

**Critical Review:**  
**Is the *Vector* algorithm superior to Point Optimized Variance Ratio (POVR) method on measures of sensitivity, specificity and test time?**

Lindsay Acton  
M.Cl.Sc (AUD) Candidate  
University of Western Ontario: School of Communication Sciences and Disorders

This critical review evaluates the available literature on two automatic auditory brainstem response screening tools: the current standard of care in Ontario and an alternative method of detection on measures of sensitivity, specificity and test time. Studies reviewed include three within groups repeated measures designs and one ascertainment study. Overall, the literature suggests that all methods reviewed provided relatively high levels of sensitivity and specificity, and test time was significantly decreased for the more recent generations of both methods reviewed. Issues relating to collecting and reporting on sensitivity and specificity data for infant hearing screening programs are also discussed, specifically, the caveats of interpreting reported results of sensitivity and specificity.

### ***Introduction***

With the advent of mass infant hearing screening programs, there is a great motivation to find more cost effective means of detecting hearing loss in the neonatal population. One method of reducing the costs associated with mass hearing screening is by maximizing sensitivity and specificity in the tests we use, while decreasing the required test time. This impetus has led to various methods of classifying screening ABR data based on statistical criterion and algorithms.

### ***Objectives***

The primary objective of this review was to evaluate the literature on the current standard of care in Ontario's Infant Hearing Program (IHP) for automatic auditory brainstem hearing screening technology, Point Optimized Variance Ratio (POVR) and to compare this to an alternative method known as the *Vector* algorithm on measures of sensitivity, specificity and test time.

### ***Methods***

#### Search Strategy

A computerized database search of Scopus, PubMed, and CINAHL was completed using the following search strategy:

(Auditory Brainstem Response) OR (Brainstem Auditory Evoked Potential) OR (ABR) AND (automatic) OR (screening) AND (infant) OR (neonate) AND (Hearing loss) OR (hearing impaired).

Results of this search strategy failed to yield relevant publications and therefore further search strategies were undertaken.

After consultation with faculty, key papers on the topic of newborn hearing screening were identified and led to the identification of other relevant resources through the reference lists as well as online searches of the current ABR screening equipment used by the Ontario IHP.

#### Selection Criterion

Studies included in this critical review were required to be evaluations of the current standard of care in the Ontario IHP for screening ABR as well as evaluations of an alternative screening method. This included Fsp and the more recent alteration of Fsp, Point Optimized Variance Ratio (POVR) as well as studies evaluating another ABR screening algorithm known as the Nottingham ABR screener, and the more recent version of this method, the *Vector* algorithm. Studies included were required to evaluate these screening methods in terms of sensitivity, specificity and test time.

## Data Collection

Studies included in this review consisted of an ascertainment study and three within groups repeated measures studies. One of the studies was an unpublished non-peer reviewed White paper document. It was necessary to include this non-peer reviewed source in order to appropriately review the relevant literature on the current standard of care (POVR).

## **Results**

### ***Fsp and POV***

The current standard of care in the Ontario Infant Hearing program is the POV method, a refinement of the Fsp method first described by Don et al, 1984 and Eberling and Don, 1984. More recently, Norton, S., Gorga, M., Widen, J., Folsom, R., Sininger, Y., Cone-Wesson, B., Vohr, B., Mascher, K., Fletcher, K. (2000) evaluated the performance of the Fsp screening ABR for identifying hearing impairment. They evaluated 4911 “at risk” infants with a 30 dB nHL click stimulus and compared these results with a gold standard of Visual Reinforcement Audiometry (VRA) completed at 8 to 12 months corrected age. The goal of the study was to follow all at risk infants, regardless of their neonatal screening results. Sixty four percent of the at risk target population that was initially screened returned for VRA testing at eight to twelve months of age and 95.6% of this group was able to successfully complete VRA testing. ROC curves were then compiled for the results of this study.

Results of the data from this study indicate that the area under the ROC curve corresponding to the gold standard VRA at 8 to 12 months corrected age was  $A=.87$ . This corresponds to a high hit rate (sensitivity) and a low false positive rate (specificity). The authors comment that the specificity results are likely reliable due to the large sample of infants who were confirmed to have normal hearing upon “gold standard” VRA testing. They do however comment that caution should be exercised when interpreting sensitivity results due to the small number (56) of hearing impaired infants present in their sample. Despite

their efforts to test a group with a higher incidence of hearing loss, the “at risk” group did not yield a significant number of children with hearing loss and therefore the estimates of sensitivity are limited. Test time was also measured and found to be on average approximately 141.41 seconds.

Raviv, G., Sininger, Y., Murphy, K. (2000) studied the addition of Point Optimized Variance Ratio (POVR) to the Fsp algorithm as clinically used in the commercially available ABAer Automated Brainstem Auditory Evoked Response System. This system is currently used in the Ontario IHP. Two hundred and eighty newborns were screened bilaterally with a 35 dB nHL click stimulus. Two conditions were utilized to evaluate the sensitivity and specificity of this screening method. In one condition, the stimulus was presented to the infant with the transducer in place, while recording the ABR. In the second condition, the ABR was recorded, with no stimulus present, to simulate hearing loss.

Results indicate that with a POV critical value of 3.1 to achieve a Pass result, sensitivity is 1.0 and specificity is approximately 0.95. Test time for this method was found to be approximately 65 seconds per ear. This result suggests a decrease in test time by more than half when compared to previous Fsp methods described in the Norton et al (2000) study. This decrease in test time may be attributable to the fact that this algorithm allows most of those with a pass result to be identified and passed quickly while a refer result will continue to allow greater test time, confirming the absence of a response. This time savings when compared to previous methods may decrease the cost of running a mass screening program significantly when considering the number of infants screened annually.

### ***Nottingham ABR Screener & Vector***

Mason, S., Davis, A., Wood, S., Farnsworth, A. (1998) evaluated the field sensitivity and specificity of the Nottingham automatic ABR screener. 6983 at risk neonates were tested with this screening method at 5 centers in the UK. The authors used an ascertainment study of the audiological records from these 5 centers over a

period of 6 years to evaluate the children who were found to have a hearing loss later in life.

Results of this study revealed that 201 children who were born between the dates of the study were found to have bilateral hearing thresholds of at least 50 dB HL. Of these 201 children, 51 had been screened using the automatic ABR with a stimulus of 50 dB nHL. An examination of the data from these children's screening results revealed that sensitivity was .90 and specificity was found to be .82. The authors noted that due to the inclusion of suspected progressive hearing losses in their results, specificity estimates are likely lower than they would be had these data points been left out. Test time was reportedly approximately 20 minutes for both ears including set up time.

The method of using an ascertainment study is an improvement over other methods of estimating sensitivity and specificity in that it provides a more applicable estimation based on a population of interest, "at risk infants". However, there are some problems with this type of research. Although this study is considered an ascertainment study, it only followed those children identified as "at risk". This is some improvement over some simulation methods of estimating real world test performance, however, since only the "at risk" infants were followed it is not possible to extend the results to the broader population of all infants undergoing hearing screening.

Keohane, B., Mason, S., and Baguley, D. (2004) evaluated the *Vector* algorithm for use in an infant hearing screening program according to sensitivity, specificity as well as test time. The *Vector* algorithm is a reconfiguration of the Nottingham screener method tested in the large multi center study by Mason et al. (1998) described above. The original Nottingham screener made use of cross correlation and variance analysis of two averaged waveforms to detect auditory brainstem response. The screening method described in this paper also employs cross correlation and variance analysis, but with the added step of a comparison with a vector template. This template is required to confirm the presence of a response and therefore decreases test time significantly for infants with normal hearing.

Sixty infants were screened with the *Vector* algorithm, under two conditions to evaluate test sensitivity and specificity of this method. Infants were tested normally, as well as in a simulated hearing loss condition. Hearing loss was simulated by removing the transducer from the infant's ear during testing. One non-stimulated (transducer removed) and at least two stimulated (normal condition with transducer present, clicks at 35 dB nHL) tests were completed for each infant tested. This resulted in 464 tests, 120 non-stimulated and 344 stimulated conducted on 120 ears. This design allowed for the test order to be altered for each consecutive infant. The study was also conducted blind. The manufactures and the clinical authors were not present during any of the testing to ensure independence of data collection

Results indicate that test sensitivity was greater than .99 and specificity was .87. Test times were found to be 1.5 minutes when an ABR is present and 4.5 minutes in the absence of an ABR. This greater test time when the ABR is absent is expected as the priority in this method is for unequivocal detection of an ABR response rather than verifying the absence of a response (Keohane et al, 2004). If it is unclear during data collection whether the response is present, the data collection continues.

### *Discussion*

Studies evaluating the POVR and the *Vector* algorithm have provided evidence that both of these tools provide relatively high levels of sensitivity and specificity, while improving on their previous generations' test times. It is important however to outline and comment on some of the caveats of the research that is currently available to support these methods.

The first important factor to consider is that it is difficult to obtain comparable estimates of sensitivity and specificity. ABR screening technologies are configured based on several test criteria that are selected by researchers and clinicians and have an effect on the resulting sensitivity and specificity of the devices. A very simple example of this is test stimuli. Some studies used a test stimulus of 30 dB nHL, whereas others used a stimulus of 50 dB nHL. This variance

will lead to a great difference in the number of infants identified as “refer” based on these arbitrary numbers. The screening tools that use a stimuli of 30 dB nHL, will likely have a greater sensitivity value, however may also report lower specificity when compared to the 50 dB nHL test. This factor and the many other criteria that may be altered in various screening methods have an unpredictable effect on sensitivity and specificity data and make it difficult to compare the studies to one another.

Also, a true measure of sensitivity may only be achieved if all children who are screened in a program are followed up with gold standard testing. There are two problems that arise with this requirement. First, this would mean screening and following up with a large number of babies to achieve precise estimates of sensitivity with small confidence intervals (Hyde, 2009, personal communication). With estimates of the number of babies with hearing loss between 1 and 2 per 1000 births, this would mean screening and following up with a prohibitive number of infants.

Also, gold standard methods of identifying hearing loss may not be performed until months after the initial screening, and this complicates matters further. Late-onset or progressive hearing losses become confounding factors and may cloud the true sensitivity of a screening test. Thus, it appears that although the studies critically evaluated in this review paper may inform us that they are all “good” tests, it is not possible to perform a solid comparison between the resulting measures of sensitivity and specificity currently obtained (Hyde, 2009). Based on the problems of the methods used to identify sensitivity and specificity estimates, it is not currently possible to identify one test as superior to another.

It is however, possible to comment about the various methods’ test time. Both the original Fsp and Nottingham hearing screener provided estimates of test time that were significantly longer when compared to both of their more recent counterparts, POVR and *Vector*. The POVR method was the shortest with estimates of approximately 65 seconds per ear on average, (including refer results which were slightly longer and pass results which were slightly shorter) and

the *Vector* algorithm reported a test time of 90 seconds per ear when a response is found and 4.5 minutes when no response is found. In this respect, results suggest that the POVR method requires significantly shorter test time than the *Vector* algorithm. However, it is important to reiterate that due to the current methods of studying sensitivity and specificity and the differences in the way it is measured, it is not possible to make strong statements about which test is superior.

### ***Conclusion and Clinical Implications***

Overall, all of the methods critically evaluated in the present paper provide relatively high estimates of sensitivity and specificity. However, due to limitations in interpreting and comparing sensitivity and specificity data for reasons of test configuration, small sample size and confounding factors such as progressive hearing loss, it is not possible to make any definitive statements about which method is superior in terms of sensitivity and specificity. The current standard of care in Ontario, POVR appears to provide excellent sensitivity and specificity values as well as significantly shorter test time when compared to other tests. Thus, the evidence reviewed supports the continued use of the POVR in the Ontario IHP. The question of selecting an efficient and valid test is still of paramount importance and more research will likely continue in this area.

Clinical implications for the validation of screening tools that increase sensitivity and specificity while decreasing test time have great importance for the provision of screening programs such as mass infant hearing programs. Savings for the health care system in terms of time and resources is of great interest to address budget concerns of such mass screening programs. By decreasing the cost of these programs, this may allow infant hearing screening programs to become more widespread across the world and may allow more resources to be spent in other aspects of the management of those identified with hearing loss.

### ***References***

Keohane, B., Mason, S., Baguley, D. (2004). Clinical evaluation of the vector algorithm for neonatal hearing screening using automated

auditory brainstem response. *The Journal of Laryngology and Otology*. 118, 112-116.

Don, M., Elberling, C., Waring, M. (1984). Objective detection of averaged auditory brainstem responses. *Scand Audiol*. 13 (4), 219-228.

Elberling, C., Don, M. (1984). Quality estimation of averaged auditory brainstem responses. *Scand Audiol*. 13 (3) 187-197.

Norton, S., Gorga, M., Widen, J., Folsom, R., Sininger, Y., Cone-Wesson, B., Vohr, B., Mascher, K., Fletcher, K. (2000). Identification of neonatal hearing impairment: evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem

response test performance. *Ear and Hearing*. 21(5). 508-528.

Mason, S., Davis, A., Wood, S., Farnsworth, A. (1998). Field sensitivity of targeted neonatal hearing screening using the Nottingham ABR screener. *Ear and Hearing*. 19(2). 91-102.

Raviv, G., Sininger, Y., Murphy, K. (2000). ABAer automated brainstem auditory evoked response system an innovative approach to infant hearing screening. (White Paper)