



HHS Public Access

Author manuscript

Int J Audiol. Author manuscript; available in PMC 2019 September 01.

Published in final edited form as:

Int J Audiol. 2018 September ; 57(SUP4): S89–S98. doi:10.1080/14992027.2017.1417644.

Clinical Trials, Ototoxicity Grading Scales, and the Audiologist's Role in Therapeutic Decision Making

Kelly A. King, Au.D., Ph.D. and **Carmen C. Brewer, Ph.D.**

Audiology Unit, Otolaryngology Branch, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD, USA

Abstract

Objectives: Define clinical trials and adverse event (AE) monitoring from the perspective of the audiologist. Rationalize the importance of audiology's involvement before, during, and after monitoring. Identify strengths and weaknesses in toxicity grading scales, and discuss factors that may influence these.

Design: Literature involving commonly cited grading scales used to capture ototoxicity is reviewed. Current regulations and language associated with clinical trial implementation and AE monitoring are described. Personal observations based on a variety of clinical populations are drawn from years of experience developing and employing ototoxicity monitoring protocols in a complex medical setting.

Results: Six commonly used grading scales for ototoxicity are systematically reviewed for strengths and weaknesses. Necessary considerations that inform selection of grading scales are presented. A review of and historical context for clinical trial development and AE monitoring is provided.

Conclusions: The audiologist's role in therapeutic decision making goes beyond collection of the audiogram. Clear communication to stakeholders in ototoxicity monitoring is paramount, and toxicity grading scales are one tool to facilitate this exchange. Various factors should be considered in advance of selecting the most appropriate scale to capture hearing loss, and no scale is without limitation.

Keywords

Ototoxicity; Ototoxicity monitoring; Clinical trial; Ototoxicity grading scale; Adverse event

Introduction

The ability of the audiologist to meaningfully communicate their results to the necessary stakeholders in the context of global and personalized therapeutic decision-making is principal in nature to the purpose of ototoxicity monitoring. One tool available for these purposes is the use of grading scales, which aim to operationally capture when ototoxicity

occurs and, in many cases, the degree of impairment. These scales provide a measure of objectivity and consistency in the interpretation of data, and typically use a metric that is more approachable to the non-audiologist than the varied and numerous data points captured by the audiogram.

On an individual level, ototoxicity grading scales allow a referring medical team (when guided by the audiologist) to easily and quickly evaluate whether a change in hearing has occurred, if that change is related to the intervention in question, and whether that change is likely to impact daily living, requires therapeutic referral, or both. Without the consultation of an audiologist to assess the validity of the data and interpret it in the context the patient's profile, these scales in isolation are far less meaningful. Nonetheless, use of grading scales offers uniformity in how ototoxicity is reported across clinicians and departments.

Globally, there is an unmet need for consistency in reporting ototoxic effects in large clinical cohorts so that data can be effectively synthesized to improve clinical care (Neuwelt and Brock 2010; Chang and Chinosornvatana 2010). Literature on many ototoxic agents reveals a wide-ranging incidence of associated otopathology. For example, the incidence of ototoxicity from cisplatin-based chemotherapies for head and neck cancers ranges from 17 to 88%, depending, in part, on how hearing loss is defined (Schmitt and Page 2017). This ambiguous rate of occurrence limits the ability to prognosticate risk for patients and a lack of clear and consistently-defined outcomes across cohorts hampers efforts to determine efficacy of potentially otoprotective interventions.

Why Monitor? Going Beyond the Audiogram

Much of the conversation surrounding ototoxicity monitoring involves the *how*. There are a number of excellent existing resources that address the current state of evidenced-based monitoring, including several in this Supplemental Edition (Brooks and Knight 2017; Garinis et al. 2017; Konrad-Martin et al. 2017). The aim herein, however, is a recapitulation of the *why*. The primary aim of an ototoxicity monitoring program (OMP) is to ensure the early identification of hearing loss (Konrad-Martin et al. 2014; Brooks and Knight 2017). This information can, at times, prevent functional hearing loss by allowing for alternative therapies or by influencing drug prescribing procedures; specifically, smaller or less frequent doses, or interruption or suspension of treatment altogether. Monitoring for ototoxicity can also lead to the provision of care and support for the patient and the family (Konrad-Martin et al. 2014). In this role, audiologists counsel regarding the signs and symptoms of hearing loss, recommend re/habilitation when necessary, and allow for informed therapeutic decision making. This latter purpose is critical, and yet often overlooked. For a patient to meaningfully participate in their own care and make informed decisions about treatment, they must have an understanding of what their hearing loss means in the context of their current lives and the lives they hope to return to at the completion of therapy. *The role of the audiologist to inform and care for patients and families is necessary whether or not an alternative therapy exists.* Finally, monitoring takes place in order to evaluate drug safety and sometimes efficacy, particularly in the domain of clinical trials.

The commonality amongst all of these goals is communication. Whether explaining to the patient the need for monitoring, which may improve compliance and allow for early detection, or capturing for the referring physician the difference between a 30 dB decline in hearing at 8 kHz versus 2 kHz, or outlining the initial ototoxic profile of a new drug for regulatory agencies, these scenarios involve conveying information that is meaningful to various stakeholders. This requires a kind of professional code switching, shaping clinical data into a language that can best be consumed by the recipient. Jargon should be avoided and the presentation should be contextualized to the unique individual or situation.

The remainder of this report will focus on audiology's role in clinical trial development and implementation, emphasizing the use of grading scales as the main metric for communication. The principles and *a priori* considerations discussed should be applicable to most ototoxicity monitoring programs, whether grading scales are used or not. The authors' experience with clinical trial development and implementation is based almost exclusively in the United States (U.S.) and, therefore, applicable U.S. regulatory institutions and procedures will be highlighted. All case examples and data presented were collected at the National Institutes of Health Clinical Center in Bethesda, MD, and were done so via protocols approved by an institutional review board (IRB), and following ascertainment of informed consent or assent (when applicable).

Clinical Trials

Studies designed to evaluate the safety and efficacy of new medical interventions in humans are designated clinical trials. Typically, clinical trials involve interventions aimed at improving detection, diagnosis, management, treatment, or prevention of disease. They represent the initial efforts to study the effect in question in humans. They are often preceded, sometimes by years, of work using *in vitro* and *in vivo* models in the laboratory. Once a proof-of-concept is established during preclinical work, the most promising studies move to human cohorts in the form of clinical trials. In the U.S., the Food and Drug Administration (FDA) determines if there is sufficient evidence to justify trial of a new medical intervention in humans. This is achieved through a process in which the developer applies to the FDA for an investigational new drug (IND) designation. The FDA is responsible for regulating clinical trials, and works to inform and protect patients who choose to participate in them. Ultimately, it is the FDA that determines whether a new medical intervention is safe and effective to use, and that benefits outweigh potential risks. Counterparts to the FDA exist around the world and function in a similar capacity (e.g., Australian Therapeutic Goods Administration, Health Canada, European Medicines Agency, Japanese Pharmaceutical and Food Safety Bureau, Saudi Food and Drug Authority).

Clinical trials are conducted in different phases that vary by scale and scope. Phase 1 studies are generally first-to-human trials when a new intervention is initially examined. Recruitment is intentionally kept small and the focus of these studies is to establish the safety profile of the intervention and gather early information regarding the appropriate therapeutic window, whereby the maximum benefit is achieved with the least degree of toxicity or side effects. The information learned during a phase 1 clinical trial lays the groundwork for developing later phases. Phase 2 studies are an extension of phase 1 work:

they are typically not large enough to determine if the intervention is working, but they further determine what side effects may exist and help to guide researchers in refining their experimental questions to design future experiments. Phase 3 clinical trials are large scale (e.g., hundreds to thousands of participants) and intended to determine efficacy and monitor for adverse events (AEs). These studies are also longer in duration than earlier phases and are better suited to identify side effects that may have gone undetected, or that may only occur after extended exposure. The final phase, 4, occurs after a drug or device has been approved by the FDA for use, and surveils after-market safety (Food and Drug Administration, 2017).

Adverse Events

When an unintended or undesirable experience (e.g., sign, symptom, abnormal laboratory test, disease) occurs in a patient exposed to a medical product, it is labeled an AE. Such events can be expected or unexpected, temporary or permanent, and can range from mild to fatal (National Cancer Institute 2010).

Government agencies (e.g., the FDA) rely on a common language that can identify these occurrences uniformly across clinical trials, protocols, and study sites to determine the safety of new products. Human subjects research requires approval and oversight by an IRB, the members of which need to understand the impact of AEs on multiple organ systems. Similarly, consistent language needs to be used to document AEs within medical records, and to facilitate accurate designation of AEs in scientific research. Reporting of such events is mandated through federal regulations.

In 1982, the National Cancer Institute (NCI) developed the Common Toxicology Criteria (CTC) (National Cancer Institute 1982), later named the Common Terminology Criteria for Adverse Events (CTCAE), for use in reporting and summarizing treatment-related AEs across studies and IND reports to the FDA, and for use in publications. The CTCAE became the worldwide standard dictionary for reporting acute AEs in cancer clinical trials and since has been translated into several languages. The most recent version, CTCAE 4.03 (National Cancer Institute 2010), improved alignment of standardized terminology with the international- and clinically-validated Medical Dictionary for Regulatory Activities, known as MedDRA. In addition to its use in clinical trials, the CTCAE also serves as standardized terminology to document the occurrence and seriousness of AEs in the medical record and scientific reports. While the CTCAE covers multiple organ systems, a similar overarching scheme is used to assign grades based on predicted or observed impact to the patient. These grades range from 1, assigned to mild AEs for which intervention is not indicated, to 4, which is assigned to AEs with life threatening consequences, and grade 5 which documents an AE-associated death. Application of the CTCAE to audiologic data is covered in the subsequent section.

Defining Ototoxic Change

Ototoxicity grading scales have been developed largely as instruments for consistent and accessible communication of audiometric test results. While it would be ideal for professional stakeholders to become audiologically literate, the value of simple and

categorical assessments to convey information across stakeholders regardless of prior familiarity with hearing data cannot be underestimated.

Despite the inherent usefulness and availability of a number of ototoxicity grading scales, most clinicians do not use (or consistently use) these scales in clinical OMPs. Consider the audiology report that describes a *10–15 dB decline in hearing at 6 and 8 kHz bilaterally*. What does this mean to the managing medical team? Such a statement is probably not useful on its own. Consider, also, the variability in reported incidence of ototoxicity across studies (Schmitt and Page 2017; Konrad-Martin et al. 2017). Some of this heterogeneity is attributed to variations in disease, dosing and treatment schemes, methods of administration, co-administration of concurrent ototoxic agents or agents that potentiate ototoxicity (e.g., radiation), patient age, and other patient-related variables. How ototoxicity is captured and defined, however, remains a significant and troubling contribution to the inconsistencies between preclinical and clinical data and across patient cohorts.

Ideally, identification of ototoxicity includes determination of pre-treatment hearing thresholds. Knowing that a hearing loss exists prior to treatment provides data necessary to determine whether post-treatment hearing status reflects a treatment-related decline or pre-existing hearing loss. Reliance on a subjective report of change in hearing or the use of age- and sex-matched normative data are insufficient techniques to accurately determine if an ototoxic change occurred. For example, only 10% of the patients shown in Figure 1A had hearing thresholds worse than those predicted by the 95th percentile for their age and sex on a post-treatment audiogram. When these data are re-examined in the context of a baseline hearing test (Figure 1B), clinically-significant changes in hearing occurred in more than twice as many patients. This would have gone unrecognized in the absence of a baseline hearing test. Importantly, the presence of significant hearing loss at a pre-treatment baseline may impact counseling and help contextualize risk for the patient and managing medical team.

ASHA Criteria for Ototoxicity

Before the severity of a decline in hearing can be qualified, it is necessary to determine what constitutes a significant change in hearing. One widely-used set of rules developed specifically for ototoxicity monitoring was established by an ad hoc committee of the American Speech Language and Hearing Association (ASHA 1994). These criteria define the following as a significant change in hearing: 20 dB decline in hearing at any single test frequency, *or* a 10 dB decline at two adjacent frequencies, *or* loss of response at maximum audiometer outputs for three consecutive frequencies where there was previously measurable hearing. Additionally, these changes need to be confirmed on a follow-up test (ASHA 1994). The ASHA guideline stresses the importance of including extended high frequencies in an identification paradigm in order to facilitate early identification of ototoxicity. These criteria are binary and conservative. As such, they are an excellent starting point but they quickly exhaust their utility to quantitatively describe the degree of toxicity, as shown in Figure 2.

Ototoxicity Scales

A number of ototoxicity grading scales were developed to distinguish nominal changes in hearing with a minor functional impact from substantial decline necessitating intervention(s). Of the currently available ototoxicity scales, some require a baseline in order to calculate change and others consider functional hearing status only. Still others were established specifically for pediatric populations. Here, several of the more commonly employed scales are highlighted, although the list is not exhaustive. The intention is to provide a sample of existing scales that vary in their approach to capturing ototoxicity, and highlight some benefits and limitations of each. The reader is referred to Crundwell, Gomersall and Baguley (2016) for a comprehensive and detailed review of 13 ototoxicity classification systems employing pure tone thresholds as the outcome measure. These authors address the strengths and weaknesses of scales that use absolute thresholds as compared to those based on changes from baseline thresholds, intended patient populations, functional significance, and application of the scales in clinical settings.

Common Terminology Criteria for Adverse Events (CTCAE)

The initial CTC included grading of hearing loss in combination with tinnitus into categories based on broad terminology that lacked specificity and objectivity (National Cancer Institute 1982). Subsequent versions, including the most recent CTCAE v4.03 (Table 1) (National Cancer Institute 2010), have benefited from input from audiologists and otologists in establishing criteria for ear-related changes. The resulting grading schema includes criteria for adults enrolled in an ototoxicity monitoring program, adults not enrolled in a monitoring program (i.e., absent baseline examination), and pediatric patients. When a baseline is available, definitions for grade changes consider the amount of pure tone shift, number of involved frequencies (adults) or the lowest frequency at which change was observed (pediatric), and frequency range for testing up to 8000 Hz. The CTCAE does not include provisions for changes in the extended high frequency range. Although progression between grades 1, 2, and 3 are finely distinguished, grade 3 represents a broad range of potential hearing threshold shift, which can limit sensitivity to further functional change (see Figure 3).

In lieu of absolute threshold shift, this scale also considers subjective complaints as well as the need for hearing aids and cochlear implants. While this addresses a functional impact of hearing loss, it adds a subjective element to grading that many not be consistent from clinician to clinician (Gurney and Bass 2012). How is hearing aid candidacy defined? How is the patient who was a hearing aid candidate at baseline treated when they have minimal threshold shifts, equivalent to a grade 1 or 2, that intensify their need for therapeutic intervention? The reader is invited to consider Figure 5 as a pre-treatment audiogram, and how a 15–25 dB decline in hearing might impact this individual.

The CTCAE is a descriptive scale meant to capture AEs associated with the use of a medical treatment or procedure. Its utility when pre-existing disease exists. Patients with pre-existing hearing loss who are enrolled in OMPs are at risk for having a change in hearing go undocumented, as in the case with scales focused on specific high frequency changes, or underappreciated by referring physicians who may not recognize that minimal changes in

the face of pre-existing disease can be a tipping point into a functional deficit. Is the scale being used to objectively quantify cases of ototoxicity related to a given intervention? Or is the purpose to communicate change in function at early and clinically significant stages in order to guide informed therapeutic decision making? For many clinicians and researchers employing these scales, the answer winds up being both. And in such cases, which purpose trumps the other? These questions and their answers create an inherent ambiguity in many of these scales when applied to patients with pre-existing disease. The CTCAE has been developed, seemingly, with both of these caveats in mind, rendering it a flexible but imperfect tool.

Adult Ototoxicity Scales

TUNE Scale: Theunissen and colleagues (2014) designed the TUNE scale for use with adult populations in an effort to develop an ototoxicity grading scale with greater applicability to everyday life, including speech understanding and sound quality (e.g., nature and music appreciation). This scale considers patient complaint, threshold shift, absolute threshold, and thresholds for the extended high frequencies of 8, 10, and 12.5 kHz. Grades 1 and 2 are determined by threshold shifts from a baseline of 10 and 20 dB, respectively, for the pure tone average (PTA) of 8, 10, and 12.5 kHz (1a, 2a) and the PTA of 1, 2, and 4 kHz (1b, 2b). Additionally, to acknowledge the significance of tinnitus or difficulty hearing in the absence of a threshold shift, subjective complaints are assigned grade 1a. In contrast, grades 3 and 4 are assigned based on absolute thresholds for the 1, 2, 4 kHz PTAs of 35 and 70 dB HL, respectively, when these hearing levels occur as a *de novo* finding. The cut-point of 35 dB HL was selected as an indicator of the level at which there would be a 50% loss of speech intelligibility at conversational levels based on the count-the-dots version of the Articulation Index (Mueller and Killion 1990). Consequently, grades 3 and 4 could be useful for providing an indication of when aural rehabilitation may be indicated, whereas grades 1 and 2 are more aligned with early detection of ototoxic changes. It remains unclear how to grade the patient with a pre-existing PTA of 35 dB HL that progresses to 50 dB HL on a post-treatment test. Is this a grade 1b? It meets the change criteria. Is this a grade 3? It is not a *de novo* hearing loss as specified by the grade 3 category, hence, it does not technically meet the stated criteria.

Pediatric Ototoxicity Scales—Development of grading scales specific to the pediatric population was largely motivated by concern for the unique listening needs of children. While an adult may be able to tolerate a mild high frequency hearing loss, this is not the case for children who are actively developing speech, language, and social skills, and expected to function in acoustically challenging classroom settings (Knight, Kraemer, and Neuwelt 2005; Brooks and Knight 2017). Use of a scale that does not take frequency or age into account may underestimate the functional impact of hearing loss on pediatric patients (Knight, Kraemer, and Neuwelt 2005). All three of the pediatric scales described below consider the functional impact of hearing loss, do not require a baseline audiogram, and do not provide guidance for grading ototoxic effects when there is a known pre-existing hearing loss.

Brock Scale: The first and most widely-used pediatric-specific ototoxicity scale was designed by Penelope Brock, a pediatric oncologist, and colleagues (Brock et al. 1991). As the scale was developed, considerations included the practical difficulties in obtaining a full audiogram at all frequencies in a child who may be too ill or fatigued to fully cooperate, and the potential for fluctuant middle ear disease that primarily affects the low frequencies. This scale is based on absolute hearing thresholds, and not change from a baseline. It has four grades, uses 40 dB HL as a boundary level differentiating significant from non-significant changes. It considers the frequencies involved, giving more weight to hearing loss at the mid-frequencies than the high frequencies, such that a 40 dB HL threshold limited to 8 kHz is classified as grade 1, and a 40 dB HL threshold at 2 kHz and above is grade 3 (Table 1).

Chang Scale: Chang and Chinosornvatana (2010) noted the deleterious impact of minimal hearing loss for children and the need for a pediatric scale capable of capturing the functional significance of ototoxicity. They modified the Brock scale to include both 20 and 40 dB HL cut-offs, added the interoctave frequencies of 3 and 6 kHz to achieve greater alignment with clinical interpretations, and included 12 kHz to increase sensitivity for identifying early hearing changes (Table 1). This corresponds with the frequencies at which ototoxic hearing loss most often appears in its initial stages. This scale added sub-grades (1a & 1b, and 2a & 2b) in recognition that a 25 dB hearing loss in the mid-frequencies may be more disadvantageous than a 45 dB hearing loss above 4kHz. Chang stressed the need to measure bone conduction when the tympanogram is abnormal or when there has been a change in hearing to ensure that middle ear dysfunction is not a confounding factor (Chang 2011). While the finer detail of the Chang scale may increase sensitivity, it is complicated to apply and requires additional threshold data that may be difficult to obtain in an ill or uncooperative patient.

Boston SIOP Scale: The SIOP grading system was developed by a working group of international stakeholders with expertise in ototoxicity and was initially presented at the 2010 Congress of the International Society of Pediatric Oncology (SIOP) (Brock et al. 2012). The developers adapted the concepts of previous pediatric scales to achieve a grading system that is simple to understand and apply, sensitive to ototoxic changes with a focus on the high frequencies, and functionally relevant. It takes into account the possibility of fluctuating middle ear disease common in children, and requires bone conduction thresholds when there is abnormal tympanometry or a clinical suspicion of a conductive component to a hearing loss. The scale is based on absolute thresholds and also uses cut-offs of 20 and 40 dB HL with more weight, and higher ototoxicity grades, given to hearing loss in the mid-frequencies than the high-frequencies (Table 1). It is designed to be applied at the end of a treatment trial for the purpose of identifying and comparing incidence and severity of hearing loss across clinical trials.

Sensitivity, Reliability, and Validity of Ototoxicity Grading Scales

The success and utility of any ototoxicity grading scale depends on the scale's sensitivity, validity, and reliability. The ASHA definition of ototoxicity is inherently designed to capture small changes in hearing that just exceed clinically-accepted test-retest variability (5–10 dB). Scales that include subjective complaints and extended high frequency thresholds are

more likely to result in a classification of ototoxicity than those that consider standard frequency pure tone thresholds only. Conversely, scales that use absolute thresholds of 40 dB HL as the cut-point for ototoxicity identification will identify fewer cases as having ototoxic hearing loss than a scale that uses 20 dB HL as the defining level.

Sensitivity of pediatric grading scales in detecting *any* ototoxicity was initially addressed by Knight and colleagues (2005) in a comparison of the ASHA definition with the CTCAE v3 and Brock scales in a group of children treated with cisplatin. ASHA and CTCAE v3 had similar sensitivity to any hearing loss (both 61%), while the Brock scale was less sensitive (40%). Subsequently, Landier et al. (2014) observed a similar prevalence of *any* hearing loss detection across ASHA, Brock, CTCAE v3, and Chang scales in a group of 333 children and young adults with a neuroblastoma after treatment with cisplatin only or a combination of cisplatin and carboplatin following one (64–71%) or two (86–90%) exposures. In another pediatric cohort of 37 children with medulloblastoma who were treated with craniospinal radiation and cisplatin, the SIOP scale was more sensitive than the Chang scale to *any* change in hearing, identifying 74 and 66%, respectively (Bass et al. 2014). More recently, Knight and colleagues (2017) compared the ASHA, Brock, CTCAE v3, and SIOP scales in a large, multinational cohort of 284 children and young adults treated for the first time with a cisplatin-containing regimen. Sensitivity in detecting *any* ototoxicity was comparable for SIOP (55%), ASHA (56%), and CTCAE v3 (51%), while it was slightly lower for Brock (40%).

In a comparison between outcomes of four ototoxicity scales in 319 adult patients treated with chemo-radiation or radiation therapy alone for head and neck cancer, the prevalence of ototoxicity was rank ordered, lowest to highest, as CTCAE v4.0, ASHA up to 8 kHz, TUNE, and ASHA up to 12.5 kHz (Theunissen et al. 2014). As expected, scales that included high frequency testing above 8000 Hz (TUNE and ASHA up to 12.5 kHz) were the most sensitive to identification of ototoxicity.

To evaluate the validity of a grading scale, it is necessary to consider the sensitivity of the scale to a functional significant hearing loss. When considering only a clinically significant hearing loss of grade 3 or worse in a group of children, the CTCAE v3 was more sensitive (25%) than the Brock scale (19%) (Knight, Kraemer, and Neuwelt, 2005). In another pediatric cohort, the SIOP and CTCAE v3 were comparable in their rates of assigning ototoxicity grade 3 and above (22% and 18%, respectively) whereas the rate for the Brock scale was 8% (Knight et al. 2017). In a group of children exposed to cisplatin whose hearing loss warranted hearing aid referral, the Brock scale graded only 49% as severe, whereas the Chang and CTCAE v3 graded 91% and 100%, respectively, in the severe category (Landier et al. 2014). The Chang scale was more specific in identifying and differentiating among those children whom audiologists referred for hearing aid evaluation and FM systems than the Brock and CTCAE v3 scales (Chang 2011), whereas the SIOP and Chang scales were equally sensitive (35%) in identifying those with hearing loss sufficient to warrant hearing aid use (Bass et al. 2014).

While a sensitive scale is desirable, this must be balanced against the need for specificity to avoid false positive test results. Theunissen and colleagues (2014) defined false positives as

a higher ototoxicity grade at the time of the last treatment as compared to follow up testing several weeks after completing treatment, which ranged from 12% for the TUNE scale, 11% for CTCAE, 3% for ASHA up to 12.5 kHz and 0% for ASHA up to 8 kHz in a group of adult patients. Similarly, in a pediatric group, false positive findings defined as identification of ototoxicity at one time point during the course of monitoring followed by no ototoxicity on a subsequent evaluation, occurred at rates of 7.4% for ASHA, 6.7% for SIOP, 4.6% for CTCAE v3, and 2.1% for Brock (Knight et al. 2017). This highlights the need for a confirmatory test following first detection of ototoxic changes.

Multi-institutional clinical trials depend on consistent interpretation of data across settings and providers. Knight et al. (2017) compared inter-rater reliability in a large clinical trial between examining audiologists at test sites and two centrally located audiologists. Agreement between the examining and centrally located audiologist in detecting any ototoxicity ranged from 91% for the Brock scale to 87% for CTCAE v3, and 84% for ASHA criteria. When identification of ototoxicity severity was compared, agreement between reviewers was 85% for the Brock scale as compared to 69% for CTCAE v3 (Knight et al. 2017).

Other pitfalls encountered in ototoxicity monitoring in multi-institutional clinical trials may lead to variability in the quality and completeness of data submitted to a central reviewing agency (Landier et al. 2014). These include failure to obtain a baseline audiogram when the scale requires one, and missing data for scales requiring specific frequencies. These audiograms are considered “unevaluable” and do not effectively contribute to establishing a safety profile. Notably, these pitfalls are not unique to the clinical trial setting, and are also common barriers to meaningful monitoring in a clinical setting. It is necessary to engage frontline clinical care providers, and to do so early on, to ensure timely and accurate collection of necessary data in both clinical trial and clinical care settings.

Selecting or Developing an Ototoxicity Scale

In selecting or developing a grading scale for a particular population or application, several factors should be considered *a priori*, for an individual, patient population, or in the development of a clinical trial.

- Is the scale sensitive to the predicted ototoxic hearing loss? The majority of ototoxic agents cause hearing changes in the high frequencies, and these changes may appear first in the extended high frequency range above 8000 Hz. Scales that include extended high frequencies, or allow for specific weighting or focus on the high frequencies may be more sensitive to early indications of ototoxicity. Conversely, if the change does not follow a typical pattern for high frequency loss, as depicted in figure 4, will the scale be effective?
- Are grading criteria clear or is there ambiguity in the definition, and how might that impact therapeutic decision making? For example, describing a hearing loss sufficient to “indicate amplification” is open to interpretation, and will change as amplification technology evolves. If grading criteria are not clearly defined, there is opportunity for inconsistent application and poor inter-rater reliability.

Moreover, clinical trials are often developed with stopping criteria, for either an individual or the trial, if an AE becomes too serious or occurs too frequently. Is the protocol written in such a way that a patient's continued participation is contingent on an ambiguous definition of toxicity? In the case of life-threatening disease, would this decision impact access to a potentially effective intervention?

- Is it preferable for the scale to specify change from the pre-exposure baseline, or to emphasize absolute threshold and functional status? Scales based on change from baseline require a pre-treatment hearing test and may not, in and of themselves, address the functional needs of the patient. Scales based on absolute hearing thresholds do not differentiate change in hearing from pre-existing hearing loss. Rather, they focus on the functional status of the patient at any given time, ignoring the amount of treatment-related hearing change. The reader is directed to figure 5 for an illustration of this scenario.
- How does pre-existing hearing loss impact use of the scale? Scales that confine the change in hearing to specific frequencies may be less useful in a population with pre-existing hearing loss. This may vary by protocol depending on whether the goals of monitoring are focused on identifying functional needs versus quantifying toxicity.
- Are the guidelines intended for adults, children, or both? Should pediatric scales be sub-divided into those applying to children who are still developing speech and language skills and those aimed toward older post-lingual children (Chang 2011)? The impact of minimal hearing loss on a child who has emerging speech and language and who functions and learns in acoustically challenging environments such as a classroom is greater than the impact of a minimal hearing loss on an adult (Littman, Magruder, and Strother 1998; Brooks and Knight 2017).
- Does the scale include provisions for grading when there is incomplete or suprathreshold data? For example, a pediatric or very ill adult patient may not provide a full audiogram, give true threshold responses, or tolerate earphones necessitating reliance on minimum response levels obtained during sound field testing at just two frequencies (Brooks and Knight 2017). How can limited information be incorporated into a grading scale, or drive test strategy to ensure that the most important data are collected first?
- Should there be guidelines for grading ototoxicity based on otoacoustic emissions or auditory brainstem response (ABR) derived thresholds? While pure tone thresholds are the current gold standard for ototoxicity monitoring, it may not be possible to obtain these data at each visit due to health status or other factors affecting ability to cooperate. Is it legitimate to substitute ABR thresholds for behavioral thresholds in grading ototoxicity? The authors developed an ABR-derived AE scale (Table 2), modeled after the CTCAE, meant to capture and segregate minimal change in hearing from functionally significant change, for use in populations who require ABR threshold assessment in whom the use of AE monitoring is necessary. This scale has yet to be validated. To date, it has

been used it to monitor hearing safety across multiple phases in a clinical trial. Otoacoustic emissions (OAEs) may prove more challenging in that they do not estimate threshold, but they do afford an opportunity to document high frequency change and may be more sensitive to early identification of ototoxicity (Littman, Magruder and Strother 1998; Brooks and Knight 2017). Absent OAEs in the setting of transient or permanent changes in middle ear function further complicates their consistent application and contribution as a monitoring tool.

- How should conductive hearing loss be factored into ototoxicity grading? On the one hand, middle ear effusion unrelated to treatment may cause conductive hearing loss and a higher ototoxicity grade. Should bone conduction thresholds supplant air conduction thresholds in this case? On the other hand, cranial radiation in conjunction with cisplatin (chemo-radiation) is a common therapeutic regimen for some cancers. The effects of radiation on hearing is varied, and there may be a resultant conductive or mixed hearing loss (Gurney and Bass 2012). In this scenario, should the conductive component be ignored or factored into grading?
- Above and beyond use in clinical trials, how should ototoxicity grading drive clinical decision-making? While clinical trials standardly identify stopping criteria based on toxicity scales, the same is not true in routine clinical practice. Application of these scales in a clinical setting may be a useful way to communicate hearing changes and assist the conversation regarding the significance of the hearing loss to the patient and managing physician. Ultimately, the decision to continue or change treatment is based on a number of factors, including available treatment options and overall patient health status. Nonetheless, the use of ototoxicity scales can make hearing data more accessible and facilitate therapeutic decision making.
- Do studies of putative otoprotectants need stricter monitoring criteria than those provided by grading scales? Keeping in mind that some ototoxicity scales have grades that span wide ranges, have a subjective element, reduce data to ordinal numbers, and do not include the extended high frequencies, it is possible that grading systems may miss or obscure effects of otoprotectants. In this case, more finely tuned analysis (e.g. high frequency pure tone thresholds or pure tone averages) will better capture protective effects, and ototoxicity grading scales may be used for supplemental analyses.

Conclusions

Audiologic monitoring for ototoxicity is not a routine session in which hearing thresholds are established and reported in isolation; it is a purpose-driven consultation with multiple goals and stakeholders. These goals include early identification of hearing changes, communication with the patient and family, prevention or mitigation of functional hearing loss, and establishing and monitoring of drug safety and efficacy.

Clinical trials are the vehicle by which we translate basic science into human applications in order to improve health and reduce disease. They inform clinical practice on the front lines of medicine, in part, by establishing the balance of toxicity and benefit for new therapeutic interventions. The severity of disease and the availability of alternative therapies drive how we tolerate the exchange of safety for efficacy. Grading scales serve a critical need in the successful fulfillment of a clinical trial as a tool to uniformly monitor AEs. There are, however, distinct advantages to more widespread use of standardized grading scales beyond their application in clinical trials. These include consistency in the interpretation of data and greater simplicity of the metric relative to the entire audiogram. Each of the scales, perhaps inherently, offers benefits and limitations, which can vary by population and setting. Ultimately, grading scales applied in isolation do not carry sufficient meaning about progression, clinical impact, or clear candidacy for re/habilitation.

Hearing data are complicated: they involve a wide range of test frequencies, multiple transducers and techniques, and stem from a bilateral, heterogeneous sensory system. Furthermore, the seasoned clinician is well aware that identical audiograms from two patients can impact individual lives in widely different ways. This means that capturing and contextualizing risk and toxicity for an individual or a cohort is challenging, which may be the reason that a uniform system to convey this information has remained elusive. Moreover, while the current emphasis in defining toxicity relies almost exclusively on pure-tone hearing thresholds, additional effects on hearing, such as speech in noise, remain largely unexamined and, potentially, overlooked.

Audiology falls at the intersection of scientific evidence and clinical circumstance in the process of therapeutic decision making. Audiologists are uniquely suited to inform patients as they establish their own preferences to guide these decisions. The audiologist's role before, during, and after ototoxic intervention is dynamic and important.

Acknowledgements:

Funding for this work was supported by the Intramural Research Program of the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Department of Health and Human Services (NIH intramural grant DC000064 to CCB). The authors are grateful to Marilyn Dille, Dawn Konrad Martin, Katharine Fernandez, and Nicole Schmitt for their careful review and feedback.

Acronyms and Abbreviations:

AE	Adverse Event
ADL	Activities of Daily Living
ASHA	American Speech Language Hearing Association
ABR	Auditory Brain stem Response
FDA	Food and Drug Administration
IND	Investigational New Drug
IRB	Institutional Review Board

MeDRA	Medical Dictionary for Regulatory Activities
NCICTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
OAEs	Otoacoustic Emissions
OMP	Ototoxicity Monitoring Program
U.S.	United States

REFERENCES

- American Speech-Language-Hearing Association (ASHA). 1994 “Guidelines for the audiologic management of individuals receiving cochleotoxic drug therapy.” *ASHA* 36: 1–19. [PubMed: 7993382]
- Bass JK, Huang J, Onar-Thomas A, Chang KW, Bhaga SP, Chintagumpala M, Bartels U, et al. 2014 “Concordance between the Chang and the International Society of Pediatric Oncology (SIOP) ototoxicity grading scales in patients treated with cisplatin for medulloblastoma.” *Pediatric Blood and Cancer* 61: 601–605. [PubMed: 24504791]
- Brock PR, C Bellman S, Yeomans EC, Pinkerton CR, and Pritchard J, J. 1991 “Cisplatin ototoxicity in children: a practical grading system.” *Medical and Pediatric Oncology* 19: 295–300. [PubMed: 2056973]
- Brock PR, Knight KR, Freyer DR, Campbell KC, Steyger PS, Blakley BW, Rassekh SR, et al. 2012 “Platinum-induced ototoxicity in children: a consensus review on mechanisms, predisposition, and protection, including a new International Society of Pediatric Oncology Boston ototoxicity scale.” *Journal of Clinical Oncology* 30: 2408–2417. [PubMed: 22547603]
- Brooks B, and Knight K 2017 “Ototoxicity monitoring in children treated with platinum chemotherapy.” *International Journal of Audiology*. doi:10.1080/14992027.2017.1355570
- Chang KW 2011 “Clinically accurate assessment and grading of ototoxicity.” *Laryngoscope* 121: 2649–2657. [PubMed: 22109766]
- Chang KW, and Chinosornvatana N 2010 “Practical grading system for evaluating cisplatin ototoxicity in children.” *Journal of Clinical Oncology* 28: 1788–1795. [PubMed: 20194861]
- Crundwell G, Gomersal P, and Baguley DM 2016 “Ototoxicity (cochleotoxicity) classifications: a review.” *International Journal of Audiology* 55: 65–74. [PubMed: 26618898]
- Food and Drug Administration (FDA). 2017 The Drug Development Process: Clinical Research. Accessed 1 July 2017. <https://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm>
- Garinis A, Kempf A, Tharpe AM, Weitkamp J-H, McEvoy C, and Steyger P 2017 “Monitoring neonates for ototoxicity.” *International Journal of Audiology* doi: 10.1080/14992027.2017.1339130.
- Gurney JG, and Bass JK 2012 “New International Society of Pediatric Oncology Boston Ototoxicity Grading Scale for pediatric oncology: still room for improvement.” *Journal of Clinical Oncology* 30: 2303–2306. [PubMed: 22547588]
- ISO. 2000. ISO 7029–1, Acoustics--statistical distribution of hearing thresholds as a function of age. Geneva: International Organization of Standardization.
- Knight KR, Chen L, Freyer D, Aplenc R, Bancroft M, Bliss B, Dang H, et al. 2017 “Group-wide, prospective study of ototoxicity assessment in children receiving cisplatin chemotherapy (ACCL05C1): a report from the Children’s Oncology Group.” *Journal of Clinical Oncology* 35: 440–445. [PubMed: 27937095]
- Knight KR, Kraemer DF, and Neuwelt EA 2005 “Ototoxicity in children receiving platinum chemotherapy: underestimating a commonly occurring toxicity that may influence academic and social development.” *Journal of Clinical Oncology* 23: 8588–8596. [PubMed: 16314621]
- Konrad-Martin D, Poling G, Garinis A, Ortiz C, Hopper J, O’Connell-Bennett K, and Dille M 2017 “Applying U.S. national guidelines for ototoxicity monitoring in adult patients: perspectives on

patient populations, service gaps, barriers and solutions.” *International Journal of Audiology*. 1–16, doi: 10.1080/14992027.2017.1398421.

- Konrad-Martin D, Reavis KM, McMillan G, Helt WJ, and Dille M 2014 “Proposed comprehensive ototoxicity monitoring program for VA healthcare (COMP-VA).” *Journal of Rehabilitation, Research and Development* 51: 81–100. [PubMed: 24805896]
- Landier W, Knight K, Wong FL, Lee J, Thomas O, Kim H, Kreissman SG, et al. 2014 “Ototoxicity in children with high-risk neuroblastoma: prevalence, risk factors, and concordance of grading scales--a report from the Children’s Oncology Group.” *Journal of Clinical Oncology* 32: 527–534. [PubMed: 24419114]
- Littman TA, Magruder A, and Strother DR 1998 “Monitoring and predicting ototoxic damage using distortion-product otoacoustic emissions: pediatric case study.” *Journal of the American Academy of Audiology* 9: 257–262. [PubMed: 9733234]
- Mueller H, and Killion MC 1990 “An easy method for calculating the articulation index.” *Hearing Journal* 43: 1–4.
- National Cancer Institute (NCI). 2010 “Common Terminology Criteria for Adverse Events (CTCAE, v 4.03.)” NCI, National Institutes of Health, Department of Health and Human Services Accessed 21 August 2017. https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-4_QuickReference_8.5x11.pdf
- National Cancer Institute (NCI). 1982, “Common Toxicity Criteria.” Accessed 21 August 2017 https://www.ucdmc.ucdavis.edu/clinicaltrials/StudyTools/Documents/NCI_Toxicity_Table.pdf
- Neuwelt E, Brock P 2010 “Critical need for international consensus on ototoxicity assessment criteria.” *Journal of Clinical Oncology* 28: 1630–1632. [PubMed: 20194840]
- Schmitt NC, and Page BR 2017 “Chemoradiation-induced hearing loss remains a major concern for head and neck cancer patients.” *International Journal of Audiology*. doi: 10.1080/14992027.2017.1353710.
- Theunissen EA, Dreschler WA, Latenstein MN, Rasch CR, van der Baan S, de Boer JP, Balm AJ, et al. 2014 “A new grading system for ototoxicity in adults.” *Annals of Otology, Rhinology and Laryngology* 123: 711–718.

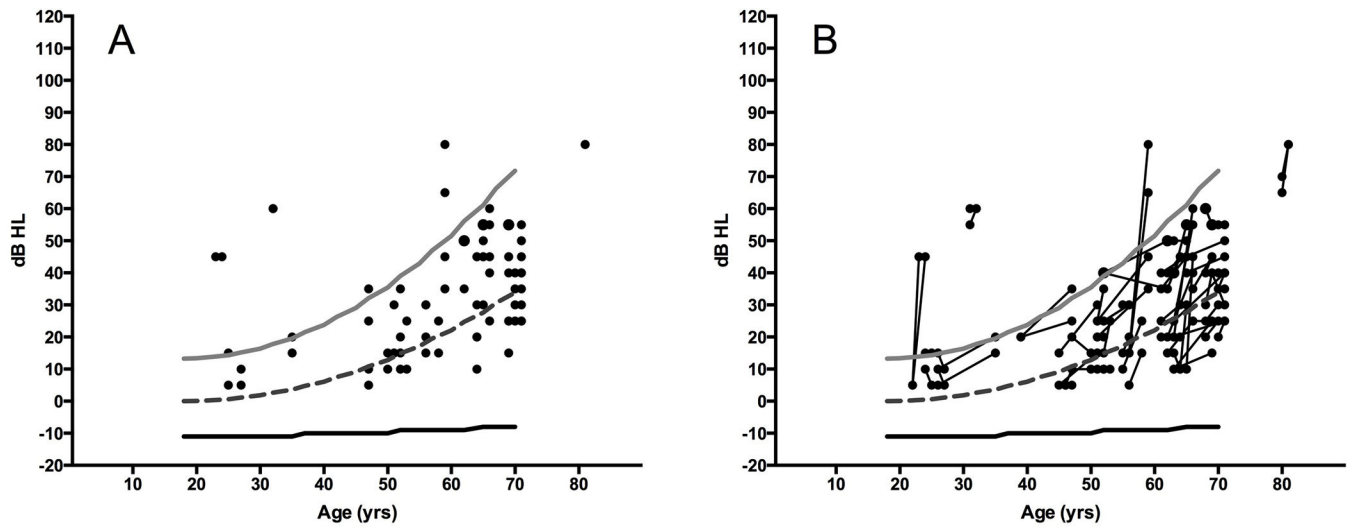


Figure 1.

Hearing sensitivity in females being treated with the aminoglycoside, amikacin, most commonly for mycobacterium infection or cystic fibrosis. Circles represent ear-specific thresholds at 4 kHz. Lines represent sex and age-matched normative data (ISO, 2000); light grey is the 95th percentile, dashed dark grey is the 50th percentile, and black is the 5th percentile. Left panel (A) thresholds obtained at the end of audiometric monitoring reveal that 10% of ears fall outside the normative range of hearing. However, when change in hearing over time is considered, right panel (B), over twice as many ears showed change (>10 dB) in hearing. Over half of these cases would not have been identified as having ototoxic change if normative ranges alone were used.

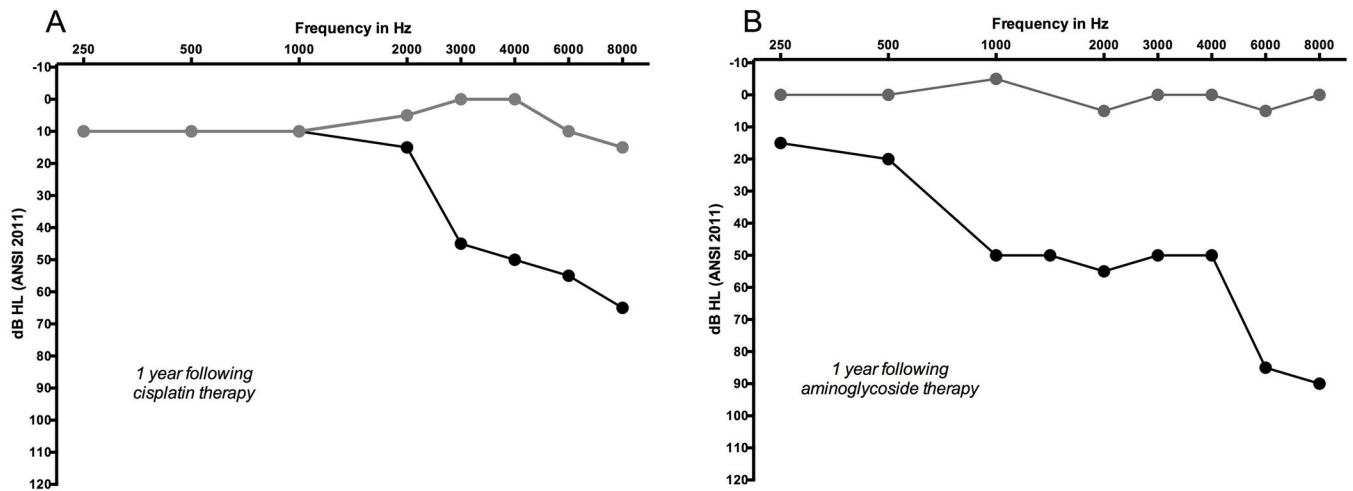


Figure 2.

Two case examples of decline in hearing sensitivity from ototoxicity. Panel A shows decline in hearing one year after cisplatin chemotherapy, and panel B shows decline one year after exposure to the aminoglycoside, amikacin. Baseline pre-exposure hearing levels are represented by grey circles and black circles represent thresholds following therapy. The amount of change and range of frequencies affected is notably different between the two cases, and yet ASHA criteria for ototoxicity treats both cases the same; affirming, yes, ototoxicity occurred but making no other distinction. In both cases the change in hearing was sensorineural (bone conduction data not plotted) and bilateral, however, only a single ear from each patient is shown.

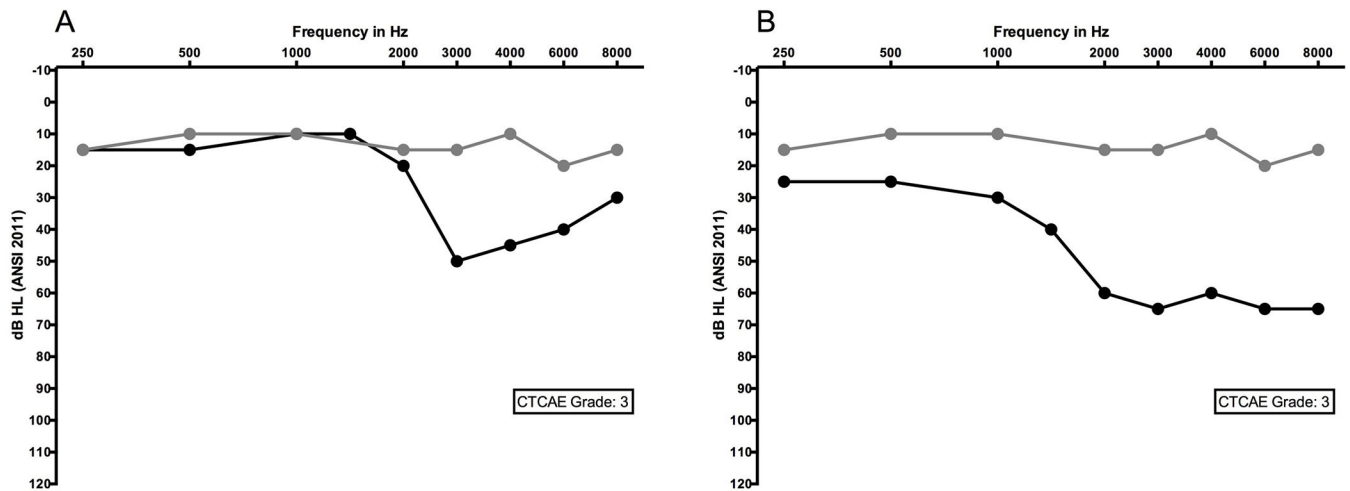


Figure 3.

Two audiograms documenting ototoxic change in the same individual. Panel A shows an early and clinically significant change from an ototoxic agent; Panel B shows later change in hearing in the same person after continued exposure. Both audiograms meet criteria for a CTCAE version 4.03 grade 3, despite the fact that one (B) represents significantly more change in hearing and a predicted increase in functional severity with the inclusion of 2 kHz compared to the other (A). Baseline pre-exposure hearing levels are represented by grey circles and black circles represent thresholds during the course of therapy. In both examples, the change in hearing was sensorineural (bone conduction data not plotted) and bilateral, although only a single ear is shown.

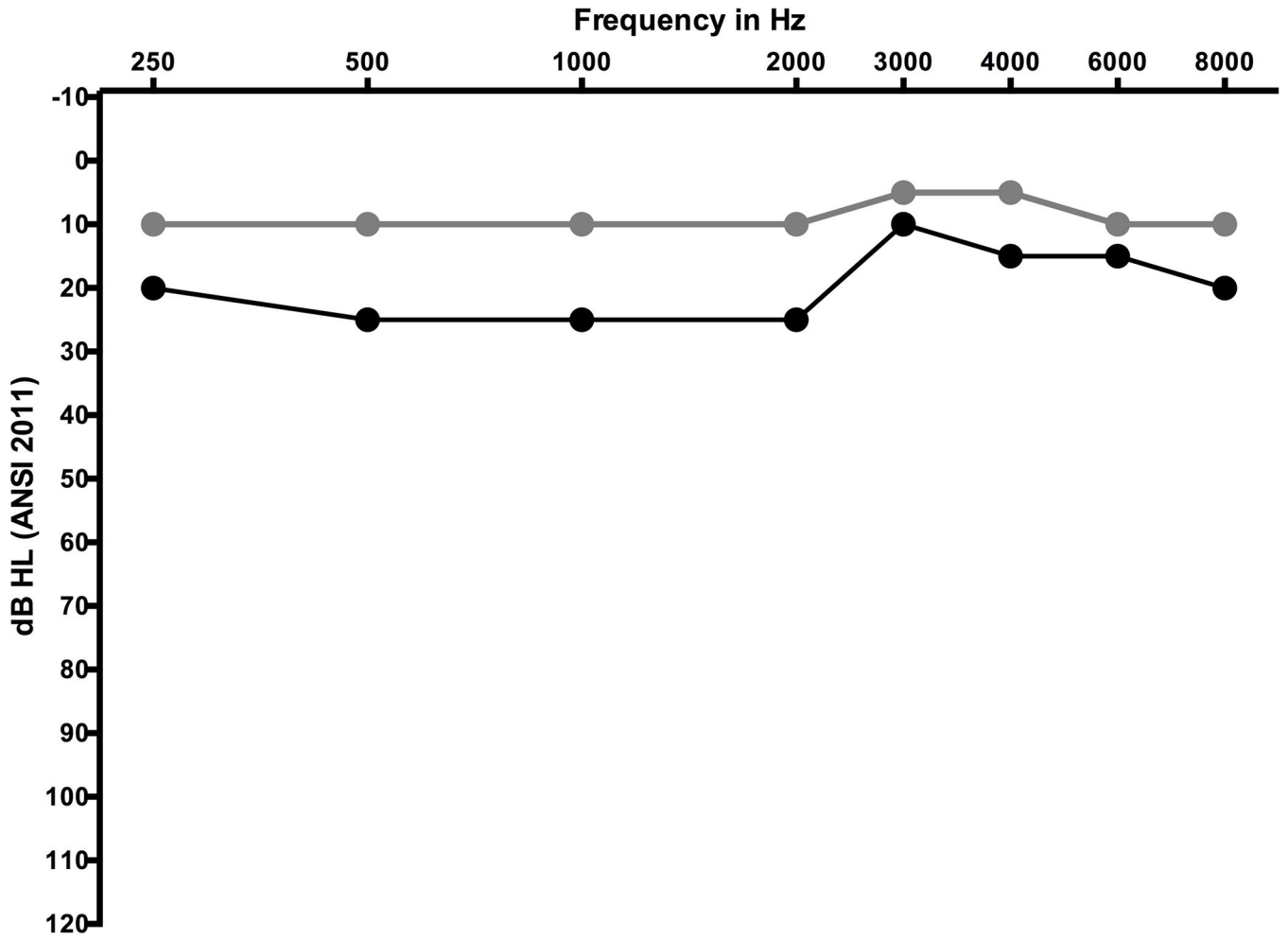


Figure 4. Baseline audiogram representing air conduction hearing thresholds from one ear of an adolescent female prior to exposure to a potential ototoxic medication. Both the SIOP and Brock scales do not account for pre-existing hearing loss; this audiogram would be graded a 3 on both scales prior to any ototoxic exposure.

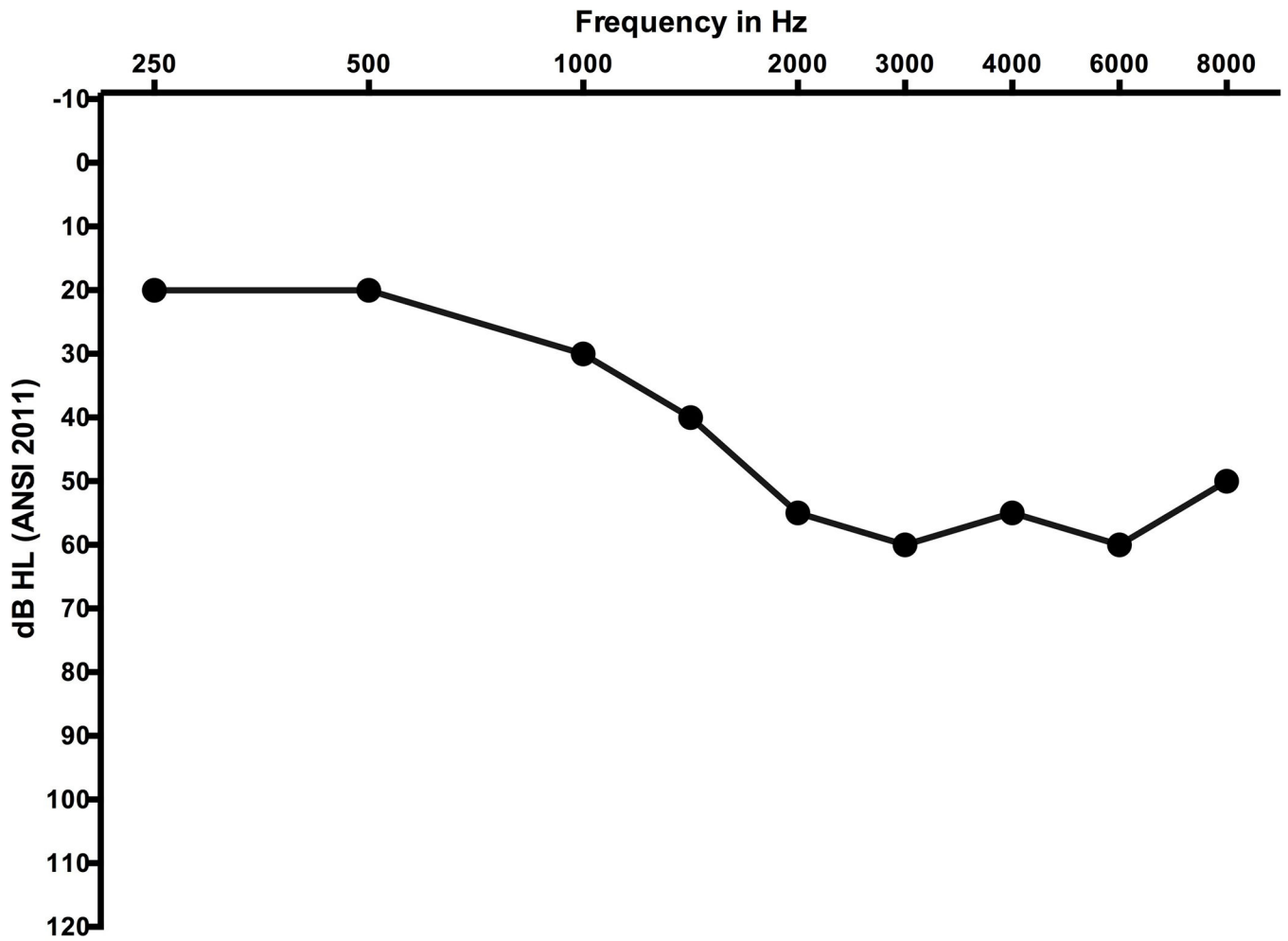


Figure 5.

Baseline (grey circles) and follow up (black circles) audiogram from an adolescent female undergoing high dose therapy with the loop diuretic, furosemide (Lasix). Ototoxic grading scales that emphasise high-frequency change in hearing (e.g. CTCAE version 4.03 paediatric version) would not be sensitive to capturing this significant decline that occurred early in the course of treatment. The change in hearing was sensorineural (bone conduction data not shown) and bilateral, although data from only a single ear is shown.

Table 1.

Otototoxicity Classifications and Grading Scales

Scale	Otototoxicity Classification Parameters
Adult and Pediatric Application	
ASHA	Based on change from baseline. Defines ototoxic change as binary (yes/ no) based on threshold changes from baseline; 10 dB change from baseline at 2 consecutive frequencies, or, 20 dB change at 1-frequency, or loss of response where one was previously obtained.
CTCAE V4.03	<p>Adult enrolled in a monitoring program. Grade 1: 15–20 dB change at avg of 2-contiguous frequencies in at least one ear; Grade 2: >25 dB change at avg of 2 contiguous frequencies in at least one ear; Grade 3: >25 dB change at avg of 3 - contiguous frequencies or therapeutic intervention indicated in at least one ear; Grade 4: bilateral decrease in hearing to >80 dB HL at 2 kHz & above; non-serviceable hearing.</p> <p>Adult not enrolled in a monitoring program. Grade 1: subjective change in hearing in absence of documented hearing loss; Grade 2: hearing loss but hearing aid or intervention not indicated; Grade 3: hearing loss with hearing aid or intervention indicated; Grade 4: bilateral decrease in hearing to >80 dB HL at 2 kHz & above; non-serviceable hearing.</p> <p>Pediatric. Grade 1: threshold shift >20 dB at 8 kHz in at least one ear; Grade 2: threshold shift >20 dB at 4 kHz & above in at least one ear; Grade 3: hearing loss sufficient to indicate therapeutic intervention, including hearing aids, or threshold shift >20 dB at 3 kHz & above in at least one ear, or additional speech-language services indicated; Grade 4: Audiologic indication for cochlear implant and addition speech-language services indicated.</p>
Adult Application	
TUNE	Grade 0: no hearing loss; Grade 1a: threshold shift 10 dB at 8–10-12.5 kHz avg or subjective complaints in absence of threshold shift; Grade 1b: 10 dB threshold shift at 1–2-4 kHz avg; Grade 2a: threshold shift 20 dB at 8–10-12.5 kHz avg; Grade 2b: threshold shift 20 dB at 1–2-4 kHz avg; Grade 3: threshold 35 dB HL at 1–2-4 kHz avg <i>de novo</i> ; Grade 4: threshold 70 dB HL at 1–2-4 <i>de novo</i> . Apply to each ear.
Pediatric Application	
Brock	All grades based on absolute threshold. Grade 0: <40 dB HL at all test frequencies; Grade 1: 40 dB HL at 8 kHz only; Grade 2: 40 dB HL at 4 kHz & above; Grade 3: 40 dB HL at 2 kHz & above; Grade 4: 40 dB HL at 1 kHz & above.
Chang	All grades based on absolute threshold. Grade 0: 20 dB HL at 1, 2 & 4 kHz; Grade 1a: 40 dB HL at any freq 6–12 kHz; Grade 1b: >20 & <40 dB HL at 4 kHz; Grade 2a: 40 dB HL 4 kHz & above; Grade 2b: >20 & <40 dB HL at any freq below 4 kHz; Grade 3: 40 dB HL at 2 or 3 kHz & above; Grade 4: 40 dB HL at 1 kHz & above.
SIOP	All grades based on absolute threshold, specifies SNHL. Grade 0: 20 dB HL at all freq; Grade 1: >20 dB HL above 4 kHz; Grade 2: >20 dB HL at 4 kHz & above; Grade 3: >20 dB HL at 2 or 3 kHz & above; Grade 4: >40 dB HL at 2 kHz & above.

Avg=average; freq=frequency; SNHL=sensorineural hearing loss.

Table 2.

Proposed ABR-derived adverse event schema. This application requires data collected using air-conducted tone burst stimuli from 0.5 to 4 kHz and assumes normal middle ear function.

Adverse Event Grade	ABR Findings (.5–4 kHz assessment)
1	Threshold shift >10 to 20 dB at 4 kHz in at least one ear
2	Threshold shift >20 dB at 4 kHz in at least one ear
3	Threshold shift >20 dB at 2 kHz in at least one ear
4	Absolute thresholds >80 dB eHL from 1–4 kHz in both ears (not previously present at baseline)

eHL: estimated hearing level