The World Health Organization's hearing-impairment grading system: an evaluation for unaided communication in age-related hearing loss

Larry E. Humes

Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN, USA

Abstract

This review evaluated the data from five datasets having pure-tone thresholds and functional measures of speech communication from relatively large groups of older adults to evaluate the validity of the proposed new World Health Organisation (WHO) hearing-impairment grading system, referred to here as WHO-proposed. This was a review of studies identified from the literature having both pure-tone audiometry and functional measures of speech communication from relatively large samples of older adults. Three population or population-sample datasets and two clinical datasets were identified with access provided to de-identified data for five of these six studies. As the WHO-proposed hearing-impairment grade progressed from "normal" to "severe" (insufficient data from older adults were available for the "profound" category), each step in this progression led to a significant difference in functional communication relative to the preceding step. Cohen's *d* effect sizes were moderate to very large between each successive step on the WHO-proposed hearing-impairment grading scale, with some exceptions for the step from "normal" to "mild/slight" grades. The WHO-proposed hearing-impairment grading system, recently developed through expert opinion and adopted by WHO, is validated here with evidence from studies of functional communication in older adults.

Keywords

Aging; hearing loss; World Health Organisation (WHO); severity; communication

The WHO HI-grading system

In 1991, the World Health Organisation (WHO) convened an informal working group on the prevention of deafness and hearing-impairment (WHO 1991). One of the key goals was to attempt to standardise the way in which severity of hearing loss was defined. This was critical to the gathering of evidence around the world regarding the prevalence of impaired hearing and deafness. Prior to developing plans to address a healthcare problem, one must be

able to determine the pervasiveness of the problem, as well as population factors that might impact prevalence.

The resulting hearing-impairment grading system, referred to here as WHO HI grade, has since been revised and a new grading system for hearing-impairment proposed (Stevens et al. 2013; Olusanya, Neumann, and Saunders 2014). The WHO-proposed HI grade system appears in Table 1, along with the presumed functional consequences for communication associated with each hearing-impairment grade. The new grade system is referred to here as WHO-proposed HI grade to distinguish it from the older and long-standing WHO HI grade system. The alternative WHO-proposed HI grade system attempts to accomplish three improvements to the original WHO HI grade system: (1) included a hearing-impairment category for unilateral hearing loss (<20 dB HL in the better ear and 35 dB HL in the worse ear for the four-frequency pure-tone average, 4fPTA); (2) lowered the onset of "mild" hearing-impairment grade from 26 to 20 dB HL (better-ear 4fPTA); and (3) added hearing-impairment categories of moderate, moderately severe, severe and profound at subsequent 15 dB HL steps, such that moderate began at 35 dB HL, moderately severe at 50 dB HL, severe at 65 dB HL, and profound at 80 dB HL (Table 1). The WHO-proposed HI grade system was developed by a 21-member Expert Group on Hearing Loss convened as a part of the WHO's Global Burden of Disease project (Stevens et al. 2013).

Armed with the definitions of impaired hearing from either the WHO HI grade or the WHO-proposed HI grade system, epidemiologists could then go about the task of determining the prevalence and incidence of impaired hearing on a world-wide basis. In addition, the influence of key variables, such as age and gender, could also be determined (e.g., Stevens et al. 2013). Once prevalence was established, strategies to reduce or eliminate impaired hearing among the world's population could then be mapped out for future evaluation (e.g., Olusanya, Neumann, and Saunders 2014).

Two considerations are noteworthy regarding both the original WHO HI and the alternative WHO-proposed HI grade systems. First, each was established by expert opinion from the consensus of a panel of 14–21 international experts, rather than via available evidence. When the WHO HI grade system was established in 1991 there was little or no evidence available to develop such a system. Second, each grading system is based on an average measure of hearing loss based on pure-tone thresholds. Since the publication of the original WHO HI grading system in 1991, WHO (2001) redefined the very basis of impairment and its consequences. Within this framework, pure-tone audiometry may be a reasonable metric for the impairment to the bodily structures or functions associated with hearing, but it may not be reflective of the associated impact on a *person's activities* or *participation* in society, two key components of the WHO (2001) model. As shown in Table 1, progressively severe deficits in communication function are implied by the WHO-proposed HI grade system and were in the prior WHO HI grade system as well. Evidence from functional measures of speech communication is needed, however, to establish the validity of the communication deficits associated with the severity of pure-tone hearing loss.

In this review, we attempt to fill this void in our knowledge regarding the validity of the recently adopted WHO-proposed HI grade system. Given the high prevalence of age-related

hearing loss (ARHL) worldwide (WHO 2012), we focussed on the application of the WHO-proposed HI grade system to this burgeoning group of adults with impaired hearing. (It should be noted that the WHO HI grade systems, old and proposed, were designed to apply to all ages, from young children through older adults.). We obtained access to various datasets with a reasonable volume of data and which contained not only pure-tone thresholds, but also at least one functional measure of communication performance, for older adults. Given the relative scarcity of such data, we considered both self-report measures of hearing-related function and direct measures of speech-recognition performance as functional measures of communication.

Datasets examined and methods

We have evaluated the application of the WHO-proposed HI grade system to the data from three samples of the general population, two from the United States and one from Australia, and two clinical samples from the Department of Veterans Affairs (VA) Medical Centers in the United States. Two of the general-population datasets were population studies, one of the residents of Beaver Dam, Wisconsin, who participated in the Epidemiology of Hearing Loss Study (EHLS2; Cruickshanks et al. 1998), and one of the residents of the Blue Mountains region west of Sydney, Australia (Golding et al. 2004). The third general-population data-set was from the 2011–2012 National Health and Nutrition Examination Survey (NHANES). NHANES is a survey that combines interviews and physical examinations for a nationally representative sample of about 5000 Americans each year.

The two VA datasets represented reasonably large samples from clinical populations sampled across the US. One study, Wilson (2011), included participants primarily from Upper East Tennessee whereas the other, Williams-Sanchez et al. (2014), included participants from Florida, Tennessee, and California.

For four of the five datasets, all except EHLS2, we were able to obtain access to the de-identified raw data for analyses. For the EHLS2, a collaborator from that study provided the analyses shown below.

For each dataset, we calculated the four-frequency pure-tone average (4fPTA) by obtaining the means of the thresholds at 500, 1000, 2000 and 4000 Hz for each ear, then selected the better-ear 4fPTA for each participant. Ear-specific and better-ear 4fPTAs were used to then assign a WHO-proposed HI grade as follows: (1) normal, 19.50; (2) slight/mild, 19.51–34.5; (3) moderate, 34.51–49.5; (4) moderately severe, 49.51–64.5; (5) severe, 64.51–80.5; and (6) profound, 80.51 dB HL. Those meeting the WHO-proposed definition of unilateral hearing loss, 4fPTA <20 dB HL in the better ear and 35 dB HL in the worse ear, were removed prior to analysis as unilateral hearing-impairment is not a typical characteristic of age-related hearing loss. Across all five studies, fewer than 1.5% of each study sample met this definition of unilateral hearing loss and were excluded from further analyses.

The resulting WHO-proposed HI grade then became the primary independent variable used in ensuing analyses of variance with the dependent variables being the functional measures of hearing available. Prior to examining the effects on functional measures of hearing, age

differences between WHO-proposed HI grades were evaluated. Because we are interested in the effects of the grade of hearing-impairment on communication performance for older adults, age was a covariate in all analyses of variance. Significant effects of WHO-proposed HI grade were followed up by post hoc *t* tests for pairwise comparisons where appropriate. Given the relatively large samples, even arithmetically small differences between WHO-proposed HI grades may be statistically significant. To get a better idea of the practical significance of between-grade differences, we also calculated Cohen's *d* to get an estimate of the effect size as one progressed to each successive step on the WHO-proposed HI grade system.

Results

Population studies or samples

Blue Mountains dataset—The Blue Mountains Hearing Study is a population-based study of hearing loss that has been described in detail elsewhere (Golding et al. 2004). For our purposes, the dataset analysed was comprised of 2899 individuals ranging in age from 49–99 years, 1647 of whom were female. The total of 2899 represents those remaining after elimination of any with unilateral hearing loss. There were low numbers for the two most severe WHO-proposed HI grades; $N = 27$ "severe" and $N = 13$ "profound." These two grades were eliminated from the subsequent analyses, leaving 1425 individuals in the "normal" group, 882 in the "mild/slight" group, 353 in the "moderate" group, and 107 in the "moderately severe" group. These four groups comprised 99.1% of the full dataset.

The Hearing Handicap Inventory for the Elderly, Screening version (HHIE-S; Ventry and Weinstein 1982, 1983) was the lone self-report functional measure of hearing available in the Blue Mountains dataset. Figure 1 provides the means and standard deviations for the HHIE-S for each of the three WHO-proposed HI grades examined. A univariate analysis of variance, controlling for age, found significant effects of WHO-proposed HI grade ([$F_{(3,2762)} = 344.9$, $p < .001$] with all Bonferroni-corrected post hoc pair-wise comparisons being significant ($p < .001$).

Several ear-specific measures of speech-recognition or speech-identification were also obtained with headphones from most of the Blue Mountains participants (Golding et al. 2004). The results for two such measures, both using the Synthetic Sentence Identification with Ipsilateral Competing Message (SSI-ICM; Speaks, Karmen, and Benitez 1967), are provided in Figure 2, with the top panel showing results for a relatively high presentation level and the bottom panel for a moderate presentation level. The message-to-competition ratio used throughout for the closed-set speech-identification SSI-ICM testing was 0 dB. Means and standard deviations are shown for each ear and each WHO-proposed HI grade. The percent-correct scores were transformed into rationalised arcsine units (RAU; Studebaker 1985) prior to analyses. Data for each ear and presentation level were analysed separately via univariate analyses of variance, each adjusted for age. For the right ear, results obtained at both presentation levels showed significant effects of WHO-proposed HI grade [lowest $F_{(3,2163)} = 13.1$, $p < .001$], but the post hoc testing of pair-wise comparisons found that the "normal" and "mild/slight" groups did not differ significantly ($p = 1.0$) for the moderate presentation level. The other pair-wise comparisons for the scores from the right

ear differed significantly ($p < .001$) with performance declining as WHO-proposed HI increased. The pattern of results observed for the analyses of the data from the left ear differed considerably from the right ear, especially at the moderate test level. Although a significant effect of WHO-proposed HI grade was observed [$F_{(3,2162)} = 3.0$, $p < .05$; age adjusted] at this presentation level, the lone pair-wise comparison that even approached significance ($p = .09$) was the contrast between "mild/slight" and "moderate" groups. No other paired comparisons were found to differ significantly for this ear and presentation level. At the higher presentation level in the left ear, SSI scores varied inversely with WHO-proposed HI grade such that higher grades had lower scores. Post hoc paired comparisons at this presentation level showed that all paired comparisons differed significantly except for the difference between the "moderate" and "moderately severe" groups.

Finally, two measures of speech recognition in quiet were obtained using AB monosyllabic words (Travers 1990). Means and standard deviations for the left and right ears on this open-set recognition task appear in Figure 3 for the three WHO-proposed HI grades. The top panel depicts scores obtained at about 90 dB SPL, on average, whereas the bottom panels shows data for an average presentation level of about 70 dB SPL. For both ears and both presentation levels, the effects of WHO-proposed HI grade on open-set speech-recognition in quiet were significant [lowest $F(3,2443) = 127.2$, $p < .001$; age adjusted] and all pair-wise comparisons differed significantly as well ($p < .001$).

Epidemiology of hearing loss study (EHLS2) dataset—The EHLS2 dataset represents the data from a population study of the community of Beaver Dam, Wisconsin (Cruickshanks et al. 1998; Wiley et al. 1998). Again, most of the individuals in this dataset had WHO-proposed HI grades of 0, 1, 2, or 3; "normal," "mild/slight," "moderate," or "moderately severe." These three hearing-impairment groups represented 97–99% of the data available, depending on the dependent measure being examined. For these four WHO-proposed HI grades and after elimination of those with unilateral hearing loss, data were available for 2613 individuals ranging in age from 52.8 to 97.0 years ($M = 68.8$ years; $SD = 9.6$ years), 1515 (58.0%) of whom were female. The WHO-proposed HI grades differed significantly in age and sex and these variables were used as covariates when examining the effects of WHO-proposed HI grade on the communication measures described below.

Figure 4 shows the means and standard deviations for the self-report HHIE-S as a function of WHO-proposed HI grade. A significant effect of WHO-proposed HI grade was observed [$F_{(3,2607)} = 274.8$, $p < .001$] and all post hoc pair-wise comparisons were significant ($p < .05$). Smaller subsets of participants in EHLS2 also completed measures of open-set speech recognition for monosyllables presented under headphones either in quiet ($N = 924$ for left ear and $N = 1666$ for right ear) or noise ($N = 842$ for left ear and $N = 1572$ for right ear). The speech stimuli were NU-6 monosyllables spoken by a female talker and the noise was a competing sentence spoken by a male talker (Wilson et al. 1990). Figure 5 shows the means and standard deviations for the quiet (top) and noise (bottom) conditions. The speech presentation levels were a moderate sensation level (36 dB) relative to the pure-tone threshold at 2000 Hz to minimise the effects of high-frequency inaudibility on performance (Wiley et al. 1998). The effects of WHO-proposed HI grade on performance in quiet and noise were significant for both right and left ears [lowest $F(3,836) = 101.5$, $p < .001$] with

performance decreasing as the severity of the hearing-impairment increased. All pair-wise comparisons of WHO-proposed HI groups differed significantly ($p < .05$) for the word-recognition scores in noise. For word recognition in quiet, only the performance of the "normal" and "slight/mild" groups failed to differ significantly from one another ($p > .05$).

NHANES dataset—The data from the 2011–2012 NHANES survey (CDC 2013) included pure-tone thresholds from both ears and a series of questions about hearing ability in quiet and noise, as well as one assessing frustration with communication problems. As noted above, given a focus on age-related hearing loss in this review, we truncated this dataset by restricting the national sample to those 50 years of age and older. Those with unilateral hearing loss, as defined above, were also eliminated. This resulted in data for 1402 individuals ranging in age from 50 to 69 years ($M = 58.8$ years; $SD = 5.6$ years), 694 (49.5%) of whom were female. Of the 1402 individuals, only 13 were classified as "moderately severe" on the WHO-proposed HI grade system and these data were excluded from analysis. The remaining 1389 (99.1%) were included with 991, 331, and 67, classified as "normal," "mild/slight," and "moderate," respectively.

Two questions in the 2011–2012 NHANES questionnaire were relatively straightforward to evaluate. One (AUQ100) simply asked, "How often do you find it difficult to follow a conversation if there is background noise, for example, when other people are talking, TV or radio is on, or children are playing?" Possible responses (and point values assigned) were: always (1); usually (2); about half the time (3); seldom (4); and never (5). Similarly, question AUQ110 asked, "How often does your hearing cause you to feel frustrated when talking to members of your family or to friends?" The response alternatives were the same as above for AUQ100. The scoring for these two questions was reversed so that lower scores meant less frequent problems.

The general condition of hearing and hearing in quiet were assessed in a series of questions that were pooled here by summing the responses across the relevant set of five questions. In all five questions, responses were assigned points such that lower points indicated less hearing difficulty or better condition of hearing. The questions summed in this case were AUQ054 [general condition of hearing with responses from excellent (1) through Deaf (6)], AUQ060 [hear a whisper across a quiet room; yes (1), no (2); skip to AUQ100 if yes], AUQ070 [hear a normal voice across a quiet room; yes (1), no (2); skip to AUQ100 if yes], AUQ080 [hear a shout from across a quiet room; yes (1), no (2); skip to AUQ100 if yes], and AUQ090 [hear if spoken loudly to in better ear; yes (1), no (2); skip to AUQ100 if yes]. In this way, those with worse hearing would have more "no" responses to each of the questions above and their total score would be higher for listening in quiet than someone with better hearing.

Figure 6 provides the means and standard deviations for each set of questions from the 2011–2012 NHANES dataset as a function of WHO-proposed HI grade. Univariate analyses of variance, age adjusted, found significant effects of WHO-proposed HI grade for all three of the questionnaire scores [lowest $F(2,1385) = 36.5$, $p < .001$) and all post hoc pair-wise group differences were significant ($p < .05$). Self-reported difficulty communicating in quiet

and in noise, as well as the frustration experienced with such difficulty, increased with each increase in WHO-proposed HI grade from "normal" to "moderate."

Clinical samples

VA dataset—In the foregoing review of population studies and samples, it was apparent that the bulk of the data available were for those with WHO-proposed HI grades of "normal" and "slight/mild," with sufficient data available in most case for the "moderate" grade as well. By examining clinical samples of older adults with ARHL, it is likely that we would be able to obtain sufficient data for more severe WHO-proposed HI grades. This proved to be true for the VA dataset from the Wilson (2011). Data were available, after excluding those with unilateral hearing loss as defined by WHO-proposed HI, for 3357 individuals, the majority of which (36.3%) had a "moderate" WHO-proposed HI grade. Only 31 individuals (0.9%), however, had a "severe" WHO-proposed HI grade and these data were excluded, leaving 99.1% of the data intact for subsequent analyses. When segregated by WHO-proposed HI grade, there were 442 (13.2%), 1,098 (32.7%), 1,218 (36.3%), and 568 (16.9%) in the "normal," "mild/slight," "moderate," and "moderately severe" WHO-proposed HI grades, respectively. The ages ranged from 20–93 years, with a mean of 62.3 years ($SD = $ 12.8 years) and 90% of the sample was 46 years of age or older. Gender was not recorded in the dataset, but Wilson (2011) noted that, for the same period over which these data were obtained, the population of the audiologic clinic from which they were obtained included 1.5% who were female.

The top panel of Figure 7 provides the HHIE-S data from this VA dataset as a function of WHO-proposed HI grade. In comparison with the HHIE-S data from the population studies in Figures 1 and 4, the self-reported hearing difficulties for the clinical sample are much greater, as expected. The univariate analysis of variance, controlling for age, revealed a significant effect of WHO-proposed HI grade on HHIE-S scores [$F_{(3, 2725)}$ 161.6, $p < .001$). Post hoc pair-wise Bonferroni-corrected comparisons among the WHO-proposed HI groups indicated that each group differed significantly from the others ($p < .05$) such that perceived hearing handicap increased as WHO-proposed HI increased.

The bottom panel of Figure 7 shows the results obtained for two other self-report items obtained from a subset of the individuals in this dataset ($N = 1133$). The two questions addressed the person's difficulty understanding speech in quiet or noisy situations. For quiet, the question was: "When listening to a conversation in quiet without your hearing aids, how difficult is it for you to understand what the speaker is saying? On a scale of 1–10, 1 means no difficulty understanding and 10 means extreme difficulty." For the noise question, the words "in quiet" above were replaced by "in a noisy background"; all else was the same. There were insufficient data for the "severe" WHO-proposed HI grade for responses to these two questions. The remaining data in the bottom panel of Figure 7 were analysed via two separate univariate analyses of variance, each controlling for age. Significant effects of WHO-proposed HI grade were observed for the quiet condition [$F_{(3,1128)} = 44.2$, $p < .001$; age adjusted] and the noise condition [$F_{(3,1125)} = 53.4$, $p < .001$; age adjusted]. Post hoc pair-wise Bonferroni-corrected comparisons among groups revealed that each group differed significantly ($p < .05$) from the others with perceived difficulty in both quiet and noise

increasing as WHO-proposed HI grade increased. As is apparent in the figure, communication difficulty was greater in noisy backgrounds than in quiet for all three WHO-proposed HI grades.

The Wilson (2011) VA dataset also included scores for open-set word recognition in quiet and in noise. The means and standard deviations for the RAU-transformed percent-correct scores for the quiet testing appear in the top panel of Figure 8. The data were obtained under earphones for the right and left ears separately so the ear-specific WHO-proposed HI grade was used to analyse the data, rather than the better-ear WHO-proposed HI grade which has been used in all prior analyses for self-report measures, such as the HHIE-S. In addition to having ear-specific word-recognition scores, scores were obtained at two presentation levels: (1) low presentation level of 80 dB SPL for those with a three-frequency (500, 1000, 2000 Hz) pure-tone average 40 dB HL or 90 dB SPL for all other participants; and (2) a high presentation level, 24 dB above the low level (either 104 or 114 dB SPL). Four separate univariate analyses of variance were performed, each corrected for age, to examine the effect of WHO-proposed HI grade on word-recognition performance in quiet. For all four measures of word recognition in quiet (left/right × low/high), significant effects of WHO-proposed HI grade were observed [lowest $F_{(4,3004)} = 326.2$, $p < .001$; age adjusted]. For the low presentation level in the right ear, all post hoc pairwise Bonferroni-corrected comparisons among WHO-proposed HI groups were also significant ($p < .001$). For the other three conditions, right ear at the higher presentation level and the left ear at both presentation levels, all pair-wise comparisons between WHO-proposed HI grades were significant ($p < .001$), except for the difference between "normal" and "mild/slight" grades ($p > .10$). In general, as hearing loss severity progressed from "normal" to "severe," word-recognition performance in quiet declined in both ears and for both presentation levels, with several exceptions noted for the difference between "normal" and "mild/slight" grades.

The bottom panel of Figure 8 shows the means and standard deviations for the adaptive Words-in-Noise (WIN; Wilson 2003) test from the Wilson (2011) dataset. The words used in the WIN are a subset of the same monosyllabic words used in quiet; those word stimuli found to yield sufficiently homogeneous performance to be used in an adaptive test procedure. In this adaptive procedure, a competing background of multi-talker babble is mixed with the monosyllabic word stimuli in the same ear and the speech level needed to achieve 50% correct performance in that background is established. This yields the speech-to-noise ratio corresponding to 50% correct. This measure is plotted in the bottom panel of Figure 8 as a function of WHO-proposed HI grade with all measures again being ear-specific given earphone presentation of the speech stimuli. Two univariate analyses of variance, one for the data from each ear and both corrected for age, were performed and both revealed significant effects of WHO-proposed HI grade [$F_{(4,3246)} = 445.0$, $p < .001$; $F_{(4,3253)} = 460.9$ $p < .001$; left and right ear, respectively]. Post hoc pair-wise Bonferroni-corrected comparisons among groups showed that each group differed significantly ($p < .001$) from all other groups. As the severity of the hearing-impairment increased, an increasingly greater speech-to-noise ratio was needed to achieve 50% recognition accuracy.

VA dataset—The other VA dataset examined here (Williams-Sanchez et al. 2014) obtained pure-tone thresholds and a variety of unaided speech-recognition measures from 693

veterans (673 males) having a mean age of 65.2 years (SD = 13.1 years). Six cases were removed that met the WHO-proposed definition of unilateral hearing loss. Ages ranged from 20 to 90 years with about 90% of the participants being 50 years of age or older. Ear-specific thresholds and speech-recognition scores were obtained and analysed. Very few data (N = 9) were available for ears with a WHO-proposed HI grade of "profound" and those cases were also eliminated from further analyses. Data for the remaining 678 adults (97.8% of the dataset) were retained for subsequent analyses.

The top panel of Figure 9 shows the data for NU-6 monosyllabic words presented in quiet at a moderate sensation level designed to maximise word-recognition performance for unaided listening. Percent-correct scores have been transformed into RAUs for this figure and in subsequent analyses. Univariate analyses of variance, controlling for age, revealed a significant effect of WHO-proposed HI grade on word-recognition in quiet for the right ear [$F_{(4,668)}$ = 102.2, p < .001] and left ear [$F_{(4,663)}$ = 101.2, p < .001. Post hoc pair-wise Bonferroni-corrected comparisons among WHO-proposed HI grades revealed that each group differed significantly (p < .01) from all other groups except for "normal" versus "mild/slight" grades for both ears (p > .10).

The middle panel presents means and standard deviations for the WIN speech-to-noise ratios (SNRs) yielding 50% correct recognition of monosyllables in a multitalker babble background for each of the WHO-proposed HI grades. Age-corrected univariate analyses of variance found the effect of WHO-proposed HI grade to be significant for the right [$F_{(4,515)}$ = 96.1, p < .001] and left [$F_{(4,505)}$ = 82.1, p < .001] ears. Subsequent post hoc pair-wise Bonferroni-corrected comparisons were found to be significant (p < .02), except for the difference between "normal" and "mild/slight" grades for the WIN thresholds from the left ear (p = .73). In general, as WHO-proposed HI grade increased from "normal" to "severe," the SNR needed to achieve 50%-correct word recognition also increased with the one exception for the left ear as noted. The actual SNRs obtained in this VA dataset are very similar to those shown previously in Figure 8 (bottom panel) from the larger Wilson (2011) VA dataset.

Another adaptive ear-specific speech-in-noise test was administered by Williams-Sanchez et al. (2014). This test was the National Hearing Test (NHT; Watson et al. 2012) which makes use of triple digits in noise to establish the SNR required for 50% correct digit recognition. The three-digit sequences were presented in a background of steady-state speech-shaped noise. The means and standard deviations for the SNRs obtained from 678 participants on the NHT are plotted in the bottom panel of Figure 9 as a function of WHO-proposed HI grade. An age-corrected univariate analysis of variance revealed significant effects of WHO-proposed HI grade for the right ear [$F_{(4,672)}$ = 67.9, p < .001] and the left ear [$F_{(4,667)}$ = 56.8, p < .001]. Follow-up Bonferroni-corrected pair-wise comparisons indicating that most WHO-proposed HI groups differed significantly (p < .03) from all other groups with following exceptions: the "normal" and "mild/slight" groups did not differ significantly (p = 1.0) for either ear and, for the left ear, the "moderately severe" and "severe" groups did not differ significantly (p = .09). As with the other adaptive speech-in-noise test, the WIN (Figure 9, middle panel), as the WHO-proposed HI grade increases from "mild/slight" to "severe," a

significantly better SNR is needed to achieve 50%-correct recognition of the three-digit sequences in noise.

Cohen's d

Given the relatively large sample sizes in the various datasets reviewed above, it is possible for small group differences to emerge as statistically significant differences. Recognizing this limitation, Cohen (1988) proposed a metric, *d*, to better interpret "effect sizes." Using this metric, which has since been identified as Cohen's *d*, *d* values of 0.2, 0.5 and 0.8 have been labelled as "small," "medium," and "large" to aid in interpretation of group differences.

Figure 10 shows the distributions of Cohen's *d* values across the datasets examined in this review. Because some obvious differences in distributions were observed for different types of functional communication assessments, each panel displays the distribution for a different type of measure: speech-in-noise scores (top), speech in quiet scores (middle), and self-report measures and ratings (bottom). Most of the observed effect sizes when progressing through successive steps in the WHO-proposed HI grade scale are considered moderate, large or very large. Only the speech-in-noise scores have several Cohen's *d* values that might be considered small (about 0.2). Overall, however, the distribution of the self-report or questionnaire items is skewed toward lower *d* values than the other two functional measures of communication with "moderate" effect sizes predominating.

Discussion

This review examined the validity of the WHO-proposed HI grade system across three relatively large population or population-sample studies and two comparably large clinical studies that included measures of pure-tone hearing thresholds and some type of functional assessment of communication. There are many more studies, some considerably larger, that have pure-tone thresholds available, but no functional assessments of communication. There is at least one very large dataset, the UK Biobank dataset (Moore et al. 2014), that has functional communication data, identical to the triple-digit NHT discussed above, but pure-tone thresholds are not available. Because the WHO-proposed HI grades are defined in terms of the four-frequency pure-tone average (4fPTA) in the better ear, pure-tone thresholds from both ears and for at least those four frequencies are needed. There was at least one other published study of which we were aware that had both pure-tone thresholds and functional measures of communication, specifically triple digits in noise, from a relatively large sample of older adults (Koole et al. 2016), but we were unable to gain access to those data for this review.

Of course, there may be additional studies with the key measures needed for evaluation but were not identified by the author for inclusion in this evaluation. Given the strong consistency of the evaluations for the five datasets examined here, however, there is good confidence in the conclusions drawn. The primary conclusion from this review is that there is good validity for the WHO-proposed HI grade system as established by evidence from relatively large population and clinical studies. Specifically, there are significant changes in functional communication among older adults as the classification of WHO HI grade progresses from "mild/slight" through "severe" grades. Most often, moreover, the significant

changes in a dataset were evident for each of several functional measures available from that sample. There were insufficient data for those with "profound" hearing loss to evaluate the validity of that grade. However, there are few older adults with profound hearing loss in their better ear.

When datasets included measures of speech recognition in quiet or in noise, these were always ear-specific measures obtained under headphones. As a result, the WHO-proposed HI grade was also ear-specific when evaluating the validity of the grading system for these speech-recognition measures. Thus, we were unable to examine the validity of using the better-ear 4fPTA within the WHO-proposed HI grading system. However, given the predominance of bilaterally symmetrical hearing loss among older adults and in the datasets examined, large differences in performance between ears are not expected. Further, binaural measures of speech-recognition performance, preferably in a sound field, would be needed to evaluate the validity of defining the WHO-proposed HI grade in terms of better-ear 4fPTA, but such speech-recognition measures are very difficult to obtain on a large-scale basis due to the special testing environment and equipment required beyond what is already in place for ear-specific pure-tone audiometry.

The progression from "mild-slight" through "severe" WHO-proposed HI grades for older adults was not just statistically significant from one grade to the next, but also manifested medium to large Cohen's d effect sizes in most cases. As shown previously in Figure 10 (top), only 16% (13 of 81) of d values would be considered by Cohen (1988) to be "small" (between 0.2 and 0.399 in Figure 10) and most of these were observed for speech-in-noise scores. Further, of the 13 small Cohen's d values, 9 (69%) involved the difference between the WHO-proposed HI grades of "normal" and "mild/slight." Nonetheless, the nine occurrences of small d values represent 31% of all d values computed between the "normal" and "mild/slight" hearing-impairment grades for the studies in this review. That is, even here, the vast majority of d values (69%) represented medium or large effect sizes. Given the somewhat blurred lines between the "normal" and "mild/slight" WHO-proposed HI grades, however, it is probably appropriate that WHO defines "disabling" hearing-impairment as beginning with the "moderate" WHO-proposed HI grade.

This review provides evidence from relatively large population and clinical samples that validates the WHO-proposed HI grade system for its application to older adults with ARHL, Extensions to other populations differing in age or in the nature of the hearing-impairment from those examined here require separate evaluations. Nonetheless, the WHO-proposed HI grade system appears to offer a valid grading of the severity of communication difficulties in the large and growing population of older adults, including those with ARHL.

Acknowledgements

References

Center for Disease Control and Prevention (CDC). 2013 "National Health and Nutrition Examination Survey, 2011–2012." https://wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2011

Cohen J 1988 Statistical Power Analysis for the Behavioral Sciences. London:Routledge.

Cruickshanks KJ, Wiley TL, Tweed TS, Klein BE, Klein R, Mares-Perlman JA, and Nondahl DM. 1998 "Prevalence of Hearing Loss in Older Adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study." American Journal of Epidemiology 148 (9): 879–886. doi:10.1093/oxfordjournals.aje.a009713. [PubMed: 9801018]

Golding M, Carter N, Mitchell P, and Hood LJ. 2004 "Prevalence of Central Auditory Processing (CAP) Abnormality in an Older Australian Population: The Blue Mountains Hearing Study." Journal of the American Academy of Audiology 15 (9): 633–642. doi:10.3766/jaaa.15.9.4 [PubMed: 15575337]

Koole A, Nagtegaal AP, Homans NC, Hofman A, Baatenburg de Jong RJ, and Goedegebure A. 2016 "Using the Digits-in-Noise Test to Estimate Age-Related Hearing Loss." Ear and Hearing 37 (5): 508–513. doi:10.1097/AUD.0000000000000282 [PubMed: 26871876]

Moore DR, Edmondson-Jones M, Dawes P, Fortnum H, McCormack A, Pierzycki RH, and Munro KJ. 2014 "Relation between Speech-in-Noise Threshold, Hearing Loss and Cognition from 40–69 Years of Age." PLoS One 9 (9): e107720. doi:10.1371/journal.pone.0107720 [PubMed: 25229622]

Olusanya BO, Neumann KJ, and Saunders JE. 2014 "The Global Burden of Disabling Hearing Impairment: A Call to Action." Bulletin of the World Health Organization 92 (5): 367–373. doi:10.2471/BLT.13.128728 [PubMed: 24839326]

Speaks C, Karmen JL, and Benitez L. 1967 "Effect of Competing Message on Synthetic Sentence Identification." Journal of Speech Language and Hearing Research 10 (2): 390–396. doi:10.1044/jshr.1002.390

Stevens G, Flaxman S, Brunskill E, Mascarenhas M, Mathers CD, and Finucane M. 2013 "Global Burden of Disease Hearing Loss Expert Group. Global and Regional Hearing Impairment Prevalence: An Analysis of 42 Studies in 29 Countries." European Journal of Public Health 23 (1): 146–152. doi:10.1093/eurpub/ckr176 [PubMed: 22197756]

Studebaker GA 1985 "A 'Rationalized' Arcsine Transform." Journal of Speech and Hearing Research 28 (3): 455–462. doi:10.1044/jshr.2803.455 [PubMed: 4046587]

Travers A 1990 AB Word Lists: NAL Protocols. Chatswood, NSW, Australia: National Acoustic Laboratories.

Ventry IM, and Weinstein BE. 1982 "The Hearing Handicap Inventory for the Elderly: A New Tool." Ear and Hearing 3 (3): 128–134. doi:10.1097/00003446-198205000-00006 [PubMed: 7095321]

Ventry IM, and Weinstein BE. 1983 "Identification of Elderly People with Hearing Problems." ASHA 25 (7): 37–42.

Watson CS, Kidd GR, Miller JD, Smits C, and Humes LE. 2012 "Telephone Screening Tests for Functionally Impaired Hearing: current Use in Seven Countries and Development of a US Version." Journal of the American Academy of Audiology 23 (10): 757–767. doi:10.3766/jaaa.23.10.2 [PubMed: 23169193]

Wiley TL, Cruickshanks KJ, Nondahl DM, Tweed TS, Klein R, and Klein BE. 1998 "Aging and Word Recognition in Competing Message." Journal of the American Academy of Audiology 9 (3): 191–198. [PubMed: 9644616]

Williams-Sanchez V, McArdle RA, Wilson RH, Kidd GR, Watson CS, and Bourne AL. 2014 "Validation of a Screening Test of Auditory Function Using the Telephone." Journal of the American Academy of Audiology 25 (10): 937–951. doi:10.3766/jaaa.25.10.3 [PubMed: 25514447]

Wilson RH 2003 "Development of a Speech-in-Multitalker-Babble Paradigm to Assess Word-Recognition Performance." Journal of the American Academy of Audiology 14 (9): 453–470. [PubMed: 14708835]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Wilson RH 2011 "Clinical Experience with the Words-in-Noise Test on 3430 Veterans: Comparisons with Pure-Tone Thresholds and Word Recognition in Quiet." Journal of the American Academy of Audiology 22 (7): 405–423. doi:10.3766/jaaa.22.7.3 [PubMed: 21993048]

Wilson RH, Zizz CA, Shanks JE, and Causey GD. 1990 "Normative Data in Quiet, Broadband Noise, and Competing Message for Northwestern University Auditory Test No. 6 by a Female Speaker." Journal of Speech and Hearing Disorders 55 (2): 244–778. [PubMed: 2329786]

World Health Organization. 1991 "Report of the Informal Working Group on Prevention of Deafness and Hearing Impairment: Programme Planning. WHO/PDH/91.1." Geneva, Switzerland: WHO.

World Health Organization. 2001 International Classification of Functioning, Disability and Health. Geneva, Switzerland: WHO.

World Health Organization. 2012 "WHO Global Estimates on Prevalence of Hearing Loss: Mortality and Burden of Diseases and Prevention of Blindness and Deafness." Accessed at: www.who.int/pbd/deafness/WHO_GE_HL.pdf
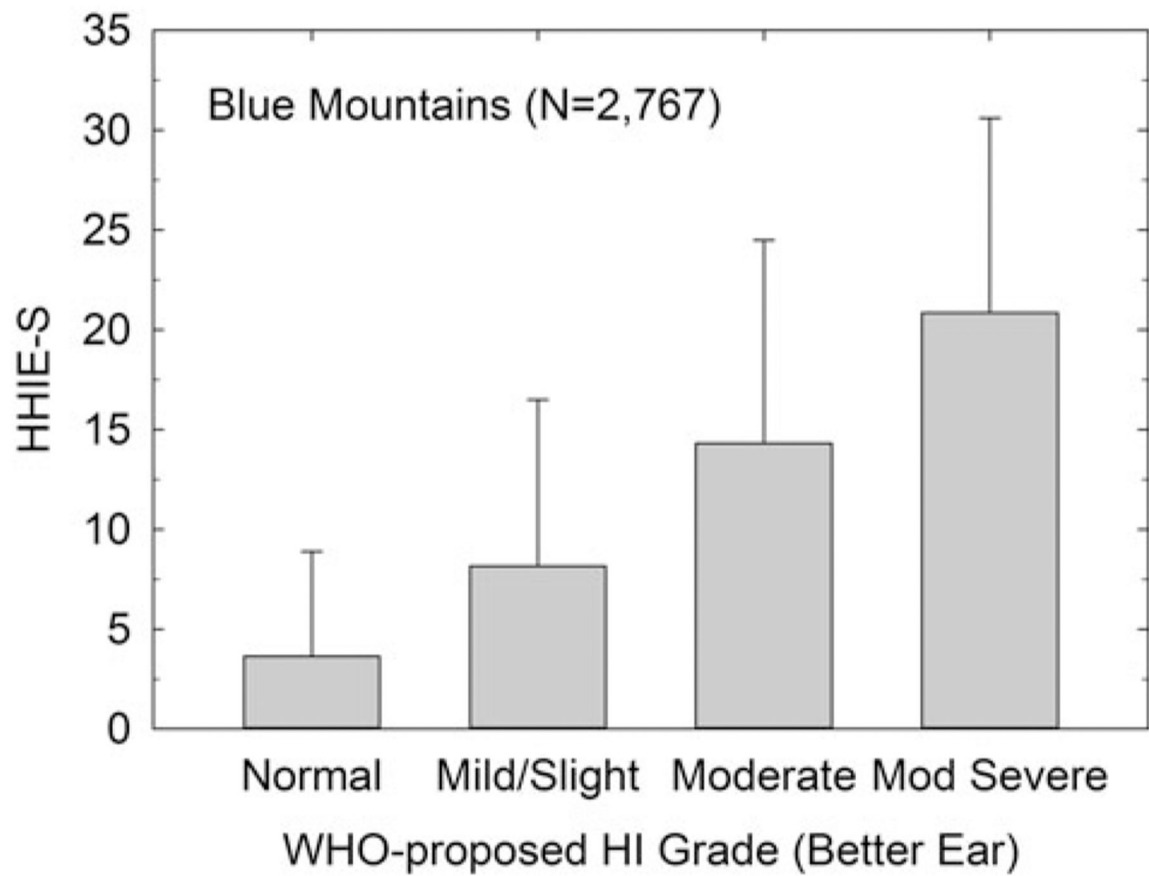
Figure 1.
Means and standard deviations for the HHIE-S as a function of WHO-proposed HI grade for the better ear from the Blue Mountains Hearing Study.
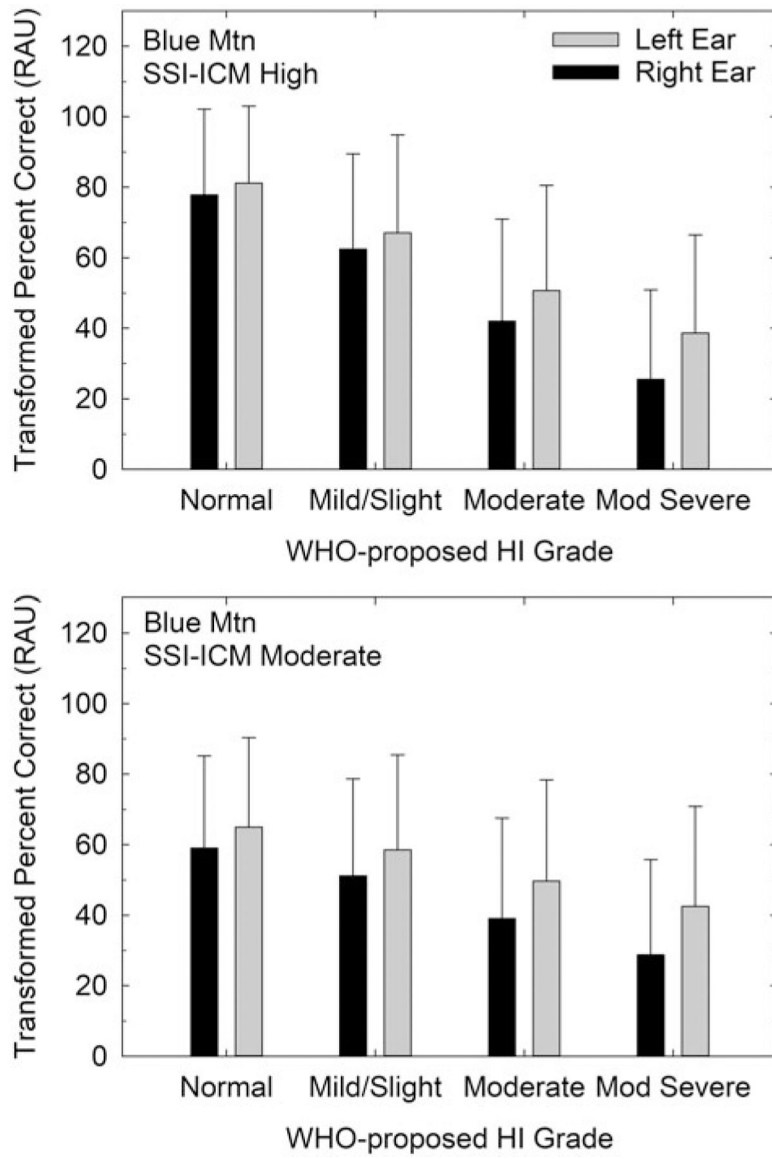
Figure 2.
Means and standard deviations for the SSI-ICM speech-identification test administered at
high (top) or moderate (bottom) presentation level plotted as a function of WHO-proposed
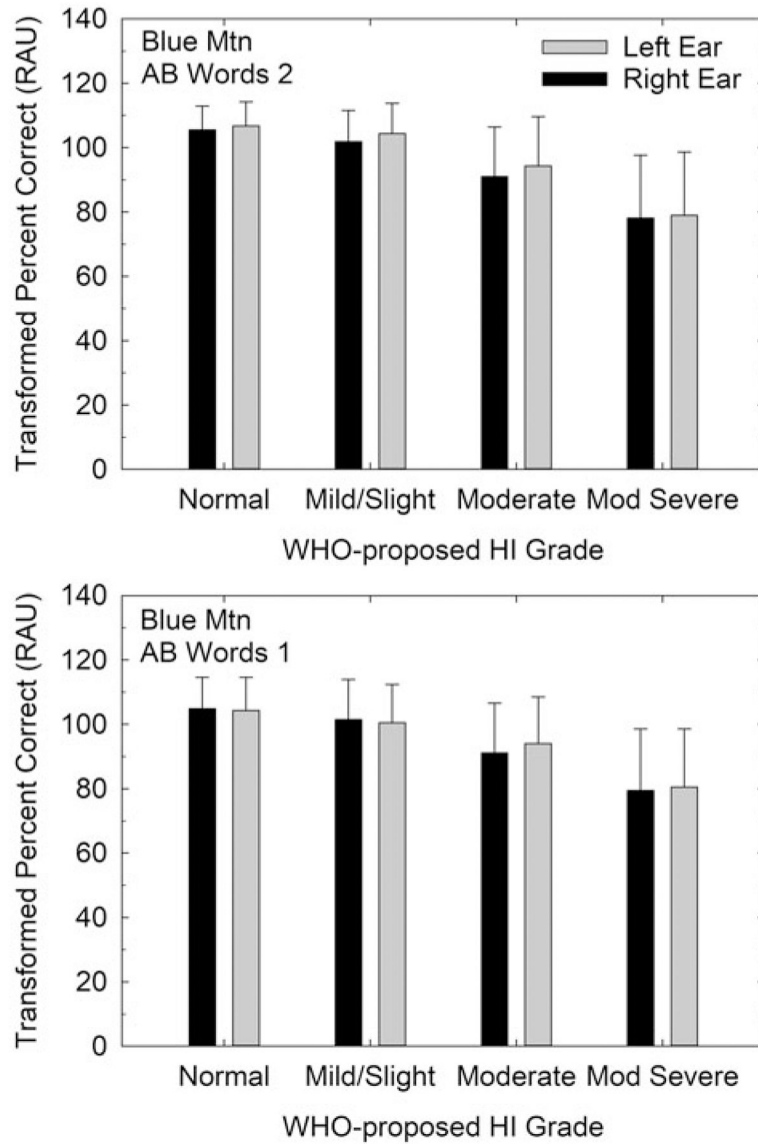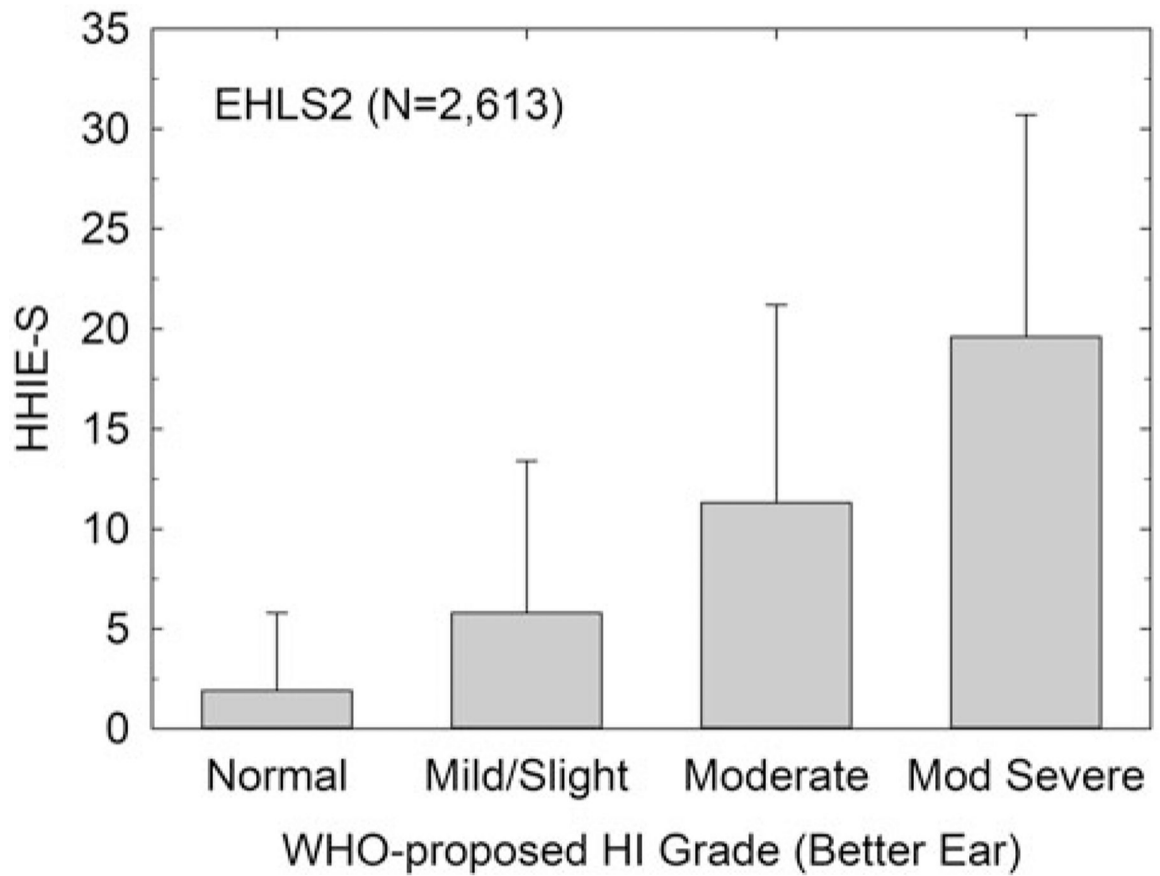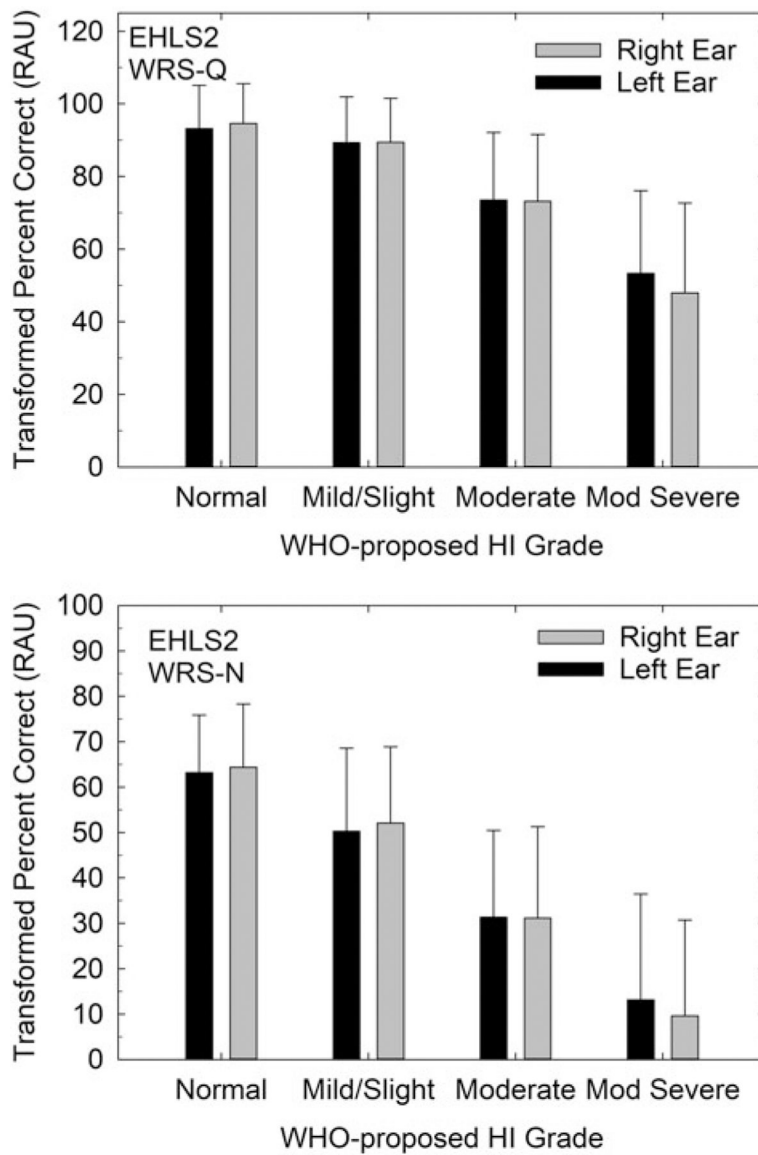HI grade for that same ear. Data for left (grey) and right (black) ears shown separately.

Figure 3.
Means and standard deviations for the AB word-recognition test administered at high (top) or moderate (bottom) presentation level plotted as a function of WHO-proposed HI grade for that same ear. Data for left (grey) and right (black) ears shown separately.

Figure 4.
Means and standard deviations for the HHIE-S as a function of WHO-proposed HI grade for the better ear from the Epidemiology of Hearing Loss Study (EHLS2).

Figure 5.
Means and standard deviations for the NU6 word-recognition test administered in quiet (top) or in competing speech (bottom) as a function of WHO-proposed HI grade for that same ear. Data for left (black) and right (grey) ears shown separately.
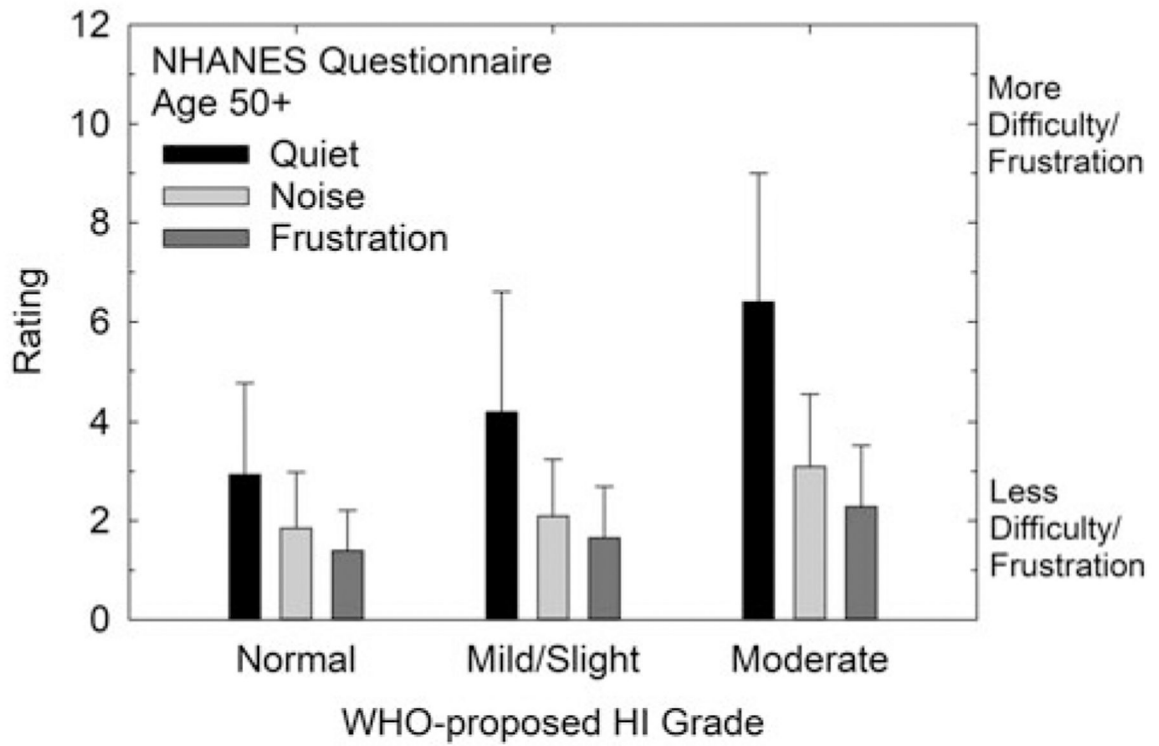
Figure 6.
Means and standard deviations for the rated difficulty or frustration in response to NHANES questionnaire items addressing hearing plotted as a function of WHO-proposed HI grade for the better ear.
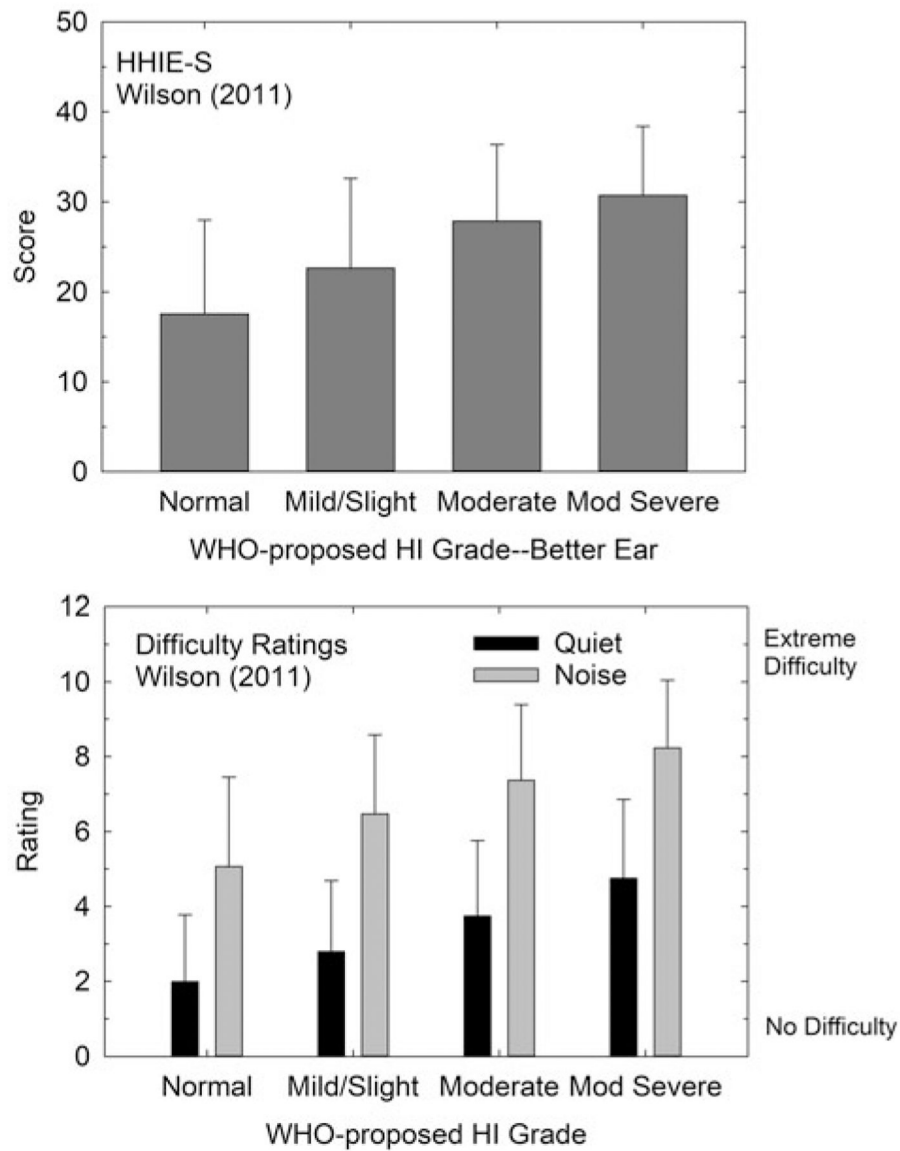
Figure 7.
Means and standard deviations for self-report data from Wilson for the HHIE-S (top) or the rated difficulty (bottom) experienced in quiet (black bars) or noise (grey bars) as a function of WHO-proposed HI grade for the better ear.
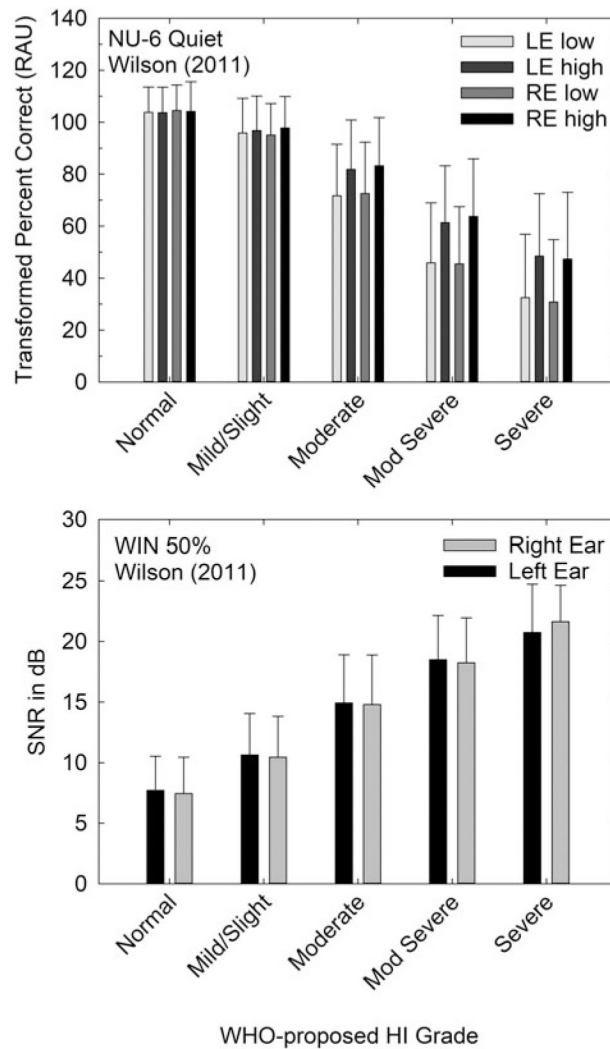
Figure 8.
Means and standard deviations for the NU6 word-recognition test administered in quiet (top) or adaptively in competing speech as the WIN test (bottom) as a function of WHO-proposed HI grade for that same ear. In the top panel, ear-specific scores are shown for each ear (LE, RE) and at two presentation levels (low, high). In the bottom panel, data for left (black) and right (grey) ears shown separately.
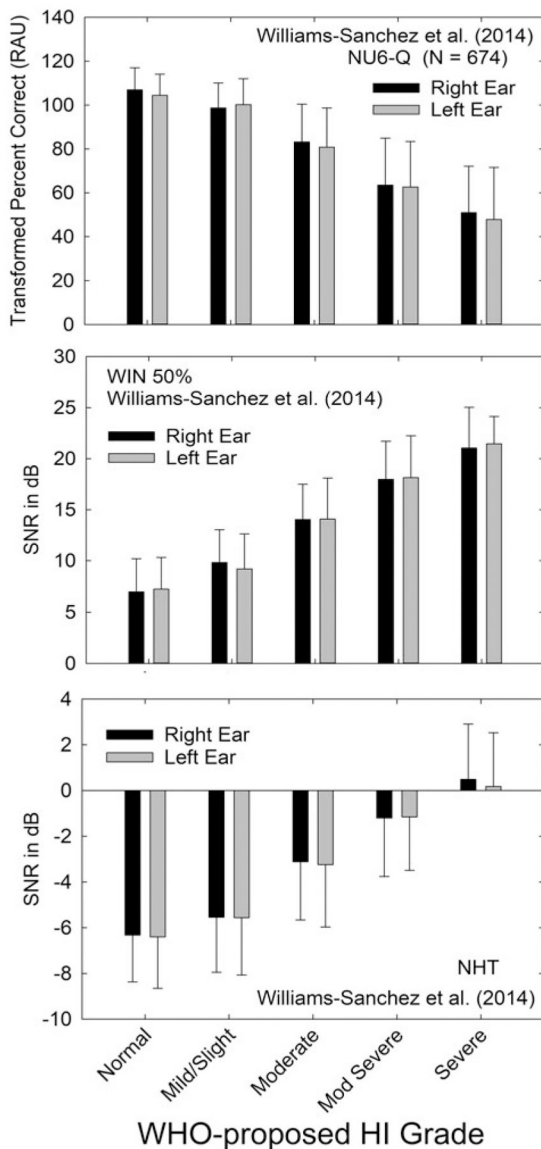
Figure 9.
Means and standard deviations for the NU6 word-recognition test administered in quiet (top) or adaptively in competing speech as the WIN test (middle) and for triple digits assessed adaptively in noise as the NHT (bottom) as a function of WHO-proposed HI grade for that same ear. Scores are plotted separately for right (black bars) and left (grey bars) ears.
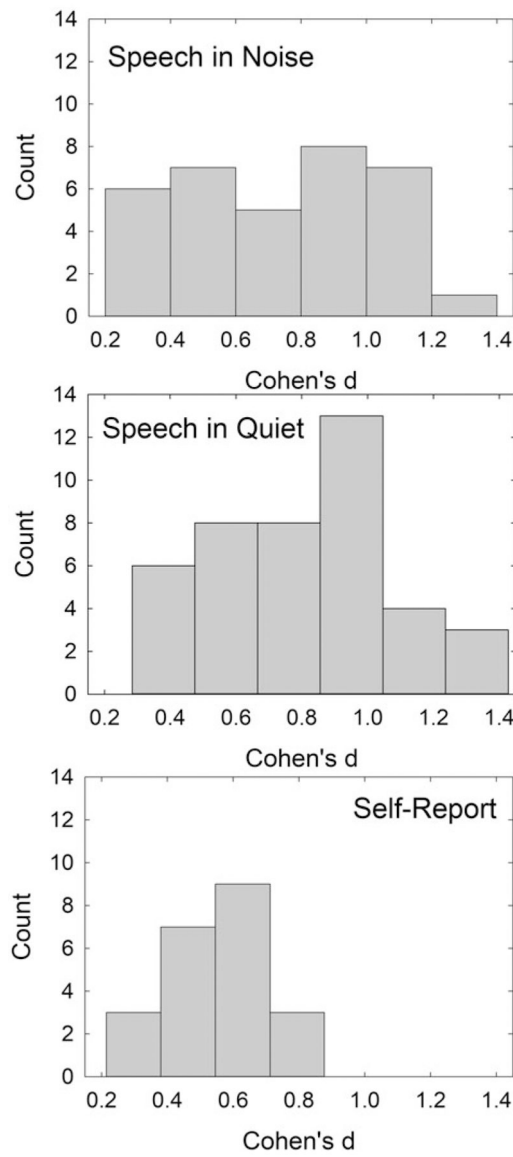
Figure 10.
Histograms showing the distribution of Cohen's *d* values for differences between successive steps on the WHO-proposed HI grading scale for speech-in-noise scores (top), speech-in-quiet scores (middle) and self-report measures of hearing handicap or difficulty (bottom).

Table 1.

WHO-proposed grades of hearing-impairment and presumed functional consequences.

Grade and corresponding audiometric ISO value [a]	Performance in Quiet and Noise [b]
0-No impairment, better than 20 dB	No or very slight hearing problems.
1-Mild 20–34 dB	No problems in quiet but may have real difficulty following conversation in noise.
2-Moderate 35–49 dB	May have difficulty in quiet hearing a normal voice and has difficulty with conversation in noise.
3-Moderately severe 50–64 dB	Needs loud speech to hear in quiet and has great difficulty in noise.
4-Severe, 65–79 dB	In quiet, can hear loud speech directly in one's ear, and, in noise, has very great difficulty.
5-Profound impairment, 80–94 dB	Unable to hear and understand even a shouted voice whether in quiet or noise.

[a]The audiometric dB HL ISO values are averages of values at 500, 1000, 2000, 4000 Hz for the better ear.

[b]From Stevens et al. (2013).