Meta-analysis of Cochlear Implantation Outcomes Evaluated With General Health-related Patient-reported Outcome Measures

*Theodore R. McRackan, *Michael Bauschard, *Jonathan L. Hatch, *Emily Franko-Tobin,
*Harris Richard Droghini, †Craig A. Velozo, *Shaun A. Nguyen, and *Judy R. Dubno

*Department of Otolaryngology–Head and Neck Surgery; and †Division of Occupational Therapy, Medical University of
South Carolina, Charleston, South Carolina

Objective: Determine the change in general health-related quality of life (HRQOL) after cochlear implantation and association with speech recognition.
Study Design: Meta-analysis.
Methods: Search was performed following the PRISMA statement using PubMed, Medline, Scopus, and CINAHL. Studies on adult cochlear implant (CI) patients measuring HRQOL before and after cochlear implantation were included. Standardized mean difference (SMD) for each measure and pooled effects were determined. A meta-analysis of correlations was also performed between all non-disease-specific patient-reported outcome measures (PROMs) and speech recognition after cochlear implantation.
Results: Twenty-two articles met criteria for meta-analysis of HRQOL improvement, but 15 (65%) were excluded due to incomplete statistical reporting. From the seven articles with 274 CI patients that met inclusion criteria, pooled analyses showed a medium positive effect of cochlear implantation on HRQOL (SMD = 0.79). Subset analysis of the HUI-3 measure showed a large effect (SMD = 0.84).

Nine articles with 550 CI patients met inclusion criteria for meta-analysis of correlations between non-disease specific PROMs and speech recognition after cochlear implantation. Pooled analysis showed a low correlation between non-disease-specific PROMs and word recognition in quiet ($r = 0.35$), sentence recognition in quiet ($r = 0.40$), and sentence recognition in noise ($r = 0.32$).
Conclusion: Although regularly used, HRQOL measures are not intended to measure nor do they accurately reflect the complex difficulties facing CI patients. Only a medium positive effect of cochlear implantation on HRQOL was observed along with a low correlation between non-disease-specific PROMs and speech recognition. The use of such instruments in this population may underestimate the benefit of cochlear implantation. **Key Words:** Cochlear implant—Cochlear implantation—Patient reported outcome measure—Quality of life—Speech recognition.

Otol Neurotol **39:**29–36, 2018.

Cochlear implantation is the gold standard for treatment for individuals with severe to profound bilateral sensorineural hearing loss. With rising health care costs, much focus has been recently placed on the effectiveness of treatments, especially surgical procedures such as cochlear implantation. Since 2002, the National Institutes of Health has been focusing considerable attention towards patient-reported outcomes to ensure treatments are improving outcomes that are important to patients and, thus, providing significant benefit (1).

Quality of life (QOL) patient-reported outcome measures (PROMs) are commonly used to determine the impact of an intervention on an individual's life. These PROMs can be subdivided into two major categories—general health and disease-specific. The former are generalizable instruments that are meant to be applied to large, diverse populations to evaluate overall QOL or an individual construct. In contrast, disease-specific instruments are typically validated for a particular population that share a common health condition or disability. General health-related quality of life (HRQOL) PROMs are the most commonly used instruments for economic analysis to determine the cost effectiveness of a particular treatment through measurement of total health (2,3). The importance of using QOL PROMs has been strengthened recently by an increased emphasis by the Center for Medicare and Medicaid Services (CMS) and the Food and Drug Administration (FDA) requirement to report these data (4,5).

In contrast to the growing importance of patient-reported outcomes, open-set speech recognition scores

measured in quiet and in noise continue to be the gold standard for assessing outcomes and benefit in adult CI recipients. In evaluating the literature, there does not appear to be a strong relationship between speech recognition ability and patient self-report (6–9). Reasons for these discrepancies are likely twofold. First, the complex communication, social, and emotional situations that CI users experience may not be fully represented by word or sentence recognition alone. Second, the manner in which cochlear implantation improves QOL likely extends well beyond improvements in speech recognition. Due to the routine use of word and sentence recognition scores in reporting CI outcomes and the increasing importance of reporting QOL PROMs, it is important to systematically determine the extent to which these measures correlate.

Our previous work has shown a very large effect size of cochlear implantation on QOL when measured with hearing and CI-specific PROMs (10). For the current study, we sought to determine how this impact compares when using HRQOL PROMs that are routinely reported in the literature, but were not developed for or validated on individuals with hearing loss. This is an important comparison as the health usefulness and economic analyses of cochlear implantation are typically calculated using HRQOL PROMs. This study will help determine their usefulness in the CI population. A second meta-analysis assessed the association of HRQOL PROMs with speech recognition ability.

MATERIALS AND METHODS

Search Methods

Literature search was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (11). The PubMed, Scopus, and OVID/Medline databases were independently searched by two authors for the search terms: ''cochlear implant'', ''cochlear implantation'', ''quality of life'', and ''patient-reported outcome measures''. This search identified 1,281 articles, of which, 591 were unique articles after eliminating duplicates (Fig. 1). These 591 articles were first reviewed by abstract, which eliminated 360 articles. There were 231 articles remaining, which underwent full-text review for inclusion (one article satisfied criteria for both analyses). Disputes regarding the inclusion of a study were mediated with a third author to reach consensus.

Letters to the editor, abstracts, book chapters, case reports, and articles not published or translated in English were excluded. There were no date range limitations on date of publication. Studies with patients younger than 18 years old in the cohort were excluded. Analyzing PROMs in pediatric patients involves many different factors than adults, thus we limited the scope of this study to adult patients (12).

The last time point available for each study was used for data collection. Data reported in graphical plots were not extracted for meta-analysis unless numerical data were available. We attempted to contact authors if we could not extract complete data from their publication; some authors provided additional data to allow inclusion of their study in our analysis.

Data Extraction

When selecting articles for meta-analysis of HRQOL improvement, studies meeting the following inclusion criteria
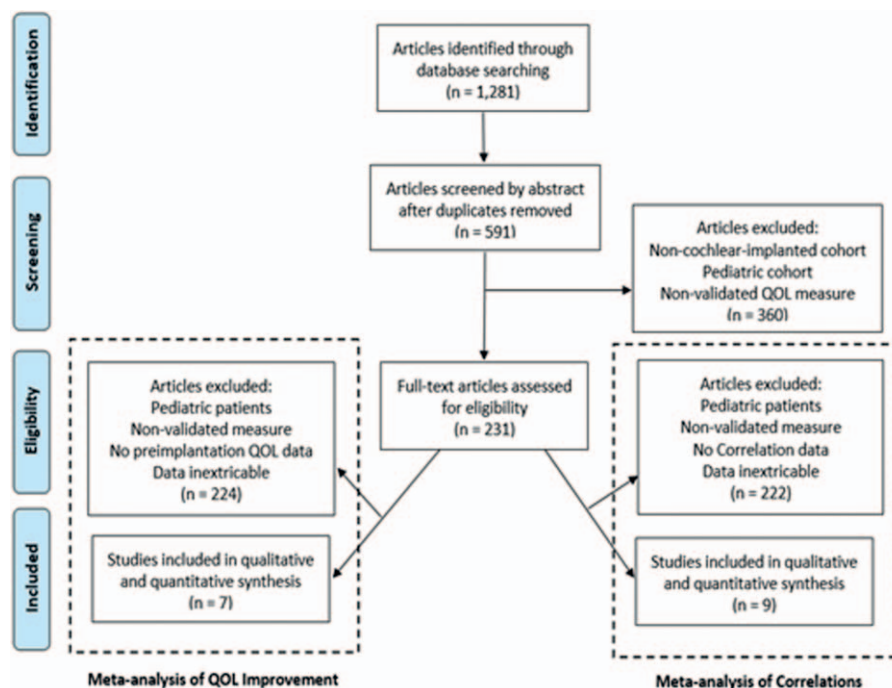

FIG. 1. Literature review process flowchart. Literature review process utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) search method.

Otology & Neurotology, Vol. 39, No. 1, 2018

were ultimately selected: assessment of HRQOL in an adult CI cohort before and after surgery (or in a post-treatment cohort versus a control cohort); sample size, mean, and standard deviation available for PROM data; and follow-up of at least 6 months. Two authors independently obtained data from articles including: year of publication, author, number of patients, patient demographics, and HRQOL PROM scores.

When selecting articles for meta-analysis of correlations, studies meeting the following inclusion criteria were used: correlation values of speech recognition scores versus any general PROM in an adult cohort after cochlear implantation; complete data available (sample size and Pearson or Spearman correlation values); and postoperative follow-up of at least 3 months. Two authors independently obtained data from articles including: author, year of publication, number of patients, patient demographics, speech recognition measure used, and correlation values. Studies included in this analysis did not require reporting of preimplantation PROM data. Rather, this analysis aimed to evaluate the correlation of speech recognition ability and each respective PROM at the latest available time point after implantation.

Data reported in graphical plots were not extracted unless numerical values were published. We contacted authors to obtain complete details of results in the event of incomplete data to allow inclusion of their study. PROMs using a reverse scale (reduction in scores represent improved QOL) had standardized mean difference (SMD) values multiplied by negative one for analysis. Level of evidence for each selected article was evaluated with the Oxford Center for Evidence-Based Medicine (13).

Meta-analysis of HRQOL Improvement Statistical Method

Meta-analysis of included studies evaluating the impact of cochlear implantation on HRQOL with a continuous measure (comparison of means and standard deviations between pre-implantation and post-implantation) was performed with Cochrane Review Manager (RevMan) version 5.3 (Nordic Cochrane Centre, Cochrane Collaboration, 2011, Copenhagen, Denmark). Fixed-effects and random-effects models were used in this study. Under the fixed-effects model, it is assumed that all studies come from a common population, and that the effect size as measured through SMD is not significantly different among the different trials. This assumption is tested by the heterogeneity test or I^2 statistic. If this test yields a low probability value ($p < 0.05$), then there is a high likelihood the fixed-effects model is invalid and the random-effects model is more appropriate. The random-effects model incorporates both the random variation within the studies and the variation

between the different studies (14). The random-effects model provides a more conservative estimate (i.e., a wider confidence interval), but the results from the two models typically agree when there is no heterogeneity. When heterogeneity was present, the random-effects model was the preferred model.

Potential publication bias was evaluated by visual inspection of the funnel plot and Egger's regression test, which statistically examines the asymmetry of the funnel plot (15). For this analysis, the null hypothesis was that there is no difference between preimplantation and postimplantation PROM scores. Analysis was performed on two subsets of PROMs: the Health Utilities Index 3 (HUI-3) HRQOL PROM and other HRQOL PROMs. Data are presented as SMD (95% confidence interval). The following thresholds were used for subjective assessment of effect size: 0.2—small effect, 0.5—medium effect, and 0.8—large effect (16).

Meta-analysis of Correlations Statistical Methods

A meta-analysis of correlations was performed for correlations between speech recognition and HRQOL PROMs using MedCalc 17.2 (MedCalc Software, Ostend, Belgium). Heterogeneity testing was performed as previously described. Each study was weighted according to the number of patients included. MedCalc uses the Hedges–Olkin method for calculating the weighted summary correlation coefficient under the fixed-effects model, using a Fisher Z transformation of the correlation coefficients (17). Under the random-effects model, the heterogeneity statistic is incorporated to calculate the summary correlation coefficient (19). For this analysis, the null hypothesis was that speech recognition ability and PROM scores do not correlate. The following thresholds were used for subjective assessment of correlation values (r): 0 to 0.3, negligible; 0.3 to 0.5, low; 0.5 to 0.7, medium; 0.7 to 0.9, high; 0.9 to 1.0, very high (18,19).

RESULTS

Meta-analysis of HRQOL Improvement

Twenty-two articles met many of the inclusion criteria for outcomes analysis, but 15 (68%) were excluded due to incomplete statistical reporting, leaving seven articles for analysis (Table 1). From the seven articles, 274 patients were included in the analysis with 100% of patients having published sex data (46% men, 54% women). The mean age across all study cohorts ranged from 49 to 62 years. Pooled analyses showed a medium positive effect of cochlear implantation on HRQOL (SMD = 0.79 [0.39–1.19]). Subset analysis of the

TABLE 1. *Articles included in meta-analysis of HRQOL improvement*

Article	Level of Evidence	Cohort Age Mean ± SD (Range)	Male%/ Female%	Follow-Up Time (mo)
Damen et al. (6)	3	49.6 ± 10.9	54/46	≥12
Arnoldner et al. (20)	4	62 (18–89)	51/49	≥11
Klop et al. (21)	4	54.7 ± 15.7	34/66	12
Palmer et al. (22)	3	56.0 ± 15.4	46/54	12
Hawthorne et al. (25)	4	49 ± 13	47/53	6
Mo et al. (24)	4	57.6 ± 14.5 (28–82)	44/56	12
Krabbe et al. (23)	4	51 ± 16	47/53	≥12

Articles satisfying inclusion criteria for meta-analysis of HRQOL improvement. Level of evidence, cohort age mean, SD, and range (if available), male/female cohort ratio, and follow-up time in months. HRQOL indicates health-related quality of life; SD, standard deviation.
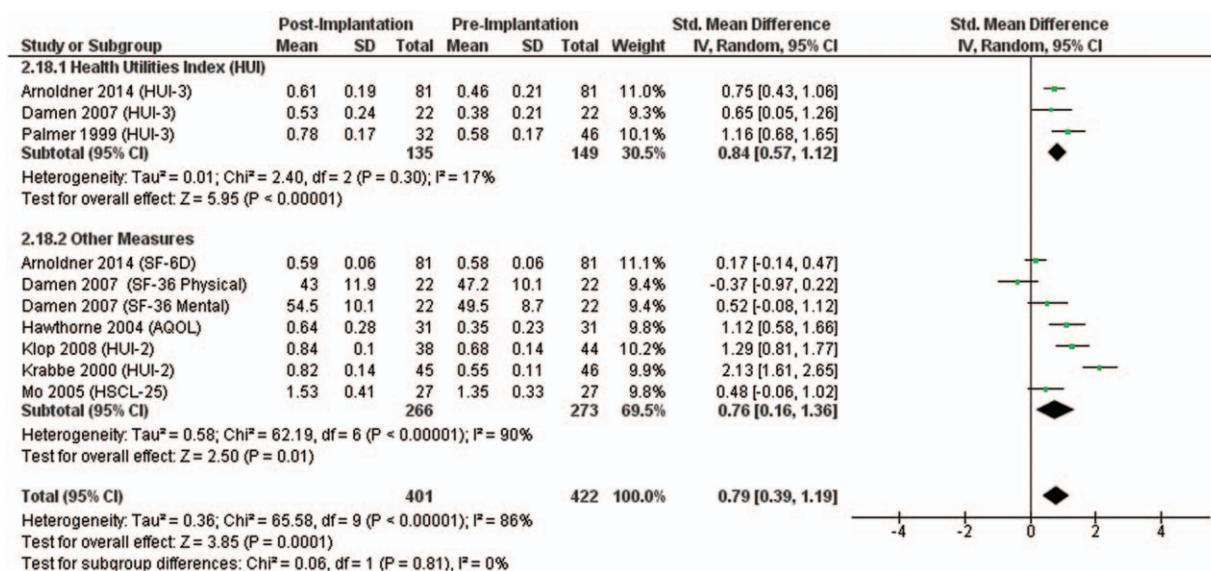
FIG. 2. Forest plot of meta-analysis of HRQOL improvement. Forest plot of HRQOL PROMs including subset analysis of Health Utilities Index 3 (HUI-3) and other HRQOL measures. AQOL indicates assessment of quality of life; HSCL-25, Hopkins symptom checklist-25; IV, inverse variance; SF-36, 36-item short form; SF-6D, 6-item descriptive system.

HUI-3 showed a large positive effect (SMD = 0.84 [0.57 −1.12]), whereas subset analysis of other HRQOL PROMs showed a medium effect (SMD = 0.76 [0.16−1.36]) (Fig. 2). To investigate the presence of publication bias, inspection of the funnel plot of effects calculated from individual studies was performed. According to funnel plots and the Egger's test, there was no indication of publication bias ($p = 0.272$) among the set of studies included in this meta-analysis.

Five studies in our analysis used the HUI PROM (6,20−23). Both the HUI-2 and HUI-3 are used to evaluate QOL (3). The HUI-2 classification system uses seven domains (sensation, mobility, emotion, cognitive, self-care, pain, fertility) whereas the HUI-3 classification uses nine alternative domains (vision, hearing, speech, walking, dexterity, emotion, cognition, pain). The HUI-3 hearing domain directly ascertains hearing function by asking patients if they can perceive what others are saying in various settings. For example, the HUI-3 asks patients if they are: ''Able to hear what is said in a conversation with one other person in a quiet room, without a hearing aid, but unable to hear what is said in a group conversation with at least three other people even with a hearing aid''. The HUI-2 sensation domain collects similar but more basic information regarding hearing function while combining it with other senses. For example, the HUI-2 asks if patients can ''See, hear, or speak with limitations even with equipment''. All five included studies utilizing the HUI found significant improvement in the hearing or sensation domain for CI patients. Only one study using the HIU-3 found improvement in the speech domain for CI patients (22). Additionally, two of the five studies of utilizing either HUI found significant improvement in the emotion domain for CI patients (20,23). No other domains except

those mentioned above showed significant improvement in the five included studies utilizing the HUI.

Damen et al. (6) used the 36-item short form (SF-36) and found no significant change in the physical summary or mental summary for CI patients. Specifically, significant improvement for CI patients was found only in the mental health subdomain of the mental summary. Arnoldner et al. (20) used the SF-6D conversion measure to find a total usefulness score of the SF-36 measure; this study found a significant improvement in the mental health and social functioning subdomains for CI patients, but overall there was no detectable effect of cochlear implantation on HRQOL (SMD = 0.17 [−0.14−0.47]).

Mo et al. (24) used the 25-question Hopkins Symptom Checklist (HSCL-25), a HRQOL PROM that gauges anxiety and depression in patients, finding nearly twice the improvement in the depression domain compared with the anxiety domain for CI patients but overall had no detectable effect of cochlear implantation on HRQOL (SMD = 0.48 [−0.06−1.02]). The Assessment of QOL (AQOL) HRQOL PROM was used by one study (25) finding a large positive effect on HRQOL of cochlear implantation (SMD = 1.12 [0.58−1.66]); this 15-question measure has one question specifically asking regarding hearing function, whereas other questions ask regarding associated ailments of poor health including emotional and social problems.

Meta-analysis of Correlations

Nine articles met criteria for inclusion in this analysis with none being excluded due to incomplete statistical reporting (Table 2) (6,7,9,26−31). A total of 550 patients were included with 41% of the patients having published sex data (45% men, 55% women). The mean age across all studies ranged from 36.8 to 63.4 years.

TABLE 2. *Articles included in meta-analysis of correlations*

Article	Level of Evidence	Speech Recognition Measure	Cohort Age Mean ± SD (Range)	Male %/ Female %	Follow-Up Time (Months)
Calvino et al. (27)	4	NS	52.8 ± 14.0	45/55	≥6
Damen et al. (6)	3	Words in quiet: Antwerp-Nijmegen, NVA	49.6 ± 10.9	54/46	≥12
Francis et al. (9)	4	Words in quiet: NS Sentences in quiet: CID	63.4 ± 8.6 (50 - 80)	NA	≥6
Hirschfelder et al. (31)	4	Words in quiet: Freiburg monosyllables Sentences in noise: HSM noise	50.2 ± 14.4 (21–72)	36/64	≥12
Knutson et al. (26)	4	Words in quiet: NU-6 Sentences in quiet: Iowa Sentence Test	51.8 ± 14.3 (24–70)	46/54	54
Kumar et al. (7)	4	Sentences in quiet: BKB/CUNY	36.8 (18–68)	NA	12
Sanchez-Cuadrado et al. (28)	4	NS	60 (24–85)	46/54	≥6
Vermeire et al. (30)	4	Words in Quiet: NVA	58	NA	≥4
Vermeire et al. (29)	4	Words in Quiet: NVA	62 (40–78)	NA	≥3

Articles satisfying inclusion criteria for meta-analysis of correlations. Level of evidence, speech recognition measure used; cohort age (mean, standard deviation, and range); male/female percentages; and follow-up time in months. BKB indicates Bamford–Kowal–Bench; CID, Central Institute for the Deaf; CUNY, City University of New York; HSM, Hochmair Schulz Moser; NS, not specified; NU-6, Northwestern University Auditory Test Number Six; NVA, Dutch Audiological Society.

Low pooled correlations were found between HRQOL PROMs and the three categories of speech recognition testing: word recognition in quiet ($r = 0.35$ [0.25–0.45]), sentence recognition in quiet ($r = 0.40$ [0.30–0.49]), and sentence recognition in noise ($r = 0.32$ [0.19–0.44]) (Table 3, Fig. 3). Subset analysis of results with psychological PROMs showed low correlations with word recognition in quiet ($r = 0.41$ [0.28–0.53]) and sentence recognition in quiet ($r = 0.45$ [0.32–0.56]). Subset analysis of HRQOL PROMs showed low correlations with word recognition in quiet ($r = 0.33$ [0.19–0.46]), sentence recognition in quiet ($r = 0.34$ [0.18–0.48]), and sentence recognition in noise ($r = 0.32$ [0.19–0.44]).

Hirschfelder et al. (31) used the SF-36 and found negligible correlations of the mental ($r = 0.09$) and physical summaries ($r = 0.25$) with word recognition in quiet, a negligible correlation of SF-36 physical summary with sentence recognition in noise ($r = 0.18$), and a low correlation with SF-36 mental summary with sentence recognition in noise ($r = 0.32$). On analysis of the SF-36 subdomains, this study found low correlations between word recognition in quiet and physical functioning ($r = 0.40$) and vitality ($r = 0.44$), and sentence recognition in noise with vitality ($r = 0.50$) and mental health ($r = 0.38$) subdomains. No correlations with speech recognition scores were found with respect to emotional role functioning, social functioning, general health perception, pain, and physical role functioning.

Vermeire et al. (29) found the highest correlations among HRQOL PROMs in our study ($r = 0.83$), correlating the Glasgow Benefit Inventory (GBI) and word recognition in quiet. This finding was an outlier as the three other studies correlating GBI versus word recognition in quiet found correlations of 0.25, 0.28, and 0.34 (28–30). The outlier study also used the smallest sample size (n = 24) resulting in a large confidence interval (0.64–0.92).

Only two other studies found an overall medium correlation or higher ($r \geq 0.50$). Francis et al. (9) found a correlation of 0.55 for both word recognition in quiet

TABLE 3. *Meta-analysis of correlations results*

	r	95% CI	I^2	p
Subtotal: HRQOL				
Word recognition in quiet	0.330	0.191–0.456	64.39%	0.0027
Sentence recognition in quiet	0.335	0.180–0.475	57.53%	0.0949
Sentence recognition in noise	0.323	0.193–0.442	0.00%	0.4163
Subtotal: psychological				
Word recognition in quiet	0.413	0.282–0.530	0.00%	0.7943
Sentence recognition in quiet	0.445	0.316–0.557	0.00%	0.9919
Sentence recognition in noise	NA	NA	NA	NA
Total				
Word recognition in quiet	0.353	0.252–0.445	51.46%	0.0110
Sentence recognition in quiet	0.397	0.299–0.487	0.00%	0.5126
Sentence recognition in noise	0.323	0.193–0.442	0.00%	0.4163

Pooled correlation values (r and 95% confidence interval [CI]) and heterogeneity statistics (I^2 and p) for meta-analysis of correlations. HRQOL indicates health-related quality of life; NA, not applicable.