Audiogram Notches in Noise-Exposed Workers

Peter M. Rabinowitz, Deron Galusha, Martin D. Slade, Christine Dixon-Ernst, Kanta D. Sircar, and Robert A. Dobie

Objectives: **Diagnostic criteria for noise-induced hearing loss include the audiometric notch, yet no standardized definition exists. This study tested whether objective notch metrics could match the clinical judgments of an expert panel.**

Design: **A panel of occupational physicians, otolaryngologists, and audiologists reviewed audiograms of noise-exposed workers. In a two-sample process, the panel judged whether a notch was present and whether hearing loss had progressed in a notch pattern. Quantitative notch metrics were compared against expert decisions.**

Results: **At least five of six experts agreed about notch identification in 71 and 72% of the cases in the two samples, and agreement about notch progression was 61 and 67%. Notch depth and professional specialty appeared to affect notch judgments. Despite this variability, a notch metric showed excellent agreement with expert notch consensus in each sample (94.7 and 96.6%; kappa = 0.88 and 0.92).**

Conclusions: **Audiogram notch metrics can agree with expert clinical consensus and assist in the surveillance of noise-exposed workers.**

(Ear & Hearing 2006;27;742–750)

High-frequency hearing loss caused by excessive noise is one of the most prevalent occupational conditions. In their role as professional supervisors of the audiometric component of hearing-conservation programs, audiologists and physicians routinely review the audiograms of noise-exposed workers. The United States Occupational Safety and Health Administration (OSHA) regulations require that when a surveillance audiogram detects a worsening in hearing thresholds from baseline of at least 10 dB average for the frequencies 2, 3, and 4 kHz (a standard threshold shift, or STS [OSHA, 1983]), a clinician must determine whether the hearing loss appears to be work related (i.e., caused by occupational noise exposure). If the loss is found to be work related, steps need to be taken to reduce further workplace noise exposure for the individual. A key step in this process is determining whether the hearing loss is typical of noise-induced hearing loss (NIHL) as opposed to other causes, such as aging. If the loss is judged not to be noise related, the emphasis shifts towards finding a medical explanation for the loss.

In making a diagnosis of NIHL, the clinician considers multiple factors, including the history of occupational and recreational noise exposure as well as whether a medical explanation for the loss exists (American College of Occupational and Environmental Medicine [ACOEM], 2003). One aspect of the evaluation is a review of the audiogram pattern. NIHL in a young person typically creates a notch in the high frequencies of the audiogram. Because noise damage often begins at the higher frequencies of 3, 4, or 6 kHz, where the ear is more susceptible to noise (ACOEM, 2003), a noise notch typically means thresholds at 3, 4, and/or 6 kHz that are substantially worse than hearing thresholds at lower frequencies (0.5 and 1 kHz) and at 8 kHz (where a recovery is said to take place). Figure 1 depicts an audiogram showing this notch. The overall audiogram shape is concave upward.

In contrast to the notch that noise creates in the audiogram, the audiogram in pure age-related hearing loss (ARHL, or presbycusis; based on surveys of persons not exposed to occupational noise) is typically down-sloping with progressively worsening thresholds in higher frequencies (Coles, Lutman, & Buffin, 2000). Figure 2 depicts an audiogram showing this presbycusis pattern. Often, there is a curvature to the audiogram that is convex upward.

In middle-aged and older people who have had noise exposure, the effects of presbycusis and noise may overlap.

Despite general acceptance of these noise- and age-related qualities of an audiogram, there are no accepted standards for describing or quantifying audiometric notches. In the absence of such criteria, individual clinicians may come to different conclusions about whether an audiogram has such a notch, and this may influence their clinical decisions about whether a case of hearing loss is attributable to noise exposure or other causes. In one study, a panel of three clinicians reviewing audiograms agreed less than 50% of the time about whether an audiogram had a notch consistent with noise exposure (ACOEM, 2003).

Because many audiograms are currently stored and analyzed electronically using commercially available software, there would be advantages to having a validated, objective method to identify notches. A number

Yale Occupational and Environmental Medicine Program, Yale University School of Medicine, New Haven, Connecticut (P.M.R., D.G., M.D.S., K.D.S.); Alcoa Inc., Pittsburg, Pennsylvania (C.D.-E.); and University of California–Davis, Davis, California (R.A.D.).

0196/0202/06/2706-0742/0 • Ear & Hearing • Copyright © 2006 by Lippincott Williams & Wilkins • Printed in the U.S.A.

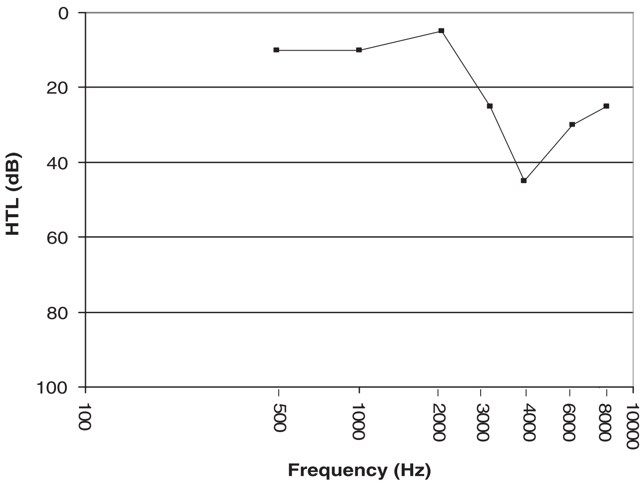

Figure 1. The notch in the audiogram of NIHL.

of objective notch definitions have been published. These include a set of notch criteria developed for medicolegal evaluation of adult patients (Coles, Lutman, & Buffin, 2000) and a method of identifying notches in a large dataset of adolescent audiograms (Niskar, Kieszak, Holmes, et al., 2001). More complex modeling has included a parabolic approximation (Gates, Schmid, Kujawa, et al., 2000) and an algebraic curve-fitting technique (Cooper & Owen, 1976). To date, there are no published reports of comparative evaluations of notch metrics or comparisons of such metrics with the clinical judgment of an expert panel.

In addition to determining whether the audiogram has a notch consistent with noise exposure, the clinician may be faced with an individual whose baseline audiogram already displayed a notch and who has then had further deterioration of hearing. The clinician may find it useful to attempt to judge whether the audiometric notch had worsened (i.e., NIHL) or whether most of the recent progression was happening in an age-related (down-sloping) pattern. Once again, there is no standardized method to evaluate this change in audiometric configuration over time.

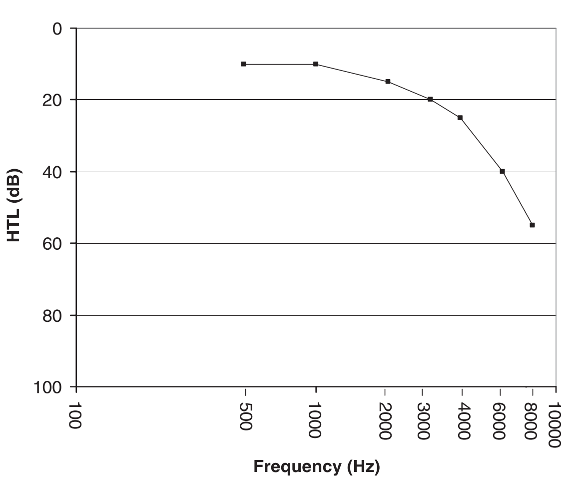

Figure 2. The down-sloping audiogram in presbycusis.

The objective of this study was to determine whether a group of experts could agree about the presence of audiometric notching and notch progression, and whether any objective metrics could achieve significant agreement with the judgments of the expert panel.

METHODS

Study Population

Alcoa Inc. maintains a database of audiometric tests performed as part of company hearing-conservation programs. Under Alcoa health and safety protocols, individuals working in jobs in which the 95th percentile of measured noise levels (8-hr time-weighted average) is 85 dBA or greater undergo yearly surveillance audiometry, and individuals exposed to between 82 and 84 dBA have audiograms performed at least every 3 yr. Pure-tone audiograms are performed using sound booth environments that meet OSHA standards for audiometric testing. For the past 20 yr, the centralized database of audiograms has been developed under the direction of a certified audiologist who directs the corporation's hearing-conservation program and who is also a certified industrial hygienist. There have been ongoing efforts to ensure quality of the testing and data entry. An anonymous set of these industrial audiograms has been assembled for research purposes (Rabinowitz, Slade, Dixon-Ernst, et al., 2003) with the approvals of the Human Investigation Committee of Yale University School of Medicine and Alcoa's Occupational and Environmental Health Advisory Committee.

The audiometric records of individuals with at least three audiograms since 1982 were analyzed to identify cases for review by the expert panel. The first test appearing in the dataset after 1982 established an individual's initial baseline. An individual was judged to have experienced a STS if an audiogram showed a 10-dB or greater shift from baseline (without age correction) for the average of hearing thresholds at 2, 3, and 4 kHz in either ear and if this shift was confirmed on a subsequent test. SAS v 8.02 (SAS Institute, Cary, NC) was used to randomly select for expert review an equal number of left- and right-ear STS audiograms from the more than 1500 STS cases identified in the dataset.

Review of Audiograms

For each case, the audiometric results were plotted as a graphic audiogram of hearing threshold versus frequency using Microsoft Excel, with an equal distance between octave frequencies on the x-axis that was equivalent in magnitude to the 20-dB hearing-threshold level on the y-axis. Two sets of audiograms were created. In the first set, only the 10-dB-shift audiogram was displayed. In the second set, both the baseline and the 10-dB-shift audiogram were displayed. The audiogram review sheet also noted the age of the individual at the time of STS and the number of years between baseline the audiogram and STS.

A voluntary independent six-person panel consisting of two otolaryngologists, two occupational medicine physicians, and two audiologists reviewed the audiograms. These individuals were selected based on their expertise in the field of NIHL. All were current or former council members of the Council for Accreditation in Occupational Hearing Conservation, a national certifying body for occupational audiometric technicians. None of the members of the panel were involved in the development of candidate notch criteria, had any affiliation with Alcoa Inc., or were aware of the results of any notch statistics or the opinions of other reviewers as they made their individual judgments. The expert reviewers scored each case as to whether or not the audiogram, at the time of the 10-dB shift, seemed to have a noise notch (i.e., notching consistent with NIHL). In a separate analysis, reviewers examined the progression of hearing loss between the baseline and the 10-dB-shift audiogram and assessed whether the progression was consistent with a notching pattern of NIHL.

Definition of Expert Cases

Cases in which at least five observers agreed a notch was present were considered, for the purposes of this study, to be expert noise-notch cases. Cases in which at least five observers did not feel there was a notch were classified as expert non-notch cases. Similarly, cases in which at least five observers agreed there was or was not a notching pattern in the progression of the audiometric hearing loss were classified as either expert notch progression or expert non-notch progression.

Notch Statistics Tested

Notch index • A notch index was defined as the difference between the pure-tone average of thresholds at 2, 3, and 4 kHz and the average of thresholds at 1 and 8 kHz in the same ear. When the audiogram between 1 and 8 kHz is a straight line (assuming the usual logarithmic axes for both frequency and intensity), the notch index equals zero. If the thresholds for 2, 3, and 4 kHz lie above the line connecting the 1- and 8-kHz thresholds, as is common in the convex pattern of ARHL, the notch index would be less than zero. If the thresholds at 2, 3, and 4 kHz lie below the line, as would be expected for a NIHL pattern, the notch index is greater than zero. The choice of the average of thresholds at 2, 3, and 4 kHz as the first term in the notch index was based on the following assumptions:

a. Thresholds at 6 khz are less valid and less reliable than those at lower frequencies.
b. Noise-induced changes at 2 kHz are both substantial and important in cases of long exposure at high levels, and may be a site for further progression of NIHL.
c. OSHA has defined a STS based on 2, 3, and 4 kHz, so many hearing-conservation programs already are calculating this average.

The choice of the 1- and 8-kHz average as the second term involves spacing the anchor points exactly an octave above and below the STS region, with the result that the notch index describes the average deviation of the audiogram (excluding 6 kHz) from the straight line connecting 1 and 8 kHz.

Coles et al. notch criteria • Coles et al. (2000) have published criteria for identification of an audiometric notch for use in medicolegal diagnosis of NIHL. This was defined as a high-frequency notch where the hearing threshold at 3, 4, and/or 6 kHz is at least 10 dB greater than at 1 or 2 kHz and at least 10 dB greater than at 6 or 8 kHz. Because of distortion at 6 kHz, an adjustment would be necessary if certain earphone types were used.

Niskar et al. notch criteria • Niskar et al. (2001) developed a set of criteria similar to those of Coles et al. (2000) for use in identifying NIHL in the audiograms of adolescents tested in the National Health and Nutrition Evaluation Survey. This definition requires a notched audiogram to have

a. Hearing-threshold level values at 0.5 and 1 kHz $\leq$15 dB
b. Worst (i.e., greatest value) threshold at 3, 4, or 6 kHz at least 15 dB worse than the worst threshold value at either 0.5 or 1 kHz
c. A hearing threshold at 8 kHz at least 10 dB better than the worst threshold at 3, 4, or 6 kHz.

Notch-progression statistic • To assess whether the change in the audiogram configuration over time occurred predominantly in a notch pattern, the notch index (described above) for the baseline audiogram was subtracted from the notch index calculated for the follow-up (STS) audiogram to give a notch-progression statistic $NP = (NI_{STS}) - (NI_{baseline})$. It was expected that NIHL would produce a notch-progression statistic greater than zero and that predominantly age-related changes would result in a notch-progression statistic less than zero.

Stages of Review

The expert review took place in two stages separated by a 1-yr period. In the first stage, the expert panel evaluated 80 STS audiogram cases (40 right ear and 40 left ear), randomly chosen as above, for the presence of noise notching or notch progression. These 80 audiograms represented 80 unique individuals; no individual contributed both a left- and right-ear audiogram to the sample. Consensus cases were identified (five of six of the experts agreeing as to notch or notch-progression status), and overall agreement between observers and a multiple-observer kappa statistic were calculated. In a separate analysis, the agreement between consensus cases identified by the expert panel and those defined by objective metrics (see below) was analyzed with a kappa statistic. For the notch index and notch-progression metrics, a range of whole-integer cutoff values were tested to find the point at which the kappa statistic for agreement with the judgment of the experts was maximized.

In the second stage of the review, the expert panel, at a different time, evaluated 80 additional randomly chosen STS audiogram cases (40 left ear and 40 right ear) and 20 audiogram cases (10 left ear, 10 right ear) picked randomly from the first set. Again, the new cases represented 80 unique individuals. The panel was not informed of the presence of previously judged audiograms in the second set. Consensus cases were identified among the set of 80 new audiogram cases, and overall agreement and multiple-observer kappa values were determined. The cutoff values for notch index and notch-progression statistic that had been determined using the first set were then applied to the analysis of the second audiogram set. In addition, the consistency of individual experts in terms of notch judgment was assessed using a kappa statistic to compare judgments between the first and second reviews.

Statistical Analysis

For each audiogram and audiogram pair as defined above, the different notch criteria and the notch-progression statistic were calculated using SAS v 8.02 (SAS Institute, Cary, NC). Descriptive statistics were calculated on the demographic variables. For the 57 consensus cases identified in the first set and the 58 consensus cases identified in the second set of audiogram cases, the results of objective notch analysis were compared with the consensus expert judgment. To assess the possibility of observer bias in notch judgments, Cochran's Q_A statistic was calculated and compared against a χ^2 distribution with 5 degrees of freedom to estimate whether the six experts had similar classification probabilities. Based on this result, the appropriate reliability kappa statistic was calculated to assess overall agreement between six observers. A prevalence-adjusted and bias-adjusted statistic was used where appropriate.

Additional Analysis of Expert Judgment Behavior

Because experts often disagreed regarding the presence or absence of a notch, we wondered whether individual reviewers might have been looking at the same audiometric features and applying different criteria cutoff values to judge an audiogram as having a notch. We therefore performed a post hoc analysis to examine, for each of the six experts, which of several cutoff values for the notch index had the greatest agreement with that individual's notch judgments, and whether this cutoff value varied among experts. In addition, because it was possible that the degree of audiometric recovery at 8 kHz was a discriminating factor for individual notch judgments, we created an 8-kHz recovery (8 kR) variable, which was defined as the difference between the worst of the three thresholds at 3, 4, and 6 kHz and the threshold at 8 kHz. (In a classic noise notch, 8 kR is positive, and in the classic audiogram of ARHL, 8 kR is negative.) We again compared different cutoffs for this variable with each expert's notch judgments. Finally, to determine whether the age of the worker whose audiogram was being reviewed was a determining factor in expert judgments, we compared expert agreement for subjects younger than 40 yr of age with those older than 40 yr.

RESULTS

Characteristics of Audiogram Cases

Table 1 shows the characteristics of the audiograms in the first and second study samples. The mean age of the cases in both samples was greater than 40 yr. The average number of years between baseline and STS audiograms was approximately 5 yr in the first sample and greater than 7 yr in the second. At the time of STS, the mean of the average hearing-threshold level at 2, 3, and 4 kHz exceeded 25 dB in both samples.

Results of Expert Judgment

In the first sample, at least five of the six experts agreed that 38 (47.5%) of the 80 audiograms were expert notch cases and that 19 (23.8%) of the audiograms were expert non-notch cases. Therefore, 57/80 (71.3%) of the cases were identified as consensus cases. The proportion of audiograms identified as notch cases varied substantially between individual

TABLE 1. Characteristics of audiogram cases

Characteristic	First sample			Second sample		
	Total (N = 80)	Left ear (n = 40)	Right ear (n = 40)	Total (N = 80)	Left ear (n = 40)	Right ear (n = 40)
Age at time of 10-dB shift	45 (8.7)	45 (8.8)	45 (8.7)	41.8 (7.8)	43.1 (7.3)	40.5 (8.1)
Years to 10-dB shift	4.9 (2.3)	5.1 (2.2)	4.8 (2.4)	7.6 (4.1)	8.5 (4.3)	6.7 (3.7)
Hearing threshold at 2, 3, and 4 KHz at time of 10-dB shift	32.4 (13.4)	33.7 (15.3)	31.2 (11.1)	31.2 (13.6)	32.9 (15.4)	29.5 (11.4)

Values are mean (SD).

experts (range, 37.5 to 76.3%). In the second sample, there were 58 consensus cases (72.5%), of which 38 (47.5%) were notch cases and 20 (25.0%) were non-notch cases. Again, there was substantial variation between experts (36.3 to 83.8%). The distribution of the agreement between experts in both samples produced a significant Cochran's Q_A statistic ($p < 0.0001$), indicating observer bias. In other words, at least one reviewer had a significantly different approach to the notch judgment compared with the other five observers. There appeared to be differences between experts according to background: the rate of notch identification for the two audiologists was 74%, versus 57% for the occupational physicians and 55% for the two otolaryngologists. The audiologists determined a notch progression to be present 77% of the time, whereas the occupational physicians made this determination for 63% of audiograms and the otolaryngologists for 53% of the cases. Figure 3 shows examples of audiograms judged by 100%, 50%, or none of the experts to have a significant notch.

For the judgment of notch progression in the pattern of the audiogram from baseline to STS, at least five of the six experts agreed that 33 (41.3%) of the 80 audiograms in the first sample progressed in a notch pattern, and at least five also agreed that 16 (20%) did not. In the second sample, the experts agreed that 45 of the 80 (56.3%) cases progressed in a notch fashion, and nine (11.3%) did not. Once again, the distribution of this agreement between experts produced a significant Cochran's Q_A statistic ($p < 0.0001$), indicating significant observer bias.

Notch Analysis

Table 2 shows the results of notch analysis for all audiograms reviewed by the expert panel. The mean notch index at baseline was 5.3 dB for the first sample of audiograms and 1.6 dB for the second sample. At the time of STS occurrence, the mean notch index had increased to 8.9 dB in the first sample (a notch progression of 3.6 dB) and 7.8 in the second sample (for a notch progression of 6.1 dB). In both samples, the mean notch-progression statistic was greater in the left ear compared with the right.

Compared with the criteria of Niskar et al. (2001), the published notch criteria of Coles et al. (2000) detected higher percentages of notches in both the first sample (79 versus 55%) and the second sample (72 versus 45%).

Individual Consistency of Experts

When members of the expert panel judged the same 20 audiograms on two different occasions, the agreement for notch judgments ranged from 55 to 90% among the experts, with a range of kappa values of 0.17 to 0.79 (median 0.49). For notch-progression judgments, the consistency of individual judgments showed even more variability, with agreement rates ranging from 50 to 90% and kappa values ranging from 0 to 0.80 (median 0.28).

Correlations between Expert Consensus and Objective Notch Measures

Analysis of the first sample of audiograms showed that the best agreement between the 57 expert consensus notch cases and the calculated notch index occurred with a notch index cutoff value of 2 dB. Notches defined by this cutoff value displayed an overall agreement of 87.7% (kappa 0.71) with the notch case consensus decisions of the expert panel. Using the same cutoff value, the notch index produced an agreement of 79.3% and a kappa statistic of 0.55 with the 58 consensus cases from the second sample. These results are shown in Table 3.

In the first sample, the notch criteria of Niskar et al. (2001) displayed the same overall agreement as the notch index (87.7%) of the expert notch consensus cases, but a slightly higher kappa (0.75). Like the notch index metric, the criteria of Niskar et al. (2001) showed slightly poorer agreement with consensus notch cases from the second sample (84.5%, kappa = 0.69). The notch criteria of Coles et al. (2000) showed the highest agreement with expert consensus cases: 94.7% (kappa 0.88) in the first sample and 96.6% (kappa = 0.92) in the second sample.

Table 3 also shows the degree of agreement between the notch-progression statistic and expert

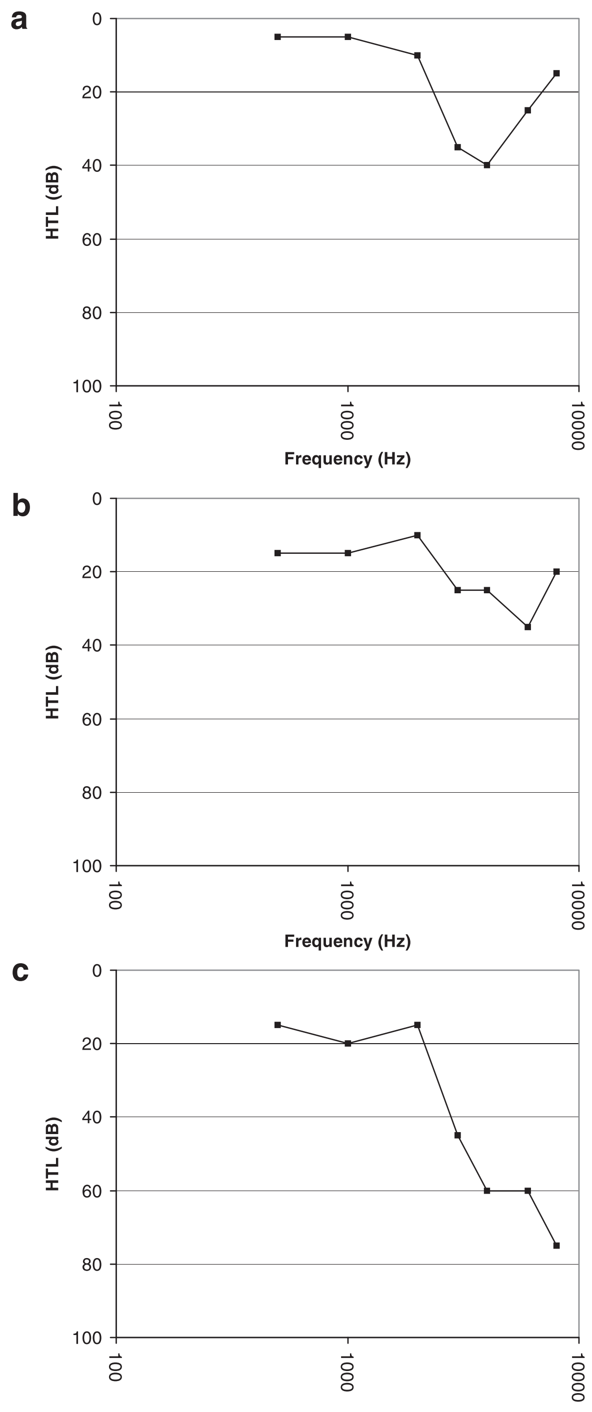

Figure 3. Examples of audiograms producing different degrees of expert agreement. a) Audiogram with 100% expert agreement that a noise notch was present. b) Audiogram with 50% agreement regarding presence of a noise notch. c) Audiogram with 100% agreement regarding absence of a noise notch.

consensus regarding notch-progression cases. For the first sample, a notch-progression statistic cutoff of more than 2 dB produced an overall agreement of 67.4% (kappa = 0.32) with expert consensus. Applying this cutoff value to the audiograms in the second set demonstrated an overall agreement of 68.5% and a kappa value of only 0.07. The prevalence- and bias-adjusted kappa value for this agreement was 0.37.

Reasons for Expert Notch Judgment

When different cutoffs for the notch index were compared with individual expert judgments of notches, wide interobserver variation was found. For two of the six experts, a notch index cutoff value of 5 dB or greater produced the greatest correlation, two other raters agreed best with a 10-dB cutoff, and one rater each agreed best with a 20- or 25-dB notch index cutoff value. The rate of best overall agreement between individual judgments and the notch index cutoffs ranged from 68 to 83%. Although the 8 kR had a better range of overall agreement with individual experts (79 to 91%), there was still wide variation in cutoffs. Two experts agreed best at a 5-dB cutoff value of 8 kR, two demonstrated best agreement with a 10-dB cutoff, and two had the best correlation with a 20-dB cutoff.

A stratified analysis by age group (younger than 40 yr versus 40 yr or older) found no significant differences by age in likelihood of judging an audiogram as notched (data not shown).

Discussion

The results of this study indicate that expert reviewers can agree on whether a significant noise notch is present in an audiogram. Among six expert reviewers, there was agreement of at least five of the six experts about the notch status of an audiogram in 115 out of the 160 audiograms (72%). At least five of the six experts also demonstrated agreement as to whether an individual's hearing loss had progressed in a notch pattern in 103 out of the 160 audiogram pairs (64%). On the other hand, this means that in 36% of cases, our experts were divided 4:2 or 3:3 on the issue of notch progression.

At the same time, analysis of the expert panel's judgments consistently indicated individual biases affecting the ability of the panel to demonstrate agreement. The analysis of expert judgment behavior showed that individual cutoffs for a notch index or an 8 kR varied widely. Some raters apparently needed to see more 8 kR than others did before declaring a notch to be present. By contrast, the age of the subject was not a predictive factor. Because the amount of recovery at 8 kHz is an indicator of the depth of the notch, our findings are similar to those of McBride et al. (2001), who found that three audiogram reviewers differed in their valuation of notch depth. In addition, the repeat analysis of 20 audiograms at two separate times showed that some experts were not

TABLE 2. Notch metric calculations for audiogram cases

Case criteria	First sample Total (N = 80)	First sample Left ear (n = 40)	First sample Right ear (n = 40)	Second sample Total (N = 80)	Second sample Left ear (n = 40)	Second sample Right ear (n = 40)
Baseline notch index (mean (SD))	5.3 (12.5)	3.9 (14.9)	6.9 (9.5)	1.6 (10.9)	2.1 (11.7)	1.1 (10.2)
Standard threshold shift notch index (mean (SD))	8.9 (13.7)	8.2 (14.7)	9.5 (12.7)	7.8 (14.6)	10.3 (17.0)	5.2 (11.3)
Notch progression (mean (SD))	3.6 (9.3)	4.4 (8.7)	2.8 (9.8)	6.1 (11.2)	8.2 (13.4)	4.1 (8.1)
Niskar notch (n (%))	44 (55.0)	21 (52.5)	23 (57.5)	36 (45.0)	18 (45.0)	18 (45.0)
Coles notch (n (%))	63 (78.8)	31 (77.5)	32 (80.0)	58 (72.5)	31 (77.5)	27 (67.5)

internally consistent in their judgments about either the presence of a notch or of notch progression.

These results suggest that although detection of notches could be clinically useful, experts may often disagree—even with their own prior judgments—about individual audiograms. This lack of consistency about a potentially meaningful judgment provides a strong argument for the development of objective measures that could be consistently applied to the assessment of audiometric patterns and the diagnosis of NIHL.

Fortunately, the results of this study indicate that a number of different metrics show promise as objective notch measures. A simple notch index showed good agreement with the judgments of the expert panel, and a set of notch criteria developed for use in the evaluation of adults with possible NIHL showed excellent agreement. We argue that, with current knowledge, it seems more appropriate to use such objective measures in the evaluation of industrial audiograms rather than to rely solely on individual clinical judgment.

It is not surprising that the Coles et al. (2000) criteria correlated best with clinician judgment, because they essentially quantify what physicians and audiologists are taught in training: a noise notch requires better hearing at both lower frequencies and at 8 kHz than at the notch frequencies (ACOEM, 2003). The Niskar notch criteria, although perhaps suitable for the very young population for which they were initially used, require excellent hearing at 0.5 and 1 kHz (many middle-aged men fail to meet this criterion) and a larger difference (15 dB compared with the 10d B required by Coles et al. [2000]) between the lower frequencies and the notch frequencies. Thus, the Niskar criteria can be considered rather strict. Indeed, all of the cases meeting these criteria were also expert consensus notch cases, whereas all of the Niskar errors were false negatives (cases in which experts agreed a notch was present despite failing to meet the Niskar et al. [2001] criteria).

The notch index was more than 2 dB in some cases that were considered by the experts not to be notches. Thresholds in the 2- to 4-kHz range can be below the line connecting 1 and 8 kHz (resulting in a positive notch index) in the absence of a notch; Coles et al. (2000) refer to this as a bulge (or concavity) in the audiogram and consider it to be a sign of mixed effects of aging and noise. The notch index was also less than 2 dB in many cases that were considered by the experts to be notches. This typically occurred with isolated 6-kHz notches, or with 4-kHz notches when hearing was relatively good at 2 and 3 kHz (data not shown).

The results of this study, although still preliminary, clearly demonstrate the potential value of systematic comparisons between objective notch measures and expert judgment. Therefore, although one of the metrics tested achieved a remarkable agreement of approximately 95% with the expert consensus case decisions, it is possible that another metric could perform even better. Some of the problems in notch estimation (e.g., upward-sloping au-

TABLE 3. Agreement between expert consensus and notch metrics

Notch metric	First-sample consensus cases (n = 57) Observed agreement (%)	First-sample Kappa	Second-sample consensus cases (n = 58) Observed agreement (%)	Second-sample Kappa	Prevalence- and bias-adjusted statistic
Notch index >2	87.7	0.71	79.3	0.55	—
Notch criteria from Niskar et al., 2001	87.7	0.75	84.5	0.69	—
Notch criteria from Coles et al., 2000	94.7	0.88	96.6	0.92	—
Notch-progression >2	67.4	0.32	68.5	0.07	0.37

diograms in lower frequencies or notches that begin their downward slopes at different corner frequencies) might require the use of more sophisticated curve fitting (Niskar, Kieszak, Holmes, et al., 2001; Gates, Schmid, Kujawa, et al., 2000) for best agreement with expert assessments. Such curve fitting might also be less sensitive to testing variability at a given frequency.

Although this panel of experts showed a reasonable degree of agreement about the presence of a notch in an audiogram, they agreed less as to whether the progression of hearing loss from baseline followed a notch pattern. Significant observer bias was detected in both determinations, implying that the observers appeared to differ consistently in their readiness to label an audiogram as having a noise notch or a notch pattern of progression. This may reflect individual variability and/or systematic differences in training regarding NIHL. Although we found that audiologists were more likely than physicians to identify notches or notch progression, the small number of experts in our panel makes it impossible to reach conclusions about interspecialty differences.

This study used the agreement of five out of six experts to define an expert consensus case of noise notching or notch-pattern audiogram progression. Because there remains no tissue biopsy or other technique to clearly differentiate noise effects from those of presbycusis or other causes of sensorineural hearing loss in humans, NIHL remains a clinical diagnosis. This diagnosis is based on careful review of noise-exposure history, review of serial audiograms, physical examination of the ear, and elimination of other possible causes of hearing loss. In the absence of any well-accepted standard criteria, the use of expert agreement to identify probable cases of NIHL therefore represents a reasonable approach to compare and validate any proposed objective diagnostic tools, such as an audiometric notch definition. Although the six observers in this study achieved a higher degree of agreement than has been previously reported for notch interpretations, it is possible that a different or a larger panel of experts might reach different conclusions regarding diagnosis. It is therefore advisable that any proposed statistic be evaluated in multiple settings and with a larger number of experts achieving consensus about the diagnosis. Such work could lead to the development of an industry-wide standard.

Another approach to the validation of objective diagnostic tools is to apply them to audiograms derived from the International Organization for Standardisation's (1990) standard set of models of ARHL and NIHL (Dobie, R., 2005). Unfortunately, the International Organization for Standardisation's (1990) models exclude 8 kHz, making it impossible to use this approach to test any of the notch statistics used in this study.

This study reviewed audiograms from only male industrial workers, whose mean age was older than some worker populations and who already showed some degree of notching with hearing thresholds at baseline. An element of presbycusis in this older population could have affected the performance of the metrics. The audiograms were identified without the use of age correction. This could have produced audiograms of individuals at earlier stages of hearing loss than would be seen if age correction had been used. It is therefore possible that the expert judgment and objective notch statistics would perform differently in the setting of age-corrected audiograms. Additionally, this study relied on audiograms performed in the industrial setting, which may differ from those performed in an audiologist's office. Despite these limitations, quantitative methods can provide a useful supplement to the subjective judgment of a clinician reviewing a case of hearing loss.

The objective notch metrics used in this study relied on the frequency of 8 kHz. Currently, many hearing-conservation programs and regulatory standards (e.g., the OSHA hearing-conservation standard) do not require testing at this frequency. If notch definitions such as the ones tested in this article are to gain wider application, it will be necessary to include routine testing at 8 kHz for hearing-conservation audiometric surveillance.

Acknowledgments

Dr. Mark Cullen and Dr. Prakash Nadkarni provided helpful critique of the methods used in this investigation. The authors are grateful to Alcoa Inc. for the research use of the company's audiometric and industrial hygiene database. Funding was provided by grants from CDC/NIOSH: K01-0H00173-01 and R01 OH07724-01. The authors thank the expert panel for their valuable efforts.

Dr. Rabinowitz, Mr. Galusha, and Mr. Slade provide consultant services to Alcoa Inc. The authors declare that they have no competing interests.

Address for correspondence: Peter M. Rabinowitz, M.D., M.P.H., Associate Professor of Medicine, Yale Occupational and Environmental Medicine Program, Yale University School of Medicine, 135 College Street, Suite 392, New Haven, CT 06510. E-mail: Peter.Rabinowitz@yale.edu.

Received October 28, 2005; accepted May 22, 2006.

References

American College of Occupational and Environmental Medicine. (2003). ACOEM evidence-based statement: noise-induced hearing loss. *Journal of Occupational and Environmental Medicine*, *45*:579–581.

Coles, R. R., Lutman, M. E., & Buffin, J. T. (2000). Guidelines on the diagnosis of noise-induced hearing loss for medicolegal purposes. *Clinical Otolaryngology and Allied Sciences*, *25*, 264–273.

Cooper, J. C., & Owen, J. H. (1976). Audiologic profile of noise-induced hearing loss. *Archives of Otolaryngology*, *102*, 148–150.

Dobie, R. A. (2005) Estimating noise-induced permanent threshold shift from audiometric shape: The ISO-1999 model. *Ear and Hearing*, *26*, 630–635.

Gates, G. A., Schmid, P., Kujawa, S. G., Nam, B, & D'Agostino, R. (2000). Longitudinal threshold changes in older men with audiometric notches. *Hearing Research*, *141*, 220–228.

International Organization for Standardisation. (1990). *1999: acoustics: determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment.* Geneva. International Organization for Standardisation.

McBride D. I. Williams, S. (2001). Audiometric notch as a sign of noise induced hearing loss. *Occupational and Environmental Medicine*, *58*, 46–51.

Niskar, A. S., Kieszak, S. M., Holmes, A. E., Esteban, E., Rubin, C., & Brody, D. J. (2001). Estimated prevalence of noise-induced hearing threshold shifts among children 6 to 19 years of age: the Third National Health and Nutrition Examination Survey, 1988-1994, United States. *Pediatrics*, *108*, 40–43.

Rabinowitz, P. M., Slade, M., Dixon-Ernst, C., Sircar, K., & Cullen, M. (2003). Impact of OSHA final rule–recording hearing loss: an analysis of an industrial audiometric dataset. *Journal of Occupational and Environmental Medicine*, *45*, 1274–1280.

United States Occupational Safety and Health Administration. (1983). 1910.95 CFR Occupational Noise Exposure: Hearing Conservation Amendment (Final Rule). *Federal Register*, *48*:9738–9785.