informa
healthcare

International Journal of Audiology 2015; Early Online: 1–10

International Journal of Audiology

Original Article

Distribution characteristics of normal pure-tone thresholds

Robert H. Margolis*, Richard H. Wilson†, Gerald R. Popelka‡, Robert H. Eikelboom#,$,¶, De Wet Swanepoel#,$,¶ & George L. Saly*

*Audiology Incorporated, Arden Hills, Minnesota, USA, †James H. Quillen VA Medical Center, Mountain Home, Tennessee, USA, ‡Department of Otolaryngology, Stanford University, Stanford, California, USA, #Ear Science Institute, Subiaco, Australia, $Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa and ¶Ear Sciences Centre, School of Surgery, The University of Western Australia, Nedlands, Australia

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

The British Society of Audiology

The International Society of Audiology

NAS NORDIC AUDIOLOGICAL SOCIETY

Abstract

Objective: This study examined the statistical properties of normal air-conduction thresholds obtained with automated and manual audiometry to test the hypothesis that thresholds are normally distributed and to examine the distributions for evidence of bias in manual testing. *Design:* Four databases were mined for normal thresholds. One contained audiograms obtained with an automated method. The other three were obtained with manual audiometry. Frequency distributions were examined for four test frequencies (250, 500, 1000, and 2000 Hz). *Study sample:* The analysis is based on 317 569 threshold determinations of 80 547 subjects from four clinical databases. *Results:* Frequency distributions of thresholds obtained with automated audiometry are normal in form. Corrected for age, the mean thresholds are within 1.5 dB of reference equivalent threshold sound pressure levels. Frequency distributions of thresholds obtained by manual audiometry are shifted toward higher thresholds. Two of the three datasets obtained by manual audiometry are positively skewed. *Conclusions:* The positive shift and skew of the manual audiometry data may result from tester bias. The striking scarcity of thresholds below 0 dB HL suggests that audiologists place less importance on identifying low thresholds than they do for higher-level thresholds. We refer to this as the *Good enough bias* and suggest that it may be responsible for differences in distributions of thresholds obtained by automated and manual audiometry.

Key Words: Audiometry; automated audiometry; pure-tone thresholds; hearing; normal hearing; hearing test; air conduction; threshold; bias

The inventions of the telephone[1] by Alexander Graham Bell in 1876 and the triode vacuum tube by Lee DeForest in 1906 were the critical developments that gave rise to the pure-tone audiometers that are in use today. Although there were many earlier, ingenious non-electric and electric devices, it was the advent of transducers and oscillator circuits stemming from the inventions of the telephone and vacuum tube that enabled the production of calibratable pure-tone signals. As these devices came into use in laboratories, hospitals, and medical practices, procedures for measuring pure-tone thresholds began to evolve. Early descriptions by Fletcher and Wegel (1922), Jones and Knudsen (1924), and Bunch (1943) described, without a great deal of detail, the determination of auditory threshold, defined by Bunch as "the faintest sound which the listener can hear, not when he is reading the newspaper or enjoying a nap, but when his attention is focused on that particular sound" (p. 45). A more detailed description of our current method was provided by Hughson and Westlake (1944) and the technique is often referred to as the Hughson-Westlake method. Although their description is far more detailed than those that preceded it, the method leaves much to the discretion of the tester. The scientific basis of the method was provided in the classic article of Carhart and Jerger (1959) and the method was enshrined in the first standard for manual pure-tone audiometry (ANSI S3.21–1978).

A parallel development was occurring in the field of psychoacoustics during the evolution of pure-tone audiometry. Because the measurement of auditory sensitivity was critical to the development of models and theories of auditory function, precise methods for threshold measurements were necessary. Rigorous psychophysical procedures were developed that were administered manually at first but later could be controlled by computer. These methods provided consistency, controlled sources of bias, and resulted in measurements that facilitated mathematical modelling of auditory function. These methods are referred to as adaptive psychophysical methods. See Leek (2001) for a review.

Although adaptive psychophysical procedures are based on a more scientific approach than methods used for pure-tone audiometry, the "Hughson-Westlake Method" has never been validated against the more rigorous methods. In this report and a subsequent one we examine statistical characteristics of pure-tone thresholds for evidence of bias in manual pure-tone threshold measurements, and

Correspondence: Robert H. Margolis, Audiology Incorporated, 4410 Dellwood St., Arden Hills, MN 55112, USA. E-mail: rhmargo001@gmail.com

(Received 25 September 2014; accepted 19 March 2015)

ISSN 1499-2027 print/ISSN 1708-8186 online © 2015 British Society of Audiology, International Society of Audiology, and Nordic Audiological Society
DOI: 10.3109/14992027.2015.1033656

RIGHTSLINK

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

Abbreviations

AMTAS	Automated method for testing auditory sensitivity
ANSI	American National Standards Institute
HL	Hearing level
RETSPL	Reference equivalent threshold sound pressure level
UM	University of Minnesota
VA	Department of Veterans Affairs

compare the threshold distributions to those obtained with a more rigorous adaptive psychophysical procedure.

Margolis et al (2015, unpublished) examined frequency distributions of air-bone gaps that revealed evidence of tester bias in manual audiometry. Bone-conduction thresholds are particularly vulnerable to bias because the tester typically has expectations of the results from previous audiograms, thresholds at other frequencies, other test results such as tympanometry and otoscopy, and case history. Sackett (1979) described 35 sources of bias in research and refers to effects of prior knowledge of test results as *diagnostic suspicion bias*. Air-conduction thresholds may be less subject to tester bias because they are usually measured before bone-conduction thresholds but the tester may have knowledge from prior audiograms, case history, and other sources. Automated audiometry is not affected by such prior knowledge and is not subject to diagnostic suspicion bias.

Studebaker (1967) offered that the air-bone gap is a normally-distributed variable with the combined variance of air-conduction thresholds and bone-conduction thresholds, both of which are also normally-distributed variables. Margolis et al. (2015) indicated that Studebaker was correct with regard to thresholds obtained with a rigorously-controlled adaptive psychophysical method but frequency distributions of air-bone gaps obtained with manual audiometry were not normally distributed, probably due to tester bias.

The current study was prompted by the authors' observations that pure-tone thresholds of normal-hearing young adults are seldom less than 0 dB HL. We explored the possibility that bias on the part of the tester could account for this observation. A comparison of frequency distributions of air-conduction thresholds obtained by conventional manual audiometry and a bias-free, adaptive, automated method could shed light on biasing effects in pure-tone audiometry. In this report we examine the assumption that air-conduction thresholds of normal-hearing listeners are normally distributed by analysis of a database obtained with automated audiometry and three clinical databases that were obtained with manual audiometry.

The databases

The databases analysed for this report contained de-identified data with no protected health information and the proper approvals to use the databases for research were obtained. The databases were mined for normal-hearing thresholds based on the criteria shown in Table 1. The number of subjects remaining in each database after exclusion criteria were applied is shown in Table 2 for each test frequency. Thresholds at four test frequencies were analysed: 250, 500, 1000, and 2000 Hz. Thresholds at 4000 and 8000 Hz were not analysed to reduce influences of age and noise exposure.

The Busselton healthy ageing study

Frequency distributions of normal thresholds are based on audiograms obtained in the course of the Busselton Healthy Ageing Study, a detailed survey of the health of up to 4000 residents in the Shire of Busselton, Western Australia (see Swanepoel et al, 2013, and Hunter et al, 2013, for descriptions of the project). All non-institutionalized participants (born between 1946 and 1964) listed on the electoral roll ($n = 6690$) and residing in the Shire are eligible to participate. Enrollment into the study is randomized with 10% of the target sample drawn and recruited at a time. Data from the first 2023 participants were included in this study before exclusion criteria were applied. Subjects ranged in age from 45 to 66 years (mean = 56 years) at the time of testing.

Pure-tone thresholds were obtained with an automated method of pure-tone audiometry (AMTAS) that has been validated against manual audiometry performed by expert audiologists (Margolis et al, 2007, 2010, 2011; Margolis & Moore, 2011; Eikelboom et al, 2013; Mahomed et al, 2013; Swanepoel et al, 2013). AMTAS is a single-interval, forced-choice adaptive psychophysical procedure. The method was implemented on a computer-controllable audiometer (Madsen Conera) that delivered the pure-tone stimuli by circumaural earphones that have excellent ambient noise attenuation (Sennheiser HDA 200). Audiograms automatically were categorized by severity, configuration, site of lesion, and bilateral symmetry by a classification system (AMCLASS) that has been validated against the categories assigned by expert judges (Margolis & Saly, 2007, 2008a,b).

Table 1. Inclusion criteria for normal thresholds.

Busselton	*UM*	*VA*	*Stanford*
Normal by AMCLASS	18–30 Years of Age Normal by AMCLASS	18–30 Years of Thresholds ≤ 20 dB HL	Age 19–30 Years Thresholds ≤ 20 dB HL Normal otoscopy Static admittance > 0.4 mmho

Table 2. Descriptive statistics and sample sizes (n) for thresholds at each test frequency and for all frequencies combined. The skewness values are for composite distributions collapsed across the four test frequencies.

	Frequency (Hz)					
	250	*500*	*1000*	*2000*	*All frequencies*	*Skewness*
Busselton						
Mean	4.2	5.6	5.1	4.7	4.9	0.15
SD	6.2	6.4	6.5	7.4	6.6	
n	1172	1172	1172	1172		
UM						
Mean	8.0	5.7	6.0	4.0	5.9	0.45
SD	6.2	5.6	5.0	5.7	5.8	
n	1990	1994	1993	1994		
VA						
Mean	10.6	10.2	9.7	9.2	10.0	− 0.03
SD	5.6	6.0	5.7	6.3	5.9	
n	74 085	76 617	76 707	74 547		
Stanford						
Mean	8.9	9.3	8.5	9.0	8.9	0.35
SD	5.6	5.7	5.4	5.4	5.2	
n	764	750	729	711		

RIGHTSLINK

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

The University of Minnesota Hospital database (UM)

A database of audiograms obtained in the University of Minnesota Hospital Audiology Clinic (UM) was described by Margolis and Saly (2008a). Audiograms were obtained by licensed audiologists in the normal course of clinical evaluations. A small subset was obtained by supervised graduate students. The largest source of referrals was the Ear, Nose, and Throat Clinic which is in adjacent space. Other patients were referred from other clinical units, within and outside the Hospital and some were self-referred. Many of these patients were seen in conjunction with a hearing-aid dispensing program.

The VA database

A database of audiograms from 1,000,001 veterans was created by the Quality: Audiology and Speech Analysis and Reporting (QUASAR) Audiogram Module that is used by VA facilities across the U.S. The database is archived at the Denver Acquisition and Logistics Center which is the unit of the VA through which hearing aids and associated devices are procured and dispensed to veterans. Audiometric data are submitted to the database at the discretion of individual clinics. Cases were unselected with regard to hearing loss characteristics and demographics but were mostly male with hearing losses that were predominantly sensorineural. The mean age was 66.7 years (SD = 14.7 years). The majority of patients were at some stage of the hearing-aid fitting process. This database has been mined in conjunction with other research projects (Wilson & McArdle, 2013, 2014).

The Stanford University database

The Stanford University Medical Center audiology clinical service is part of a comprehensive tertiary care hospital clinic with outpatient and inpatient referrals from the Neurotology Division, other Otolaryngology divisions, other specialty services including Oncology, Internal Medicine, and Occupational Health, external sources including independent otolaryngology and audiology practices, and self-referrals by patients with predominantly age-related hearing loss. The database includes audiometric records from 13,180 sequential patient visits between April 20, 2010, and November 8, 2013. The data were collected from patients seen in the normal course of diagnostic evaluations provided by Stanford licensed audiologists. Multiple audiograms from the same patient were included as well as some pure-tone audiometric data ($<12\%$ of the database) from sources outside of Stanford but entered into the electronic medical record system for purposes of audiologic management. The database includes tympanometry and otoscopy results that provided valuable information for excluding subjects with middle-ear pathology.

Definition of normal hearing

The databases provided different information that was used to exclude subjects with hearing loss so the operational definitions of normal hearing were slightly different. Inclusion criteria are shown in Table 1. The threshold ≤ 20 dB HL criterion was implemented separately at each test frequency. The other inclusion criteria were implemented on a subject basis.

Subjects in the Busselton study were included if the AMCLASS severity category was "normal" and the site of lesion category was not "conductive" (which can occur if air-conduction thresholds are in the normal range and there is one or more air-bone gaps). A designation of "normal" is based on a complex set of rules that were empirically found to maximize agreement with a panel of expert judges (Margolis & Saly, 2007). Because these subjects ranged in age from 45–66 years, an age criterion could not be used. Instead, an age correction was applied to the mean Busselton thresholds based data from Corso (1959).

Subjects in the UM were included based on the same criterion used for the Busselton data, a "normal" AMCLASS determination. In addition, only subjects aged 18–30 years were included.

Subjects in the VA database were included if their ages were between 18 and 30 years. Thresholds exceeding 20 dB HL were excluded on a frequency by frequency basis.

Subjects in the Stanford database were included if their ages were between 18 and 30 years. Thresholds exceeding 20 dB HL were excluded on a frequency by frequency basis. Because the Stanford database includes otoscopy and tympanometry results, subjects with possible middle-ear involvement could be excluded based on these measurements. Cases with low static admittance at 226 Hz (<0.4 mmho) and abnormal otoscopic findings were excluded.

The rationale for excluding thresholds exceeding 20 dB HL is as follows. Each distribution can be viewed to be the sum of two separate distributions, one comprised of normal-hearing subjects and one comprised of subjects with hearing loss. The distribution of normal thresholds is expected to be normal in form (Studebaker, 1967) with a mean that is close to the reference equivalent threshold sound pressure levels that are in the standards, that is, 0 dB HL. The standard deviation associated with normal thresholds is typically about 5–6 dB (Wilbur & Goodhill, 1967, Tables 2 &3; Weissler, 1968, Table VIII; Lawton, 2005, Table 1). Based on these assumptions, a threshold exceeding 20 dB HL is more than three standard deviations above the mean and likely from a different distribution, namely, the distribution of abnormal thresholds.

Results

Means and standard deviations of thresholds at all test frequencies are summarized in Table 2. Figures 1 to 4 show frequency distributions at all frequencies and composite distributions collapsed across test frequencies.

The Busselton database

Frequency distributions of normal pure-tone thresholds from the Busselton database are shown in Figure 1. When the thresholds at all frequencies are combined, they are well described by a normal distribution with a mean of 4.9 dB HL and a standard deviation of 6.6 dB. The best-fit normal distribution was corrected for age based on age-related normal thresholds reported by Corso (1959). A 3.4-dB shift was applied, shown as the dashed curve in the lower panel of Figure 1. This age-corrected composite distribution represents the model against which the results from the other databases are compared.

The University of Minnesota Hospital (UM) database

The frequency distributions of normal thresholds from the UM database are shown in Figure 2. The composite distribution resembles a normal distribution, shifted about 4 dB toward higher thresholds relative to the age-corrected Busselton distribution with a slight positive skew.

RIGHTSLINK

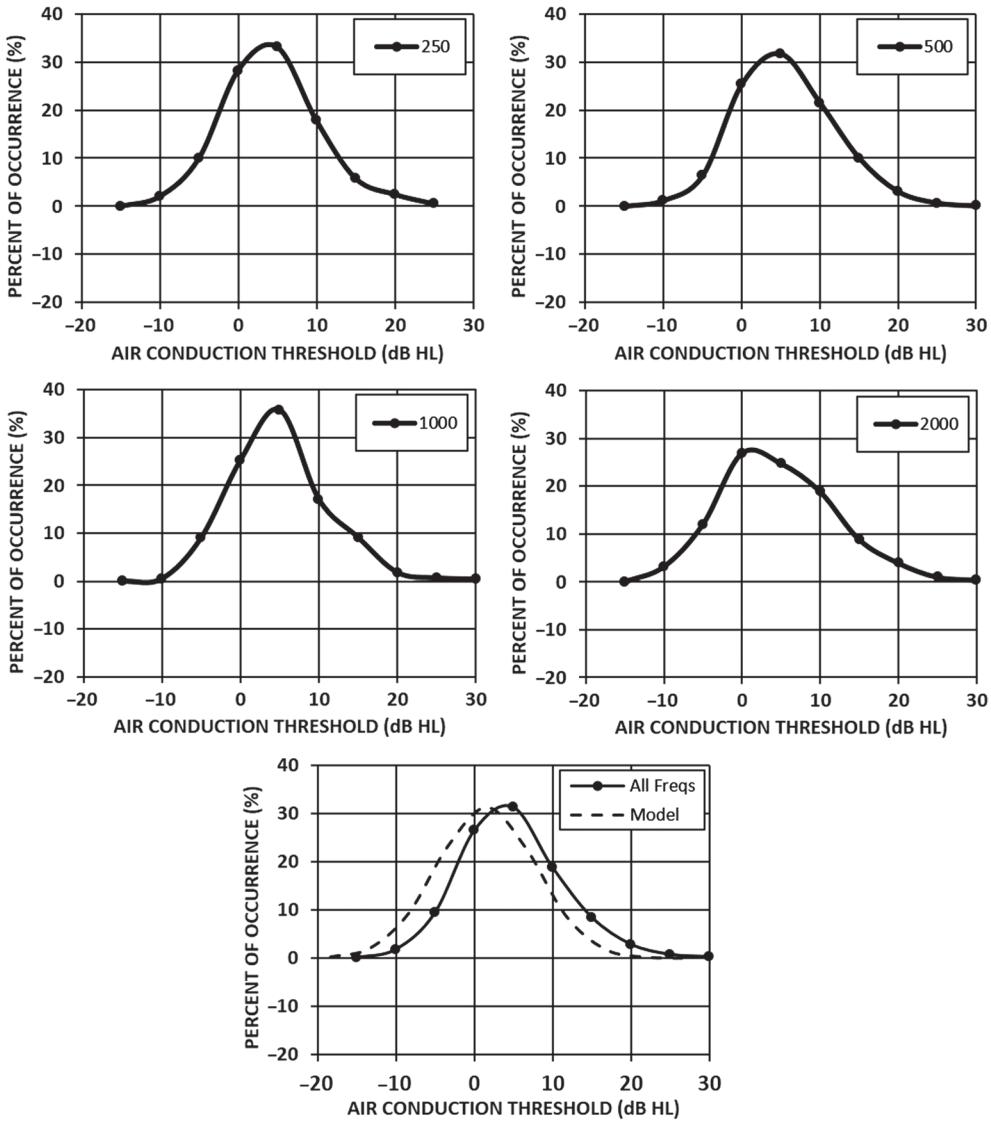

Figure 1. Distributions of normal thresholds from the Busselton database. The solid curve in the lowest panel is the composite distribution collapsed across test frequency. The dashed curve is the best-fit age-corrected normal distribution. Thresholds were obtained with automated audiometry.

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

The VA database

The frequency distributions of normal thresholds from the VA database are shown in Figure 3. The distributions are shifted about 9 dB toward higher thresholds relative to the age-corrected Busselton distribution. There is no apparent skew in the VA distributions.

Stanford University database

The frequency distributions of normal thresholds from the Stanford University database are shown in Figure 4. The results are very similar to the UM distributions with the composite distribution resembling a normal distribution shifted about 7 dB toward higher thresholds relative to the age-corrected Busselton distribution with a slight positive skew.

Composite distributions

Figure 5 shows composite frequency distributions for all test frequencies for each database. The dashed curve in each panel is the age-corrected composite distribution from the Busselton data. That distribution is normal in form with a mean of 1.5 dB HL and a standard deviation of 6.6 dB. A 3.4-dB age correction has been applied based on the mean difference in thresholds at the four test frequencies for 43–49 year-olds and 18–25 year-olds from Corso (1959). The means and standard deviations of each distribution are given in Table 2.

The shift toward higher thresholds relative to the age-corrected Busselton data is evident in each panel of Figure 5. The VA data appear to be normal in form and the UM and Stanford functions show a slight positive skew as indicated in Table 2.

Discussion

Relation to reference-equivalent threshold sound pressure levels (RETSPLs)

Average normal thresholds obtained in the samples of clinical audiograms analysed in this report do not show the expected prevalence

RIGHTSLINK

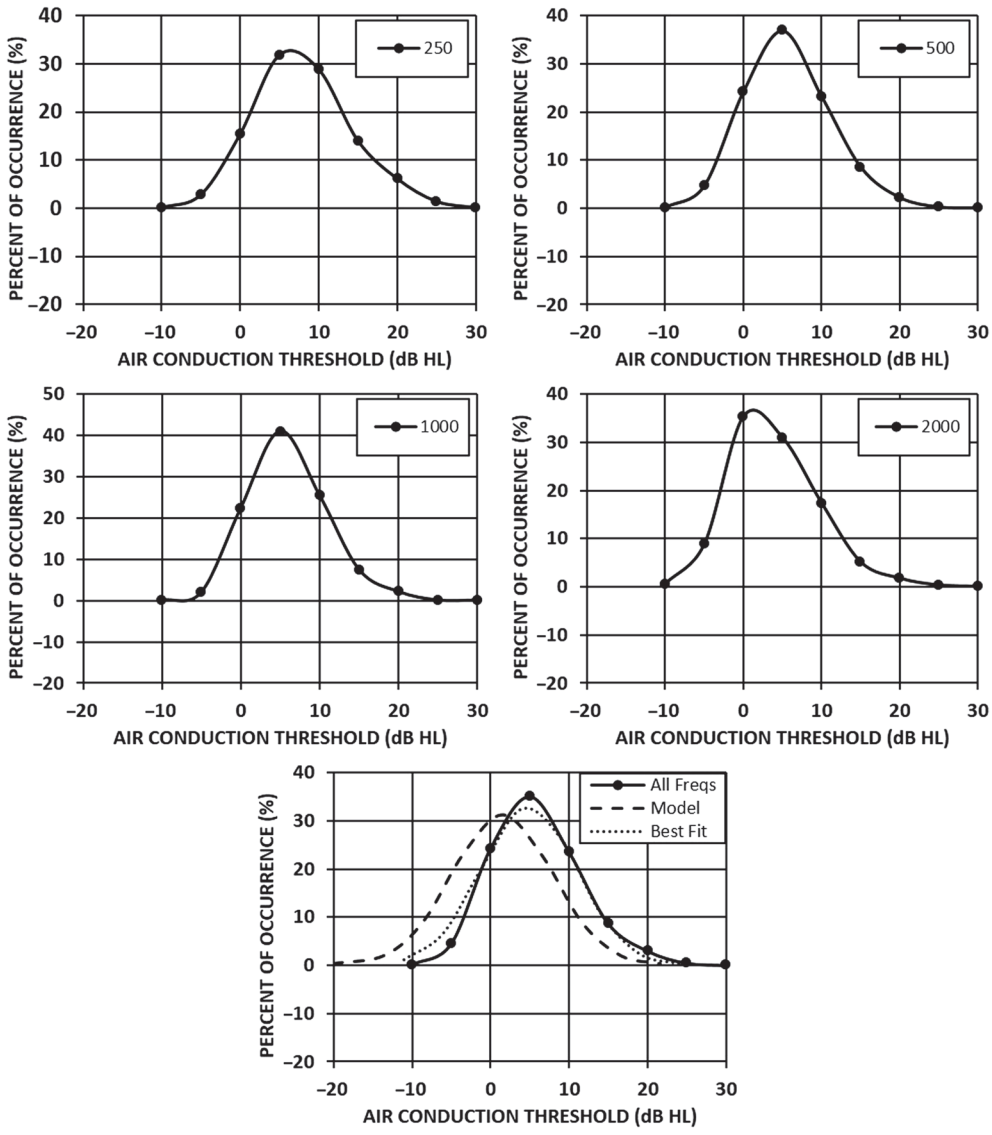

Figure 2. Distributions of normal thresholds from the UM database. The lowest panel is the composite distribution collapsed across test frequency. The solid curve is fit with a curved line connecting the data points. The dotted curve is the best-fit normal distribution. The dashed curve is the best-fit age-corrected distribution from the Busselton database. Thresholds were obtained with manual audiometry.

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

of thresholds of 0 dB HL and below. This is evident in Figures 2–4 where mean thresholds range from 5–10 dB HL and are rarely below 0 dB HL. However, average normal thresholds from the Busselton study, when corrected for age, are close to the RETSPLs specified in the audiometer standards (ANSI S3.6-2010; ISO 389.8-1998). We suggest that the differences between mean thresholds obtained with the manual procedure and the mean thresholds obtained with the automated procedure are, at least in part, due to tester bias.

Evidence of bias

Sackett (1979) described bias in research as "any process at any stage of inference which tends to produce results or conclusions that differ systematically from the truth." In Sackett's framework one of the seven stages of research in which bias can occur is "in measuring exposures or outcomes" (p. 60). In our study of statistical properties of air-bone gaps we argued that one of Sackett's bias types, *Expectation bias,* influences bone-conduction thresholds as a result of prior knowledge of the tester that leads to expectations of what the results should be.

Two features of the differences between automated and manual air-conduction threshold distributions suggest that tester bias affects manual audiometry – the higher mean thresholds and the positively skewed distributions.

The higher mean thresholds obtained with manual audiometry may reflect an indifference to very low thresholds by testing audiologists. When subjects respond at a low level that the audiologist interprets as an indication of normal hearing, there is little incentive to accurately measure the lower threshold. This form of bias does not align with any of the 35 forms of research bias described by Sackett (1979). Perhaps the *Good Enough bias* is an apt description.

The skewness of the UM and Stanford composite frequency distributions can be seen in the lower panels of Figures 2 and 4. The dotted line in each figure is the best fit normal distribution fit to the

RIGHTSLINK

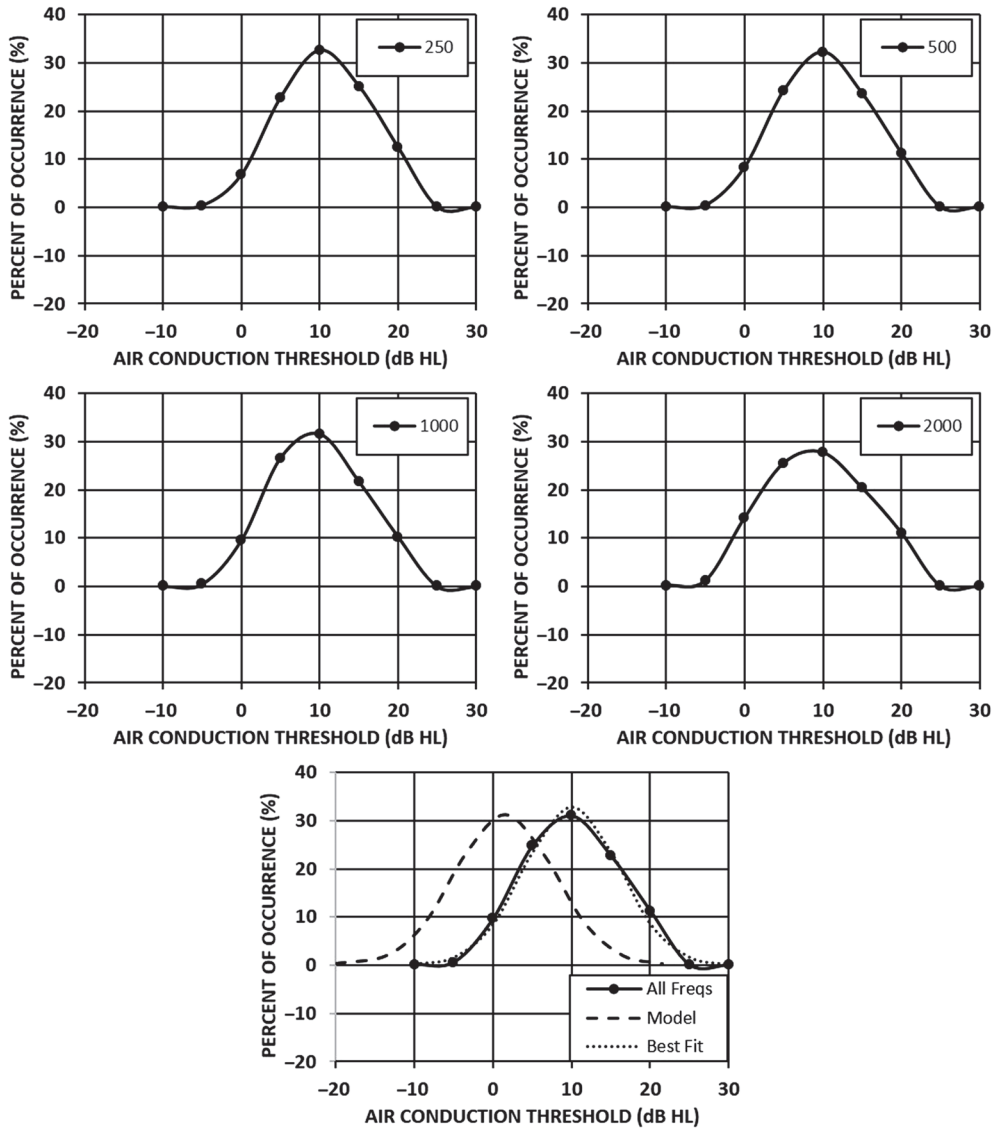

Figure 3. Distributions of normal thresholds from the VA database. The lowest panel is the composite distribution collapsing across test frequency. The solid curve is fit with a curved line connecting the data points. The dotted curve is the best-fit normal distribution. The dashed curve is the best-fit age-corrected distribution from the Busselton database. Thresholds were obtained with manual audiometry.

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

composite data and shifted slightly (1 dB in Figure 2, and 2 dB in Figure 4) to align the peaks of the curves to facilitate an examination of the shapes. The steeper slope on the negative side of the solid curves relative to the positive side of the curves is an indication of positive skew. This positive skew can be quantified as follows.

The skewness of a frequency distribution is given by

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{S} \right)^3 \quad (1)$$

where n is the number of cases, x is the value for each case, $\bar{x}$ is the mean, and s is the standard deviation. (See Doane & Seward, 2011.) A skewness value of 0 indicates a symmetrical (unskewed) distribution. Negative values indicate negative skew, i.e. over-representation of negative values. Positive values indicate a positive skew i.e. over-representation of positive values. The standard error of skewness is given by

$$SE_S = \sqrt{(6/n)} \quad (2)$$

where n is the number of cases (Tabachnick & Fidell, 1996).

The 90% confidence interval (95th %ile – 5th %ile) is given by

$$90\%\ \text{C.I.} = \bar{x} \pm 1.96\ SE_S. \quad (3)$$

Skewness values and confidence intervals for the composite distributions from each database are shown in Figure 6. Skewness values for the Busselton and VA databases are close to zero suggesting symmetrical distributions that are normal in form. The UM and Stanford distributions are characterized by positive values indicating that the distributions are characterized by positive skews. The overlapping

RIGHTSLINK

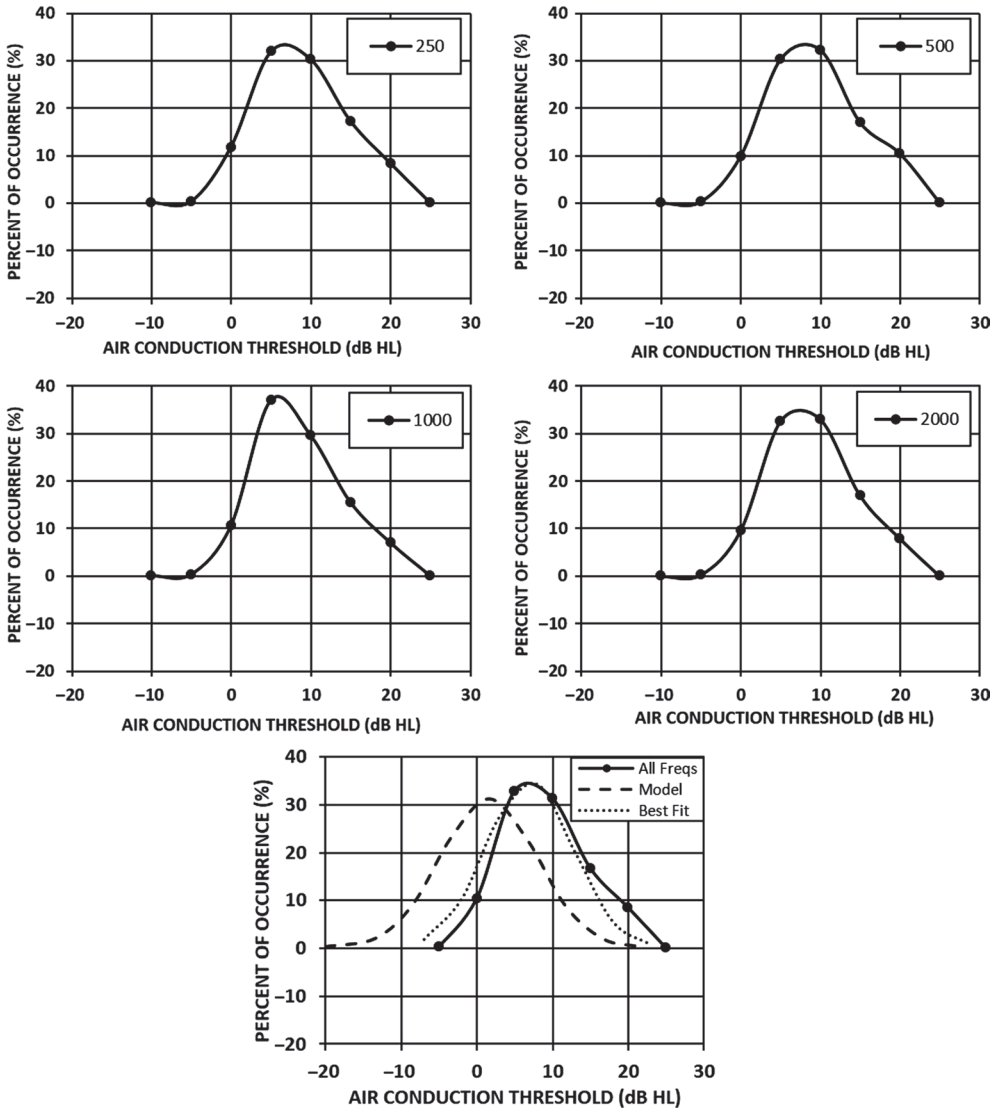

Figure 4. Distributions of normal thresholds from the Stanford database. The lowest panel is the composite distribution collapsed across test frequency. The solid curve is fit with a curved line connecting the data points. The dotted curve is the best-fit normal distribution. The dashed curve is the best-fit age-corrected distribution from the Busselton database. Thresholds were obtained with manual audiometry.

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

confidence intervals for the UM and Stanford distributions indicate that the skewness values are not significantly different. The positive skew is consistent with the interpretation of tester bias against low threshold values.

By redistributing thresholds toward higher values, this bias would be expected to decrease the variance (see Fortmann-Roe, 2012 for a discussion). This is evident in the smaller standard deviations for the distributions obtained by manual audiometry relative to that for the Busselton data (Table 2). The absence of bias in the VA distribution may result from the higher thresholds. Because low thresholds rarely occur the *Good Enough bias* is not often invoked.

It is interesting to speculate about how tester bias creeps in to the hearing test process. Certainly audiologists are not trained to use a procedure that does not accurately measure auditory threshold. Perhaps it creeps in due to the exigencies of the clinical setting with its emphasis on efficiency and throughput. If this is the case, one would expect the biasing effects to be quite variable across clinics and audiologists. Training programs should educate audiology students about the nature of bias in clinical measurement and how to avoid it. An unbiased measurement procedure would result in more confidence in small differences that may reflect changes over time and in small air-bone gaps that could reflect small but clinically significant middle-ear involvement.

Contribution from ambient noise

Ambient room noise elevates thresholds, particularly for normal-hearing listeners. Because ambient noise tends to be weighted toward lower frequencies, low frequency thresholds are most vulnerable. The influence of room noise is affected by the ambient sound attenuation provided by the earphone that is used for audiometry. An analysis of testable ranges in various test rooms with various earphones revealed that when using supra-aural earphones, thresholds as low as − 10 dB HL at 1000 Hz and above can be measured in typical single-wall and double-wall clinic sound rooms but at lower frequencies the lower limit of the testable range could be limited to − 5 or

RIGHTSLINK

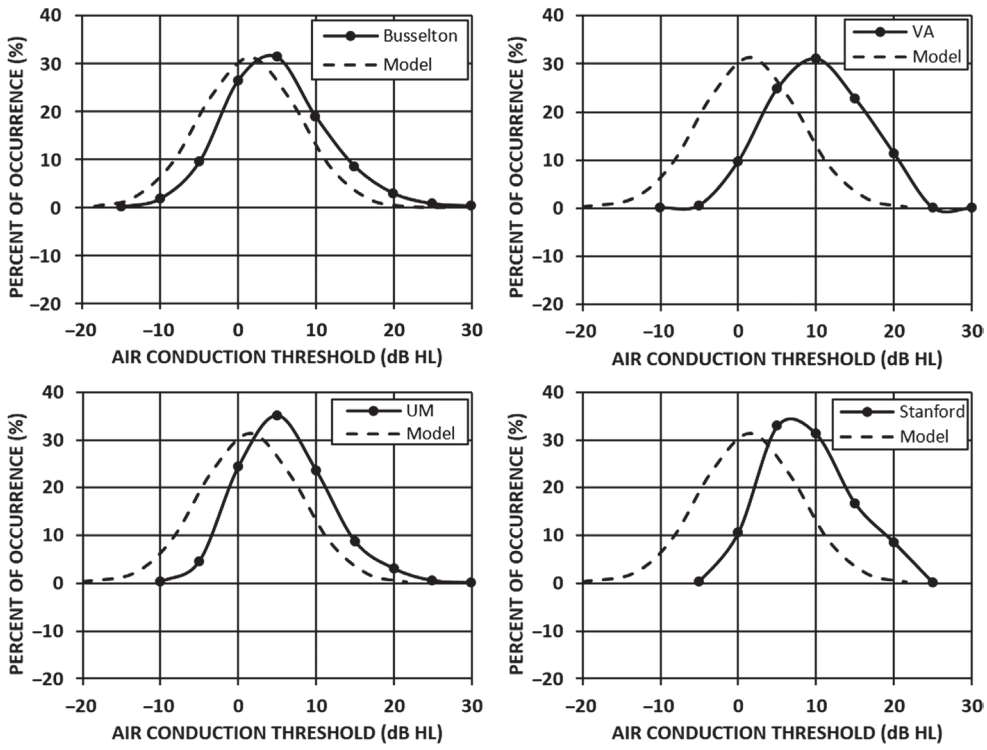

Figure 5. Composite distributions of normal thresholds from each database. The solid curves are the distributions collapsed across four test frequencies. The dashed curve in each panel is the age corrected best-fit normal distribution for the Busselton data.

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

0 dB HL (Margolis & Madsen, 2015, unpublished). Although the ambient noise levels and transducers are not known for the manual audiometry databases analysed in this report, ambient noise could be a contributor to 250- and 500-Hz results but probably not to the results at 1000 and 2000 Hz.

The slight threshold elevations in the Busselton data that we attribute to age may have diminished the effect of ambient noise on low thresholds. It is possible that the distribution of thresholds obtained with AMTAS for a younger population may be skewed by a greater influence of ambient noise. A study of younger normal-hearing subjects would reveal the effect of room noise on distribution characteristics.

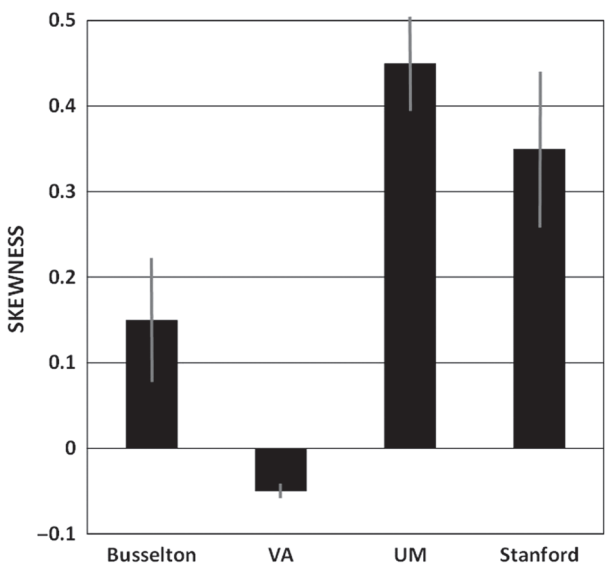

Figure 6. Skewness values for the composite distributions shown in Figure 5. The vertical lines are 90% confidence intervals.

Implications for audiometer design

The paucity of thresholds below 0 dB HL and the apparent *Good Enough bias* that it suggests is an indication that testing audiologists feel there is little clinical information of value to be gained by testing at those low levels. The audiometer standards (ANSI S3.6-2010; IEC 60605-1) specify performance requirements for four audiometer types, from the most complex diagnostic instruments (Type 1) to screening devices (Type 4). The required minimum hearing level for all types is − 10 dB HL or lower. It is clear from the data in Figures 1 to 4 that audiologists seldom present signals below 0 dB HL. For screening audiometers (Type 4) the requirement to provide levels below 0 dB HL does not seem to be justified. A reasonable approach would be to require levels as low as 0 dB HL and let manufacturers and the marketplace decide if lower levels should be provided.

Summary and Conclusions

Four audiometric databases were analysed to study the frequency distributions of the air-conduction thresholds of normal-hearing listeners. One of the databases was acquired with a bias-free automated procedure (AMTAS). Three were acquired with manual audiometry procedures as used in routine clinical assessment. The results suggest the following conclusions:

1. The frequency distribution of thresholds of normal-hearing 45–65 year olds tested as part of a community-based investigation and obtained with an unbiased automated method is well

RIGHTSLINK

described by a normal distribution with a mean of 4.9 dB HL and a standard deviation of 6.6 dB. When corrected for age, the mean thresholds are within 1.5 dB of the RETSPLs provided in the audiometer standards.

2. The frequency distributions of normal thresholds obtained from two academic health centers are well-described by normal distributions shifted by about 5 dB and show a slight positive skew. These characteristics are consistent with a bias on the part of the testing audiologist that undersamples responses to low-level stimuli ($<$ 0 dB HL). This *Good Enough bias* reflects the low importance placed on measuring thresholds at lower levels.
3. The frequency distribution of normal thresholds from the VA database is well described by a normal distribution shifted by about 10 dB and is normal in form. The normal form of the distribution indicates that the *Good Enough bias* does not occur because lower levels are seldom tested.
4. The paucity of thresholds below 0 dB HL calls into question the requirement in the audiometer standards that all audiometers must provide minimum levels of $-$ 10 dB HL or that audiologists modify their procedures to eliminate the *Good Enough bias*.
5. Audiologists should be aware of the risk of bias in behavioral testing and encouraged to use unbiased methods.

Acknowledgements

Portions of this work were supported by grant RC3DC010986 from the National Institutes of Deafness and Other Communication Disorders. The Rehabilitation Research and Development Service of the U.S. Department of Veterans Affairs supported this work through the Auditory and Vestibular Dysfunction Research Enhancement Award Program (REAP) and a Senior Research Career Scientist award to the second author. Appreciation is extended to Kevin Quitmeyer, Pam Urrutia, and Mary Ann Blumenthal from the Denver Acquisition and Logistics Center who extracted the VA archived data. Dr. Michael Hunter and his team at the Busselton Health Study provided assistance in collecting the Busselton data. The contents of this report do not represent the views of the Department of Veterans Affairs or the United States Government.

Note

1. Although Bell is widely credited with the invention of the telephone, a case has been made for Antonio Meucci (1808–1889) for his "*telettrofono*" which was submitted to the U.S. Patent Office in 1871. See Meucci, 2010.

Declaration of interest: AMTAS intellectual property is owned by Audiology Incorporated (AI) in which the first and last author have commercial interests. That intellectual property may be incorporated into commercial products. The other authors report no conflicts of interest.

References

ANSI S3.21. 1978. Methods for manual pure-tone threshold audiometry. American National Standards Institute, New York.

ANSI S3.6. 2010. Specification for audiometers. American National Standards Institute, New York.

Bunch C.C. 1943. *Clinical Audiometry*. St. Louis: C.V. Mosby.

Carhart R. & Jerger J.F. 1959. Preferred method for clinical determination of pure-tone thresholds. *J Speech Hear Dis*, 24, 330–345.

Corso J.F. 1959. Age and sex differences in pure-tone thresholds. *J Acoust Soc Am*, 31, 498–507.

Doane D.P. & Seward L.E. 2011. Measuring skewness: A forgotten statistic. *J Stat Educ*, 19, 1–18.

Eikelboom R.H., Swanepoel D.W., Motakef S. & Upson G. 2013. Clinical validation of the AMTAS automated audiometer. *Int J Audiol*, 52, 342–349.

Fletcher H. & Wegel R.L.1922. The frequency sensitivity of normal ears. *Proc National Acad Sciences*, 8, 5–8.

Fortmann-Roe S. 2012. www.http://scott.fortmann-roe.com/docs/BiasVariance.html, retrieved 2 March 2015.

Hughson W. & Westlake H.D. 1944. Manual for program outline for rehabilitation of aural casualties both military and civilian. *Trans Amer Acad Ophthalmology Otolaryngology*, 48, 1–15.

Hunter J.A., Straker L., Beilby J., Bucks R., Davis T. et al. 2013. Rationale, design and methods for a community-based study of clustering and interactions of chronic disease processes and their effects on ageing: *The Busselton Healthy Ageing Study*. BMC Public Health. 13,936. Published online Oct 8, 2013. doi: 10.1186/1471-2458-13-936.

IEC 60645-1. 2012. Electroacoustics - Audiometric Equipment - Part 1: Equipment for pure-tone audiometry. Geneva: International Electrotechnical Commission.

ISO 389.1. 1998. Acoustics - Reference zero for the calibration of audiometric equipment - Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones. Geneva: International Organization for Standardization.

ISO 389.8. 1998. Acoustics - Reference zero for the calibration of audiometric equipment - Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones. Geneva: International Organization for Standardization.

Jones I.H. & Knudsen V.O. 1924. Functional tests of hearing. *Laryngoscope*, 34, 673–686.

Lawton B.W. 2005. Variation of young normal-hearing thresholds measured using different audiometric earphones: Implications for the acoustic coupler and the ear simulator. *Int J Audiol*, 44, 444–451.

Leek M.R. 2001. Adaptive procedures in psychophysical research. *Percep Psychophys*, 63, 1279–1292.

Mahomed F., Swanepoel D.W., Eikelboom R.H. & Soer S. 2013. Validity of automated threshold audiometry: A systematic review and meta-analysis. *Ear Hear*, 34, 745–752.

Margolis R.H., Eikelboom R.H., Johnson C., Ginter S.M., Swanepoel D.W. et al. 2013. False air-bone gaps at 4 kHz in listeners with normal hearing and sensorineural hearing loss. *Int J Audiol*, 52, 526–532.

Margolis R.H., Frisina R. & Walton J.P. 2011. AMTAS® automated method for testing auditory sensitivity: II. Air conduction audiograms in children and adults. *Int J Audiol*, 50, 434–439.

Margolis R.H., Glasberg B.R., Creeke S. & Moore B.C.J. 2010. AMTAS automated method for testing auditory sensitivity: Validation studies. *Int J Audiol*, 49, 185–194.

Margolis R.H. & Madsen B. 2015. The acoustic test environment for hearing testing. *J Amer Acad Audiol*, in press.

Margolis R.H. & Moore B.C.J. 2011. Automated method for testing auditory sensitivity: III. Sensorineural hearing loss and air-bone gaps. *Int J Audiol*, 50, 440–447.

Margolis R.H. & Saly G.L. 2007. Toward a standard description of hearing loss. *Int J Audiol*, 46, 746–758.

Margolis R.H. & Saly G.L. 2008a. Distribution of hearing loss characteristics in a clinical population. *Ear Hear*, 29, 524–532.

Margolis R.H. & Saly G.L. 2008b. Asymmetrical hearing loss: Definition, validation, prevalence. *Otol Neurotol*, 29, 422–431.

Margolis R.H., Saly G.L., Le C. & Laurence J. 2007. Qualind™: A method for assessing the accuracy of automated tests. *J Am Acad Audiol*, 18, 78–89.

Margolis R.H., Wilson R.H., Popelka G.R., Eikelboom R.H., Swanepoel D.W. et al. 2015. A statistical model of the air-bone gap: Evidence of bias in manual audiometry. *Int J Audiol*, submitted for publication.

Meucci S. 2010. *Antonio and the Electric Scream*. Boston, USA: Brandon Books.

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

RIGHTSLINK

Sackett D.L. 1979. Bias in analytic research. *J Chron Dis*, 32, 51–63.

Studebaker G.A. 1967. Intertest variability in the air-bone gap. *J Speech Hear Disord*, 32, 82–86.

Swanepoel D., Eikelboom R.H., Friedland P.F., Atlas M.D. & Hunter M. 2013. Self-reported hearing loss in baby boomers from the Busselton Health Study: Audiometric correspondence and predictive value. *J Am Acad Audiol*, 24, 514–521.

Tabachnick B.G. & Fidell L.S. 1996. *Using Multivariate Statistics (3rd ed.)*. New York: Harper Collins.

Weissler P. 1968. International standard reference zero for audiometers. *J Acoust Soc Am*, 44, 264–275.

Wilber L.A. & Goodhill V. 1967. Real ear versus artificial mastoid methods of calibration of bone-conduction vibrators. *J Speech Hear Res*, 10, 405–416.

Wilson R.H. & McArdle R. 2013. Characteristics of the audiometric 4000 Hz notch (744 553 veterans) and the 3000, 4000, and 6000 Hz notches (539 932 veterans). *J Rehabil Res Dev*, 50, 111–132.

Wilson R.H. & McArdle R. 2014. A treatise on the thresholds of interoctave frequencies: 1500, 3000, and 6000 Hz. *J Am Acad Audiol*, 25, 171–186.

Int J Audiol Downloaded from informahealthcare.com by University of Minnesota on 05/04/15
For personal use only.

RIGHTSLINK