amount of spoken language experience that the child received. Again, we do not intend to argue that these factors are unrelated in principle to speech and language outcomes; we merely failed to find effects of these variables in our sample.

**Correlations With Other Long-Term Outcomes**

It is clear that there is a great deal of variability among speech intelligibility scores. We next wanted to better understand the extent to which the variability observed in speech intelligibility was associated with the variability in other long-term outcome measures we had on these CI users. In Table 6, we correlated speech intelligibility scores with several other long-term speech and language outcome variables as well as with other nonlinguistic outcome measures.

An important pattern observed in Table 6 is that the other cognitive outcome measures that correlate with speech intelligibility scores all have a linguistic component. Individuals with high speech intelligibility also tend to exhibit high scores on other speech-language-related tasks, including sentence perception, vocabulary size, and other language abilities, as well as forward digit span, a processing task in which working memory demands are supported by linguistic ability. Further, speech intelligibility scores were not related to performance on cognitive tasks that do not involve as strong a linguistic component. These tasks include the backward digit span, in which the working memory demands of the task are not consistent with temporal sequencing and forward linearization of language, and nonverbal IQ. Variability in speech intelligibility scores may be influenced by the same factors that affect other speech and language outcome measures because the knowledge and linguistic abilities necessary for good speech intelligibility substantially overlap with the knowledge and linguistic abilities necessary to perform these other language processing tasks. However, we failed to find any relation between speech intelligibility and nonlinguistic cognitive outcomes. This finding suggests that speech intelligibility skills measured in playback experiments are dependent on and possibly influence a variety of other speech and language outcomes, whereas nonlinguistic cognitive outcomes may be unrelated to, or develop at least somewhat independently of, these core foundational language abilities.

**Table 6. Correlations between speech intelligibility scores and other speech, language, and cognitive outcomes.**

<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking duration</td>
<td>57</td>
<td>−.420</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>HINT-Quiet</td>
<td>62</td>
<td>.739</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>PPVT (standard score)</td>
<td>63</td>
<td>.409</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>CELF (standard score)</td>
<td>60</td>
<td>.370</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Forward digit span</td>
<td>63</td>
<td>.342</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Backward digit span</td>
<td>63</td>
<td>−.055</td>
<td>&gt; .6</td>
</tr>
<tr>
<td>Nonverbal IQ</td>
<td>63</td>
<td>.017</td>
<td>&gt; .8</td>
</tr>
</tbody>
</table>

**Note.** HINT-Quiet = Hearing in Noise Test–Quiet; PPVT = Peabody Picture Vocabulary Test; CELF = Clinical Evaluation of Language Fundamentals.

**Discussion**

The results of the present study on speech intelligibility outcomes in a large, unrestricted sample of long-term CI users replicate earlier findings reported by Tobey et al. (2011) in their investigation of long-term speech intelligibility outcomes of CI users. One of the most important findings from this long-term outcome study of speech intelligibility is that the speech intelligibility scores for most of the CI users were, overall, very high. CI users are about 90% intelligible, as assessed by the McGarr Intelligibility Task, with about half the sample of CI users falling within 1 SD of the mean of the NH controls and about two thirds of the sample of CI users falling within five percentage points of the controls. This high level of speech intelligibility suggests that long-term expressive language outcomes of many profoundly deaf individuals (about two thirds of the current sample) who use CIs are, on average, quite good and overlap with the range of intelligibility scores observed for age- and nonverbal IQ-matched NH controls. Although most CI users in this study were well within the range of NH controls on speech intelligibility, about one third of the sample fell below the NH controls. One long-term goal of our research program is to understand the sources of this variability and explain the nature of individual differences in speech and language outcomes in CI users (Beer et al., 2014; Kronenberger et al., 2013; Pisoni et al., 1999; Ruffin et al., 2013; Svirsky et al., 2000).

As in previous studies of speech intelligibility of deaf children with CIs, the present analyses suggest that almost half the variability in speech intelligibility in these long-term CI users can be accounted for by a few global demographic factors: age of implantation, communication mode, and family income, of which the latter two have been shown to be proxies for the amount of speech and language input a child receives (communication mode: Geers et al., 2003; Geers & Sedey, 2011; Miyamoto et al., 1994; Pisoni et al., 1999; Svirsky et al., 2000; family income: Fernald et al., 2012; Hart & Risley, 1995; Hoff, 2003; Weisleder & Fernald, 2014).
These findings are not surprising because they have been identified as important factors that contribute to language learning and language proficiency in NH children (Fernald et al., 2012; Hart & Risley, 1995; Hoff, 2003; Huttenlocher et al., 1991; Weisleder & Fernald, 2014). The fact that these very same experiential factors account for variability in language production outcomes in CI users tested here, and in other studies in the literature, suggest important foundational commonalities between the language learning processes of prelingually implanted CI users and those of NH children. These commonalities in speech perception and production may have important implications for the development of novel interventions and strategies that will ensure optimal language outcomes in all CI users.

The results of the present study also suggest that many of the same factors that underlie variability in speech perception and receptive language skills are also associated with variability in speech intelligibility and expressive language skills. The three factors that we found that contribute to variability in speech intelligibility also contributed to the development of speech perception and language skills. In addition, we uncovered age cohort effects similar to those reported in our previous investigation of long-term speech and language receptive outcomes (Ruffin et al., 2013). These cohort effects reflect the fact that medical and audiological criteria for CI candidacy and age of implantation have changed significantly over the past two decades. The differences observed between the three age cohorts in our sample document clinically important and meaningful differences between samples of CI users of different ages who were implanted at different time periods since CIs first became widely available as a medical intervention for profound hearing loss. Specifically, the higher average age of implantation in the older CI cohort in our sample seems to account for the relationship between age cohort and language outcomes. Ruffin et al. (2013) offered a similar conclusion. They reported that the negative relationship between speech perception and years of CI use can be accounted for by systematic demographic differences, including age of implantation between the oldest and younger cohorts in this long-term sample.

The age cohort effects obtained in the present study as well as in Ruffin et al. (2013) suggest close links between long-term outcomes in speech and language perception and speech production. It is likely that similar underlying causal factors contribute to variability in both domains of spoken language processing. These cohort effects also contribute significantly to our understanding of the factors that underlie CI users’ ability to produce intelligible speech. The present study is one of very few to investigate long-term speech intelligibility outcomes in CI users, and the broad age range of our sample allows us to detect these cohort effects that previous work has not been able to detect because their samples did not allow for these analyses. This ability to detect influences of demographic variables that other studies have not been able to detect represents a significant finding that contributes to our understanding of the development of language abilities in prelingually deaf CI users.

In this sample of long-term CI users, almost 50% of the variability in speech intelligibility could be accounted for by three factors: age of CI implantation, communication mode, and family income. Previous work has shown that each one of these three factors is a reliable proxy for the amount of spoken language input that an NH child receives from his or her caretakers (Fernald et al., 2012; Hart & Risley, 1995; Hoff, 2003; Weisleder & Fernald, 2014). These findings are consistent with a large body of previous work identifying demographic factors that contribute to speech and language outcome variability in CI users as well as research investigating language variability in NH populations (Fernald et al., 2012; Hart & Risley, 1995; Hoff, 2003; Huttenlocher et al., 1991; Weisleder & Fernald, 2014). These factors also suggest that, as with NH children, active language experiences and activities through speaking and listening are extraordinarily important for future language outcomes. It is also possible that the highly degraded and underspecified speech signal that CI users receive through their implants causes CI users to need greater exposure to speech and more repetition of spoken utterances to reliably learn to abstract the same regularities and patterns that NH children learn (Pisoni & Cleary, 2004). The fact that nearly 40% of the variability in speech intelligibility can be accounted for by either communication mode or family income suggests that variability in spoken language experience and basic learning processes may play important roles in explaining the mechanisms underlying a large portion of the observed variability in language outcomes in CI users. This suggests that maximizing the amount of spoken language that a prelingually deaf child is exposed to early on after receiving a CI could be important for optimizing language outcomes in those children. This is exactly the intervention strategy advocated recently by several developmental scientists who study language outcome gaps in CI users (Sacks et al., 2014) as well as family income-based language outcome gaps in NH populations (Hart & Risley, 1995; Leffel & Suskind, 2013; Weisleder & Fernald, 2014).

Another means toward optimizing language outcomes suggested by this analysis is to implant prelingually deaf children with CIs at as young an age as possible (Archbold et al., 2008; Fryauf-Bertschy et al., 1997; Nikolopoulos et al., 1999; Niparko et al., 2010; Osberger et al., 1993; Pisoni et al., 1999; Ruffin et al., 2013; Sharma et al., 2002; Tye-Murray et al., 1995). However, early implantation is not always possible, and in some situations in which etiology or other factors prevent early implantation, the child may be at higher risk for poorer language outcomes. Of course, many of the long-term CI users in this study received their implants later in childhood, and many CI users who received their implants even after 36 months of age exhibited speech intelligibility scores well within the range of scores obtained from the NH controls. In fact, as shown in Table 3, there were even several children who were implanted after 36 months of age who displayed speech intelligibility scores within the normal range. These later-implanted CI users, however, exhibited a higher standard deviation in their speech intelligibility.
scores. The 24 CI users who were implanted after the age of 36 months had a mean sentence intelligibility score of 84.2% and a standard deviation of 14.4%, whereas the 23 earliest implanted CI users (through 24.3 months) had a mean of 92.6 and a standard deviation of only 4.9%. Thus, later age of implantation is not necessarily associated with poorer speech intelligibility, and the difference in mean intelligibility may be better attributed to more variable long-term outcomes among children implanted after the age of 3 years. Maximizing the amount of spoken language that a child is exposed to and active interactions with caregivers may be particularly important for children who receive their CI later in life because although later implantation may put a child at higher risk for poorer language outcomes, it does not guarantee it (Ruffin et al., 2013). This finding has already been incorporated into current clinical practice, as nearly all early-identified deaf children are now implanted before the age of 2 or 3 years (Kral & O’Donoghue, 2010; O’Donoghue & Pisoni, 2014; Osberger, Zimmerman-Phillips, & Koch, 2002).

Cochlear implantation is currently the standard of care for profoundly deaf children. As CIs become more common and children with CIs enter adulthood, it is critically important to understand more about the late effects of CIs after long-term use and to assess the long-term speech and language outcomes of prelingually, profoundly deaf children who have learned spoken language via CIs. Our findings on speech intelligibility are consistent with those of previous studies of long-term speech and language outcomes in CI users and suggest that age of implantation and language exposure are two significant predictors for future language success. Future work should be directed at identifying early risk factors and developing new directions for novel interventions for CI users who show early delays and disturbances in speech and language outcomes.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grants R01DC009581 and T32DC000012, awarded to the third author.

References


Montag et al.: Speech Intelligibility After Long-Term CI Use


