Ototoxicity risk assessment combining distortion product otoacoustic emissions with a cisplatin dose model[a)]

Marilyn F. Dille[b)] and Garnett P. McMillan
National Center for Rehabilitative Auditory Research, VA Medical Center, 3710 SW US Veterans Hospital Road, Portland, Oregon 97239

Kelly M. Reavis
Department of Otolaryngology/Head & Neck Surgery, Irvine Medical Center, University of California, Building 56, 101 The City Drive South, Orange, California 92868

Peter Jacobs, Stephen A. Fausti, and Dawn Konrad-Martin
National Center for Rehabilitative Auditory Research, VA Medical Center, 3710 SW US Veterans Hospital Road, Portland, Oregon 97239

(Received 4 February 2010; revised 2 July 2010; accepted 6 July 2010)

An objective method for identifying ototoxic hearing loss among patients receiving cisplatin is necessary since the ability of patients to take a behavioral test may change over the course of treatment. Data from 56 monitoring visits by 19 Veterans taking cisplatin were used to identify combinations of distortion-product otoacoustic emission (DPOAE) metrics and ototoxicity risk factors that best identified ototoxic hearing loss. Models were tested that incorporated DPOAE metrics generated statistically using partial least-squares analysis. Models were also tested that incorporated *a priori* DPOAE change criteria, such as a minimum DPOAE level shift of 6 dB. Receiver Operating Characteristic analysis was used to compare the accuracy of these models. The best performing model incorporated weighted combinations of pre-treatment hearing, cumulative cisplatin dose and DPOAE metrics that were determined using partial least-squares and evaluated over a quarter octave range near each subjects' high frequency DPOAE limit. Using this model and the DPOAE recording methods described herein, the chance of ototoxic hearing change can be determined at any given observed change in DPOAE level. This approach appears to provide an accurate and rapid ototoxicity risk assessment (ORA) that once validated can be used clinically.
[DOI: 10.1121/1.3473693]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Ri, 43.64.Bt [BLM] Pages: 1163–1174

I. INTRODUCTION

A cornerstone of treatment for many cancers is the administration of cisplatin, an antineoplastic chemotherapeutic agent. In adults, cisplatin is used to treat a variety of cancers, including bladder, testicular, ovarian, head and neck, lung, and cervical cancer and is higly effective. However, a dose-limiting side effect of cisplatin administration is hearing impairment, termed ototoxicity.

Ototoxicity can present unilaterally or bilaterally of varying degrees or as an aggravation of an existing hearing loss. The amount of the ototoxic hearing change is dependent primarily on cisplatin dose, but there is marked individual variability related to other subject risk factors (Fischel-Ghodsian *et al.*, 1993; Forge and Schacht, 2000). Left untreated, ototoxic hearing changes can be associated with depression, anxiety, decreased participation in communication opportunities and/or stress on intimate relationships (Kochkin and Rogin, 2000). Further, hearing changes can lead to decreases in health literacy at a time when health-related decisions are vitally important (Dalton *et al.*, 2003; Amalraj *et al.*, 2009). Despite major impacts on quality of life, hearing loss tends to be overlooked by the sufferer, as well as under-treated by health professionals, particularly when it coincides with a disease that threatens a patient's general health (Durrant *et al.*, 2005).

Early detection of ototoxic hearing loss provides the medical team an opportunity to reduce dosages, change treatment regimens and/or change to lesser toxic medication to mitigate hearing changes. If hearing changes are unavoidable, early detection provides the audiologist an opportunity to educate the patient and the family regarding hearing loss, its effects and management.

Traditionally, early detection is achieved by the monitoring of pure-tone thresholds at each chemotherapy patient-visit. However this has drawbacks. The main drawback is that more than one-third of patients who receive ototoxic medications are, at some point in treatment, unable to be tested using behavioral techniques (Fausti *et al.*, 1991, 1992) either because of the effects of the disease or the side-effects of treatment. There is a clear need for an objective monitor-

[a)]Portions of this work were presented at the American Auditory Society Meeting, Scottsdale, Arizona, March 2009 and by invitation at the American Academy of Audiology AudiologyNOW! Meeting, Dallas, Texas, April 2009 and the National Center for Rehabilitative Auditory Research 2009 NCRAR International Conference, Portland, OR.

[b)]Author to whom correspondence should be addressed. Electronic mail: marilyn.dille@va.gov

ing strategy for all patients and, in particular, for those who cannot be tested using traditional behavioral means.

Distortion-product otoacoustic emission (DPOAE) testing holds promise as an excellent objective ototoxic monitoring technique. It is generally accepted that DPOAE generation depends on the physiological status of the outer hair cells, which are the auditory receptor cells damaged first by most ototoxic drugs (Hodges and Lonsbury-Martin, 1998; Alam *et al.*, 2000; Laurell *et al.*, 2000; Mukherjea *et al.*, 2006). Because DPOAEs are likely by-products of the same mechanism that provides the extreme threshold sensitivity and frequency selectivity that characterizes the normal auditory system, DPOAEs have been used clinically to identify the presence of hearing loss in hard-to-test populations. DPOAEs are diminished in ears with mild to moderate hearing losses up to approximately 50–60 dB HL, and are typically absent with more severe hearing losses (Gorga *et al.*, 1996, 1997). These relationships also support the application of DPOAEs for monitoring ototoxicity. Moreover, several reports show that DPOAEs can be successfully monitored even in older patients in whom normal pre-exposure hearing cannot be assumed (Ress *et al.*, 1999; Reavis *et al.*, 2008).

Previously, DPOAE level differences have been noted between pre- and post-ototoxin exposure in both adults and children (Ress *et al.*, 1999; Stavroulaki *et al.*, 2002; Knight *et al.*, 2007). Some studies have compared rates of change for DPOAEs and pure tone thresholds in exposed populations. In general, comparatively fewer changes were found using pure tone thresholds measured at conventional audiometric frequencies (Ress *et al.*, 1999; Knight *et al.*, 2007). However, similar proportions of exposed patients had changes in DPOAEs and ultra-high frequency hearing thresholds (Ress *et al.*, 1999; Stavroulaki *et al.*, 2002; Knight *et al.*, 2007; Reavis *et al.*, 2008). Previous work suggests then, that prospective testing using DPOAE may detect ototoxic changes in auditory function, but it does not provide a basis for interpreting DPOAE changes in relation to a gold standard measure of ototoxicity. Interpretation of results from an ototoxicity monitoring test requires an understanding of the overall test accuracy.

Just as DPOAEs have previously been compared to the audiogram in order to determine their ability to identify hearing impairment, changes in DPOAE level can be compared to hearing changes that meet gold standard criteria for ototoxic hearing change, such as those criteria recommended by ASHA (1994). A drawback of this approach is that it assumes any disagreement between DPOAE and hearing results is a diagnostic error associated with the DPOAE measurements, which may not be true. Audiometric testing is done using relatively coarse (5 dB) steps and clinical thresholds are associated with test-retest variability in the 5 to 10 dB range depending on test frequency. There is also the potential for DPOAE changes to signal subclinical changes in auditory function that do not yet cause a hearing change. However, advantages of this approach include a comparison of true positive rates and false positive rates for various DPOAE metrics and cut-off criteria. Perhaps more importantly, this approach can be used to determine the probability that any DPOAE change is arising from an ear with changed hearing.

There is evidence that suggests DPOAEs at widely spaced frequencies can give an inaccurate picture of the corresponding audiogram because constructive and destructive interactions can occur between the distortion and reflection source emissions comprising the ear canal recorded DPOAE (Mauermann and Kollmeier, 2004). DPOAEs collected with smaller step-sizes (fine-structure paradigm) and then smoothed may lead to better correspondence with behavioral hearing change.

The goal of the current study was to develop a diagnostic method that uses changes in DPOAE levels collected with fine-frequency step sizes to distinguish ears with ototoxic hearing shift from those with stable hearing. A diagnostic method was developed, based on weighted combinations of several DPOAE metrics and a Dose-Hearing model which incorporates pre-treatment hearing and cisplatin dose that best identified ototoxic hearing shift in Veterans taking cisplatin. For comparison, we also evaluated test accuracy for a single DPOAE criterion value, a 6 dB change in DPOAE level, because this value is frequently suggested for DPOAE ototoxicity monitoring applications (Franklin *et al.*, 1992; Roede *et al.*, 1993; Beattie *et al.*, 2003). We hypothesized that DPOAE test performance would be optimized by a multivariate approach and by applying statistical methods to select candidate DPOAE metrics.

II. METHODS

A. Sample

Subjects receiving cisplatin for the treatment of cancer were recruited from the Portland Veteran Affairs Medical Center. A list of patients prescribed cisplatin generated by the Chemotherapy Unit was used to identify potential subjects for the study. Patients' electronic medical charts were reviewed for the purpose of subject recruitment from October, 2007 through June, 2009. Inclusion criteria included: a) cognitively and physically able to participate, b) ability to provide reliable behavioral responses; c) hearing no worse than 70 dB HL at and below 4 kHz; d) no active or recent history of middle-ear disorder, Meniere's disease, or retrocochlear disorder; and e) willingness to participate in the study. All subjects were consented to participate in the study following the guidelines of the medical center's Institutional Review Board and were compensated for their time.

B. Testing schedule

Subjects completed a battery of tests at the baseline session and during follow-up visits that included questions regarding tinnitus and vertigo onset or changes, otoscopy, immittance testing, behavioral audiometry, and DPOAE testing. Out of time considerations, DPOAE testing was done in one ear only chosen either as the better hearing ear or, in the case of symmetrical hearing, by coin toss. Baseline testing was performed within 24 h of initial treatment with cisplatin. Subsequent monitoring visits were completed within 24 h of each chemotherapy treatment. The total number of patient visits (PV) and intervals between visits varied across subjects since treatment regimens depend on cancer type, patient health and other medical factors. Additionally, testing was

performed at one month after cessation of treatment. Ototoxicity monitoring protocols seek to identify presence or absence of hearing change at each patient visit. Since the VA Medical Center in Portland, OR is regional, often Veterans returned here only on the day of treatment.

C. Behavioral audiometry measurements

The gold standard for hearing change in this study was determined by serial pure tone threshold monitoring. Pure-tone thresholds were obtained using the modified Hughson-Westlake technique (Carhart and Jerger, 1959). Baseline (pre-exposure) thresholds were measured from 2–20 kHz using a Virtual Corporation, Model V320 audiometer and modified Koss Pro/4X Plus earphones. The audiometer and earphones were calibrated twice each month. Detailed descriptions of the set-up including calibration have been described elsewhere; see Fausti *et al.*, 1979.

A behavioral sensitive range for ototoxicity, SRO_{BEH}, was identified for each ear from the baseline pure-tone thresholds (2–20 kHz). The upper bound of the SRO_{BEH} was defined as the highest frequency at which a threshold could be obtained using a pure-tone signal of 100 dB SPL or less. The pure-tone thresholds of the six lower adjacent frequencies in 1/6-octave steps were then obtained. Thus, seven frequencies constituted the behavioral SRO_{BEH}, which was the frequency range tested at all monitoring visits. If a hearing change was noted within the SRO_{BEH}, then full frequency (2–20 kHz) testing was done.

Behavioral hearing change was assessed relative to the SRO_{BEH} measured at baseline. Presence or absence of behavioral hearing change was based on published clinical guidelines (ASHA, 1994) and includes: a) $\geq$20 dB change at any one test frequency; b) $\geq$10 dB change at any two consecutive test frequencies; or c) loss of response at three consecutive test frequencies where responses were previously obtained. Using these criteria, a binary indicator for presence or absence of hearing change was constructed for each post-baseline PV in the sample. This binary indictor is the gold standard against which all candidate objective measures are compared.

D. Distortion-product otoacoustic emission measurements

DPOAEs were collected using custom software [Otoacoustic Emission Averager, EMAV; Boys Town National Research Hospital; (Neely and Liu, 1994)] run on a PC. The software utilized a CardDeluxe digital signal processing board (Digital Audio Laboratories) to generate stimuli and record responses. The primary frequencies (f_1 and f_2, where $f_1<f_2$) were separately digitized, converted to analog voltages, passed through custom headphone buffers to two earphones (Etymotic Research, ER-2) and delivered to the sealed ear canal. The probe also contained a low-noise microphone (Etymotic Research, ER-10B+) to record responses. The signal was sampled at a rate of 32 kHz, amplified 20 dB by the ER-10B+ pre-amplifier, digitized in 64-ms time windows, and stored in two interleaved buffers, which were averaged in the time domain. The DPOAE level at $2f_1$-f_2 was estimated from a Fast Fourier transform of the grand average of the two response buffers ([A+B]/2). The noise level was estimated at the DPOAE frequency from the A–B spectrum. Measurement-based stopping rules were used, such that at each f_2 frequency, averaging stopped when the noise floor was <-26 dB SPL or after 32 s of artifact-free averaging, whichever occurred first. The system was electrically calibrated annually according to the EMAV manual.

DPOAEs as a function of f_2 frequency were measured first in seven half-octave steps with f_2 ranging from 2–14 kHz using a fixed primary frequency ratio $f_1/f_2=1.22$. The levels of the f1 and f2 primaries were $L_1=L_2=65$ dB SPL. The two highest half-octave steps (i.e., a one octave range) that elicited DPOAEs at +6 dB signal to noise ratio defined the individualized SRO_{DP}. DPOAEs as a function of f_2 frequency were then measured across this highest octave in 1/48-octave steps with f_2 ranging over the SRO_{DP} and sweeping from high to low frequencies. Thus the SRO_{DP} comprised 48 DPOAE measurements. Because measurement-based stopping rules were used, testing time was longer for subjects with greater physiological noise and/or DPOAE monitoring limited to frequencies below about 2 kHz. That is, a low frequency SRO_{DP}, imposed by impaired hearing at higher frequencies, increased testing time since biological noise in DPOAE measurements is greater at the lower frequencies. Total test time for octave-range frequencies measured in fine steps ranged from 20–45 min with 45 min being the typical upward limit of tolerable testing for our patients. As a result, due to time constraints, the lower half-octave of the SRO_{DP} was not always collected, resulting in missing data. Thus, this report only describes DPOAEs measures derived from the highest half-octave with valid responses.

Both DPOAE and stimulus levels were measured at the plane of the microphone near the entrance to the ear canal. In-the-ear calibration was used to adjust voltage applied to the source transducers in order to set the SPL of f_1 and f_2 to desired values. Ear canal transfer functions obtained during in-the-ear calibration for baseline recordings were employed as target calibration spectra in order to ensure consistent probe placement across PV and thus improve test-retest reliability. Recorded DPOAE levels were smoothed using a 5-point running average at every PV in order to minimize fine structure level variations.

System distortion was estimated as the DPOAE level at $2f_1$-f_2 measured in a standard 2cc cavity (Brüel and Kjær 4153 Coupler) for the frequencies and intensity levels used in the present study. For the purpose of assessing system performance, estimates of system distortion were made weekly to ensure that system distortion remained at levels less than -20 dB SPL. In order to determine whether a DPOAE response recorded at a PV was valid, ear canal DPOAE and noise level measurements were compared to the corresponding system distortion averaged across 7 separate coupler measurements.

Specifically, data from the 7 coupler runs were converted to intensity values and means and standard deviations (SD) were calculated from them. For each ear canal measure-

ment, biological noise was converted to intensity, added to the corresponding mean coupler intensity level and the combined noise and distortion value was transformed to dB SPL. The signal to noise ratio was defined as the observed DPOAE level in dB SPL minus the back transformed sum of the subject noise and system distortion in dB SPL.

For a given stimulus condition, a DPOAE ear canal response was considered valid and present if the SNR was at least 6 dB. If the SNR was less than 6 dB and if the subject noise was less than or equal to mean system distortion plus 2 SDs, the DPOAE measure was still considered valid, i.e., the low level emission was considered present, interpretable, and the measurement value was used in the analyses. If the SNR was less than 6 dB and the subject noise was greater than the mean system distortion plus 2 SDs, the DPOAE measure was set to missing, i.e., the measurement was considered uninterruptable and was not used in the analyses.

For the analyses, the highest frequency tested was $f=1$, followed by $f=2$ for the next highest 1/48-octave step, and so forth. Change in DPOAE level was computed as

$$\Delta\mathrm{OAE_f} = (\mathrm{Baseline\ \ DPOAE_f}) - (\mathrm{PV\ \ DPOAE_f}),$$
$$\mathrm{f} = 1,2,\ldots,24 \quad (1)$$

so that positive values indicate a decrease in DPOAE level at step f and negative values indicate an increase. Cisplatin is expected to cause a positive $\Delta\mathrm{OAE_f}$, corresponding to a response decrement.

E. Data analysis

The purpose of this analysis is to develop a scoring function, denoted $\mathrm{R_i}$, for the $\mathrm{i^{th}}$ PV that best distinguishes PVs with a hearing change from those without a hearing change:

$$R_i = \sum\nolimits_K w_k M_{ik}, \quad k = 1,2,\ldots,K, \quad (2)$$

where $\mathrm{M_{ik}}$ is one of K measurements taken on the $\mathrm{i^{th}}$ PV, and $\mathrm{w_k}$ are weights assigned to each metric. A single DPOAE criterion value was initially selected, a 6 dB change at any DPOAE test frequency, and denoted throughout this analysis as the '6 dB method'. Using this notation, the 6 dB method has K=1 metric, such that

$$M_{ik} = \begin{cases} 1 & \max(\Delta\mathrm{OAE}_f) \geq 6\ \mathrm{dB} \\ 0 & \text{otherwise} \end{cases} \quad \text{with} \quad \mathrm{w_k} = 1. \quad (3)$$

The 6 dB method thus simply assigns a risk score, $\mathrm{R_i}$, of 0 or 1 to each PV based solely on the largest observed $\Delta\mathrm{OAE_f}$.

The larger goal was to define $\mathrm{M_{ik}}$ more generally, including DPOAE summary metrics as well as patient characteristics and aspects of the cisplatin treatment regimen.

There were three stages to this analysis. 1) Identify candidate scoring functions, each distinguished by the metrics included in K, for diagnosing hearing change. Partial least-squares regressions were used as well as some relatively simple summary metrics to identify candidate scoring functions in stage 1. 2) Determine the weights, $\mathrm{w_k}$, to assign to each metric so that PVs with a hearing change have higher values of $\mathrm{R_i}$ than PVs without any hearing change. Logistic regression was used to establish the weights in each candidate scoring function in stage 2. 3) Compare empirically the accuracy of each scoring function against the gold standard of ASHA-significant hearing change. Receiver Operating Characteristic (ROC) curve analysis was used to assess accuracy in stage 3. An advantage to this approach is that scoring functions determined *a priori* (such as the 6 dB method) can be objectively compared against alternative scoring functions. Simon and others (Radmacher *et al.*, 2002; Simon, 2005a) provide an accessible introduction to the data analytic program used here. A schematic outline of the approach used here is shown in Fig. 1, and is described in the remainder of this section.

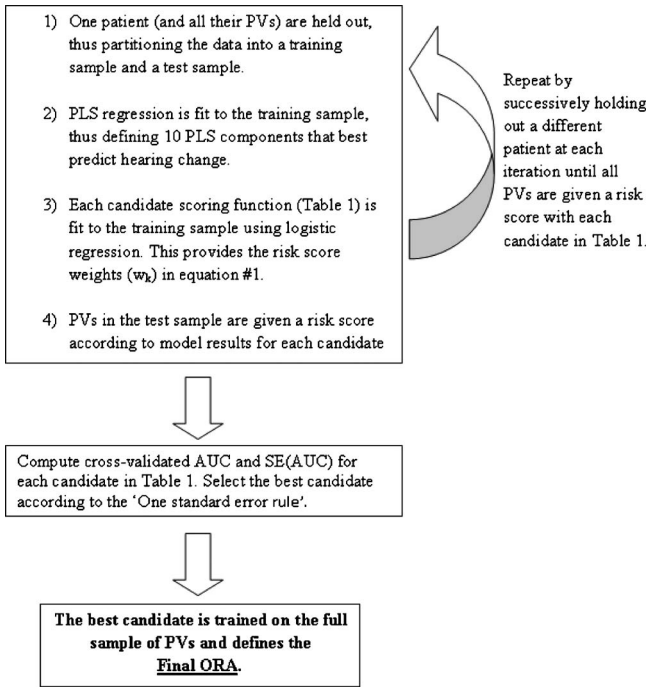

FIG. 1. Diagram of Leave One Out Cross-Validation (LOOCV) analysis from which the area under the receiver operating characteristic curve (AUC) and standard error of the AUC were computed from Table I and from which the best candidate scoring function for the Ototoxicity Risk Assessment (ORA) was determined.

1. Identify candidate scoring functions

In a previous study among Veterans (Reavis *et al.*, 2010), pre-treatment hearing (specifically, a subject's average pure tone threshold in the behavioral SRO frequencies, $\mathrm{SRO_{BEH}}$) and cumulative cisplatin dose (in mg) were found to be associated with risk of an ASHA-significant hearing shift during treatment. Therefore, prior to DPOAE analysis, a baseline model relating ASHA-significant hearing change to treatment regimen and patient features was developed using logistic regression, denoted as the Dose-Hearing model. Correlation among repeated measurements on a patient both within a monitoring appointment and at successive monitoring appointments was modeled using generalized estimating equations (Fitzmaurice *et al.*, 2004) in order to adjust for these correlations. Other potential patient and treatment risk factors for hearing loss were also tested in the same model. It

TABLE I. Candidate scoring functions and number of metrics contained within each function compared in this analysis. Each scoring function was evaluated using both the highest half-octave and highest quarter-octave DPOAE fine structure measurements in 1/48-octave steps.

No. of metrics (K)	Metrics included in the candidate ORA
1	Dose-hearing model
2	6 dB method+dose-hearing-model
2	Sum ΔOAE_f+dose-hearing model
2	Maximum ΔOAE_f+dose-hearing model
2	Mean ΔOAE_f+dose-hearing model
2	PLS component 1+dose-hearing model
3	PLS components 1–2+dose-hearing model
4	PLS components 1–3+dose-hearing model
5	PLS components 1–4+dose-hearing model
6	PLS components 1–5+dose-hearing model
7	PLS components 1–6+dose-hearing model
8	PLS components 1–7+dose-hearing model
9	PLS components 1–8+dose-hearing model
10	PLS components 1–9+dose-hearing model
11	PLS components 1–10+dose-hearing model

was found that patient age, cancer type, cancer stage and location at diagnosis, single dose volume or dose level of cisplatin, concomitant medications, and concurrent radiation therapy did not significantly alter the risks of hearing change in this sample. However, mean thresholds from the pre-treatment SRO_{BEH} and cumulative cisplatin dose were again found to be statistically important variables ($p<0.05$). More specifically, the Dose-Hearing model is defined for the standardized (mean=0; SD=1), pre-treatment SRO_{BEH} average pure-tone threshold in dBHL (B_j) and the standardized log cumulative cisplatin dose in mg (L_{ij}), as the average log(odds) of hearing change for the j^{th} subject at the i^{th} PV:

$$\text{Log(odds of hearing change}_{ij}) = -0.24 + 0.84 \cdot L_{ij} - 1.28 \cdot B_j - 1.04 \cdot B_j \cdot L_{ij}. \quad (4)$$

The DPOAE analysis begins with this Dose-Hearing model embedded within it.

There were 24 DPOAE level measurements taken during each PV at 1/48-octave steps in the highest half-octave of the SRO_{DP}. Simple summaries of change in OAE level were computed, including mean ΔOAE_f, maximum ΔOAE_f, and sum ΔOAE_f. These, in combination with the Dose-Hearing model, are among the candidate scoring functions to be compared in this analysis (Table I). The 6 dB method was also considered alone and with the Dose-Hearing model (Table I).

An alternative approach to simple summary measures is to use partial least-squares (PLS) regression to generate optimal DPOAE summary measures. PLS is suited to situations where there are many measures that are highly correlated such as, DPOAE data. As such PLS constructs new explanatory variables, denoted "components," which are linear combinations of the ΔOAE_f that best predict hearing change. In this approach, each component is defined such that it has maximum covariance with the observed hearing changes (i.e., how much these two variables change together), and is uncorrelated with previous components (Barker and Rayens, 2003). The first component accounts for the most variance in the ΔOAE_f, the second component for the second most variance, and so on. 10 uncorrelated PLS components were estimated from the sample of PVs.

Next, each uncorrelated component was sequentially added to the base Dose-Hearing model yielding 10 candidate scoring functions for assessment. Thus, the scoring functions that include the 1^{st} component, the 1^{st} and 2^{nd} components, the 1^{st}, 2^{nd}, and 3^{rd} components, and so on, constituted the remaining scoring function candidates under consideration. The set of all 15 scoring function candidates considered (Dose-Hearing model, 1 scoring function with the 6 dB method, 3 scoring functions with simple summary metrics, and 10 scoring functions with PLS components) is provided in Table I.

2. Determine the weights of each scoring function

The weights, w_k, for each combination of metrics in Eq. (2) were estimated by logistic regression. Specifically, the log-odds of hearing change at each PV were modeled as a linear function of the metrics listed in Table I. Regression coefficients from the fitted model correspond to the w_k, so that R_i in Eq. (2) is equivalent to the estimated log-odds of hearing change at each PV. Separate logistic regression models were fit for each candidate model listed in Table I.

3. Find the most accurate scoring function

Each of the scoring functions described in Table I, and their weighting schemes established using logistic regression, were compared using Receiver Operating Characteristic (ROC) curve analysis. The ROC is a plot of the true positive rate against the false positive rate for different cut-offs of the scoring function, and is the basis of most diagnostic test evaluations. The true positive rate and the false positive rate describe the accuracy of a particular candidate scoring function. The true positive rate is the proportion of PVs with a hearing change that are correctly diagnosed using the scoring function, and the false positive rate is the proportion of PVs without a hearing change that the scoring function incorrectly diagnoses with a hearing change. The true positive rate and the false positive rate depend on the cut-off risk score above which a PV would be diagnosed with hearing change. The true positive rate can be arbitrarily increased by lowering the cut-off point, but this comes at the cost of increasing the false positive rate.

The accuracy of each candidate model was succinctly estimated using the area under the ROC curve (AUC). The AUC estimates the average true positive rate over the domain of false positive rates. Higher AUC are associated with overall more accurate diagnostic methods that correctly identify more PVs with hearing change and have comparatively few false positive diagnoses. The AUC was estimated using an analog to the Wilcoxon-Mann-Whitney U-statistic (Hanley and McNeil, 1982; Obuchowski, 1997). Since most subjects were observed during several monitoring visits, some degree of correlation was anticipated among the ΔOAE_f across monitoring appointments. Estimates of the AUC under these circumstances that are based on the U-statistic are

correct, but the standard error of the estimated AUC is incorrect (Obuchowski, 1997). Therefore, the non-parametric estimator suggested by Obuchowski (1997) to compute the standard error of AUC was used and denoted as SE(AUC).

The accuracy of any diagnostic method that is applied to the same sample from which the risk score weights were derived will always be overly optimistic. A scoring function candidate that works well to construct the scoring function on this data set might perform poorly in a separate sample of cisplatin patients. A common approach for obtaining nearly unbiased estimates of the diagnostic accuracy is leave-one-out cross-validation [LOOCV; (Simon, 2005a; Hastie *et al.*, 2009)]. This is a computational algorithm whereby each patient is successively excluded from the training data set, thus partitioning the data into a test sample, which includes the omitted subject's PVs, and a "training" sample composed of all remaining PVs. The risk score weights for all candidate metrics in Table I are determined from the training sample, and each is then used to predict the excluded patient's risk of hearing change at each PV. The procedure is iterated by leaving out a different patient at each step until all PVs are assigned a risk score according to each candidate in Table I. Nearly unbiased cross-validated ROC curves, AUCs, and SE(AUC) of each candidate in Table I are then computed from the risk scores. Note that these estimates are 'nearly unbiased' (as opposed to unbiased) because LOOCV is a sample re-use algorithm, which always induces a certain degree of bias. Molinaro *et al.* (2005) use simulations and data examples to show that LOOCV provides the smallest bias among a variety of sample re-use algorithms.

A test that is rapid is preferable, but only insofar that accuracy is not sacrificed. Therefore, the half-octave analysis was repeated by using only DPOAE frequencies restricted to the highest quarter-octave. Given the same level of accuracy, the quarter-octave model would be preferable to the half-octave model since it would take half as long to implement in a real world clinical setting.

More complex models tend to be more accurate than simpler models when evaluated within the sample used to develop the model, but may not generalize to other samples. This is because complex models tend to 'adapt' to the idiosyncrasies of the training data sets, which may not represent other, independent samples. Accordingly, model reduction is necessary to enhance generalizability. Model reduction techniques are commonly used in standard statistical practice, but most are unsuitable for the current analysis. The PLS models are non-nested, so methods based on the likelihood ratio are inappropriate. Metrics often proposed for selecting among nonnested models, such as Akaike's Information Criterion (AIC), are also inappropriate because all of the models listed in Table I are fit to the same sample and are, therefore, correlated. Furthermore, these reduction methods are based on the likelihood, which is ill defined in the LOOCV setting. Because the likelihood is conditional on the fitted model, which is different at each LOOCV iteration, the AIC or likelihood ratio statistics are incorrect. In light of the fact that formal ranking and testing methods are unavailable for general problems such as that in the current study, researchers in machine learning advocate model reduction according to the 'One Standard Error Rule' (Hastie *et al.*, 2009). The best model is the scoring function with the smallest number of metrics (K) that is within one SE(AUC) of the most accurate scoring function. Put another way, the simplest model that is statistically indistinguishable from the best model is preferred. Also, scoring functions using the quarter-octave fine structure that are within one SE(AUC) of the most accurate scoring function are preferable to scoring functions using half-octave fine structure for reasons noted. Once selected according to these criteria, the best scoring function, hereafter called the "Ototoxicity Risk Assessment" (ORA), was trained on the entire sample, and constitutes the best method among those considered for diagnosing hearing change during a follow-up PV.

III. RESULTS

Patients were recruited into the study over a 17 month period. One-hundred twenty three patients were identified from pharmacy lists and chart reviews as receiving cisplatin. Fifty-six (45.5%) of these patients met inclusion criteria for the study. Of these, 36 (64%) agreed to participate in the study. Of the 36 patients who agreed to participate, 19 (55.6%) passed the screening physical examination and provided two or more total visits for use in the analysis. Refusals to participate or complete the protocol requirements were primarily due to the time commitment, inconvenience and/or discomfort associated with going to the research laboratory for testing after chemotherapy treatment was administered. This underscores the need for objective and portable measures of hearing.

Table II summarizes the sample used in this analysis. Fifty-six post-baseline PVs from 19 subjects contributed data. The average number of post-baseline PVs per subject was 2.9, ranging from 1 to 12 PVs. Twenty-three of the 56 PVs (41.1%) had an ASHA-significant hearing change. The sample was generally composed of older veterans (mean age 62.6; range 51–79 years). Baseline average pure tone threshold in the SRO frequencies was 69.9 dB SPL and ranged from 43.6 to 86.7 dB SPL. The majority (n=12; 63.2%) of subjects had head and neck cancers, followed by lung (n=5; 26.3%), and one each of bladder and skin cancer. On average, the median starting cisplatin dose level was 100 mg/m^2, and ranged from 50 to 100 mg/m^2.

Figure 2 shows changes in DPOAE level within the highest half-octave of the SRO_{DP}. The dashed line indicates median ΔOAE_f among PVs with a hearing change. The solid line indicates the same among PVs without an ASHA-significant hearing change. The hatched region shows the inter-quartile range of ΔOAE_f among PVs with a hearing change, and the shaded region shows the same for PVs without a change. An indication of the utility of the DPOAE measurements is given by the degree of non-overlap in the median and inter-quartile ranges between PVs with and without a hearing change.

Recall that positive values of ΔOAE_f indicate a *decrease* in DPOAE level, while negative values indicate an increase. Several features stand out in Fig. 2. First, median ΔOAE_f is higher across all *f* among PVs with a hearing change. This is

TABLE II. Study sample patient and treatment characteristics.

		All
All	N	19
Patient ear-visits (post-baseline)	Total	56
	Mean	2.9
	Min	1
	Max	12
Visits with hearing change	N	23
	%	41.1%
Age	Mean	62.6
	Min	51
	Max	79
Baseline SRO average threshold	Mean	69.6
	Min	43.6
	Max	86.7
Cancer Location		
Bladder	N	1
	%	5.3
		All
Head/Neck	N	12
	%	63.2
Lung	N	5
	%	26.3
Skin	N	1
	%	5.3
Starting Dose Level Cisplatin (mg/m^2)	Median	100
	Min	50.0
	Max	100

promising evidence of the utility of DPOAEs for monitoring hearing change among patients treated with cisplatin. Second, the difference in the median ΔOAE_f appears greater at the highest (left edge of the horizontal axis) and lowest (right edge) steps, and smaller in the middle region. More than anything, this suggests that frequencies should be weighted differently for predicting hearing change, and underscores the potential advantage of PLS over simply averaging the ΔOAE_f since the latter assumes constant weights. Finally, the inter-quartile range shows the biggest separation at the lowest steps (f=12 to 24), suggesting that a complete half-octave of testing may be necessary to accurately predict hearing change.

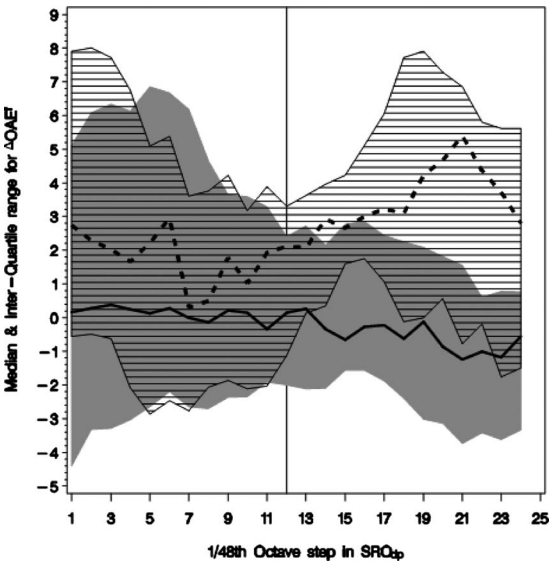

FIG. 2. Graph of median and inter-quartile range for the indexed ΔOAE_f (Baseline $DPOAE_f$ − PV $DPOAE_f$) as a function of DPOAE fine structure steps (f) ordered from highest (f=1) to lowest (f=24) frequency. The dashed line represents the mean ΔOAE_f for those subjects with an ASHA-significant change in hearing (inter-quartile range indicated with horizontal shading lines). The solid line represents the mean ΔOAE_f for those subjects with no change in hearing (inter-quartile range indicated with solid gray shading). The vertical reference line separates the upper and lower quarter octave.

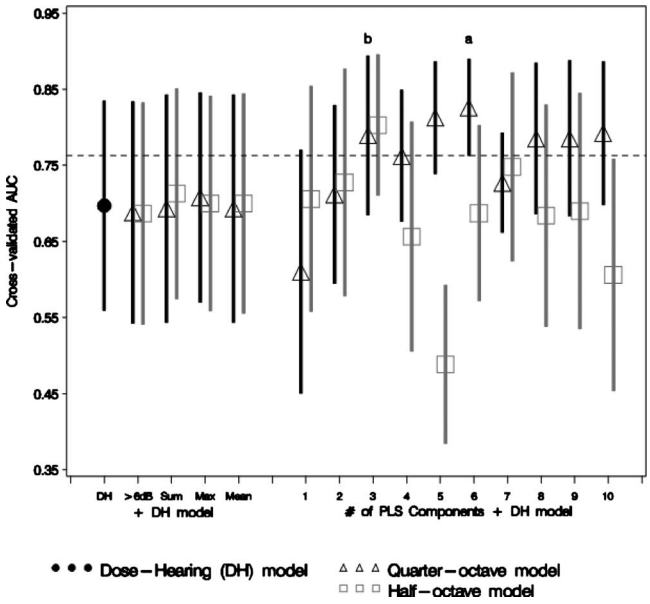

FIG. 3. Cross-validated AUC as a function of each candidate scoring functions listed in Table I. The dashed line represents AUC-SE of the most accurate model. 'a' indicates the most accurate model and 'b' indicates the preferred model according to the 'One standard error rule'. The quarter-octave model is shown in black triangles (inter-quartile range) and half octaves are shown in gray squares (±1 SE).

Observations on Fig. 2 provide a better understanding of the differences in the distribution of DPOAE fine structure between PVs with and without a hearing change. However, Fig. 2 does not adjust for the fact that some patients provided many more PVs than other, which might influence the appearance of Fig. 2. This is mitigated using the LOOCV analysis with the candidate scoring functions described in Table I. The LOOCV procedure successively holds out each biologically independent unit (i.e., the subject) while developing the predictive model. Subjects with a relatively large number of PVs cannot influence predicted hearing change on their own PVs, since that subject does not contribute to model fitting during that iteration of the LOOCV procedure.

Results of the LOOCV analysis are shown in Fig. 3. The vertical axis corresponds to the cross-validated AUC and the horizontal axis identifies the candidate scoring functions from Table I. Black triangles show results using the quarter-octave fine structure, while gray squares show results for the half-octave fine structure. Vertical bars at each symbol indicate ±SE(AUC). As a point of reference, the AUC based on the Dose-Hearing model alone, without any DPOAE monitoring, is indicated by the black dot.

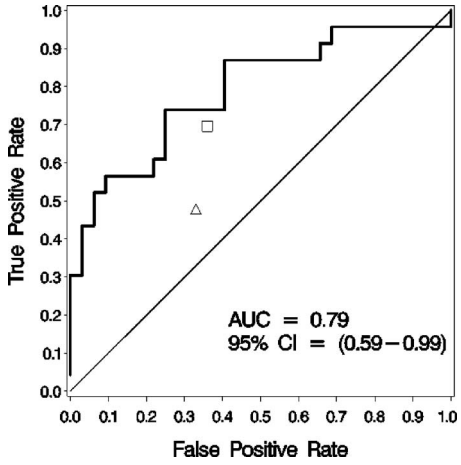

FIG. 4. Cross-validated ROC curve for the PLS components 1–3+Dose-Hearing model. The true positive and false positive rates of the '6 dB method' relying on the half-octave ('□') and quarter-octave ('△') DPOAE fine structures are also shown. Since the 6 dB method (half- or quarter-octave) is binary and an ROC curve cannot be constructed with these results, they are represented as single points on the figure.

Figure 3 shows that simple summary metrics based on changes in DPOAE fine structure offer little improvement over the Dose-Hearing model alone, which already achieves some success in identifying hearing change (AUC=0.7). However, several of the PLS models using more than two components show considerably higher accuracy than the simple Dose-Hearing model or the simple ΔOAE_f summary metrics in conjunction with the Dose-Hearing model. The most accurate model is the six PLS component model, labeled 'a', based on one quarter octave DPOAE fine structure, with a cross-validated AUC of 0.83. However, with six PLS components (and the Dose-Hearing model), this scoring function is more complex than some other candidates that have only slightly lower accuracy. The preferred scoring function is thus selected according to the 'One Standard Error Rule', that is the simplest model with an AUC that is within one standard error of the most accurate model. The dashed, horizontal line marks the AUC minus SE(AUC) of the six PLS component, quarter-octave model. The preferred scoring function is the one with the fewest components that has an AUC above the dashed reference line. According to these criteria, the preferred scoring function is the quarter-octave model using the top 3 PLS components (AUC =0.79) and is labeled 'b' in Fig. 3.

The cross-validated ROC curve for the preferred scoring function is shown in Fig. 4. The estimated AUC is 0.79 (95% confidence interval=0.59–0.99). The ROC curve shows that this is a moderately effective diagnostic method. The true positive rate rises sharply at low false positive rates, but increases more slowly with higher false positive rates. For example, supposing the risk of false positives should be no greater than 10% so that hearing change is not erroneously identified, the method can accurately detect almost 60% of PVs that have a hearing change.

Figure 4 also shows the true positive rate and false positive rate for univariate 6 dB methods utilizing DPOAE data collected in the half-octave (□) (true positive rate=0.69; false positive rate=0.36) and top quarter-octave (△) (true positive rate=0.48; false positive rate=0.33) measurement ranges. These are univariate methods that do not include the Dose-Hearing model. The 3 component PLS model is more accurate than either the simple 6 dB method using ≥+6 dB across half-octave (□) or the simple 6 dB method using ≥+6 dB across quarter-octave (△), since the ROC curve is above each of these points in Fig. 4. In particular, the simple 6 dB methods have unreasonably high false positive rates (0.33 or 0.36) at least as determined using the ASHA criteria for ototoxic hearing change as the gold standard.

TABLE III. Final ORA weights based on partial least-squares (PLS) components 1–3+Dose-Hearing model. C1, C2, and C3 denote PLS components 1–3.

Parameter	w_k	Standard error	P-value
Intercept	−0.91	0.48	0.060
Dose-hearing	0.95	0.31	0.002
C1	0.49	0.17	0.005
C2	0.82	0.35	0.018
C3	0.31	0.35	0.376

The final ORA is trained on the full sample using the Dose-Hearing (DH) model along with the top 3 PLS components, denoted C1, C2, and C3. C1 is a linear combination of the ΔOAE_f that has maximal sample covariance with the hearing change indicator. C2 is similarly constructed subject to the constraint that it is uncorrelated with C1. C3 is so constructed subject to the constraint that it is uncorrelated with C1 and C2.

The risk score weights, estimated using logistic regression, are shown in Table III. These weights are combined in the final ORA risk scoring algorithm for the i^{th} PV, such that

$$R_i = -0.94 + 0.93 \cdot DH_i + 0.49 \cdot C1_i + 0.82 \cdot C2_i + 0.31 \cdot C3_i. \quad (5)$$

The effects of ΔOAE_f on the risk score in (5) are captured by the linear combination of PLS components C1, C2, and C3, which are themselves linear combinations of the ΔOAE_f. Accordingly, the effects of the ΔOAE_f on the chances that a hearing change has occurred can be written as a single function of the ΔOAE_f. This function is shown graphically in Fig. 5. The DPOAE contribution to the final ORA risk score is equal to the weighted sum of the observed ΔOAE_f over quarter-octave at each PV, with weights corresponding to values shown on the vertical axis in Fig. 5. PVs with ΔOAE_f profiles that closely match Fig. 5 are ones that have the highest estimated risk of a hearing change. Thus, the highest risk of an ASHA-criteria hearing change occurs in PVs that show large degradations in DPOAEs at the highest frequencies, followed by improvement in the middle frequencies. The R_i are on a log-odds scale, which might not be useful clinically. Instead, the estimated risk, or probability, that a patient has a hearing change at a particular PV can be computed using the inverse logit transformation

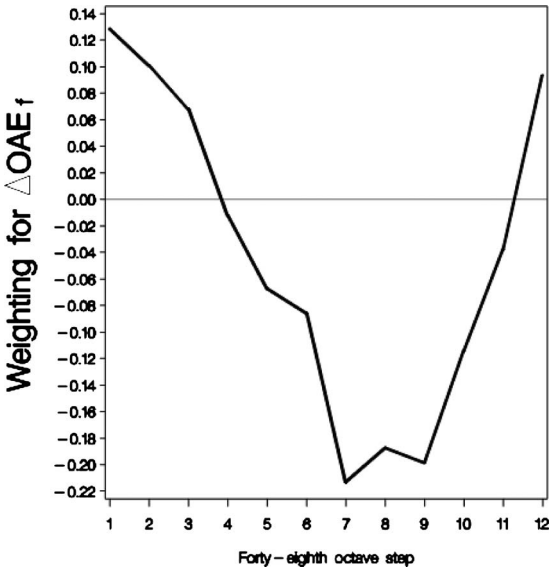

FIG. 5. ΔOAE_f weights used in the final ORA as a function of the highest 12 frequencies measured in $1/48^{th}$ octave steps. Positive weights represent a reduction in the DPOAE level re: baseline DPOAE levels while negative weights represent an increase in emission level.

$$\text{Probability of Hearing Change} = \frac{\exp^{R_i}}{1+\exp^{R_i}}. \quad (6)$$

Probabilities close to zero indicate little likelihood that the patient has had a hearing change at that monitoring appointment, and values close to one indicates an almost certain ASHA-criteria hearing change.

Histograms of the ORA scores for the 56 PVs in our sample are shown in Fig. 6. The top panel shows the scores of PVs that did not have an ASHA-criteria hearing change. The bottom panel shows scores for PVs that did have such a change. As noted earlier, an ideal diagnostic test should give risk scores for the PVs with a hearing change that are higher and well separated from the PVs without a change. This is apparent in Fig. 6. The distribution of scores on the bottom panel is centered to the right of the scores in the top panel, and none of the scores among PVs without a hearing change exceeded 1.5, where the scores on the bottom panel are concentrated.

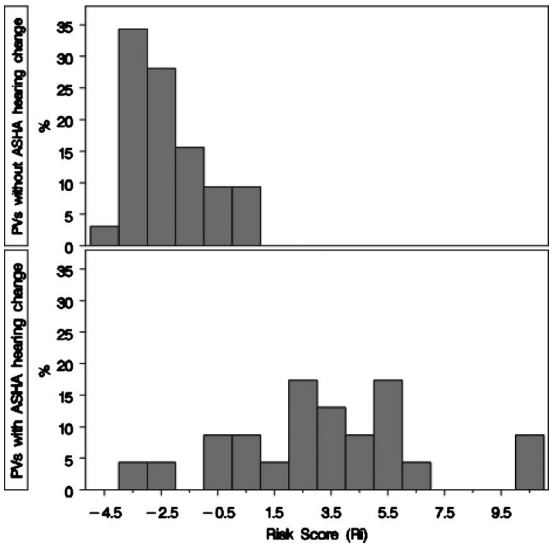

FIG. 6. Histogram of the final ORA risk scores for PVs in the study sample. Top panel shows risk scores for PVs without an ASHA-significant hearing change while bottom panel shows PVs with an ASHA-significant hearing change.

IV. DISCUSSION

The present results from Veterans receiving cisplatin chemotherapy confirm that cisplatin exposure reduces or eliminates DPOAEs. Results also show an association between changes in DPOAE levels and the presence of a clinically significant hearing change as defined by ASHA (1994). However, significant improvement in the accuracy of DPOAEs for predicting hearing change was shown to be associated with the use of weighted combinations of DPOAE metrics and two risk factors for ototoxicity, behavioral hearing thresholds prior to treatment and cumulative cisplatin dose, from a Dose-Hearing model. A multivariate approach combining these three sources of information, appropriately weighted and using DPOAE level changes evaluated over a quarter-octave range near each subjects' high frequency DPOAE limit, yielded a reasonably accurate (AUC=0.79) and rapid assessment of ototoxicity risk. The present findings suggest that once validated on a separate study sample such an ORA could be useful as part of a test battery for all patients receiving cisplatin, but particularly for those unable to take a behavioral hearing test. Although the current study considered only changes in hearing and DPOAEs between each monitoring visit and the pre-exposure baseline test, the approach was designed to both identify and monitor progression of ototoxicity in a clinical setting. This would be done by establishing a new baseline following each significant hearing change that is confirmed on a repeat test. By shifting the baseline, DPOAEs can be used to monitor the progression of ototoxic hearing loss until they are no longer recordable.

This study used $1/48^{th}$ octave step sizes. The rationale for using fine measurements and employing a smoothing algorithm was to reduce the test variability of the DPOAE measurements. We reasoned that such a method would minimize effects of dips in the DPOAE fine structure, as well as any spurious measurements, that could confound estimates of DPOAE change. Our data (Fig. 5) support that important information about weighting frequencies differently would not have been obtained using large step sizes. Having said this, it may be that larger spacing between test frequencies would be another way to decrease test time. We consider this possibility and others as ways to improve clinical performance of this method. One planned strategy for the continuation of this research is to collect other distortion product emissions which require no additional test time and might improve test performance.

A number of studies, including the present report, have supported the use of DPOAEs as objective measures of ototoxicity among patients receiving cisplatin (Ress *et al.*, 1999; Stavroulaki *et al.*, 2002; Knight *et al.*, 2007; Reavis *et al.*, 2008). Previously, most investigations into DPOAE applica-

tions for the detection of ototoxicity have utilized either statistical tests of group differences, or clinically significant differences determined *a priori* in control populations. Additionally, reports from normal hearing subjects' unexposed to ototoxic agents have suggested using a level change criteria of 6 to 9 dB based on test-retest reliability and which corresponds to the upper bound of the 95% confidence limits (Beattie *et al.* 2003; Franklin *et al.* 1992; Roede *et al.* 1993). This criterion was developed to reduce false positive rates but the corresponding specificity and sensitivity in adult cancer patients was unknown prior to this report. Evaluating the univariate 6 dB method against the behavioral gold-standard, we found unacceptably high false positive rates using the 6 dB method with only modest sensitivity. Previous work leading to the proposed 6 dB method involved studies of DPOAE level repeatability performed in healthy, normal-hearing research subjects. A DPOAE level change of approximately 6 dB was identified in a number of these studies as a change that is large relative to normal test-retest variability, would yield an estimated false positive rate of only about 5%, and therefore that would potentially indicate a real DPOAE change in an exposed ear. Test repeatability among young healthy volunteers, however, may not be representative of the current sample of adult cancer patients, many of whom had some hearing loss prior to treatment, and therefore may have had DPOAEs that were lower in level and more likely to be contaminated by noise. Alternatively, the unexpectedly high false positive rates obtained for the 6 dB method may be associated with real cisplatin-induced changes in DPOAEs at a given PV for ears that lacked corresponding ASHA-significant hearing changes.

While the second possibility could not be tested because there is no gold standard test for subclinical changes in ototoxicity, a post hoc analysis of test repeatability was performed on two control subjects using the same systems and protocols described for the current data set. These were Veteran subjects with pre-existing hearing loss who were hospitalized at the Portland VA Medical Center for either severe infection or cancer but were taking no ototoxic medications. Otherwise, they were recruited using the same exclusionary criteria as the cisplatin exposed subjects in this study. The control subjects were seen across 7 PVs over 6–12 weeks with an average of 3.5 visits yielding up to 168 possible test-retest ΔOAE_f calculations. These data are presented in Fig. 7 in the form of cumulative distributions, with cumulative percent represented on the *y* axis and test-retest differences on the *x* axis given in 2-dB bins. Separate curves are shown for the top half octave (gray line) and quarter octave (black line) ranges of recordable DPOAEs. Vertical lines indicate the proportion of observed DPOAE test-retest level changes that were 6 dB or smaller (73.2% and 54.8% for half and quarter octave ranges of data, respectively). These results correspond to an estimated false positive rate for the 6 dB method of 26.8% and 45.2% for the top half or quarter octave ranges of data, respectively. Though test-retest difference data were available from only two unexposed hospitalized Veteran patients, the false positive rates estimated empirically from these data agreed with actual DPOAE false

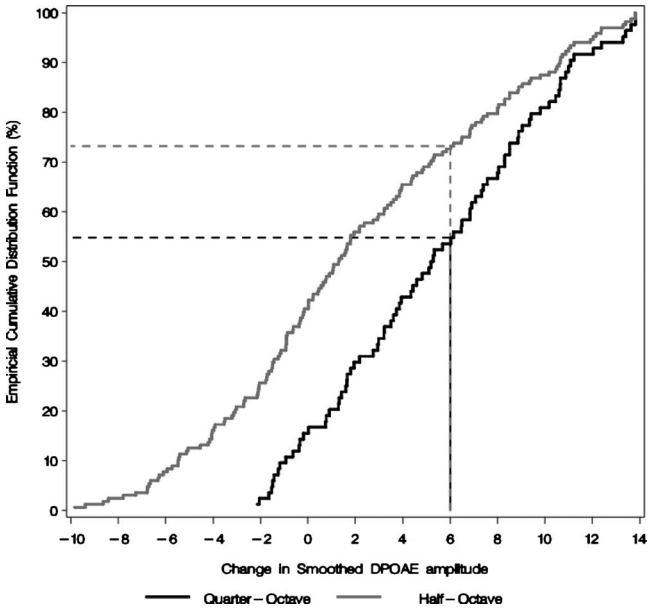

FIG. 7. Cumulative distribution functions (%) for smoothed ΔOAE_f measured over 7 visits from 2 control subjects. Gray line represents level changes over highest half octave while black line includes level changes for highest quarter octave.

positive rates determined using a gold standard measure of ototoxic hearing change among subjects administered cisplatin. Furthermore, adding the *a priori* 6 dB change criterion to the Dose-Hearing model to form a multivariate 6 dB method, did not increase test accuracy. The overall test accuracy was comparable to the Dose-Hearing model by itself which can be determined from data available at baseline without information from subsequent monitoring visits. Both univariate and multivariate 6 dB methods were inferior to alternative tests for the early detection of ototoxicity.

Analysis of other DPOAE fine-structure summary metrics developed *a priori* revealed performances that were only slightly better than the 6 dB method. Additionally, results seen in Fig. 2 highlight the need to weight DPOAE changes differently across the half-octave test range. The ototoxicity risk assessment (ORA) in which ΔOAE_f were allowed to be weighted differently (through PLS methods) and combined with the Dose-Hearing model, resulted in improvements in test performance over that achieved with either the 6 dB method or simple summary metrics. Thus, the ORA resulted in greater separation of the response distributions from ears exhibiting changed hearing as compared to those with stable hearing when compared to the more traditional approaches.

Consistent with other published reports, multivariate solutions were found to be more accurate than univariate methods. Dorn *et al.* (1999) compared single variable with multivariate methods and found that clinical decision methods were improved when multiple frequencies (measured in half octave steps) from the DPOAE were used. Gorga *et al.* (1999) provided much needed but often overlooked validation of this finding and, further, demonstrated that no additional time will be spent in data collection using multivariate methods to improve test performance.

Recently, Reavis *et al.* (2010) similarly found that a multivariate solution better predicted hearing changes among

Veterans exposed to cisplatin compared to any univariate solution. These authors investigated the ability of simple summary metrics derived from DPOAE input/output functions to predict cisplatin-induced hearing changes. In their study, DPOAE input/output functions were reduced by simple quantitative summations of the stimulus levels, recorded OAE levels, and the signal-to-noise ratios. The valid responses in the growth function were summed with equal weights. The result was a highly accurate (cross-validated AUC=0.91) multivariate solution which included measures of baseline hearing, cumulative drug dose, and the change in the summated DPOAE stimulus level. The greater test accuracy achieved by Reavis and colleagues may be related to the lower level primaries (measurements occurred down to L1/L2 of 53/35 dB SPL) compared to the equal level stimulus parameters used here (L1=L2=65 dB SPL). In prior studies, low level primaries were those most sensitive to effects of noise exposure (e.g., Sutton *et al.*, 1994). However, other studies have shown high level DPOAEs to be more useful for the purpose of ototoxicity in adult cancer patients in whom some pre-exposure hearing loss was common (Ress *et al.*, 1999). Indeed, the cross-validated AUC reported by Reavis *et al.* (2010) (0.91) falls within the 95% confidence limits of the ORA reported here (AUC 95% CI: 0.59–0.99). Thus it is difficult to state if one solution is better than another.

The shape of the weighting function (Fig. 5) provides an indication of the way in which cisplatin altered DPOAEs in subjects with ASHA-significant hearing shifts. As predicted, the highest DPOAEs able to be monitored in each subject were those that showed the greatest cisplatin-induced level decrements. While there was some inconsistency in the pattern of the observed DPOAE changes, an enhancement, or increase in the DPOAE level was often seen at the adjacent lower frequencies. Occasional DPOAE enhancement was previously observed following punctuate noise exposures in rodent models (Harding and Bohne, 2004; Howard *et al.*, 2002). Damage experimentally induced in chinchilla using a comparatively wide octave band of noise produced structural damage of pillar cells, increases in auditory brainstem responses thresholds and decreases in DPOAE levels elicited using test frequencies tuned to the damage region. In addition, an unexpected but significant enhancement of DPOAE levels was frequently observed immediately apical to the region damaged by the octave band noise, a pattern suggestive of the DPOAE damage pattern in the current study. Increments in the level of the octave band noise caused structural damage to spread and the DPOAE enhancement region to shift apically, suggesting that the effect was systematic. The reason for such an enhancement is unclear based on the current analysis, but might be related to changes in the relative contribution of multiple DPOAE source components. In the context of the current study, there may be a differential effect whereby cisplatin affects the DPOAE component arising via linear reflection from the DP characteristic frequency place more readily compared with the component arising through nonlinear distortion near f2 (Shera and Guinan, 1999). Recent evidence suggests that DPOAEs may also contain components arising from regions of the cochlea basal to the eliciting frequencies (Martin *et al.*, 2009). Any basal components might be those first impacted by the base to apical progression of cisplatin ototoxicity, and their selective removal could conceivably result in occasional DPOAE enhancement. Changes in complex interactions among multiple DPOAE components might also be expected to produce occasional frequency shifts in fine structure patterns (Deeter *et al.*, 2009). However, no clear frequency shift in fine structure patterns was observed.

Though fewer diagnostic errors were observed when using the ORA compared to *a priori* approaches, it is not perfect. Diagnostic tests never are. One potential impediment to the use of DPOAEs to monitor hearing in adult cancer patients is that the objective and behavioral measures monitored for changes may not overlap. DPOAEs monitored were often at frequencies below those that showed behavioral change. Though basal hearing thresholds have been shown to correlate with more apically generated DPOAEs (Arnold *et al.*, 1999), better performance likely would have been achieved had objective and behavioral test frequencies consistently overlapped (Reavis *et al.*, 2008). Overlapping frequency regions would be expected to contribute to better performance for both univariate and multivariate solutions. Nevertheless, the ORA accuracy observed within our sample population of pre-exposed hearing impaired Veterans was still remarkably high.

If the ORA accurately identifies cisplatin-induced hearing change in an independent validation sample, the next step is to develop a protocol for implementing the ORA at bedside. Equation (6) shows how to compute the risk that an ASHA-criteria hearing change has occurred, which can easily be implemented in a mobile DPOAE monitoring device. Based on patient characteristics, treatment progress, and DPOAE results, the device will provide the clinician an estimate of the risk that hearing change has occurred, along with a 95% confidence interval for that risk. A pass/fail result for each PV is also available once a suitable false positive rate is chosen from the validation sample ROC curve. This can be illustrated using the cross-validated ROC curve in Fig. 4. It might be good practice to use a pass/fail cut-off that minimizes the false positive rate to an acceptably small level so that the cancer treatment is not unnecessarily modified. Supposing that 10% false-positives are acceptable then a cut-off of 1.9 gives a test sensitivity of about 55% correctly identified ASHA-significant hearing changes. Using this, or other appropriate criteria, test failures could prompt the clinician to further audiological assessment. Investigations are underway to refine and validate this method for use in clinical practice.

Applying the ORA multivariate solution is simple, requires no additional test time, and could be performed immediately following DPOAE data collection at each PV resulting in a clinically applicable interpretation of DPOAE change. In view of the results of this study, the ORA would appear to have advantages over current univariate approaches used clinically, specifically greater accuracy and yields a probability of behavioral hearing change given DPOAE change. This is promising evidence for the effec-

tiveness of the ORA, even so, the ORA must be validated in an independent sample prior to being implemented clinically.

ACKNOWLEDGMENTS

Work supported by the Department of Veterans Affairs RR&D Service (Grant Nos. C4183R and C4447K) and the VA RR&D National Center for Rehabilitative Auditory Research, Portland, OR. The authors wish to thank Stephen T. Neely, D. Sc. and Boys Town National Research Hospital for the Emav System that was used to collect the DPOAE responses.

Alam, S., Ikeda, K., Oshima, T., Suzuki, M., Kawase, T., Kikuchi, T., and Takasaka, T. (**2000**). "Cisplatin-induced apoptotic cell death in Mongolian gerbil cochlea," Hear. Res. **141**, 28–38.

Amalraj, S., Starkweather, C., Nguyen, C., and Naeim, A. (**2009**). "Health literacy, communication, and treatment decision-making in older cancer patients," Oncology **23**, 369–375.

Arnold, D., Lonsbury-Martin, B., and Martin, G. (**1999**). "High-frequency hearing influences lower-frequency distortion-product otoacoustic emissions," Arch. Otolaryngol. Head Neck Surg. **125**, 215–222.

ASHA (**1994**). "Guidelines for the audiologic management of individuals receiving cochleotoxic drug therapy," American Speech-Language-Hearing Association, www.ashaorg/docs/html/GL1994-00003.html, pp. 11–19.

Barker, M., and Rayens, W. (**2003**). "Partial least squares for discrimination," J. Chemom. **17**, 166–173.

Beattie, R., Kenworthy, O., and Luna, C. (**2003**). "Immediate and short-term reliability of distortion-product otoacoustic emissions," Int. J. Audiol. **42**, 348–354.

Carhart, R., and Jerger, J. (**1959**). "Preferred method for clinical determination of pure-tone thresholds," J. Speech Hear Disord. **24**, 330–345.

Dalton, D., Cruickshanks, K., Klein, B., Klein, R., Wiley, T., and Nondahl, D. (**2003**). "The impact of hearing loss on quality of life in older adults," Gerontologist **43**, 661–668.

Deeter, R., Abel, R., Calandruccio, L., and Dhar, S. (**2009**). "Contralateral acoustic stimulation alters the magnitude and phase of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **126**, 2413–2424.

Dorn, P., Piskorski, P., Gorga, M., Neely, S., and Keefe, D. (**1999**). "Predicting audiometric status from distortion product otoacoustic emissions using multivariate analysis," Ear Hear. **20**, 149–163.

Durrant, J., Palmer, C., and Lunner, T. (**2005**). "Analysis of counted behaviors in a single-subject design: Modeling of hearing-aid intervention in hearing-aid patients with Alzheimer's disease," Int. J. Audiol. **44**, 31–38.

Fausti, S., Frey, R., Erickson, D., Rappaport, B., Cleary, E., and Brummert, R. (**1979**). "A system for evaluating auditory function from 8,000–20,000 Hz," J. Acoust. Soc. Am. **66**, 1713–1718.

Fausti, S., Frey, R., Henry, J., Olson, D., and Schaffer, H. (**1992**). "Early detection of ototoxicity using high-frequency, tone-burst-evoked auditory brainstem responses," J. Am. Acad. Audiol **3**, 397–404.

Fausti, S., Rappaport, B., Frey, R., Henry, J., Phillips, D., Mitchell, C., and Olson, D. (**1991**). "Reliability of evoked responses to high-frequency (8–14 kHz) tone bursts," J. Am. Acad. Audiol **2**, 105–114.

Fischel-Ghodsian, N., Prezant, T., Bu, X., and Oztas, S. (**1993**). "Mitochondrial ribosomal RNA gene mutation in a patient with sporadic aminoglycoside ototoxicity," Am. J. Otolaryngol. **14**, 399–403.

Fitzmaurice, G., Laird, N., and Ware, J. (**2004**). *Applied Longitudinal Analysis* (Wiley, Hoboken, NJ), pp. 291–321.

Forge, A., and Schacht, J. (**2000**). "Aminoglycoside antibiotics," Audiol. Neuro-Otol. **5**, 3–22.

Franklin, D., McCoy, M., Martin, G., and Lonsbury-Martin, B. (**1992**). "Test/retest reliability of distortion-product and transiently evoked otoacoustic emissions," Ear Hear. **13**, 417–429.

Gorga, M., Neely, S., and Dorn, P. (**1999**). "Distortion product otoacoustic emission test performance for a priori criteria and for multifrequency audiometric standards," Ear Hear. **20**, 345–362.

Gorga, M., Neely, S., Ohlrich, B., Redner, J., and Peters, J. (**1997**). "From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss," Ear Hear. **18**, 440–455.

Gorga, M., Stover, L., Neely, S., and Montoya, D. (**1996**). "The use of cumulative distributions to determine critical values and levels of confidence for clinical distortion product otoacoustic emission measurements," J. Acoust. Soc. Am. **100**, 968–977.

Hanley, J., and McNeil, B. (**1982**). "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology **143**, 29–36.

Harding, G., and Bohne, B. (**2004**). "Temporary DPOAE level shifts, ABR threshold shifts and histopathological damage following below-critical-level noise exposure," Hear. Res. **196**, 94–108.

Hastie, T., Tibshirani, R., and Friedman, J. (**2009**). *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (Springer, New York), pp. 241–245.

Hodges, A., and Lonsbury-Martin, B. (**1998**). *Hearing Management* (Singular, San Diego), pp. 269–290.

Howard, M., Stagner, B., Lonsbury-Martin, B., and Martin, G. (**2002**). "Effects of reversible noise exposure on the suppression tuning of rabbit distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **111**, 285–296.

Knight, K., Kraemer, D., Winter, C., and Neuwelt, E. (**2007**). "Early changes in auditory function as a result of platinum chemotherapy: Use of extended high-frequency audiometry and evoked distortion product otoacoustic emissions," J. Clin. Oncol. **25**, 1190–1195.

Kochkin, S., and Rogin, C. (**2000**). "Quantifying the obvious: The impact of hearing instruments on the quality of life," Hear. Rev. **7**, 6–34.

Laurell, G., Viberg, A., Teixeira, M., Sterkers, O., and Ferrary, E. (**2000**). "Blood-perilymph barrier and ototoxicity: An *in vivo* study in the rat," Acta Oto-Laryngol. **120**, 796–803.

Martin, G., Stagner, B., Fahey, P., and Lonsbury-Martin, B. (**2009**). "Steep and shallow phase gradient distortion product otoacoustic emissions arising basal to the primary tones," J. Acoust. Soc. Am. **125**, EL85–EL92.

Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emissions (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.

Molinaro, A., Simon, R., and Pfeiffer, R. (**2005**). "Prediction error estimation: A comparison of resampling methods," Bioinformatics **21**, 3301–3307.

Mukherjea, D., Whitworth, C., Nandish, S., Dunaway, G., Rybak, L., and Ramkumar, V. (**2006**). "Expression of the kidney infury molecule 1 in the rat cochlea and induction by cisplatin," Neuroscience **139**, 733–740.

Neely, S., and Liu, Z. (**1994**). *EMAV: Otoacoustic Emission Averager* (Boys Town National Research Hospital, Omaha, NE).

Obuchowski, N. (**1997**). "Nonparametric analysis of clustered ROC curve data," Biometrics **53**, 567–578.

Radmacher, M., McShane, L., and Simon, R. (**2002**). "A paradigm for class prediction using gene expression profiles," J. Comput. Biol. **9**, 505–511.

Reavis, K., McMillan, G., Austin, D., Fausti, S., and Konrad-Martin, D. (**2010**). "Distortion-product otoacoustic emission test performance for ototoxicity monitoring," Ear Hear. In press.

Reavis, K. M., Phillips, D. S., Fausti, S. J., Gordon, J. S., Helt, W. J., Willington, D., Bratt, G. W., and Konrad-Martin, D. (**2008**). "Factors affecting sensitivity of distortion-product otoacoustic emissions to ototoxic hearing loss," Ear Hear. **29**, 875–893.

Ress, B. D., Sridhar, K. S., Balkany, T. J., Waxman, G. M., Stagner, B. B., and Lonsbury-Martin, B. L. (**1999**). "Effects of cis-platin chemotherapy on otoacoustic emissions: The development of an objective screening protocol," Otolaryngol.-Head Neck Surg. **121**, 693–701.

Roede, J., Harris, F., Probst, R., and Xu, L. (**1993**). "Repeatability of distortion product otoacoustic emissions in normally hearing humans," Audiology **32**, 273–281.

Shera, C., and Guinan, J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Simon, R. (**2005a**). "Roadmap for developing and validating therapeutically relevant genomic classifiers," J. Clin. Oncol. **23**, 7332–7341.

Stavroulaki, P., Vossinakis, I., Dinopoulou, D., Doudounakis, S., Adamopoulos, G., and Apostolopoulos, N. (**2002**). "Otoacoustic emissions for monitoring aminoglycoside-induced ototoxicity in children with cystic fibrosis," Arch. Otolaryngol. Head Neck Surg. **128**, 150–155.

Sutton, L. A., Lonsbury-Martin, B. L., Martin, G. K., and Whitehead, M. L. (**1994**). "Sensitivity of distortion-product otoacoustic emissions in humans to tonal over-exposure: Time course of recovery and effects of lowering L2," Hear. Res. **75** 1–2, 161–174.