Audiometric Test Criteria in the Detection of Cisplatin Ototoxicity

Thomas H. Simpson*†
Sabina A. Schwan*‡
William F. Rintelmann*

Abstract

Audiometric criteria establishing the ototoxic effects of cisplatin (CDDP) vary widely in the literature. Normal test-retest variability in audiometric thresholds resulting from serial audiometry is likely a confounding variable in the over-reporting of ototoxic effects. This study investigated the variability of serial audiometry for a group of 21 head and neck cancer patients receiving no known ototoxic drugs in an effort to establish guidelines for determining reasonable audiometric criteria for CDDP therapy. Comparison of this group to a similar group of 31 patients receiving high dose (100+ mg/m²) bolus administration of CDDP suggests that audiometric shift criteria developed from Receiver Operating Characteristic analysis and employing multiple frequency averaging above 1000 Hz may lead to more universally accepted audiometric criteria for ototoxicity, which should in turn lead to more reliable reports of disease incidence.

Key Words: Cisplatin, ototoxicity, pure-tone thresholds, Receiver Operating Characteristic, Signal Detection Theory

Although the effects of cisplatin (CDDP) chemotherapy on the auditory system are well documented, considerable disagreement exists in the literature with respect to incidence and severity of CDDP ototoxicity. Van der Hulst et al (1988) reported incidence rates ranging from 4% to 91%, but pointed out the difficulty in comparing studies with different audiometric definitions of ototoxicity. Many studies define an ototoxic change if serial audiometry reveals a reduction of hearing acuity of 15 dB or greater at any single test frequency or a reduction of 10 dB or greater at any two test frequencies (van der Hulst et al, 1988). A more conservative criterion for ototoxic change reported by Dreschler et al (1985) is a 20 dB or greater decrement at a single frequency, 15 dB at two or more frequencies, or 10 dB or greater at four or more frequencies. Other confounding variables to determination of cisplatin ototoxicity include differing dose protocols and individual susceptibility for possible risk factors such as pre-existing hearing loss, age, and kidney function (van der Hulst et al, 1988).

To date, no study has examined systematically the performance of serial audiometry in detecting cisplatin ototoxicity by taking normal audiometric test-retest variability into account. Such a study would need to examine the incidence of hearing threshold shifts in a control or comparison group receiving no known ototoxic agents. In a study designed to simulate an aminoglycoside "control" group, Brummett and Morrison (1990) found "incidence rates" ranging from 20% to 33% in a group of 20 subjects receiving no known ototoxic agents. Their audiometric criteria for ototoxicity were a worsening of hearing of 15 dB or greater at two or more test frequencies, or 20 dB or greater at a single test frequency. These authors concluded that the incidence of aminoglycoside ototoxicity is likely over estimated in the literature for failure to account for normal variability in hearing threshold levels resulting from serial audiometry.
The purpose of the present study was to account for normal variability in serial audiometry in the assessment of audiometric test performance in the detection of CDDP ototoxicity. The study was conducted to answer the following questions:

1. What is the reliability/variability of serial audiometric thresholds in a comparison group of patients receiving no cisplatin or any other known ototoxins?
2. How do Receiver Operating Characteristic (ROC) curves compare various ototoxic shift criteria in terms of single and multiple (averaged) frequency shift combinations of different shift magnitudes?

The ROC curve is an analysis paradigm borrowed from Signal Detection Theory (SDT). For detailed discussions of this theory, refer to Swets (1964), Green and Swets (1966), Egan (1975), or Gelfand (1990). A brief explanation of terms and techniques borrowed from SDT follows. Additional explanation can be found (where appropriate) in the Analysis section of this paper.

Signal Detection Theory (SDT) has been applied in classical psychophysics to quantify and characterize subject sensitivity and response bias. SDT questions the traditional concept of threshold in favor of the detectibility of a signal in the presence of background noise. In the case of audiometry, threshold will vary depending upon the status of background noise in the test environment and also as a function of biological noise generated by the test subject. Turner (1991) adapted this concept to characterize audiometric test performance for detection of cochlear versus retrocochlear site-of lesion, and it is Turner's adaptation of SDT that provides the basic nomenclature and analysis paradigms of this study. Original discussions of this application can be found in Turner and Nielsen (1984) and Turner et al (1984a, 1984b).

The "signals" of interest in this study are ototoxic shifts due to CDDP therapy. The confounding "noise" is normal variability of hearing threshold levels resulting from serial audiometry. The purpose of serial audiometry in studies of ototoxicity is to extract these signals (true ototoxic shifts) from the noise (hearing shifts attributable to normal test-retest variability alone). In analogous terms more germane to the characterization of test performance, the purpose of serial audiometry is to detect true positive ototoxic shifts (hits) from a background of false positive hearing shifts (false alarms), which are known to occur in subjects receiving no known ototoxins. Audiometric test performance can indeed be quantified from the Hit Rate (%) and False Alarm Rate (%) and visually characterized by Receiver Operating Characteristic (ROC) curves, which plot Hit Rate as a function of False Alarm Rate.

MATERIALS AND METHOD

Subjects

Subjects were selected from a data base of head and neck cancer patients seen by the Otolaryngology and Audiology Service of Har- per Hospital from April, 1987 through Septem- ber, 1990. Normal audiometric test protocol was to perform baseline audiometry for all patients and follow-up for all patients receiving ototoxic chemotherapy. Typically, periodic audiometric testing was continued for all patients until a decision was made regarding whether or not to include chemotherapy as part of their treatment program. Some patients received surgery and/or radiation but no chemotherapy. A total of 125 patients had received at least one audiometric test. From this group, a total of 21 patients who received no cisplatin treatment over the first three serial audiograms were selected. Most of these patients never received chemotherapy, and audiometric testing did not continue past the third test. This group became the comparison group. A total of 31 subjects receiving cisplatin chemotherapy over the first three serial audiograms was selected as the cisplatin group.Subjects were eliminated from the cisplatin group if chemotherapy began prior to the baseline audiogram. Subjects were eliminated from both cisplatin and comparison groups if audiometric data indicated conductive hearing loss as evidenced by tympanometry (Type B) and air-bone gaps in excess of 10 decibels.

Slightly over half the subjects from each group reported occupations with potentially hazardous noise exposure. Demographic data for age, sex, and noise exposure potential are presented in Table 1.

Audiometric Testing

Pure-tone air- and bone-conduction thresholds were measured in the conventional frequency range by means of the Carhart and Jerger (1959) modification of the Hughson-
Table 1 Subject Demographics at Baseline

<table>
<thead>
<tr>
<th>Subject Demographics</th>
<th>Comparison Group (n=21)</th>
<th>Cisplatin Group (n=31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (yr)</td>
<td>55.1</td>
<td>57.9</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>12.0</td>
<td>11.1</td>
</tr>
<tr>
<td>Minimum Age</td>
<td>26</td>
<td>35</td>
</tr>
<tr>
<td>Maximum Age</td>
<td>77</td>
<td>93</td>
</tr>
<tr>
<td>Males</td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td>Females</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Potential Noise Exposure</td>
<td>11</td>
<td>17</td>
</tr>
</tbody>
</table>

Westlake ascending technique. Air-conduction thresholds were obtained at octave intervals from 250 to 8000 Hz, and bone-conduction thresholds were measured from 250 to 4000 Hz. The audiometric tests were administered by a staff of three audiologists with three different audiometers, but repeated tests for each patient were conducted with the same audiometric equipment. The audiometers used were a GSI 16 with TDH 50P earphones, a GSI 1701 with TDH 39 earphones, and a Beltone 2000 with TDH 50P earphones for air-conduction testing. Bone-conduction thresholds on each audiometer were obtained with Radioear B-71 bone vibrators. All threshold tests were conducted in prefabricated two-room testing suites, with the audiologist in the control room and the patient in a double-walled sound-treated test booth (IAC 1200 series). Ambient noise level in the testing rooms was sufficiently low so as not to interfere with threshold measurement, meeting the tolerances of ANSI S3.1 (1977). Calibration of the audiometers was checked weekly with appropriate equipment during the course of the study. The air- and bone-conduction systems for each audiometer remained in conformity with ANSI S3.6 (1989) and ANSI S3.26 (1981) specifications.

All subjects received three serial audiograms. The first audiogram (T1) was considered as baseline. Comparison subjects (n = 21) received their second and third audiograms (T2 and T3) with no cisplatin treatments. Mean elapsed time (in days) from T1-T2 and T1-T3 for this subject group was 80.7 (SD = 63.2) and 215.7 (SD = 114.4) days, respectively. Cisplatin subjects (n = 31) had no courses of cisplatin at baseline (T1). Mean elapsed time (in days) from T1-T2 and T1-T3 for this subject group was 48.1 (SD = 49.1) and 109.8 (SD = 87.0) days, respectively. Mean cumulative dosages (mg/m²) of cisplatin for these subjects were 307.2 (SD = 131.2) at T2 and 459.1 (SD = 153.3) at T3. A summary of testing schedules (days between serial audiograms) and cumulative dosages is presented in Table 2.

Analysis

Statistical analysis followed four major avenues: (1) determining the reliability and variability of serial audiometric data for the comparison group, (2) comparing pure-tone threshold shifts both within and between groups for single and multiple frequency shifts, (3) constructing ROC curves for retest shift populations to compare the different audiometric criteria, and (4) applying the audiometric criteria selected from ROC analysis to the study population of subjects.

Comparison group reliability was measured directly with correlation coefficients. Variability (stability) was measured by mean retest differences in audiometric thresholds, stand-

Table 2 Summary of Accumulated Dosage of CDDP and Testing Schedule

<table>
<thead>
<tr>
<th></th>
<th>T1 Baseline Accumulated Dose CDDP (mg/m²)</th>
<th>Days</th>
<th>T2 Accumulated Dose CDDP (mg/m²)</th>
<th>Days</th>
<th>T3 Accumulated Dose CDDP (mg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison Group (n=21)</td>
<td>0</td>
<td>Days</td>
<td>0</td>
<td>Days</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>80.7</td>
<td>0</td>
<td>215.7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>63.2</td>
<td>0</td>
<td>114.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Minimum</td>
<td>1</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>265</td>
<td>0</td>
<td>460</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cisplatin Group (n=31)</td>
<td>0</td>
<td>Days</td>
<td>0</td>
<td>Days</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>48.1</td>
<td>0</td>
<td>109.8</td>
<td>0</td>
<td>459.1</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>49.1</td>
<td>0</td>
<td>87.0</td>
<td>0</td>
<td>153.3</td>
</tr>
<tr>
<td>Minimum</td>
<td>8</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>190</td>
</tr>
<tr>
<td>Maximum</td>
<td>293</td>
<td>0</td>
<td>364</td>
<td>0</td>
<td>750</td>
</tr>
</tbody>
</table>
ard errors of the mean, 95 percent confidence intervals, and standard deviations of the retest difference scores. This last statistic (SD of the difference scores) was used by Dobie (1983) to characterize retest variability in audimetric data from industrial hearing conservation programs and seemed appropriate for this study.

Mean audimetric thresholds and shifts were computed for all retests, groups, and audimetric test frequencies. Within group mean differences were analyzed with multiple t-tests for correlated means (Linton and Gallo, 1975) for each individual test frequency (octave intervals from 250–8000 Hz) and averaged frequency combinations. Two-frequency combinations of interest were: (1) 250 and 500 Hz, (2) 1000 and 2000 Hz, and (3) 4000 and 8000 Hz. Three-frequency combinations of interest were: (1) 250, 500, and 1000 Hz, and (2) 2000, 4000, and 8000 Hz.

Construction of ROC curves for the retest shift populations was based upon operational definitions for hits and false alarms. A hit was operationally defined as any criterion shift in the cisplatin group, and a false alarm was operationally defined as any criterion shift in the comparison group. As this method simply analyzed threshold shifts per se as a function of single and multiple frequencies, the single-frequency shift population was 504 for the comparison group (21 subjects × 2 ears × 6 frequencies × 2 retests) and 744 for the cisplatin group (31 subjects × 2 ears × 6 frequencies × 2 retests). This shift population grew commensurately smaller with two- and three-frequency averaging. Subsequent to ROC analysis of the shift population, the “best” single-, double-, and triple-frequency criteria were selected for application to the patient population.

### Table 3 Summary of Reliability and Variability Indices for Comparison Subjects

<table>
<thead>
<tr>
<th>Test Frequency (Hz)</th>
<th>Audiometric Tests Compared</th>
<th>Mean Difference (dB HTL)</th>
<th>Standard Error of the Mean</th>
<th>Standard Deviation of Difference Thresholds</th>
<th>95% Confidence Interval</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>T1-T2</td>
<td>-1.31</td>
<td>1.40</td>
<td>6.2</td>
<td>5.8</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-2.38</td>
<td>2.16</td>
<td>9.7</td>
<td>9.0</td>
<td>0.63</td>
</tr>
<tr>
<td>500</td>
<td>T1-T2</td>
<td>-1.79</td>
<td>1.36</td>
<td>6.1</td>
<td>5.7</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-2.02</td>
<td>1.73</td>
<td>7.7</td>
<td>7.2</td>
<td>0.73</td>
</tr>
<tr>
<td>1000</td>
<td>T1-T2</td>
<td>-1.90</td>
<td>1.13</td>
<td>5.1</td>
<td>4.7</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>0.00</td>
<td>1.37</td>
<td>6.1</td>
<td>5.7</td>
<td>0.88</td>
</tr>
<tr>
<td>2000</td>
<td>T1-T2</td>
<td>-0.12</td>
<td>1.07</td>
<td>4.8</td>
<td>4.5</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-1.19</td>
<td>1.23</td>
<td>5.5</td>
<td>5.1</td>
<td>0.95</td>
</tr>
<tr>
<td>4000</td>
<td>T1-T2</td>
<td>-1.19</td>
<td>1.81</td>
<td>8.1</td>
<td>7.6</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-0.12</td>
<td>1.71</td>
<td>7.7</td>
<td>7.1</td>
<td>0.92</td>
</tr>
<tr>
<td>8000</td>
<td>T1-T2</td>
<td>-2.02</td>
<td>2.08</td>
<td>9.3</td>
<td>8.7</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-1.79</td>
<td>1.99</td>
<td>8.9</td>
<td>8.3</td>
<td>0.93</td>
</tr>
<tr>
<td>250-500</td>
<td>T1-T2</td>
<td>-1.55</td>
<td>1.27</td>
<td>5.7</td>
<td>5.3</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-2.20</td>
<td>1.88</td>
<td>8.4</td>
<td>7.8</td>
<td>0.70</td>
</tr>
<tr>
<td>1k–2k</td>
<td>T1-T2</td>
<td>-1.01</td>
<td>1.00</td>
<td>4.5</td>
<td>4.2</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-0.96</td>
<td>1.21</td>
<td>5.4</td>
<td>5.0</td>
<td>0.93</td>
</tr>
<tr>
<td>4k–8k</td>
<td>T1-T2</td>
<td>-1.61</td>
<td>1.78</td>
<td>8.0</td>
<td>7.4</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-0.95</td>
<td>1.69</td>
<td>7.6</td>
<td>7.1</td>
<td>0.93</td>
</tr>
<tr>
<td>250–500–1k</td>
<td>T1-T2</td>
<td>-1.67</td>
<td>1.09</td>
<td>4.9</td>
<td>4.6</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-1.47</td>
<td>1.63</td>
<td>7.3</td>
<td>6.8</td>
<td>0.79</td>
</tr>
<tr>
<td>2k–4k–8k</td>
<td>T1-T2</td>
<td>-1.11</td>
<td>1.37</td>
<td>6.1</td>
<td>5.7</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>T1-T3</td>
<td>-0.67</td>
<td>1.34</td>
<td>6.0</td>
<td>5.6</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Data are presented for each single- and multiple-frequency combination computed from baseline (T1) to test 2 (T2) and test 3 (T3); n = 21.
A second operational definition of hits and false alarms was necessary for application of the selected shift criteria to the patient population. Hits were defined by confirmatory shifts of 10 dB or greater (for the worse) from baseline (T1) to T2 and T3, while false alarms were operationally defined as any such shift occurring on either T2 or T3. Thus, a threshold shift occurring on T2 had to remain or become worse on T3 in order to be classified as a “hit.” Finally, the statistic A’ (Turner, 1991) was computed as a measure of audiometric test performance for each test criterion applied to the patient population.

RESULTS

Reliability of Comparison Group

Table 3 summarizes indices of reliability and variability for the comparison group (n = 21). Serial audiograms are compared from baseline (T1) to Tests 2 and 3 (T2 and T3) for each single-frequency and two- and three-frequency combinations. No significant differences were found for mean thresholds at any single- or multiple-frequency combination. Standard error and standard deviation of difference scores compare favorably with data reported by Dobie (1983), and 95 percent confidence intervals compare well with those reported by Brummett and Morrison (1990). Finally, retest correlation coefficients indicated strong, positive relationships for baselines and retests for the comparison group.

Hearing Shifts: Comparison and Cisplatin Group

Figure 1 depicts the distribution of hearing shifts at all single test frequencies for both comparison (COM) and cisplatin (CIS) groups. Hearing shifts for the COM group were essentially distributed normally, with the majority of positive and negative shifts falling within the ±5 dB criterion that traditionally has been considered to be acceptable test-retest reliability. The shift distribution for the CIS group essentially overlaps that of the COM group, with the exception that far more negative shifts equal to or greater than 20 dB occurred for the

Table 4 Percentage of Shifts Meeting All Frequency Averaging and Shift Magnitude Criteria

| Shift (dB) | ANY | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS | COM CIS |
|-----------|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| ANY       |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| COM CIS   |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| CIS       |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 25-5      |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 1-2       |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 4-8       |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 25-5-1    |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 2-4-8     |     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |

COM = Comparison group; CIS = Cisplatin group.
Cisplatin group. The data plotted in Figure 1 correspond to the ANY column from Table 4, which also contains the shift distributions calculated for all other single-and multiple-frequency shift combinations. Note that negative signs indicate a worsening of hearing levels.

Table 5 summarizes mean audiometric thresholds for all groups, audiometric frequencies, and retests. No significant differences from baseline were found for COM group retests. Significant differences were found, however, for the CIS group, with most differences found for the higher individual and grouped frequencies.

Figure 2 depicts mean shifts from baseline for COM and CIS groups. Comparison group shifts were minimal for all test frequencies. Cisplatin group shifts were minimal for frequencies below 2000 Hz but dramatic for 2000 Hz or higher.

Figures 3 to 5 provide ROC comparison of single- and multiple-frequency test criteria on the population of retest shifts. Each curve represents a single or multiple (averaged) frequency combination, while each data point on a particular curve depicts the percentage of hits and false alarms for a specific shift magnitude criterion. Data points extending to the upper left hand corner of each figure demonstrate more effective criteria, as hits are maximized while false alarms are minimized. Figure 3 compares audiometric test performance for single-frequency shifts, with the curve marked “ANY” representing a shift at any of the six

Figure 2  Mean single-frequency hearing threshold level shifts from baseline for Comparison (COM) (n = 21) and Cisplatin (CIS) (n = 31) groups at Test 2 (T2) and Test 3 (T3). Negative numbers indicate a worsening of hearing thresholds.

Figure 3 Receiver Operating Characteristic (ROC) curves comparing audiometric criteria for single-frequency shifts. Each curve represents a different single-frequency criterion. Data points on each curve represent shift magnitude criteria (−5, −10, −15, and −20 dB). The curve marked “ANY” depicts criteria for shifts occurring at any single test frequency. Hits and false alarms were operationally defined from the population of retest shifts.
frequencies examined. Figure 4 compares criteria involving multiple-frequency (two- and three-frequency average) shifts. Figure 5 gives an enhanced view of the “best” single-, double-, and triple-frequency performers. This figure focuses upon the shift magnitudes of interest of this study (i.e., -10, -15, and -20 dB).

Figure 6 applies the “best” single-, double-, and triple-frequency criteria from shift analysis to the patient population, again focusing upon the shift magnitudes of interest of this study (i.e., -10, -15, and -20 dB). Finally, the appropriate A’ statistic (Turner, 1991) appears in parenthesis adjacent to each data point in Figure 6.

DISCUSSION

These results tend to confirm reports in the literature that hearing shifts for 100+ mg/m² bolus injection dosages of CDDP cause dramatic shifts in the higher audiometric test frequencies (Kopelman et al, 1988; Laurell and Jungnelius, 1990). Of more immediate interest to this study, however, were the effects of test criteria on hit and false alarm rates.

Comparison Group Reliability

Brummett and Morrison (1990) discussed the issue of normal variability in serial audiometric testing as it relates to determining shift magnitude criteria for detecting ototoxicity. Dobie (1983) provided a similar but more detailed discussion of this issue for determining shift criteria for noise-induced hearing loss. Although the majority of hearing shifts in a control group will fall within the accepted ±5 dB criterion for retest reliability, a certain portion of the shifts will be 10, 15, or 20 dB in magnitude. The comparison group in this study yielded 70 percent of retest shifts within the ±5 dB limit. Eighteen percent of retest shifts for this group were 10 dB or greater in a negative direction, however.

Comparison group variability as measured by standard deviation of the difference scores compares favorably with those reported by Dobie (1983). It is interesting to note the rather systematic increase in these scores from T2 to T3, however, suggesting increased variability with
increased retest intervals. Dobie (1983) and a number of other investigators have suggested that audiometric variability is increased as retest time intervals are increased. It would seem reasonable to assume, therefore, that normal variability in the cisplatin group would be somewhat less than that of the comparison group, as the cisplatin group was retested in a more timely fashion than the COM group.

Comparison group variability as measured by 95 percent confidence intervals compares well with data reported by Brummett and Morrison (1990). In general, confidence intervals become larger from T2 to T3 and also increase as a function of test frequency. These results confirm reports by Dobie (1983) and Brummett and Morrison (1990) that variability increases above test frequencies of 1000 Hz. Confidence intervals also tend to increase with increased test-retest time intervals.

To summarize, all comparison group measures of reliability or variability were judged to be within acceptable limits to rule out unusual or inordinate variation from this group that would preclude comparisons to the cisplatin group.

Hearing Shifts—Signals or Noise?

The population of retest shifts for single frequencies (see Fig. 1) provides a basis for a discussion of audiometric test performance from a Signal Detection Theory (SDT) standpoint. The comparison distribution can be considered to be background noise or hearing shifts due to normal variability. The cisplatin distribution can be considered to consist of signal (otoxic shifts) plus the background noise. Visual scrutiny of the overlap of the two distributions in Figure 1 gives some indication of the difficulty in determining any single frequency predictor for maximizing the probability of a “hit” while minimizing the probability of a “false alarm.” Certainly, Figure 1 would discourage the often-used criteria of 10 or 15 dB shifts at any single test frequency.

Careful scrutiny of the data provided in Table 4, however, reveal signal or noise distributions far more favorable to the detection of otoxic shifts. In particular, when only shifts at 8000 Hz are considered, we find that 48 percent of the CIS group trials exhibited shifts of -20 dB or greater, whereas only 8 percent of the COM group exhibited similar shifts. Frequency averaging, as reported by Dobie (1983), tends to smooth out differences between the COM and CIS groups by lowering both Hit Rate and False Alarm Rate. The far right-hand columns of Table 4 illustrate this principle. Note that two-frequency averaging of shifts for 4 and 8 kHz reduces the false alarm rate to 0 percent while also reducing the hit rate to 40 percent for -20 dB shifts. Computing three-frequency average shifts for 2, 4, and 8 kHz further reduces the hit rate to 35% for -20 dB shifts, while maintaining a 0 percent false alarm rate.

The degree to which noise exposure could be a contaminating variable was judged to be minimal. Slightly over 50 percent of each group reported occupations generally associated with noise exposure potential, and no statistically significant threshold shifts were observed in the comparison group.

Single-versus Multiple-Frequency Test Criteria

Dobie (1983) found frequency averaging techniques to be superior to single-frequency shift criteria for detecting potentially noise-induced hearing loss in industrial populations. In general, single-frequency shift criteria produced an unacceptable rate of false alarms. This unacceptable number of false alarms for single-frequency shifts also was reported by Gasaway (1985) for a large military hearing conservation data base.

In general, the limitations of single-frequency shift criteria are borne out by this study. Figures 5 and 6 demonstrate systematic reductions to both hit and false alarm rate with increased frequency averaging. The question remains for the clinician, however, to determine what shift magnitudes will be accepted as “real” when balancing the costs and benefits of administering potentially otoxic medication. Ultimately, a criterion adopted for clinical use must account for acceptable hit and false alarm rates within the framework of the overall medical and communicative needs of the patient.

The statistic A' (Turner, 1991) is computed from hit and false alarm rate for each data point (in parenthesis) in Figure 6, indicating that a 20-dB shift at 8 kHz outperforms all other test criteria when applied to the patient population. These data should be interpreted with caution, however, considering the limited number of subjects involved in this study and very similar A' statistics computed for other data points employing frequency averaging. Perhaps more importantly, these data should be interpreted with caution due to operational definitions of
hits and misses, as this study was conducted
without benefit of a "gold standard" for oto-
toxicity. More data need to be examined before
this question can be answered reasonably.

It is interesting to note, however, that the
relative location in ROC space for curves repre-
senting the "best" single-, double-, and triple-
frequency performers is essentially unchanged
for two different operational definitions of hits
and false alarms (see Figs. 5 and 6). The gener-
ally higher hit rates found in Figure 6 result
from the division of the cisplatin group into
"diseased" and "nondiseased" patients based
upon confirmatory threshold shifts on T2 and
T3 and are likely more reflective of actual
audiometric test performance than Figures 3 to
5, which examine threshold shifts per se, not
the subject population.

A possible clinical application of frequency
averaging techniques may be determining ab-
breviated audiometric test protocols for pa-
tients undergoing CDDP chemotherapy. Rout-
line monitoring of the lower audiometric fre-
quencies (i.e., 1000 Hz and below) may not be
warranted unless certain magnitude shifts are
noted at 2000 Hz and above. Or perhaps some
combination of high-frequency pure-tone and
speech audiometry would be appropriate for
monitoring these patients. Many of the patients
are very ill and are unmotivated to take a full
audiometric test. Perhaps test modification
based upon ROC analysis and frequency aver-
ing techniques will increase the utility of serial
audiometry in detecting the effects of
otoxicity while making the procedure more
tolerable to the patient.

Further ROC modeling of audiometric test
criteria for ototoxicity will likely lead to more
universally accepted shift criteria, which will,
and, in turn, lead to more reliable estimates of
disease incidence or prevalence. Ultimately, more
reliable estimates of incidence will benefit cli-
icans in making decisions regarding ototoxic
therapy regimens.

Preliminary results for extended high-fre-
quency audiometry on a limited number of
patients suggests that audiometric criteria for
CDDP ototoxicity involving multiple-frequency
averaging above 8000 Hz provides the earliest
and most dramatic evidence of hearing damage.
Also, analyses on subsets of the data presented
here whereby 3000 and 6000 Hz also were
tested indicates that multiple-frequency aver-
ing of adjacent frequencies in the 3000 to
8000 Hz range may prove to be equally useful in
reliably monitoring the ototoxic effects of CDDP
therapy.

CONCLUSIONS

1. Single-frequency audiometric criteria for
otoxicity tend to increase false alarm
rates; whereas, multiple-frequency av-
erging tends to reduce false alarm rates.

2. Audiometric variability increases with
prolonged retest intervals and to some
extent with increasing test frequency.

3. Further investigation employing Re-
ceiver Operating Characteristic anal-
ysis and multiple-frequency averaging
may result in more universally accepted
audiometric criteria for ototoxicity,
which should in turn lead to more reli-
able reports of disease incidence.

Acknowledgment. This paper was presented to the
Michigan Speech-Language and Hearing Association Con-
Supported by NIH-NCI Grant PO1 CA43838.
The authors thank Drs. Frank Musiek and Robert
Turner for their helpful comments concerning this manu-
script.

REFERENCES

for Permissible Ambient Noise during Audiometric Test-
ing. ANSI S3.1-1977. New York: ANSI.

Equivalent Threshold Force Levels for Audiometric Bone

American National Standards Institute. (1989). Specifi-
cation for Audiometers. ANSI S3.6-1989. New York: ANSI.

Brummett RE, Morrison RB. (1990). The incidence of
aminoglycoside antibiotic-induced hearing loss. Arch

Carhart R, Jerger JF. (1959). Preferred method for clini-
cal determination of pure tone threshold. J Speech Hear
Disord 24:330-345.

audiometry: implications for hearing conservation pro-
grams. Laryngoscope 93:906-927.

Dreschler WA, van der Hulst RJAM, Tangs RA, Urbanus
NAM (1985). The role of high frequency audiometry in

Egan JP. (1975). Signal Detection Theory and ROC Analy-

Gasaway DC. (1985). Adopting a scheme for identifying
significant threshold shifts. In: Gasaway DC, ed. Hearing
Conservation: A Practical Manual and Guide. Englewood


