Review: Inter and Intra-Reader Agreement Among Audiologists in Reading Auditory Brainstem Response Waves

Revue : concordances entre audiologistes et chez le même audiologiste pour la lecture des ondes des potentiels évoqués auditifs du tronc cérébral

KEY WORDS

AUDITORY BRAINSTEM RESPONSES

AUDIOLOGISTS

INTER-READER AGREEMENT

INTRA-READER AGREEMENT

Maha Zaitoun
Steven Cumming
Alison Purcell

Abstract

This paper presents a review conducted to evaluate the scientific evidence regarding variability in audiologists' interpretation of auditory brainstem response (ABR) tests and to determine the factors that may affect audiologists' performance when reading ABRs. A search of the literature on Pubmed, Medline (Ovid), ScienceDirect and Google scholar yielded 4,735 articles. After culling, only six articles remained which investigated audiologists' variability in interpreting ABR, and the findings were inconsistent. Four of the six studies reported evidence that audiologists were variable when reading ABR waves, while two studies reported that audiologists were highly consistent when reading ABR waves. This conflict may be explained by the heterogeneity in the methods used in the six studies. More experienced audiologists were likely to show less variability in interpretation, but no other factors were shown to predict variability.

Abrégé

Il s'agit d'une revue de littérature effectuée pour évaluer la preuve scientifique de la variabilité dans l'interprétation, par les audiologistes, des tests de potentiels évoqués auditifs du tronc cérébral (PÉATC) et pour déterminer les facteurs pouvant affecter la performance des audiologistes à lire les PÉATC. Une recherche sur Pubmed, Medline (Ovid), ScienceDirect et Google scholar a permis de répertorier 4 735 articles. Après élagage, il n'est resté que six articles qui portaient sur la variabilité des audiologistes dans leur interprétation des ondes des PÉATC, et dont les conclusions étaient contradictoires. Quatre des six études rapportaient la preuve que la lecture des ondes des PÉATC était variable entre les audiologistes tandis que deux autres rapportaient qu'elle concordait fortement. Cette divergence peut s'expliquer par l'hétérogénéité dans les méthodes utilisées dans les six études. Les audiologistes possédant plus d'expérience auraient moins de variabilité dans leur interprétation. Aucun autre facteur n'apparaissait comme pouvant prédire la variabilité.

Maha Zaitoun
Faculty of Health Sciences,
The University of Sydney
Lidcombe, NSW, 1825
AUSTRALIA
Faculty of Applied Medical Sciences, Jordan University of Science and Technology
P.O. Box 3030.
Irbid 22110,
JORDAN

Steven Cumming,
Associate Professor
Faculty of Health Sciences,
The University of Sydney
Lidcombe, NSW, 1825
AUSTRALIA

Alison Purcell, Ph.D.
Faculty of Health Sciences,
The University of Sydney
Lidcombe, NSW, 1825
AUSTRALIA

The Auditory Brainstem Response in audiology

What is ABR?

The auditory brainstem response (ABR) is an electrical evoked potential that arises as a result of the neural activity from the auditory nerve, auditory nuclei, and the tracts of lower brainstem within the first 10-15ms after a click or tone burst stimulus (Ballachanda, Moushegian, & Stillman, 1992; Hall, 2007; Vidler & Parker, 2004). The ABR was first described by Jewett, Romano and Williston in the 1970s, when they reported recording a series of waves through the scalp after stimulating the ear with a click stimulus. Since then, the potential applications of the ABR have been explored including establishing hearing thresholds, diagnosing retrocochlear lesions, monitoring brain activity during surgical operations, and using it as a hearing screening tool for newborns (Hall, 2007; Hood, 1998)

The most frequent stimulus used in the ABR testing is the click, which is a broadband signal that can excite a wide region at the cochlea (Hood & Berlin, 1986). Click ABRs are usually used to assess auditory nerve and auditory brainstem pathway integrity and determine the hearing threshold for infants and young children when behavioral audiometric results cannot be obtained reliably. However, the click stimulus lacks frequency specificity. This means that the ABR from a click may not reflect responses for cochlear regions that are responsive to lower frequencies (Bogus, 1996). Tone burst or tone pips are alternative stimuli for ABR testing that can provide frequency specific auditory thresholds (Hood & Berlin, 1986).

The ABR elicited to clicks consists of seven waveforms; however, only waves I, III and V are identified and analyzed in clinical settings due to the variability in waves II, IV, VI and VII (Burkard, Eggermont & Don, 2007; Roeser, Valente & Hosford-Dunn, 2007). Wave I appears at approximately 1.67ms following a click stimulus and is believed to reflect afferent activity of the auditory nerve fibres from the cochlea to the auditory brainstem. Wave III is evident approximately 3.8ms following a click stimulus; is thought to arise from the cochlear nucleus, the trapezoid body, and the superior olivary complex. Wave V, arising about 5.6ms following a click stimulus, is the largest and most robust wave in the ABR. It is thought to arise from the lateral lemniscus termination in the inferior colliculus (Hall, 2006). To interpret the ABR waveforms and determine hearing thresholds, audiologists consider a number of characteristics such as amplitude, latency and reproducibility (Burkard et al., 2007; Roeser et al., 2007).

Purposes of ABR

The ABR is commonly used either to establish hearing threshold with patients for whom a behavioral response is difficult to elicit or to detect lesions or abnormalities in the auditory system (Hyde & Blair, 1981). For hearing threshold estimation purpose, wave V is usually targeted, as it is a robust wave even close to threshold, while for lesion detection purpose, all ABR waves are targeted and monitored (Hall, 2006). The ABR is a valuable tool in the test battery of estimating hearing threshold for hard to test patients such as pediatric populations and developmentally delayed patients who cannot complete the conventional behavioral audiometry tests (Burkard et al., 2007; Cornacchia, Vigliani, & Arpini, 1982; Jerger, Hayes, & Jordan, 1980). Schmulian and McMahon (2005) referred to the ABR as the most important and definitive test in assessing infants. The ABR is now widely used in young infants to obtain information about hearing threshold from each ear in the frequency range between 500-4000 Hz using both air and bone conduction stimuli (Schmulian & McMahon, 2005; Stapells, 2000). The second purpose for the ABR is lesion detection. Previous work has shown that ABR is a sensitive tool for detecting large lesions (equal or larger than 2 centimeters), but not sensitive for small lesions (equal or less than 1 centimeters) (Chandrasekhar, Brackmann, & Devgan, 1995; Gordon & Cohen, 1995; Schmidt, Sataloff, Newman, Spiegel, & Myers, 2001; Zappia, O'Connor, Wiet, & Dinces, 1997).

ABR interpretation

Because the ABR does not require any verbal or behavioral response from the patient, it has been referred to as an objective test of hearing. However, the ABR is highly dependent on the interpretation and judgment of the audiologist (Cone-Wesson, Dowell, Tomlin, Rance, & Ming, 2002; Stueve & O'Rourke, 2003) and indeed Weber referred to ABR as one of the most subjective audiometric techniques (Weber, 1983). In order to determine hearing thresholds from the ABR, audiologists use a visual inspection method by looking at waveforms to determine whether the waves are absent or present. In most cases, the same audiologist determines the latency of each wave (Weber, 1983). Despite the fact that minimum standards have been set to define the most reliable threshold, the final decision for hearing threshold estimation still depends on the interpretation of individual audiologists (Vidler & Parker, 2004). Weber (1983) suggested that the ability to read and mark waves of ABR could improve with experience, which may take a long time, particularly if the examiner is mainly "self taught". He also argued that the skills required to interpret ABR waves are often underestimated and even the most experienced and skilled audiologists may struggle to interpret some difficult cases during their career (Weber, 1983).

Decision making and reliability in the ABR interpretation

As reading the ABR results and estimating the different latencies for different waves requires perceptual judgment and decision making, it is likely that it is susceptible to a range of factors that influence human perception and judgment. The ability of one or more readers to produce the same result for the same sample under similar settings defines the reliability of a rating (Downing, 2004; Gajdosik & Bohannon, 1987). Relatively little attention has been paid to measuring or improving the reliability of the ABR.

In the present review, two types of reader reliability are discussed:

- Inter-reader reliability or agreement, the degree to which measurements performed by different readers are similar (Hayen, Dennis, & Finch, 2007).
- Intra-reader reliability or agreement, the degree to which measurements performed by the same observer are consistent over time (Hayen et al., 2007).

An understanding of the extent of variability in ABR interpretation and the factors that influence audiologists' decisions should help to minimize possible errors and improve practice standards. Therefore, this review has two aims. The first aim is to identify studies that have investigated audiologists' variability when reading ABR waves. The second aim is to identify the factors, highlighted through previous literature, which may influence audiologists' performance when reading ABR results.

Search method

Our review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) principles, which describes the minimum set of evidence base for reporting in systematic reviews and meta-analyses (The PRISMA Group, 2009). The databases used in searching data for this narrative review were PubMed, Medline (OvidSP), ScienceDirect and Google scholar. Searching included all languages and all dates until the final date of the search (December 12, 2013). The research team developed a broad list of topics and key words for searching and an experienced librarian assisted in building the search process. Details about the search strategy are available from the authors. Keyword searches for databases were carefully chosen to maximize the potential to identify relevant articles with a slight change in each database. Keywords used for searching included: "auditory brainstem response", "auditory brainstem evoked potentials", "variability", "consistency", "observer variation" and "audiology". All keywords were combined and appropriate subject heading terms were used in searching databases. The first step of the search yielded a total number of 4,735 papers. The authors excluded 470 duplicate references. Non-duplicate items that satisfied the keyword search requirements were then screened by title and subsequently by abstract. 4,101 articles were excluded based on the title alone. For example, studies that are not related to the auditory brain stem response at all and/ or that related to a different purpose for the ABR testing. A further 107 studies were excluded following an analysis of abstracts ("e.g." studies that were not related to reader variability). Six articles remained that related directly to variability among audiologists in interpretation of the ABR results (Fig.1).

Results

Overview of the literature

The search found a limited number of articles that investigated audiologists' variability in interpreting the ABR waves. Conflicting conclusions were drawn regarding audiologists' variability in reading ABR traces (Table 1). Results were analyzed according to three main aspects: inter-reader agreement; intra-reader agreement, and the factors that affect audiologists' variability when reading the ABR waves.

Inter-reader agreement

Inter-reader agreement is the extent to which two or more examiners/ raters give similar scores to an identical observation when using similar rating measurement (Gisev, Bell, & Chen, 2013). High inter-reader agreement is therefore one index of low variability in measurement. While four studies reported that audiologists show variability when reading ABR results (Gans, Del Zotto, & Gans, 1992; Naves, Pereira, Nasuto, Russo, & Andrade, 2012a; 2012b; Vidler & Parker, 2004), two studies reported that audiologists show good agreement when reading ABR waves (Olsen, Pratt, & Bauch, 1997; Pratt, Olsen, & Bauch, 1995). The difference in results may be explained by the different methods adopted to examine audiologists' variability.

Poor inter-reader agreement.

Two studies with the largest samples of audiologists provided strong evidence of variability between audiologists

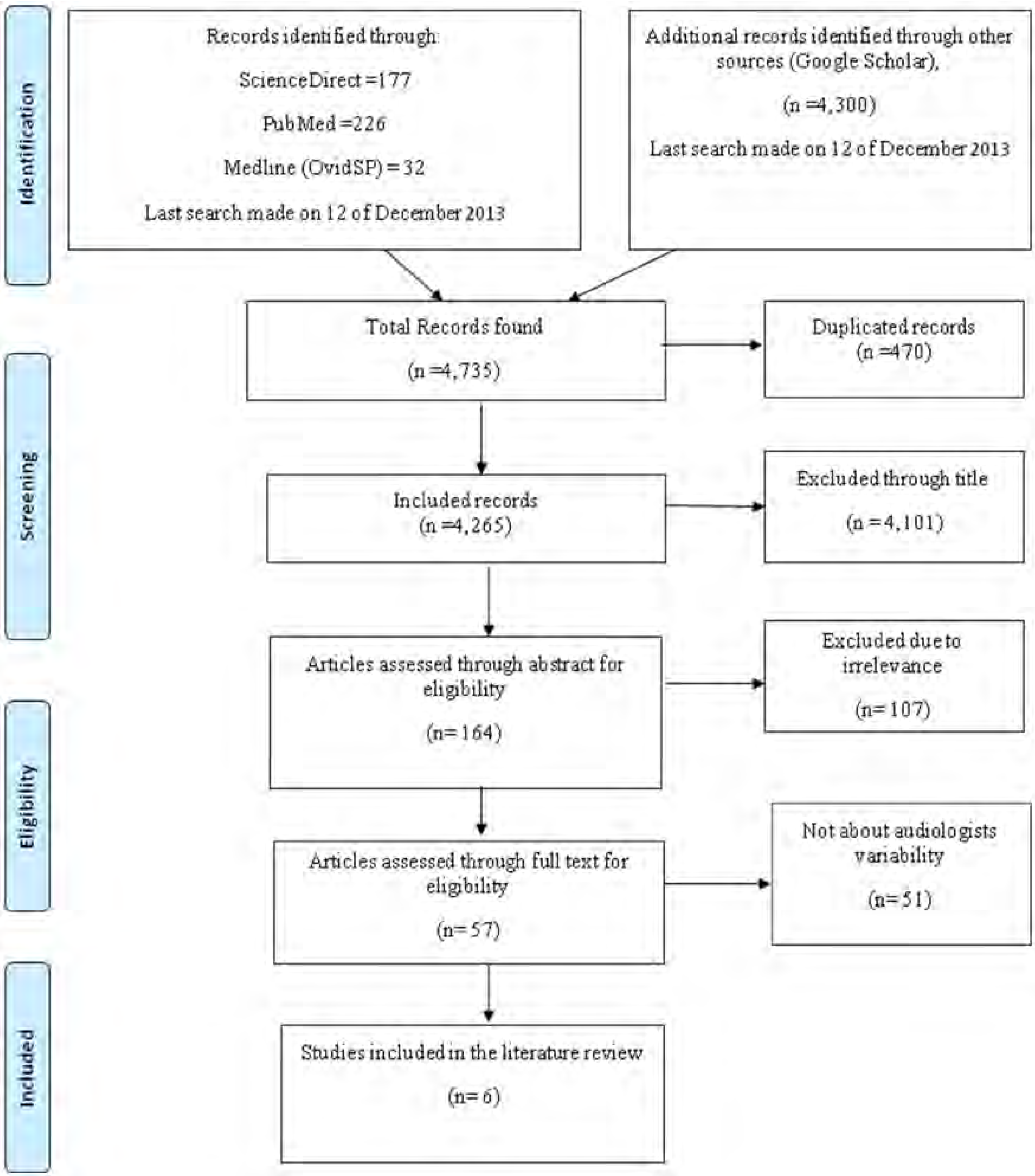

Figure 1. Selection process.

Table 1. Summary of the included studies

The study	No. of audiologists	Experience in reading ABR	Method*	Findings and factors highlighted
Gans, et al., (1992)	9	8 audiologists had experience ranging from (4-40) months with average of 8 months. One audiologist has no experience in reading ABR.	Estimation of ABR hearing threshold for 50 multi handicapped children who were tested behaviorally. Each case was presented to the audiologists twice with one week gap. Either true threshold (TT) or false thresholds (FT) information were provided. Bias occurred if audiologist sets the threshold for the same case differently according to the TT or FT.	Significant variability within and among audiologists. More experienced audiologists were more consistent in estimating ABR waves.
Pratt et al., (1995)	3	Each audiologist had 4 years experience.	Marking of the three main waves: I, III, V and their latencies for a group consisting of 63 ears, the same task was repeated six month later. Comparison for the estimated latencies was performed among the three audiologists.	86-100% consistency between audiologists. Good intra-reader agreement 87- 95% for different montages.
Olsen et al., (1997)	3	Each audiologist had 4 years experience.	Determine the latency and interpret a multi-channel ABR traces for 30 patients. Four months later, audiologists were asked to repeat the same task.	Great consistency among and within audiologists in estimating latency with difference equal or less than 0.2ms in 90% of cases.
Vidler and Parker (2004)	16	Range from 1.5-25 years Average of 8.41 years	Determine hearing threshold for 12 ABR cases designed through computer-simulation. Audiologists had control over data acquisition "e.g." determine the stimulus level and decide when to terminate the averaging of each trace.	No agreement on thresholds estimation for any of the 12 cases across the 16 audiologists. In nine of the cases, the difference between highest and lowest estimated threshold was 40dB or greater and the maximum difference was 60dB.

Naves et al., (2012)	4	Range from 3-11 years Average of 7.25 years.	Interpret ABR data for ten normal hearing adults through estimating the first five waves and their latencies at four intensity levels. Latencies values were compared between each pair of audiologists. If difference exceeded (0.1ms), indication of significant variability.	A difference of 0.1ms or more in estimated latencies between audiologists found in 40% of the cases. Audiologists with different level of experience showed well-matched results and the least experienced audiologist showed the largest discrepancy.
Naves et al., (2012a)	4	Range from (3-11 years). Average of 7.25 years.	Interpret ABR data for ten normal hearing adults through estimating the first five waves and their latencies at four intensity levels. Compared the performance between audiologists.	A difference of equal or more than 0.1ms between audiologists in estimating wave's latencies appeared in 18% of the studied sample.

* All studies used click stimulus to elicit the ABR

(Gans et al., 1992; Vidler & Parker, 2004). Both studies used hearing threshold estimation as the method of choice for assessing variability among audiologists. Audiologists were given a number of ABR cases, they were asked to read the results and estimate a hearing threshold for each case. Variability was assessed through comparing the estimated threshold by each one of the audiologists.

Gans et al. (1992) asked nine audiologists to estimate hearing thresholds for 50 multi-handicapped children and compared the estimated thresholds among audiologists. The authors reported that audiologists' estimates of hearing thresholds varied when different information was provided about the same case, suggesting variability in interpreting the traces (Gans et al., 1992). The second study asked 16 audiologists to estimate the hearing thresholds for 12 computer-simulated cases of ABR results. In this study, audiologists had control over the stimulus levels they wanted to test and the number of traces they needed to make their decision. They found no agreement on threshold estimation for any of the 12 cases across the 16 audiologists, suggesting high levels of variability. Furthermore, for 9 of the 12 cases the difference between maximum and minimum estimated threshold was 40 dB or greater (Vidler & Parker, 2004). Despite the fact that this study tried to provide audiologists with the optimum clinical conditions to perform ABR testing by using computer-simulated cases, high levels of variability were observed among audiologists.

Two additional studies with fewer numbers of audiologists suggested the presence of variability among audiologists (Naves et al., 2012a; 2012b). It should be noted that both studies used the same set of data and method as they were undertaken by the same team of researchers. However, each study reported different aims and the data were analyzed differently. Four audiologists, all whom were experienced in reading ABR (average of 7.25 years), were asked to mark the first five waves and their latencies for ABR results of 10 normal hearing adults. A comparison between the latencies values obtained by each of the four audiologists was made; a difference of equal to or more than 0.1ms among the audiologists in estimating wave latencies was considered as indicative of variability. In both studies, audiologists showed significant level of variability; differences in latency estimation among audiologists were reported in 40% of the cases (Naves et al., 2012b) and variability of equal or more than 0.1ms among audiologists in estimating waves latencies appeared in 18% of the studied sample (Naves et al., 2012a).

Good inter-reader agreement

Two of the six studies reported that audiologists show good agreement when reading ABR results (Olsen et al., 1997; Pratt et al., 1995). The two studies were conducted by the same team of researchers. Variability was investigated through comparing the latencies values obtained by each one of the audiologists for the three main waves I, III and V. A difference of equal or more than 0.2ms among audiologists in estimating wave latencies was considered as indicative of variability. Both of the studies found considerable consistency among audiologists in estimating latencies with at least 90% agreement (Olsen et al., 1997) and 86-100% agreement among readers (Pratt et al., 1995). However, the results of these studies should be interpreted cautiously as the authors themselves served as the only participants. Furthermore, both studies included audiologists who apparently work in very similar environments and with similar levels of experience (4 years) in reading ABR results, which can reduce variability and enhance agreement.

This narrative review presents a limited number of studies that investigated agreement among audiologists who read ABR results but with conflicting findings. Four studies found that audiologists show variability when interpreting the same set of ABR results and two studies, with questionable methods, found good agreement among audiologists.

Intra-reader agreement

Intra-reader agreement can be defined as the degree to which measurements performed by the same observer are consistent over time (Hayen et al., 2007). While all of the six studies examined inter-reader variability in audiologists, only three studies assessed intra-reader variability in audiologists (Gans et al., 1992; Olsen et al., 1997; Pratt et al., 1995). Conflicting results were found. While one study reported that audiologists show a poor level of consistency in reading the same set of ABR results over time (Gans et al., 1992), two other studies indicated that audiologists show a good level of consistency (Olsen et al., 1997; Pratt et al., 1995). One study assessed intra-reader agreement by asking audiologists to estimate the hearing threshold for the same set of ABR data twice over one week. The audiologists were not aware that they were assessing the same sets of data, and different information for each case in relation to the possible behavioural threshold was presented each time. Results showed that audiologists estimated a different hearing threshold for the same cases over time. In addition, giving different clues about the possible hearing thresholds was shown to affect the audiologists' decision (Gans et al., 1992). This suggests that audiologists' interpretation of ABR results was affected by factors beyond the traces itself, such as contextual information.

The other two studies assessed intra-reader agreement by asking audiologists to mark the three main waves with their latencies for the same set of ABR data twice over six months (Pratt et al., 1995) and four months (Olsen et al., 1997). Both of the studies found that audiologists show good consistency in their ABR interpretation over time with 87-95% and more than 90% respectively. However, it is not clear whether audiologists knew that they were examining the same set of ABR data on the second occasion or not. Furthermore, the results could be confounded by the fact that the authors served as the only participants in both the studies.

In summary, only a limited number of studies have investigated the intra-reader agreement of audiologists when reading ABR waves and the available studies suggest conflicting results due to their different approaches.

Factors affecting audiologists' performance in reading ABR results

Only two studies have compared years of experience with audiologists' performance in reading ABR results (Gans et al., 1992; Naves et al., 2012b). One study suggested that audiologists with more experience show greater consistency over time in reading the same set of ABR results (Gans et al., 1992). The other study found that audiologists with different levels of experience in reading ABR results showed well matched results, with the exception that the largest discrepancy was linked to the least experienced audiologist (Naves et al., 2012a). While both of the studies commented on the experience of the audiologists, different levels of experience were adopted by each of the studies. Gans et al. (1992) included nine audiologists, eight of whom had an 8 months average in reading ABR results compared to four audiologists with average of 7.25 years in reading ABR results (Naves et al., 2012a). While both studies concluded that greater experience in reading ABR waves results in more consistency over time and better agreement among audiologists, the low level of experience among participating audiologists in Gans et al. (1992) study limits the generalizability of the conclusion.

The heterogeneity in the methods used to investigate the variability across the present studies makes it difficult to understand if audiologists vary in their assessment when reading ABR results: different tasks were adopted (hearing threshold estimation or wave latency estimation); different status of hearing were examined (normal or hearing impaired); and different numbers of audiologists were included with varied levels of experience in reading ABR waves. The different definitions of variability used in the studies should also be taken into consideration (Naves et al., 2012a, 2012b; Olsen et al., 1997; Pratt et al., 1995), as different values yielded different results. Authors who defined variability as a difference of 0.2ms or more, found less evidence of variability than those who defined it as a difference of 0.1ms or more. At present, the literature shows some disagreement regarding what is an acceptable variation level with 0.1ms or 0.2ms being described as acceptable (Don, 1989; Hood, 1998; Vannier, Adam, & Motsch, 2002). A range of different statistical and descriptive criteria was used to establish reliability, ranging from linear regression (Naves et al., 2012a) to visually comparing the estimated thresholds for all cases among the audiologists (Vidler & Parker, 2004). Gans et al. (1992) compared the judges' accuracy with the level of experience using the Spearman Correlation Coefficient. Naves et al. (2012b) implemented the Bland Altman statistical method to assess inter-examiner agreement and variability in one of their papers and used the linear regression statistical method to compare the performance between each pair of audiologists in the other paper (Naves et al., 2012a).

Evaluation of methodological/ scientific evidence

An evaluation of the available published research was undertaken in three aspects; firstly, the level of evidence against the National Health and Medical Research Council (NHMRC); secondly, the quality of evidence represented by the risk of bias; thirdly, the quantity of evidence represented by the sample size of audiologist participants.

Criteria for quality of evidence suggests that the published literature provides only level III or IV evidence for variability among audiologist in ABR interpretation (National Health and Medical Research Council, 2000). Risk of bias was the second aspect in evaluating the current evidence. For this review, low risk of bias was defined as audiologists not being aware/ informed that they were assessing the same set of ABR data over time. High risk of bias was defined when the participating audiologists were the authors themselves and/ or were aware that they are assessing the same set of data. Some of the studies made this point clear by mentioning that audiologists were naive to the cases that they assessed on a second occasion but in other studies, it was not clear. The third aspect in evaluating the evidence was the sample size. For this review, using a larger number of audiologists was considered to provide better representation of variability among audiologists. None of the included studies discussed the limitation of its sample size (Table 2).

Table 2. Evaluating the evidence.

The study	Level of evidence against NHMRC	Quality of evidence (risk of bias)	Quantity of evidence (participant sample size)
Gans et al., (1992)	III or IV	Low risk of bias	12
Pratt et al., (1995)	III or IV	High risk of bias	3
Olsen et al., (1997)	III or IV	High risk of bias	3
Vidler and Parker (2004)	III or IV	Unclear risk of bias/ not applicable	16
Naves et al., (2012)	III or IV	Unclear risk of bias/ not applicable	4
Naves et al., (2012a)	III or IV	Unclear risk of bias/ not applicable	4

Discussion

Interpretation of ABR results is a highly subjective process. Different interpretations of ABR results may lead to different management plans and patient outcomes. It is important to understand the possible sources of variability and factors that may influence audiologists' decision making when reading ABR waves. This will minimize the possible errors and improve practice standard. This review identified six studies that evaluated audiologists' variability in interpreting ABR results and examined the factors that may affect audiologist's performance. There is relatively little evidence on the variability among audiologists when reading ABR results. Four of the studies suggested that audiologists interpret the same set of ABR results differently; the remaining two studies reported that there is no variation among audiologists when they interpret the same set of ABR results. However, results from these two studies may be subject to a high risk of bias, as the authors were also the only participants in their investigation. Little data on intra-reader variability in audiologists was discussed in the literature and is conflicting. Further investigation is clearly warranted. Only one factor-experience-was identified as impacting on audiologists' variability when reading ABR waves, with audiologists more experienced in reading ABR being less variable. More investigation of the role of experience on audiologists' performance in reading ABR waves is needed. Only a small number of studies investigated audiologists' variability and those studies provide a low level of research evidence. This raises the need for more investigation using better quality evidence.

Conclusion

Overall, the results suggest that there is variability in ABR interpretation both within and among audiologists. Furthermore, level of experience was the only factor highlighted as contributing to audiologists' improved consistency when reading ABR waves. However, due to methodological concerns, variable sample sizes and the use of computer-generated cases, the quality of evidence and the applicability to clinical knowledge is difficult to ascertain.

In light of the importance placed upon the ABR test in determining hearing loss in pediatric clients and its highly subjective nature, it is important to highlight the factors that may affect audiologists' decisions in determining hearing threshold. The authors are currently conducting a study investigating factors that may affect upon ABR interpretation such as, experience and expertise, variability during the day in reading ABR, and psychological factors. The anticipated findings from this project should help policy makers, clinical directors, or educational leaders to determine optimum activities, and this in turn will promote the audiologist's performance in ABR threshold estimation.

References

Ballachanda, B. B., Moushegian, G., & Stillman, R. D. (1992). Adaptation of the auditory brainstem response: Effects of click intensity, polarity, and position. *Journal of the American Acadamy of Audiology, 3*(4), 275-282.

Bogus, J. C. (1996). *The effect of varying stimulus rate and tone burst frequency on the auditory brainstem response*. (doctoral dissertation), University of South Carolina.

Burkard, R. F., Eggermont, J. J., & Don, M. (2007). *Auditory evoked potentials: Basic principles and clinical application*. Philadelphia: Lippincott Williams & Wilkins.

Chandrasekhar, S. S., Brackmann, D. E., & Devgan, K. K. (1995). Utility of auditory brainstem response audiometry in diagnosis of acoustic neuromas. *Otology & Neurotology, 16*(1), 63-67.

Cone-Wesson, B., Dowell, R. C., Tomlin, D., Rance, G., & Ming, W. J. (2002). The auditory steady-state response: Comparisons with the auditory brainstem response. *Journal Am Acad Audiol, 13*(4), 173-187.

Cornacchia, L., Vigliani, E., & Arpini, A. (1982). Comparison between brainstem-evoked response audiometry and behavioral audiometry in 270 Infants and children. *International Journal of Audiology, 21*(4), 359-363. doi: doi:10.3109/00206098209072751.

Don, M. (1989, November). *Quantitative approaches for defining the quality and threshold of auditory brainstem responses*. Paper presented at the Engineering in Medicine and Biology Society, 1989. Images of the Twenty-First Century., Proceedings of the Annual International Conference of the IEEE Engineering in.

Downing, S. M. (2004). Reliability: On the reproducibility of assessment data. *Medical Education, 38*(9), 1006-1012. doi: 10.1111/j.1365-2929.2004.01932.x

Gajdosik, R. L., & Bohannon, R. W. (1987). Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. *Physical Therapy, 67*(12), 1867-1872.

Gans, D., Del Zotto, D., & Gans, K. D. (1992). Bias in scoring auditory brainstem responses. *British Journal of Audiology, 26*(6), 363-368.

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy, 9*(3), 330-338. doi: http://dx.doi.org/10.1016/j.sapharm.2012.04.004.

Gordon, M. L., & Cohen, N. L. (1995). Efficacy of auditory brainstem response as a screening test for small acoustic neuromas. *Otology & Neurotology, 16*(2), 136-139.

Hall, J. W. (2007). *New handbook of auditory evoked responses*: Pearson.

Hayen, A., Dennis, R. J., & Finch, C. F. (2007). Determining the intra- and inter-observer reliability of screening tools used in sports injury research. *Journal of Science and Medicine in Sport, 10*(4), 201-210. doi: 10.1016/j.jsams.2006.09.002.

Hood, L. J., & Berlin, C. I. (1986). *Auditory evoked potentials*. Austin, TX: Pro-Ed.

Hood, L. J. (1998). *Clinical applications of the auditory brainstem response*. San Diego, CA: Singular Publishing Group.

Hyde, M. L., & Blair, R. L. (1981). The auditory brainstem response in neuro-otology: Perspectives and problems. *The Journal of otolaryngology, 10*(2), 117-125.

Jerger, J., Hayes, D., & Jordan, C. (1980). Clinical experience with auditory brainstem response audiometry in pediatric assessment. *Ear and Hearing, 1*(1), 19-25.

National Health and Medical Research Council. (2000). *How to use the evidence: Assessment and application of scientific evidence*. Biotext, Canberra.

Naves, K. F., Pereira, A. A., Nasuto, S. J., Russo, I. P., & Andrade, A. O. (2012a). Analysis of the variability of auditory brainstem response components through linear regression. *Journal of Biomedical Science and Engineering (JBiSE), 5*(9), 517-525.

Naves, K. F., Pereira, A. A., Nasuto, S. J., Russo, I. P., & Andrade, A. O. (2012b). Assessment of inter-examiner agreement and variability in the manual classification of auditory brainstem response. *Biomedical Engineering Online, 11*(1), 86. doi: 10.1186/1475-925x-11-86.

Olsen, W. O., Pratt, T. L., & Bauch, C. D. (1997). Consistency in Latency Measurements and Interpretation of ABR Tracings. *American Journal of Audiology, 6*(1), 57-62.

Pratt, T. L., Olsen, W. O., & Bauch, C. D. (1995). Four-channel ABR recordings: Consistency in Interpretation. *American Journal of Audiology, 4*(2), 47-54.

Roeser, R. J., Valente, M., & Hosford-Dunn, H. (2007). *Audiology: Diagnosis*. New York: Thieme.

Schmidt, R. J., Sataloff, R. T., Newman, J., Spiegel, J. R., & Myers, D. L. (2001). The sensitivity of auditory brainstem response testing for the diagnosis of acoustic neuromas. *Archives of Otolaryngology Head and Neck Surgery, 127*(1), 19-22.

Schmulian, D., & McMahon, C. (2005, Winter). Essays in Audiology: Auditory seady state responses primer. *Audiology Now,* (21), 41-46. Retrived from www.audiology.asn.au.

Stapells, D. R. (2000). *Frequency-specific evoked potential audiometry in infants.* In A sound foundation through early amplification: Proceedings of an international conference (pp. 13-31). Retrived from https://www.phonakpro.com/content/dam/phonak/b2b/Events/conference_proceedings/1st_pediatric_conference_1998/1998proceedings_2.pdf.

Stueve, M. P., & O'Rourke, C. (2003). Estimation of hearing loss in children: Comparison of auditory steady-state response, auditory brainstem response, and behavioral test methods. *American Journal of Audiology, 12*(2), 125-136. doi: 10.1044/1059-0889(2003/020)

The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-Analyses: The PRISMA statement. *Journal of Clinical Epidemiology, 62*(10),1006-1012.doi: 10.1016/j.jclinepi.2009.06.005

Vannier, E., Adam, O., & Motsch, J. F. (2002). Objective detection of brainstem auditory evoked potentials with a priori information from higher presentation levels. *Artificial Intelligences in Medicine, 25*(3), 283-301.

Vidler, M., & Parker, D. (2004). Auditory brainstem response threshold estimation: Subjective threshold estimation by experienced clinicians in a computer simulation of the clinical test. *International Journal of Audiology, 43*(7), 417-429. doi: doi:10.1080/14992020400050053

Weber, B. A. (1983). Pitfalls in auditory brain stem response audiometry. *Ear and Hearing, 4*(4), 179-184.

Zappia, J. J., O'Connor, C. A., Wiet, R. J., & Dinces, E. A. (1997). Rethinking the use of auditory brainstem response in acoustic neuroma screening. *The Laryngoscope, 107*(10), 1388-1392. doi: 10.1097/00005537-199710000-00018

Acknowledgements

The authors are thankful for the scholarship and support provided to the first author by Jordan University of Science and Technology.

Authors' Note

Correspondence concerning this article should be addressed to Maha Zaitoun, Faculty of Health Sciences, The University of Sydney, Lidcombe, NSW, 1825 AUSTRALIA. Email: mzai1104@uni.sydney.edu.au.