Published in final edited form as:

Ear Hear. 2015 ; 36(5): e251–e260. doi:10.1097/AUD.0000000000000176.

Meta-Analysis of Distortion Product Otoacoustic Emission Retest Variability for Serial Monitoring of Cochlear Function in Adults

Kelly M. Reavis¹, Garnett P. McMillan¹, Marilyn F. Dille^{1,2}, and Dawn Konrad-Martin^{1,2}

¹VA RR&D National Center for Rehabilitative Auditory Research (NCRAR), VA Medical Center, Portland, Oregon, USA

²Department of Otolaryngology/HNS, Oregon Health & Science University, Portland, Oregon, USA

Abstract

Objective—Distortion product otoacoustic emissions (DPOAEs) have long been heralded as a means to objectively monitor cochlear function and increasingly are becoming a key component in hearing surveillance programs for individuals at risk for ototoxic- and occupational noise-related hearing loss. Yet clinicians are unsure how to define clinically meaningful shifts in DPOAE level. In this study, a meta-analysis approach is used to synthesize the DPOAE level test–retest literature to construct a set of DPOAE level shift reference limits that can be used clinically to define a statistically significant emission change.

Design—The authors reviewed all published articles identified through a Medline search using the terms “Otoacoustic Emission Variability,” “Otoacoustic Emission Reliability,” “Otoacoustic Emission Repeatability,” and “Otoacoustic Emission Test Retest” restricted to DPOAEs, adults, and English language. Articles with DPOAE level data elicited by moderate stimulus levels for f_2 frequencies of 1000, 2000, 4000, or 6000 Hz were selected because these stimulus parameters were relatively well represented in the literature. The authors only included articles that reported the standard error of the measurement (SEM) or from which the SEM could be calculated. Meta-analysis was used to estimate the population mean SEM over the included studies. Models were fit separately for each f_2 primary and included days since baseline and study-specific random effects.

Results—Ten DPOAE test–retest studies met inclusion criteria for this meta-analysis. The SEM values varied widely across published studies (0.57 to 3.9 dB) and were provided for relatively short time intervals (less than 15 days on average). Time, or days since baseline, was statistically significant at higher f_2 frequencies (4000 and 6000 Hz). From the model results, 90% reference limits specific to the f_2 and elapsed time between baseline and follow-up measurements were established. Reference limits provided correspond to negative (emission decrement) and positive (emission enhancement) shifts indicative of the amount of measurement variability that, using this

Address for correspondence: Kelly M. Reavis, VA RR&D National Center for Rehabilitative Auditory Research (NCRAR), 3710 SW US Veterans Hospital Rd., Portland, OR 97239, USA. Kelly.Reavis@va.gov.

The authors have no conflicts of interest to disclose.

approach, must be tolerated as “normal” fluctuations over time. Changes larger than the reference limits are considered significant and warrant follow-up testing.

Conclusions—The meta-analysis presented provides reference limits that are appropriate for a set of specific f_2 frequencies and time intervals. The meta-analysis concerns the SEM statistic directly, so that any preferred reference limit can be computed from the results and should be predicated upon the screening application. The presumed advantage of this meta-analytic approach is increased precision relative to limits suggested by any of the individual studies included in the analysis.

Keywords

Distortion product otoacoustic emissions; DPOAE; Meta-analysis; Reference limits; Test–retest

Introduction

Distortion product otoacoustic emissions (DPOAEs) are low-level sounds emitted from the cochlea in response to two closely spaced stimulating tones (Kemp 1978). The acoustic response measured in the ear canal is the result of an active cochlear amplification process that is linked to the integrity of the outer hair cells (Brownell 1990). Damaged and missing outer hair cells result in changed, typically diminished, emissions. It is this close relation between otoacoustic emissions and cochlear status that has propelled the clinical application of DPOAEs for use in pediatric audiology, and in other difficult to test populations, such as patients receiving ototoxic medications. The clinical application of DPOAEs is well established for hearing loss screening and diagnostics in infants, children, and adults (Gorga et al. 1997; Lonsbury-Martin & Martin 2003).

Because of their widespread clinical availability as well as their ability to test outer hair cell function at higher frequencies compared with standard transient-evoked otoacoustic emissions, DPOAE testing is increasingly becoming a key component in ototoxicity monitoring programs and occupational noise-exposure monitoring programs. Furthermore, DPOAEs are believed to be sensitive to subtle cochlear damage before it becomes apparent using pure-tone threshold testing (Mulheran & Degg 1997; Katbamna et al. 1999; Stavroulaki et al. 2002; Lapsley Miller & Marshall 2007; Helleman & Dreschler 2012). Finally, DPOAEs have the distinct advantages of being fast to obtain, nonbehavioral (capable of assessing cochlear function in inattentive patients, such as newborns or critically ill patients), and obtainable by nonspecialist staff with specific training in making emission measurements.

The standard approach to serial monitoring with DPOAEs for ototoxic or noise-induced cochlear damage is to take a baseline, pre-exposure DPOAE measurement and then repeat the measurement during subsequent monitoring appointments to determine whether an alarming change in cochlear function, presumably due to damage from noise or medication, has occurred. If the difference between the baseline and monitoring measurements exceeds test–retest standards, then further testing is recommended (Konrad-Martin et al. 2012a).

Test–retest standards are determined in the same way in virtually all areas of medicine. Measurements are obtained on two or more occasions in a sample of healthy subjects. These are assembled into a statistic that estimates test–retest variability. This statistic is used to estimate the range of test–retest differences that a pre-specified percentage of otherwise healthy patients would show. This range, which can be called the shift reference interval, contains a fixed percentage of the central distribution of the reference sample. Typical clinical reference intervals cover 90 or 95% of the reference sample so that 10 or 5% of otherwise healthy people would be incorrectly referred for follow-up care. It is a clinical question as to how high a false referral rate one can accept as a trade-off for detecting more truly abnormal test–retest changes. Fortunately, any size reference interval can be computed from a published estimate of test–retest variability (McMillan et al. 2013).

Adult DPOAE test–retest variability estimates can be found in several published studies, and it is up to the clinician to choose among these for application to a particular patient. Unfortunately, this strategy fails to appreciate that each published test–retest standard is a statistic that must, by definition, vary among studies even within clinical populations. These estimates will be more or less accurate and precise depending on a multitude of factors, such as sample size, testing protocols, and range of hearing among the subjects. It is not simply a matter of picking the “best” single estimate of test–retest variability for a clinical application, but rather of synthesizing the relevant clinical literature, recognizing that combining several study results will give more accurate and precise test-retest standards than any single study can provide, in the same sense that larger samples give more accurate estimates of population parameters (such as the mean and variance) than smaller samples. The goal of this report is to synthesize published DPOAE test–retest standards for adults while accounting for variability among studies as well as possible covariate effects on test–retest standards.

The standard statistical approach to synthesizing published quantitative results is meta-analysis. Meta-analysis is a commonly used research and statistical method to combine results from multiple independent studies. By combining studies, the precision and accuracy of a measurement can be improved. Research into DPOAE test–retest variability is particularly amenable to meta-analysis because most studies apply similar measurement techniques and estimate test–retest variability using the “standard error of the measurement” (SEM; Demorest and Walden 1984). The sample SEM is given by

$$\widehat{SEM} = s \cdot \sqrt{1 - r}, \quad (1)$$

where s is the standard deviation of combined baseline and follow-up measurements and r is Pearson's correlation coefficient between baseline and follow-up measurements. In a homeostatic reference population, 90% reference limits for DPOAE shifts are given by $\pm 1.645 \cdot \sqrt{2} \cdot \widehat{SEM}$. Thus, 90% reference limits should yield a 10% total false referral rate, with 5% falling below the lower bound and 5% falling above the upper bound of the reference interval. Ninety-five percent shift reference limits are given by $\pm 1.96 \cdot \sqrt{2} \cdot \widehat{SEM}$.

Clinical application of a reference interval with upper and lower reference limits implicitly assumes that either a large negative or positive value of the DPOAE level shift is clinically alarming. This approach is preferable in the case of DPOAE shifts because it can be generally assumed that either a DPOAE level degradation or an enhancement may indicate cochlear damage worthy of follow-up. DPOAE level enhancement has been observed in serial monitoring following exposure to noise or ototoxic drugs. In particular, among adults occupationally exposed to noise, Helleman and Dreshler (2012) showed DPOAE enhancements around 3000 Hz, a frequency region below the classical audiometric noise-notch area of 4000–6000 Hz. In addition, Konrad-Martin et al. (2012b) reported that DPOAE enhancements were most common at frequencies below the damaged region identified by pure-tone hearing changes among Veterans exposed to cisplatin. Enhancement below the “audiometric edge” of the presumed damage is consistent with animal models of noise (Harding and Bohne 2004) and ototoxic exposures (Kakigi et al. 1998; Mei et al. 2009). Researchers have hypothesized that reduced activity of the medial olivocochlear system and/or removal of contributions from additional “basal sources” may play a role in generating enhancements in this context (Kakigi et al.; Martin et al. 2011). In addition, DPOAEs are thought to arise from at least two generator mechanisms, nonlinear distortion and linear reflection, which interact constructively or destructively depending on their phase relationship (Shera & Guinan 1999). Some enhancements conceivably arise when damage changes the contributions from these sources and thus their interaction in the ear canal. Even so, one might prefer a test that only screens for DPOAE decrement. In this case, only the lower reference limit is of interest, so that a lower 10% reference limit will have a 10% false referral rate. The limits are computed as before, taking into account the target total false referral rate. For this reason, a lower 10% reference limit for DPOAE decay is given by $-1.282 \sqrt{2} \cdot \widehat{SEM}$. A lower 5% reference limit for DPOAE decay is given by $-1.645 \cdot \sqrt{2} \cdot \widehat{SEM}$, and so forth. In any event, published estimates of the $\widehat{SEM}$ give the requisite reference limits.

This report provides a meta-analysis of all published DPOAE test-retest studies in adults that met our inclusion criteria. The result is a set of DPOAE level shift reference limits for serial monitoring that are expected to be more accurate and precise than estimates arising from any individual included study. We also use the published data to investigate temporal effects on DPOAE test-retest variability, and obtain results that suggest the test-retest standards should change with increasing time between baseline and subsequent testing. Serial monitoring using this or any other published DPOAE test-retest variability data must carefully consider certain caveats that may make this set of reference limits more or less desirable under different clinical situations.

Materials and Methods

Literature Review

PubMed was used to access MEDLINE in September 2013 for the search terms: “Otoacoustic Emission Variability,” “Otoacoustic Emission Reliability,” “Otoacoustic Emission Repeatability,” and “Otoacoustic Emission Test Retest.” The search strategy was limited to published English language studies in adult humans. The search was limited to

adults so that the results could be applied to DPOAE level shifts among adults exposed to noise occupationally and ototoxic medications. The search was supplemented by personal files and reference lists of identified papers. Any DPOAE study accompanied by data from which DPOAE SEM could be determined, regardless of the sample size or retest interval, was included. Other studies have characterized test–retest variability of DPOAEs by estimating the standard deviation of the amplitude difference (Roede et al. 1993) and the cumulative distribution of the amplitude difference (Dreisbach et al. 2006). To better reflect the clinical application of using DPOAEs for serial monitoring, only studies and data where the DPOAE probe was removed and placed again for subsequent testing were included. SEM across repeated measures was extracted from the study tables for use in the statistical analysis. Specifically, data were extracted for DPOAEs elicited with an f_2/f_1 ratio = 1.2 and moderate primary levels (L_1 range 55 to 75 dB SPL; L_2 range 50 to 70 dB SPL; L_1-L_2 range 0 to 10 dB) for f_2 frequencies of 1000, 2000, 4000, or 6000 Hz. SEM values have been shown previously to vary insignificantly across moderate primary levels ranging from 55 to 75 dB SPL (Franklin et al. 1992). The choice of primary f_2 frequencies was dictated by the volume of data extracted from the published literature to warrant inclusion in a meta-analysis. Other DPOAE collection parameters such as signal-to-noise ratios (SNRs) and calibration techniques were allowed to vary across studies to allow for broader applicability of the meta-analysis results. DPOAEs were collected using both custom and clinical systems. Studies restricted to click, tone-burst, transient, or stimulus-frequency evoked OAEs were excluded, as were DPOAE studies that did not provide sufficient information from which DPOAE SEM could be determined by f_2 frequency.

Data were extracted independently by two reviewers (G.P.M. and K.M.R.). Any disagreement was resolved by discussion. In total, the search yielded a unique list of 110 studies for review. Only 10 studies met inclusion criteria and are summarized in Table 1. The studies are characterized by total sample, gender inclusion, age (range, mean, or median as available), f_2 frequencies, primary levels, SNRs, and retest intervals. For retests occurring on the same day as the baseline test, the following interval values were assigned: 0.00 for immediate retests, 0.01 for retests conducted 1–20 min after baseline, and 0.04 for retests conducted 21–60 min after baseline. When the retest interval was reported as a range, the average between the minimum and maximum values was calculated and used in the analysis. Occasionally, authors reported a combined SEM value across a range of retest intervals; when this occurred, the combined SEM was used for the purpose of the meta-analysis. Sample size and SEM values from Helleman and Dreschler (2012) were provided by the publication authors.

All papers contributing to this meta-analysis excluded data that were missing at either the baseline or the follow-up test. Missing data could have resulted from patient noise or high system distortion obliterating a low-level response or from a truly absent emission. For SEM determination in the reported publications, if the DPOAE was missing at baseline in a subject, then that subject did not contribute data at that particular f_2 frequency. If the DPOAE was missing at a follow-up but was present at baseline, then that subject did not contribute to data for that follow-up at that particular frequency. McMillan et al. (2013) describe an alternative, model-based approach to estimating test–retest standards that more

efficiently uses all available data and would give more precise estimates of the relevant parameters.

Statistical Approach

We regard the published SEM estimates across studies as samples from the population of SEMs. Thus, the goal of the meta-analysis is to synthesize this collection of studies to estimate the central tendency and variability of SEM. Variability among studies can be due to the subjects sampled, measurement reliability, and measurement conditions, among many other possibilities. These sources of variation, which are broadly applicable to almost any area of medical research, are controlled explicitly via regression analysis or implicitly via mixed effects models (DerSimonian & Kacker 2007). Here, each SEM estimate is entered into the mixed model as an observation, along with an estimate of its sampling variance. Variability among studies is modeled using random effects. Thus, meta-analysis is a formal statistical analysis of published results and provides the best estimate of the overall SEM averaged across studies.

Meta-analysis was used to estimate the population mean SEM for the studies in Table 1. Specifically, for the j th estimate given in the i th study, the SEM was modeled as

$$\widehat{SEM}_{ij} \sim \text{Normal}(\mu_{ij}, s_{ij}^2), \text{ where}$$

$$\mu_{ij} = (\beta_0 + \delta_i) + (\beta_1 + \varphi_i) \cdot D_{ij}, \text{ and}$$

$$s_{ij}^2 = \frac{\widehat{SEM}_{ij}^2}{n_{ij}} \left[n_{ij} - \left(\frac{\Gamma\left(\frac{n_{ij}+1}{2}\right)}{\Gamma\left(\frac{n_{ij}}{2}\right)} \right)^2 \right]$$

(see Appendix, Supplemental Digital Content, <http://links.lww.com/EANDH/A189>).

D_{ij} denotes the number of days that the i th study reports separating pairs of measurements (see the fourth column from the right in Table 1). The μ_{ij} term is the underlying, true i th study SEM estimated by the j th result, and the s_{ij}^2 is the standard error of that estimate. The n_{ij} is the sample size associated with the j th estimate published in the i th study. In this model, the δ_i and φ_i are study-specific random effects around the population intercept β_0 and slope β_1 , respectively. β_0 is the general population SEM, presumably applicable across clinical sites, whereas β_1 is the change in SEM per day of follow-up, also generalizable across clinical sites. The slope term relates the mean SEM to days D since baseline test. In general, correlation between repeated measurements of any kind tends to decay over time. The SEM is a function of one minus the correlation between baseline and follow-up tests (Eq. 1). Therefore, if correlation decays, then $(1 - r)$ increases with time so that the SEM should increase as the number of days between baseline and follow-up increases. The random effects δ_i and φ_i are assumed independent and follow a normal distribution with mean 0 and variance τ and γ , respectively. The population mean SEM after D days between baseline and monitoring is given by. $\widehat{SEM}(D) = \beta_0 + \beta_1 \cdot D$. Models were fit separately for each f_2 primary using SAS software, PROC MIXED, v9.3.

Results

Published data contributing to the mean DPOAE SEM calculations are provided in Table 1. The 10 studies included both men and women, and ages ranged from 18 to 60 years old. Most studies enrolled subjects with normal or near-normal (thresholds < 30 dB HL) hearing (nine out of 10 studies), and half of published SEM values used data derived from both ears of each subject. Furthermore, in-the-ear calibration was used in three of the 10 studies, another four studies employed an examiner probe-fit check, and the remaining three studies did not indicate what (if any) probe-fit calibration procedures were used. Table 1 shows obvious variation among published SEM values across frequency and time. This variability is shown in Figure 1, which plots the published SEM values by time (days) and f_2 frequency. The mean test interval between test and retest was 3.6 days and ranged from 0 days to just slightly more than 15 days, and published SEM values ranged from 0.57 to 3.9 dB. Based on any single publication, one can determine the 90% DPOAE shift reference limits based on the formula $\pm 1.645 \cdot \sqrt{2} \cdot \widehat{SEM}$. For example, the SEM value reported by Beattie et al. (2003) at 4 kHz was 3.2 dB for a retest conducted about 7 days from baseline; this corresponds to a 90% DPOAE shift reference interval of ± 7.44 dB. Therefore, a DPOAE level shift of more than 7.44 dB between tests would be larger than measurement error and taken to indicate a real change in the cochlea if one were to rely exclusively on the Beattie et al. result. However, if one were to use Keppler et al. (2010), the SEM at that frequency and time frame is 0.95, less than a third that reported by Beattie et al., and corresponding to a 90% shift reference interval of 2.21 dB. Thus, a 5 dB test–retest measurement in a patient would be reasonable using Beattie et al. as the standard but would be alarming if Keppler et al. were the standard. From a statistical point of view, these are both estimates from an underlying population of SEM and are subject to sampling variability. It would be more accurate and precise to combine these and other results to estimate the true SEM at 4 kHz over a 7-day time frame using meta-analysis instead of forcing clinicians to pick one publication for test–retest standards.

Results of the meta-analysis analysis are shown in Table 2. The intercept is the estimate of the mean SEM for a retest done on the same day as the baseline test and is consistent across f_2 , varying from 1.58 to 1.76 dB. The days from baseline are estimates of the increase in SEM per day separating baseline and follow-up measurements and increased from 0.01 to 0.04 with increasing frequency. The days from baseline effect was statistically significant for f_2 at 4000 and 6000 Hz ($p < 0.05$). Although the effects of time are consistent across f_2 , the effect is not significant at 1 and 2 kHz ($p > 0.05$). As stated earlier, the mean SEM after D days between baseline and monitoring is given by $\widehat{SEM}(D) = \beta_0 + \beta_1 \cdot D$, where β_0 is the intercept estimate in Table 2 and β_1 is the days from baseline estimate. For example, the mean SEM at 4000 Hz after 10 days is $1.76 + (0.03 \cdot 10 \text{ days}) = 2.06$ dB. After 30 days, the SEM increases to $1.76 + (0.03 \cdot 30 \text{ days}) = 2.66$ dB.

Ninety percent reference limits can be computed using the equation $\pm 1.645 \cdot \sqrt{2} \cdot \widehat{SEM}$. Table 3 shows the fitted mean SEM and the calculated 90% DPOAE level shift reference limits by f_2 (columns) and days since baseline (rows). Note that no published study in Table 1 exceeds 15 days from baseline, so that all estimates outside this range (i.e., at 20 days) are

based on extrapolation. After 10 days, DPOAE level shifts exceeding 4.24 dB at 1000 Hz, 4.35 dB at 2000 Hz, 4.85 dB at 4000 Hz, or 4.55 dB at 6000 Hz would be considered clinically alarming at the 10% false referral rate. Other 90% reference limits for days not shown in Table 3 can be computed using the equation $\pm 1.645 \cdot \sqrt{2} \cdot \widehat{SEM}$. In addition, stricter (or more lax) criteria for a “significant” DPOAE level change can be calculated. Therefore, the ranges can be made larger or smaller depending on what false referral rate is deemed appropriate medically for any given screening application. For example, a stricter criterion, such as 95% shift reference limits, which will further minimize false referral rates, can be calculated using the equation $\pm 1.96 \cdot \sqrt{2} \cdot \widehat{SEM}$.

The 90% DPOAE level shift reference limits are plotted against days from baseline in Figure 2 (gray shaded region), with each study's reference limits (symbols) based on their published SEM. Least square regression lines are fit to each study's data to visually relate the model-based reference limits to those derived in each study. As expected, the individual study data (lines connecting symbols) show variability about the meta-analytic estimate (shaded region), and the temporal trends are approximated. Interestingly, some studies, particularly Wagner et al. (2008; filled squares) and Beattie et al. (2003; filled circles), at the lower f_2 show decreased SEM (and thus smaller reference intervals) over time. Some variability in these trends is to be expected and is modeled using the random slope coefficients ϕ_i in the meta-analysis model.

Discussion

Key Findings and Significance

There is a clinical need for DPOAE shift threshold criteria for serial monitoring programs. Often, clinicians and researchers report using a set DPOAE criterion for monitoring purposes derived from the results of a single published study. While useful, the values reported by individual studies were derived from relatively small samples of control subjects and values were often nonspecific to frequency and time. Ideally, DPOAE reference intervals would be determined using a very large, diverse (in terms of gender, age, and hearing loss) sample control population measured over multiple frequencies and at multiple times over the course of several months to years. This would yield the most precise, accurate, and generalizable DPOAE reference intervals for use in monitoring programs. This goal is ambitious and has yet to be accomplished. One method for overcoming such difficult to implement studies is the use of meta-analytic techniques, which combine data from multiple published studies. The aim of this study was to synthesize published DPOAE SEM values derived from adult control populations to provide a mean SEM value that can be used to determine DPOAE level shift reference limits. Applying the meta-analytic techniques, more precise and generally applicable reference limits can be computed, which is preferable to a clinician picking any one study to establish clinical DPOAE change criteria.

These are the outcomes: First, estimates of SEM ranged from as low as 0.57 to as high as 3.9 dB among published reports. Variability among published SEM values existed even within similar time intervals. An important source of variability is a lack of precision in individual studies due to small sample sizes. In this study, published SEM values were aggregated from

various studies with the primary objective to improve the precision of the overall mean SEM estimate. In the present meta-analysis of data from 10 observational studies describing test–retest differences in DPOAE level, mean SEM values and 90% DPOAE level shift reference limits were presented based on retest time interval and f_2 frequency. These limits correspond to 10% false referral rates. However, reference limits can be determined for any false referral rate from the data provided thus allowing a clinician to adjust (make larger or smaller) the reference interval depending on the screening application.

Second, the results of the meta-analysis showed DPOAE level shift reference limits increase as the time between baseline and retest increased. In addition, the published data only extends out to about 15 days post-baseline. These two features have significant ramifications for monitoring where retest intervals can extend over several months (e.g., ototoxicity monitoring) or even occasionally years (e.g., noise conservation monitoring). As a result, if one is to rely on published data only, one must extrapolate beyond the published retest intervals. Although all SEMs increased with increasing time between tests, the retest time interval was statistically significant at 4000 and 6000 Hz. Compared with lower frequencies, reference limits at higher frequencies, such as 4000 and 6000 Hz are more likely to be influenced by calibration errors (Siegel and Hirohata 1994; Siegel 2007), especially in adults with larger ear canal volumes. Traditional calibration methods are imperfect. In the ear, calibration is susceptible to standing waves in the ear canal, pressure nodes caused by the interaction between the applied (incident) stimulus waveform and the waveform reflected back from the tympanic membrane. Depending on whether the interaction is constructive or destructive at the location of the measurement microphone, the calibrated level can be higher or lower than the SPL that reaches the ear drum. If in the ear calibration were not used, the voltage applied to the loudspeakers would be fixed from patient to patient and test to test, however, the level of the signal reaching the eardrum presumably would vary even more widely (because there are many conditions in which in the ear calibration works well). Because the location of the measurement microphone depends on probe placement, even slight variations in probe placement can alter the primary levels reaching the tympanic membrane and represents a substantial source of retest variability. It is thus critically important to control for probe placement when using DPOAEs for serial monitoring of cochlear status and to implement more accurate calibration methods. One such calibration method that has proven to reduce the dependence of stimulus level within the ear canal on probe insertion depth is forward pressure level (Souza et al. 2014).

Limitations and Caveats for Use

The meta-analysis presented provides reference limits that have increased precision relative to limits suggested by any of the individual studies, and provides both negative (emission decrement) and positive (emission enhancement) shift thresholds that are appropriate for the f_2 and time interval tested. For proper interpretation and utilization of the results, it is important to understand the caveats for use.

Patient Population—An important caveat pertains to the patient population undergoing serial monitoring. For example, adult cancer patients receiving the ototoxic chemotherapeutic agent cisplatin are likely to be older, with poorer hearing than the subjects

used in published studies. Thus, one must be careful applying the shift reference limits in Table 3 to a hearing-impaired population that may have low level emissions associated with their impaired hearing. For example, Thorson et al. (2012) noted that in a sample that included some individuals with hearing loss, small amplitude emissions (defined as less than 0 dB SPL) were more variable over time compared with emissions with larger amplitudes. This observed relation between low-level emission and increase variability may be related to the DPOAE SNR. Therefore, response SNR is an important component to consider when interpreting serial DPOAE measurements because small SNRs indicate there is some influence of noise on the measured DPOAE level. Furthermore, for ototoxicity monitoring, the serial monitoring interval will often eventually exceed any published time interval. On the other hand, our meta-analysis is entirely suitable for serial monitoring of patients with cystic fibrosis receiving aminoglycosides. These are generally younger patients with good hearing who are being treated daily over a 2 to 3 week period. Results from Table 3 can be confidently used. Serial monitoring for noise induced hearing loss, particularly in industrial or military applications, tends to occur over very long periods of time (over a year), but are more often done in younger populations. Thus, the meta-analysis can be used because it is based on studies of younger subjects, although extrapolation to long time frames must be done cautiously.

Stimulus Frequency—Results of the meta-analysis are limited to the small discrete set of frequencies 1000, 2000, 4000, and 6000 Hz. Although some authors published SEM values at other frequencies, not enough data existed in the literature to be included in a formal meta-analysis. For those interested in shift threshold limits at other frequencies, refer to the above referenced articles, keeping in mind the limits should be restricted to a similar population, methodology, and time interval.

Stimulus Level—DPOAE responses contributing to the meta-analysis were restricted to those obtained at moderate stimulus levels. One would expect even greater variability among DPOAEs recorded using low-level stimuli because they will generate smaller emissions that are more susceptible to the influence of noise. However, increasing the stimulus to high-levels in an attempt to maximize the response level may decrease sensitivity to cochlear pathologies compared with low-level stimuli (Wier et al. 1988; Sutton et al. 1994; Stover et al. 1996; Katbamna et al. 1999; Johnson et al. 2010; Rao and Long 2011). Use of DPOAEs generated by moderate levels maximizes the SNR while generating a potentially more sensitive test.

Retest Interval—Published SEMs cover about a 15 day time window. Estimate of reference limits beyond that time, such as one needs for ototoxicity or noise exposure monitoring, require extrapolation. This restricts the generalization of the results to short-term serial monitoring until more data is published documenting the variability of DPOAEs over longer time intervals. If these longer time horizons are immediately necessary, cautious use of the extrapolated shift thresholds is recommended, with careful error checks in a reference sample to ensure that the 10% false referral rate is maintained.

Multiple Test Frequencies—Another important caveat is that virtually all screening reference limits are developed with the, usually implicit, assumption that patient status is dictated by the test-retest differences for a single measurement condition. If this is a reasonable assumption, then reference limits based on the meta-analysis and on the univariate statistical methods described by McMillan et al. (2013) are applicable. In practice, however, audiologists typically measure DPOAEs at several, possibly correlated, frequencies and/or levels when determining whether auditory function is stable or changed over time. However, when multiple measures are obtained on a single patient, false referral rates increase. (This problem is not unique to screening with DPOAEs, screening with tympanometry, or pure-tone testing also typically involves multiple measures on patients as do screening tests in other realms of medicine.) Establishing appropriate multivariate reference limits would help to neutralize sporadic false referral rates at any one DPOAE measured frequency, thus avoiding unnecessary follow-up visits. In screening test development, there is no commonly accepted way to handle multiple measurements across, in this instance, test frequencies. To address this type of problem, Boyd and Lacher (1982) have proposed computing multivariate reference limits in which all of the patient's test-retest differences across all metrics and stimulus conditions are compiled into a single response function. This methodology allows the clinician to compute a “ p value” as a measure of support for the notion that the patient is in some general sense abnormal compared with the reference population. A meta-analytic synthesis of published research using Boyd and Lacher's methodology would require published covariances among all pairs of f_2 primaries, which to our knowledge has never been done. In the meantime, we offer the above reference limits for a single measurement condition and when multivariate reference limits are established, the univariate reference limits here are available as a confirmatory check of the multivariate results.

Which Frequency to Monitor—In applying reference limits, it is assumed there is a single value for which comparisons are made. Clinicians are encouraged to collect DPOAEs at multiple test frequencies and levels depending on the protocol established for their clinic. Most clinicians would argue that this helps them determine the reliability of the measurements. However, an a priori choice can be made about which frequency or bands of frequencies to monitor, based on the expected damage region from an exposure as well as the baseline DPOAE results. For example, it is expected that cisplatin exposure will systematically damage the cochlea in a base to apex fashion (Brummett 1980; Komune et al. 1981; Nakai et al. 1982; Konishi et al. 1983; Schweitzer et al. 1984) and it has been demonstrated that higher frequency DPOAEs change before lower frequency DPOAEs (Reavis et al. 2011). In this situation, monitoring the highest DPOAE frequency with a valid response and accompanying reference limits is preferable. However, when monitoring for noise over-exposure, 4000 or 6000 Hz may be the best frequency because they are classically associated with the noise notch. Although 4000 and 6000 Hz may seem more likely to reflect cochlear changes, 2000 Hz, and even possibly 1000 Hz, may prove uniquely useful for capturing DPOAE enhancements associated with noise (Helleman & Dreshler 2012; Mei et al. 2009) and medication ototoxicity (Kakigi et al. 1998; Huang et al. 2005; Konrad-Martin et al. 2012b). Thus, if baseline measures preclude the ability to monitor higher frequencies, these lower frequencies may still be valuable to test.

Follow-Up for Significant DPOAE Change—Detecting significant DPOAE changes using reference limits are useful for serial monitoring. Once a change is detected (i.e., difference between baseline visit DPOAE and subsequent visit DPOAE exceeds the shift reference limit at a single frequency), the next step is usually to follow up with a more comprehensive audiometric test battery, such as an audiogram (for older children and adults) or an ABR test (for a pediatric population). DPOAE changes can be temporary and/or occur in the absence of hearing loss, which may be indicative of their sensitivity to preclinical damage. Specific protocols for their use are based on DPOAE theory, the population under test, the clinical objectives of a screening program, and known patterns of damage (see Konrad-Martin et al. 2012a, 2014, for more detailed discussions).

Clinician's Test–Retest Variability—The clinician's retest reliability is critical to successful application of these or any published DPOAE reference limits; therefore, it is recommended that each clinician or group of clinicians (if monitoring is done by more than one individual) assess their test–retest reliability for the specific protocol to be used in the serial monitoring using a statistically sound method. The general approach is to test and retest a healthy sample to determine if the observed false referral rate (the percent of ears that exceed the reference limit for a specific frequency at a specific time interval) approximates the expected 10%. Observed false referral rates much less than 10% suggests better reliability than that found in the studies contributing to the limits provided here. This could occur in situations where strict measures are employed for probe placement or when calibration measures are employed which minimize the influence of standing waves on the calibrated values, such as forward pressure level calibration (Souza et al. 2014). In this situation, narrower limits might be appropriate. Conversely, achieving false referral rates much greater than 10% would result in over-referral of healthy patients for follow-up audiometric testing. Here, a clinician might believe a patient has an alarming DPOAE change when, in fact, no change exists and wider reference limits might be more appropriate. It also may be that in this case, the clinician can make changes to their technique to improve their variability. The bottom line: clinicians should ensure that their own test–retest performance is similar to that reported here before using DPOAE change criteria based on these results.

Future Directions

New methods for optimizing the sensitivity of DPOAEs and decreasing test time are being explored by many research groups. Reference limits will need to be developed for methods proven most effective (such as DPOAE levels measured at multiple frequencies, frequency ratios, or levels, measures that incorporate DPOAE phase, use of DPOAE source separation techniques), and for other types of emissions. Work is also ongoing to improve technological limitations of clinical DPOAE systems, such as calibration, noise and system distortion, to decrease the variability of DPOAEs measured over time. For populations with good hearing (infants, children, young adults), it will be advantageous to test to higher frequencies than is now possible. In addition, there is a definite clinical need for DPOAE test–retest measurements obtained over longer time periods that better reflect the testing regimen, across more frequencies, and in control populations that better represent the

individual patient under test. Technological improvements and additional data in any of these areas will necessitate revision of reference limits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks to Drs. Helleman and Dreschler for providing their data for use in this analysis.

This research was funded by the Office of Rehabilitation Research and Development Service, U.S. Department of Veterans Affairs (Grants C7223R, C6373R, and C7113N). The contents of this manuscript do not represent the views of the U.S. Department of Veterans Affairs or the U.S. Government.

References

- Beattie RC. Distortion product otoacoustic emissions: Comparison of sequential versus simultaneous presentation of primary-tone pairs. *J Am Acad Audiol*. 2003; 14:471–484. [PubMed: 14708836]
- Beattie RC, Kenworthy OT, Luna CA. Immediate and short-term reliability of distortion-product otoacoustic emissions. *Int J Audiol*. 2003; 42:348–354. [PubMed: 14570243]
- Boyd JC, Lacher DA. The multivariate reference range: An alternative interpretation of multi-test profiles. *Clin Chem*. 1982; 28:259–265. [PubMed: 7055945]
- Brownell WE. Outer hair cell electromotility and otoacoustic emissions. *Ear Hear*. 1990; 11:82–92. [PubMed: 2187727]
- Brummett RE. Drug-induced ototoxicity. *Drugs*. 1980; 19:412–428. [PubMed: 7389589]
- Demorest ME, Walden BE. Psychometric principles in the selection, interpretation, and evaluation of communication self-assessment inventories. *J Speech Hear Disord*. 1984; 49:226–240. [PubMed: 6748618]
- DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: An update. *Contemp Clin Trials*. 2007; 28:105–114. [PubMed: 16807131]
- Dreisbach LE, Long KM, Lees SE. Repeatability of high-frequency distortion-product otoacoustic emissions in normal-hearing adults. *Ear Hear*. 2006; 27:466–479. [PubMed: 16957498]
- Franklin DJ, McCoy MJ, Martin GK, Lonsbury-Martin BL. Test/retest reliability of distortion-product and transiently evoked otoacoustic emissions. *Ear Hear*. 1992; 13:417–429. [PubMed: 1487104]
- Gorga MP, Neely ST, Ohlrich B, Hoover B, Redner J, Peters J. 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*. 1997; 18:440–455. [PubMed: 9416447]
- Harding GW, Bohne BA. Temporary DPOAE level shifts, ABR threshold shifts and histopathological damage following below-critical-level noise exposures. *Hear Res*. 2004; 196:94–108. [PubMed: 15464306]
- Helleman HW, Dreschler WA. Overall versus individual changes for otoacoustic emissions and audiometry in a noise-exposed cohort. *Int J Audiol*. 2012; 51:362–372. [PubMed: 22436020]
- Huang ZW, Luo Y, Wu Z, et al. Paradoxical enhancement of active cochlear mechanics in long-term administration of salicylate. *J Neurophysiol*. 2005; 93:2053–2061. [PubMed: 15590729]
- Johnson TA, Neely ST, Kopun JG, Dierking DM, Tan H, Gorga MP. Clinical test performance of distortion-product otoacoustic emissions using new stimulus conditions. *Ear Hear*. 2010; 31:74–83. [PubMed: 19701088]
- Kakigi A, Hirakawa H, Harel N, Mount RJ, Harrison RV. Basal cochlear lesions result in increased amplitude of otoacoustic emissions. *Audiol Neurotol*. 1998; 3:361–372. [PubMed: 9732130]
- Katbamna B, Homnick DN, Marks JH. Effects of chronic tobramycin treatment on distortion product otoacoustic emissions. *Ear Hear*. 1999; 20:393–402. [PubMed: 10526861]

- Kemp DT. Stimulated acoustic emissions from within the human auditory system. *J Acoust Soc Am*. 1978; 64:1386–1391. [PubMed: 744838]
- Keppeler H, Dhooge I, Maes L, et al. Transient-evoked and distortion-product otoacoustic emissions: A short-term test-retest reliability study. *Int J Audiol*. 2010; 49:99–109. [PubMed: 20151884]
- Komune S, Asakuma S, Snow JB Jr. Pathophysiology of the ototoxicity of cis-diamminedichloroplatinum. *Otolaryngol Head Neck Surg*. 1981; 89:275–282. [PubMed: 6787526]
- Konishi T, Gupta BN, Prazma J. Ototoxicity of cis-dichlorodiammine platinum (II) in guinea pigs. *Am J Otolaryngol*. 1983; 4:18–26. [PubMed: 6687035]
- Konrad-Martin D, Reavis KM, McMillan GP, Dille MF. Multivariate DPOAE metrics for identifying changes in hearing: Perspectives from ototoxicity monitoring. In *J of Audiol*. 2012a; 51:S51–S62.
- Konrad-Martin D, Jacobs PG, McMillan G, Reavis KM, Dille MF. Usefulness of source knowledge for detecting ototoxicity using DPOAEs. *Assoc Res Otolaryngol Abs*. 2012b; 842:294.
- Konrad-Martin D, Reavis KM, McMillan G, Helt WJ, Dille MF. Proposed comprehensive ototoxicity monitoring program for VA healthcare (COMP-VA). *J Rehabil Res Dev*. 2014; 51:81–100. [PubMed: 24805896]
- Kumar UA, Methi R, Avinash MC. Test/retest repeatability of effect contralateral acoustic stimulation on the magnitudes of distortion product otoacoustic emissions. *Laryngoscope*. 2013; 123:463–471. [PubMed: 23161444]
- Lapsley Miller, JA.; Marshall, L. Otoacoustic emissions as a preclinical measure of noise-induced hearing loss and susceptibility to noise-induced hearing loss. In: Robinette, MS.; Glatke, TJ., editors. *Otoacoustic Emissions: Clinical Applications*. 3rd. New York, NY: Thieme; 2007. p. 321–341.
- Lapsley Miller JA, Marshall L, Heller LM, Hughes LM. Low-level otoacoustic emissions may predict susceptibility to noise-induced hearing loss. *J Acoust Soc Am*. 2006; 120:280–296. [PubMed: 16875225]
- Lonsbury-Martin BL, Martin GK. Otoacoustic emissions. *Curr Opin Otolaryngol Head Neck Surg*. 2003; 11:361–368. [PubMed: 14502067]
- Marshall L, Lapsley Miller JA, Heller LM, et al. Detecting incipient inner-ear damage from impulse noise with otoacoustic emissions. *J Acoust Soc Am*. 2009; 125:995–1013. [PubMed: 19206875]
- Martin GK, Stagner BB, Chung YS, Lonsbury-Martin BL. Characterizing distortion-product otoacoustic emission components across four species. *J Acoust Soc Am*. 2011; 129:3090–3103. [PubMed: 21568412]
- McMillan GP, Reavis KM, Konrad-Martin D, Dille MF. The statistical basis for serial monitoring in audiology. *Ear Hear*. 2013; 34:610–618. [PubMed: 23563060]
- Mei L, Huang ZW, Tao ZZ. Paradoxical long-term enhancement of distortion product otoacoustic emission amplitude after repeated exposure to moderate level, wide band noise in awake guinea pigs. *J Laryngol Otol*. 2009; 123:1090–1096. [PubMed: 19607735]
- Mulheran M, Degg C. Comparison of distortion product OAE generation between a patient group requiring frequent gentamicin therapy and control subjects. *Br J Audiol*. 1997; 31:5–9. [PubMed: 9056039]
- Nakai Y, Konishi K, Chang KC, et al. Ototoxicity of the anti-cancer drug cisplatin: An experimental study. *Acta Otolaryngol*. 1982; 93:227–232. [PubMed: 7199807]
- Ng IHY, McPherson B. Test-retest reliability of distortion-product otoacoustic emissions in the 1 to 7 kHz range. *Audiol Med*. 2005; 3:108–115.
- Rao A, Long GR. Effects of aspirin on distortion product fine structure: Interpreted by the two-source model for distortion product otoacoustic emissions generation. *J Acoust Soc Am*. 2011; 129:792–800. [PubMed: 21361438]
- Reavis KM, McMillan G, Austin D, et al. Distortion-product otoacoustic emissions test performance for ototoxicity monitoring. *Ear Hear*. 2011; 32:61–74. [PubMed: 20625302]
- Roede J, Harris FP, Probst R, et al. Repeatability of distortion product otoacoustic emissions in normally hearing humans. *Audiology*. 1993; 32:273–281. [PubMed: 8216026]
- Schweitzer VG, Hawkins JE, Lilly DJ, et al. Ototoxic and nephrotoxic effects of combined treatment with cis-diamminedichloroplatinum and kanamycin in the guinea pig. *Otolaryngol Head Neck Surg*. 1984; 92:38–49. [PubMed: 6422414]

- Shera CA, Guinan JJ Jr. Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs. *J Acoust Soc Am*. 1999; 105(2 Pt 1):782–798. [PubMed: 9972564]
- Siegel, JH. Calibrating otoacoustic emission probes. In: Robinette, MS.; Glatke, TJ., editors. *Otoacoustic Emissions: Clinical Applications*. 3rd. New York, NY: Thieme; 2007. p. 403–427.
- Siegel JH, Hirohata ET. Sound calibration and distortion product otoacoustic emissions at high frequencies. *Hear Res*. 1994; 80:146–152. [PubMed: 7896573]
- Souza NN, Dhar S, Neely ST, Siegel JH. Comparison of nine methods to estimate ear-canal stimulus levels. *J Acoust Soc Am*. 2014; 136:1768–1787. [PubMed: 25324079]
- Stavroulaki P, Vossinakis IC, Dinopoulou D, et al. Otoacoustic emissions for monitoring aminoglycoside-induced ototoxicity in children with cystic fibrosis. *Arch Otolaryngol Head Neck Surg*. 2002; 128:150–155. [PubMed: 11843723]
- Stover L, Gorga MP, Neely ST, Montoya D. Toward optimizing the clinical utility of distortion product otoacoustic emission measurements. *J Acoust Soc Am*. 1996; 100:956–967. [PubMed: 8759949]
- Sutton LA, Lonsbury-Martin BL, Martin GK, Whitehead ML. Sensitivity of distortion-product otoacoustic emission in humans to tonal over-exposure: Time course of recovery and effects of lowering L2. *Hear Res*. 1994; 75:161–174. [PubMed: 8071143]
- Thorson MJ, Kopun JG, Neely ST, Tan H, Gorga MP. Reliability of distortion-product otoacoustic emissions and their relation to loudness. *J Acoust Soc Am*. 2012; 131:1282–1295. [PubMed: 22352502]
- Wagner W, Heppelmann G, Vonthein R, Zenner HP. Test-retest repeatability of distortion product otoacoustic emissions. *Ear Hear*. 2008; 29:378–391. [PubMed: 18382378]
- Wier CC, Pasanen EG, McFadden D. Partial dissociation of spontaneous otoacoustic emissions and distortion products during aspirin use in humans. *J Acoust Soc Am*. 1988; 84:230–237. [PubMed: 3411052]

Fig. 1.

Published SEM values as a function of days since baseline, with each panel representing a different frequency (1000–6000 Hz). Each symbol is the SEM of different published study. SEM, standard error of the measurement.

Fig. 2.

Publication data (*symbols*) and 90% reference limits (*shaded gray region*) by primary f_2 frequency derived from the SEM meta-analysis results presented in Table 2. Reference intervals beyond about 15 days need to be extrapolated from the model results since publication data does not extend beyond about 2 weeks. SEM, standard error of the measurement.

Table 1
Published data from 10 studies meeting inclusion criteria that contributed to estimates of mean SEM by frequency and time

Author(s)	Equipment	L1/L2 (dB SPL)	Males: Females	Sample Size	Mean Age (Range)	DPOAE Criteria	Days Between Tests	f_2 (kHz)	Ears	SEM
Beattie (2003)	GSI-60	65/55	0:62	62	23 (19–26)	10 dB SNR	0.01	1	62	2.6
							0.01	2	62	2.2
							0.01	4	62	1.4
Beattie et al. (2003)	GSI-60	65/65	0:50	50	24 (19–27)	6 dB SNR	0.01	1	50	3.2
							7.5	1	50	2.6
							0.01	2	50	2.7
							7.5	2	50	2.5
							0.01	4	50	2.1
							7.5	4	50	3.2
Franklin et al. (1992)	Custom	65/65	7:5	12	30 (19–44)	3 dB SNR	2.7	1	12	1.42
							2.7	2	12	2.26
							2.7	4	12	0.95
							2.7	6	12	1.22
							11.7	1	12	1.46
							11.7	2	12	1.29
							11.7	4	12	1.60
							11.7	6	12	1.41
Helleman and Dreschler (2012)	ILO288	75/70	60:0	60	42 (23–60)	S > 2 SD above N	8	1	28	2.1
							8	2	14	2.9
							8	4	15	3.8
							8	6	33	3.9
Kumar et al. (2013)	GSI-Audera	65/55	24:0	24	25 (18–45)	6 dB SNR	7.5	1	24	0.69
							7.5	2	24	0.86
							7.5	4	24	1.40
							7.5	6	24	0.72
Keppeler et al. (2010)	ILO288 USB II	65/55	14:15	29	22 (19–28)	S > N, averaged over several f_2 's	0	1	53	0.96
							0.04	1	55	1.36
							7	1	51	1.46

Author(s)	Equipment	L1/L2 (dB SPL)	Males: Females	Sample Size	Mean Age (Range)	DPOAE Criteria	Days Between Tests	f_2 (kHz)	Ears	SEM
Lapsley Miller et al. (2006)	ILO292	59/50	20:8	28	Female 31* (20–53) Male 26* (20–47)	S > 2 SD above N	0	2	54	0.8
							0.04		56	0.74
							7	2	52	0.97
							0	4	54	0.79
							0.04	4	56	0.66
							7	4	52	0.92
							0	6	54	0.78
							0.04	6	56	0.57
							7	6	52	0.91
							0	1	40	0.86
Marshall et al. (2009)	ILO292	59/50	32:0	32	20* (18–27)	S > 2 SD above N	0.04	1	42	0.99
							7	1	40	1.28
							0	2	45	0.76
							0.04	2	47	0.73
							7	2	46	0.83
							0	4	48	0.8
							0.04	4	50	0.67
							7	4	48	0.95
							0	6	49	0.8
							0.04	6	51	0.59
Ng and McPherson (2005)	ILO96	70/70	18:17	35	23 (19–35)	6 dB SNR at 1 kHz 12 dB SNR at 2, 4, and 6 kHz	7	6	48	0.92
							1	2	39	2.9
							1	4	35	2
							1.5	4	51	2.9
							0.01	1	61	1.71
							0.01	2	61	1.63
							0.01	4	61	2.42
							0.01	6	61	2.18
							15.4	1	61	2.94
							15.4	2	61	2.13
							15.4	4	61	2.65

Author(s)	Equipment	L1/L2 (dB SPL)	Males: Females	Sample Size	Mean Age (Range)	DPOAE Criteria	Days Between Tests	f_2 (kHz)	Ears	SEM
Wagner et al. (2008)	Custom	63/60	11:9	20	25 (21–43)	6 dB SNR	15.4	6	61	2.85
							5.9	1	40	1.44
							5.9	2	40	1.21
							5.9	4	40	1.81
							5.9	6	40	2.73
					24 (20–38)	6 dB SNR	6.9	1	40	1.01
							6.9	2	40	0.91
							6.9	4	40	1.37
							6.9	6	40	1.88

* Sample characteristics, DPOAE parameters, test interval, and SEM are presented. Mean age is presented unless otherwise noted by, in which case the median is reported. Retests occurring on the same day as the baseline test were assigned 0.00 for immediate retests; 0.01 for retests conducted 1–20min following baseline; and 0.04 for retests conducted 21–60min following baseline. When days between tests were reported as a range (e.g., retest was conducted between 2 and 14 days), the average between the minimum and maximum values was calculated (e.g., 8 days). DPOAE, Distortion product otoacoustic emission; SEM, standard error of the measurement; SNR, signal-to-noise ratio.

Table 2

Meta-analysis results

f_2	Intercept		Days From Baseline		
	Estimate	Standard Error	p	Estimate	Standard Error
1000	1.69	0.28	<0.001	0.01	0.03
2000	1.70	0.27	<0.001	0.02	0.01
4000	1.76	0.24	<0.001	0.03	0.01
6000	1.58	0.42	0.013	0.04	0.01

Intercept estimate is the mean SEM for a retest done on the same day as the baseline test. Days from baseline estimate is the increase in SEM per day separating baseline and follow-up measurements. SEM, standard error of the measurement.

Table 3
Table of mean SEM with upper and lower 90% reference limits of each f_2 primary frequency by days since baseline test

Days From Baseline	f_2 Frequency											
	1000				2000				4000			
	SEM	90% Reference Limits	SEM	90% Reference Limits	SEM	90% Reference Limits	SEM	90% Reference Limits	SEM	90% Reference Limits	SEM	90% Reference Limits
1	1.7	±3.95	1.7	±3.98	1.8	±4.16	1.6	±3.76				
2	1.7	±3.98	1.7	±4.03	1.8	±4.23	1.7	±3.85				
3	1.7	±4.02	1.7	±4.07	1.9	±4.31	1.7	±3.93				
4	1.7	±4.05	1.8	±4.11	1.9	±4.39	1.7	±4.02				
5	1.8	±4.08	1.8	±4.15	1.9	±4.47	1.8	±4.11				
6	1.8	±4.11	1.8	±4.19	2.0	±4.54	1.8	±4.20				
7	1.8	±4.15	1.8	±4.23	2.0	±4.62	1.8	±4.29				
8	1.8	±4.18	1.8	±4.27	2.0	±4.70	1.9	±4.37				
9	1.8	±4.21	1.9	±4.31	2.1	±4.78	1.9	±4.46				
10	1.8	±4.24	1.9	±4.35	2.1	±4.85	2.0	±4.55				
15	1.9	±4.41	2.0	±4.56	2.3	±5.24	2.1	±4.99				
20	2.0	±4.57	2.0	±4.76	2.4	±5.63	2.3	±5.43				

The 90% reference limits can be computed at other days not shown using the equation $\pm 1.645 \cdot \sqrt{2} \cdot \widehat{\text{SEM}}(D)$. However, results beyond about 15 days need to be extrapolated and should be used with caution. Results at 20 days, shown below, are extrapolated from the fitted model. SEM, standard error of the measurement.