A New Standardized Format for Reporting Hearing Outcome in Clinical Trials

Richard K. Gurgel, MD1, Robert K. Jackler, MD1, Robert A. Dobie, MD2,3, and Gerald R. Popelka, PhD1

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Abstract
The lack of an adequate standardized method for reporting level of hearing function in clinical trials has hampered the ability of investigators to draw comparisons across studies. Variability in data reported and presentation format inhibits meta-analysis and makes it impossible to accumulate the large patient cohorts needed for statistically significant inference. Recognizing its importance to the field and after a widely inclusive discussion, the Hearing Committee of the American Academy of Otolaryngology–Head and Neck Surgery endorsed a new minimal standard for reporting hearing results in clinical trials, consisting of a scattergram relating average pure-tone threshold to word recognition score. Investigators remain free to publish their hearing data in any format they believe is interesting and informative, as long as they include the minimal data set to facilitate inter-study comparability.

Keywords
hearing loss, reporting standard, hearing outcome, hearing results, audiogram, review, meta-analysis

Received February 15, 2012; revised June 28, 2012; accepted July 26, 2012.

Rationale
Uniformity in reporting hearing outcomes in clinical trials is critical for an accurate appraisal of the medical literature related to hearing. Without standardized guidelines for reporting hearing outcomes, data comparison between studies is severely compromised, and pooling of data in meta-analysis is not possible.1 In 1995, the Committee on Hearing and Equilibrium of the American Academy of Otolaryngology–Head and Neck Surgery (AAO-HNS) published guidelines on minimal reporting standards for evaluating hearing outcomes in acoustic neuromas, Meniere’s disease, and conductive hearing loss.2–4 None of these guidelines, however, were intended to represent a comprehensive scale of hearing outcomes.

Although the 1995 hearing standard significantly improved the quality of data reported for hearing outcomes, certain limitations remain. Any classification system that groups a heterogeneous population of patients into fixed categories will lose some data resolution in describing hearing outcomes. For example, according to the 1995 AAO-HNS system, a patient treated for an acoustic neuroma with a 51% word recognition score could be placed in the same class as a patient with a 100% word recognition score if both had pure-tone averages (PTAs) between 30 and 50 dB hearing level (HL) (both class B). The functional difference between these 2 outcomes, however, can be dramatic. The arbitrary boundaries that separate different classes of hearing are not based on validated hearing outcome data and are not sufficiently granular to describe the diversity of hearing outcomes encountered clinically.

To more accurately represent a wider breadth of hearing outcomes, a new hearing outcomes scale for clinical trials has been created. This comprehensive scale is intended for use in all forms of hearing loss—sensory, neural, conductive, and mixed.

Deliberating a New Minimal Standard for Reporting Hearing Loss
Proposals for a new reporting standard were debated and refined over a 2-year period of discussion that included consideration of extensive literature sources on measurement of hearing disability. The design chosen was shaped by input from the diverse membership of the AAO-HNS Hearing Committee.
Committee with substantial guidance from leading otological and audiological organizations, including the American Otological Society, American Neurotology Society, European Academy of Otology & Neurotology, American Academy of Audiology (AAA), and the Editorial Board of the journal *Otology & Neurotology*.

**Design Considerations for the Minimal Standard for Reporting Hearing Loss**

For all forms of hearing loss, hearing outcomes in clinical trials should include at a minimum the average air conduction pure-tone threshold (in dB HL) and a word recognition score (WRS, in %). The data should be reported on a scattergram relating average air conduction pure-tone threshold to the WRS (Figure 1).

The average pure-tone threshold data are acquired in a fairly well-standardized manner. The reported pure-tone average (PTA), calculated using 0.5-, 1-, 2-, and 3-kHz air conduction thresholds and rounded to the nearest whole number, is plotted on the y-axis of the scattergram in increasing 10-dB intervals from 0 to >91 dB from top to bottom. Hearing thresholds at additional frequencies can be included at the investigator’s discretion but are not mandatory. When the threshold at 3 kHz is missing, it is valid to interpolate a 3-kHz threshold by averaging the thresholds at 2 and 4 kHz.5-7

The Hearing Committee workgroup recognizes the large variability in how WRS data are acquired. Ideally, the WRS should be measured using a validated recording in the patient’s native language, a minimum 50-word list, and a standardized presentation level such as 40-dB sensation level or maximum comfortable loudness level. Although this ideal may be achievable in prospective clinical trials within a single institution, much variability in practice exists. At present, a degree of data heterogeneity is tolerable compared with not reporting WRS scores, but efforts are under way to work with audiological organizations to achieve greater standardization of how speech audiometry is performed. The WRS is plotted on the x-axis in 10% intervals in descending order from left to right. The WRS at presentation levels of up to 40-dB sensation level or maximum comfortable loudness, whichever is less, should be recorded before and after treatment. The WRS can be determined by live or recorded voice and should be given in the patient’s native language. A recording of a 50-word list is preferred and recommended. Authors should describe the methods by which the WRS is determined. It is also acknowledged that standard audiometry practice in many countries does not include measurement of the WRS. In studies coming from countries in which the WRS is not routinely measured, only the PTA is reported.

For any study reporting an intervention that affects hearing, both a scattergram of the baseline hearing characteristics and a posttreatment scattergram showing change in hearing should be reported (Table 1). Interventions include any pharmaceutical, surgical, or, in the future, genetic or molecular therapy that affects hearing. If multiple, sequential interventions take place in the same patient population, a posttreatment scattergram should be included for each intervention. The posttreatment scattergram should be reported with percentage change in the WRS plotted on the x-axis in ±10% intervals with 0% (no change) at the midpoint. Decibel change in PTA is plotted on the y-axis in ±10-dB intervals with 0 dB (no change) at the midpoint. The changes in the WRS represent absolute changes in the WRS score, not relative changes. For example, a patient who has a pretreatment WRS of 80% and a posttreatment WRS of 40% would be reported in the 40% class.

![Figure 1. Scattergram of pretreatment hearing results in a hypothetical sample of patients. Pure-tone averages are represented on the y-axis and word recognition scores are represented on the x-axis. Each number represents the number of patients whose audiometric data place them into a certain square.](image-url)

Table 1. Key Features of the New Hearing Reporting Standard

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Represents the first comprehensive minimal reporting standard for clinical trials in all forms of hearing loss</td>
<td></td>
</tr>
<tr>
<td>Authors are encouraged to upload their full hearing result database as an online supplement to journal publication</td>
<td></td>
</tr>
<tr>
<td>Conductive hearing loss standard (1995) remains useful as a means for describing air-bone gap</td>
<td></td>
</tr>
<tr>
<td>New standard is not intended for occupational disability or forensic use</td>
<td></td>
</tr>
<tr>
<td>Reporting cochlear implant outcomes requires additional specialized testing</td>
<td></td>
</tr>
</tbody>
</table>
change category on the plot, even though the patient has had a 50% relative change. The columns and rows for WRS and PTA, respectively, are inclusive of the 9 integers less than the label. For example, column or row “10” includes the numbers 1-10, “20” includes 11-20, etc. Figures 2 and 3 show examples of posttreatment scattergrams for a group of hypothetical patients treated for acoustic neuromas and otosclerosis, respectively. Pre- and posttreatment scattergrams should be reported at identified time intervals to be determined by individual researchers.

If authors are reporting on changes in conductive hearing loss, the previous Hearing Committee guidelines on reporting changes in air-bone gap (ABG) remain in effect. The ABG is determined as the preoperative minus the postoperative air-bone gap. This is a continuous variable from negative values (worse gap after treatment) to positive values (less gap). The committee recommends reporting the mean, standard deviation, and range of the postoperative air-bone gaps and the number of decibels of change. For convenience, investigators may wish to report summaries of the air-bone gaps in bins. The committee recommends that when bins are used, they should be constructed as follows: 0 to 10 dB, 11 to 20 dB, 21 to 30 dB, and >30 dB.

The optimal measurements of hearing outcomes in cochlear implantation are evolving. Many factors are improving hearing results in this unique population, including improved cochlear implant technology, surgical techniques, and individuals with increasingly more residual hearing receiving implants. Because of the complexity of determining the optimal testing for performance outcomes in cochlear implant patients, the current standards are not applicable to studies dealing with cochlear implantation. If acoustic hearing is preserved following cochlear implantation and is measurable by standard audiometry, these standards can be applied and the hearing results reported in pre- and posttreatment scattergrams. If the efficacy of other types of hearing devices such as BAHA or middle ear implants is judged by standard audiometric evaluation, then hearing outcome data should be presented according to new standards published in this report.

Investigators are encouraged to make available all of their audiometric data as a manuscript appendix. Peer-reviewed journals may choose to offer the option of posting these data electronically. To facilitate the electronic creation of the scattergram, a spreadsheet is available for download that will create scattergrams from raw data (see instructions for authors for steps on how to use the scattergram tool, then use the link referenced here to download the tool). Data from spreadsheets can be uploaded and will be formatted in a scattergram as a high-resolution file suitable for publication.

The workgroup discussed at length the possibility of categorizing hearing loss into a 4-, 6-, or even 8-group scale analogous to the House-Brackmann scale in widespread use for describing facial palsy. There was a strong consensus that both the 4-class categorization of the 1995 AAO-HNS acoustic neuroma scale and the 5-class 1988 Gardner-Robertson scales were seriously inadequate and ought not be used further. Under the new minimal data standard, the scattergram method of reporting results allows investigators to categorize their results in ways they believe are informative while allowing subsequent investigators to reinterpret previous data according to their preferred presentation method.
Technical Details for Implementing the Minimal Standard for Reporting Hearing Loss

Implementing the New Reporting Standard

As detailed in this report, the AAO-HNS Hearing Committee has adopted a comprehensive hearing outcome scale that would encompass all patients, from those with normal hearing to those with profound deafness. These standards represent an iterative process that took place over 2 years of deliberation. Advice from hearing experts in both otolaryngology and audiology was widely sought. The AAO-HNS Hearing Committee sought input from the AAA regarding the creation of these standards. The senior specialty societies of otology and neurotology in both the United States and Europe contributed to the creation of the new hearing outcomes scale. After much thoughtful cooperation, the current standards were agreed upon unanimously by the AAO-HNS Hearing Committee and Board of Directors. As of this writing, the journals Otology & Neurotology and Otology & Neurotology have adopted the new minimal reporting standard as a publication requirement. We plan to similarly engage the editorial boards of all major, peer-reviewed publications in Otolaryngology.

Any scale that groups a heterogeneous population into defined categories has certain limitations in accurately describing the population. Alternatively, presenting only raw data from studies can be onerous to interpret. These new standards represent a compromise that seeks to accurately describe broad patient populations in a graphical way that will be easy to interpret, yet retain detailed critical information. Although there are still limitations to this new system, it has been widely accepted as an improvement over existing standards.

The baseline hearing characteristics of any population of patients should be reported in any study dealing with hearing. For studies that evaluate any intervention that may influence hearing, both a baseline and a posttreatment scattergram should be reported. As the hearing scale does not require point-of-care calculation, a more robust and descriptive presentation of the data can be employed. It should also be emphasized that the new standards represent minimal reporting standards. Investigators are encouraged to consider and report additional hearing data if relevant and create innovative ways to assess and report hearing outcomes. Moreover, studies are needed that validate hearing outcomes with carefully controlled, patient-based responses. These types of studies may allow for the creation of objective, evidence-based boundaries that delineate different classes of hearing.

The current standards supersede previous guidelines as they relate to reporting hearing outcomes and add to the previously established standards for conductive hearing loss. Other previously established reporting standards that do not relate to hearing, such as using the House-Brackmann facial nerve scale in outcomes of acoustic neuroma treatment and the Meniere’s staging system, still remain in effect. The new standards should not be used for measuring disability in occupational hearing loss or in forensic purposes.

The 4-frequency PTA using 0.5, 1, 2, and 3 kHz was recommended in the 1979 American Medical Association (AMA) guidelines on calculating hearing handicap and continues to be recommended by the AMA. Since that time, this combination of frequencies has been evaluated extensively against other combinations using self-reported hearing disability as the gold standard, and none has shown statistical superiority using large audiometric data sets. Moreover, this combination of frequencies was included in the 1995 Hearing and Equilibrium Committee’s guidelines and has been used until the present time. To ensure comparability with previously reported data, the same frequencies have been chosen for the current standards.

It is recognized that implementing new standards requires a period of time for transition. Since the new standards use data that are routinely collected and have only recommended a new way of presenting the data, it is expected that the transition period will be relatively brief and should be considered effective immediately.

Conclusion

Until now, no universal standard has existed for reporting hearing results in clinical trials. These new standards enable more precise and comprehensive representation of hearing outcomes. They will also allow for more efficient comparability between studies and pooling of data for meta-analysis.

Acknowledgments

We recognize the members of the AAO-HNS Hearing Committee for their work in developing and approving these standards. Members include (in alphabetical order) the following: Douglas Backous, Craig Buchman, Alan Cheng, James Coticchia, James Crawford, Robert Dobie, John Dornhoff, Thomas Eby, Jose Fayad, Neil Giddings, Barry Hirsch, Tina Huang, Timothy Hullar, Jon Isaacs, Robert Jackler (Chair), Paul Kilency, Brenda Lonsbury-Martin, William Luxford, Cliff Megerian, Ted Meyer, Alan Micco, Elias Michaelides, Robert O’Reilly, Kourosh Parham, Ryan Porter, and Jess Roberts.

Author Contributions

Richard K. Gurgel, conception and design, analysis and interpretation of data, critical revision; Robert K. Jackler, conception and design, analysis and interpretation of data, critical revision; Robert A. Dobie, conception and design, analysis and interpretation of data, critical revision; Gerald R. Popelka, conception and design, analysis and interpretation of data, critical revision.

Disclosures

Competing interests: Robert A. Dobie serves as consultant on hearing loss to a number of legal, industrial, and insurance concerns but has no conflict of interest directly pertaining to this report; he is also a consultant for Merz Pharmaceuticals.

Sponsorships: None.

Funding source: None.
References