Numerical Assessment of TOAE Screening Results: Currently Used Criteria and their Effect on TOAE Prevalence Figures

J. J. J. DIRCKX, K. DAEMERS, TH. SOMERS, F. E. OFFECIERS and P. J. GOVAERTS

From the Department of Otorhinolaryngology, St. Augustinus Hospital, University of Antwerp, Belgium

Dirckx JJJ, Daemers K, Somers Th, Offeciers FE, Govaerts PJ. *Numerical assessment of TOAE screening results: currently used criteria and their effect on TOAE prevalence figures.* Acta Otolaryngol (Stockh) 1996; 116: 672–679.

The literature on neonatal hearing screening by means of oto-acoustic emissions (OAE's) presents various prevalence figures, and gives little quantitative information on the procedure used to score the recordings. If the OAE test is to be interpreted by users who do not have the opportunity to develop intuitive interpretation skills through extensive training, a clear numerical decision criterion is needed. The present paper discusses the scoring procedure used by 25 teams, which together screen 22,356 neonates annually. More than 60% of the groups involved in this study use visual interpretation of the recorded OAE response, together with numerical criteria. Amongst the teams, 21 different ways of numerical scoring are used. It is shown that for a given set of OAE recordings, prevalence varies from 61% to 90%, depending on the numerical decision criterion being applied. We conclude that at this moment no consensus exists regarding the numerical criterion to be used when assessing OAE screening results. In view of the strong effect of criteria on the outcome of OAE screening, such consensus is urgently needed, but should be based on sensitivity and specificity figures for each scoring technique. *Key words: oto-acoustic emissions, specificity, neonatal.*

INTRODUCTION

Following Kemp's first publication (1) on the recording of transient oto-acoustic emissions (TOAE's), a number of authors published figures of TOAE prevalence. Some authors agreed on an extremely high prevalence rate, up to 100% (e.g. 1–5). Others found figures which were a bit less optimistic (e.g. Stevens & Ip (6): 97%, Rutten (7): 90%, and Grandori (8): 90%), and some workers even encountered quite a low number of ears in which emissions could be recorded (e.g. Zwicker (9): 70%, and van Dijck & Wit (10): 40%). In these first years of OAE research, many different recording techniques were used and the recording apparatus was not standardised, which can be one reason for the large range of prevalence figures reported. Measurements were also done in different populations, and were not focused on neonatal groups. The number of measured ears was quite small, usually less than 50 (except Bonfils et al. (5): 131, and van Dijck & Wit (10): 210).

Following the pioneering work on OAE's, figures became available on OAE recordings in large (> = 100) neonatal populations. For example, Elberling et al. (11) and Johnsen et al. (12) reported a prevalence of 100% in groups of respectively 100 and 200 neonates, Bonfils et al. (13) found 98% prevalence in a group of 100 neonates. More recently, Kok et al. (14) measured a prevalence rate of 93.4% in a group of 1,036 healthy new-borns, and Stevens et al. (15) measured a prevalence of 95.9% in a population of 1,367 babies. Both Kok et al. and Stevens et al. based their OAE scoring on "visual interpretation" of the response.

Many reports are the result of a research project, and figures have therefore not been obtained in daily clinical practice where recording time is limited and a noise free environment is not always available. Such factors will inevitably have a negative result on OAE screening outcome.

In a recent 5 year field study, Meredith et al. (16) found a prevalence figure of only 72.3% in a population of 772 babies. The screen was targeted on neonates with one or more risk factors. Follow-up confirmed hearing loss in 2.2% of the cases, so the OAE screening gave a falls positive result in 25.5% of the cases. The decision criterion for OAE presence was described by the authors as: "correlations . . . along with signal to noise measure and subjective assessment of the waveforms", without any numerical values being specified.

At this moment it is unclear whether any consensus exists about the criterion to be used by the clinician to decide upon TOAE pass or fail. Salomon et al. (17) have shown, for linear clicks recorded in 378 neonatal ears, that a correlation of more than 0.7 is present in only 80.4% of the cases, while correlation of more than 0.5 is present in 90.7% of the cases. Changing the criterion for required correlation to give a pass on OAE testing from 0.5 to 0.7 will therefore change the prevalence by 10%. The lack of information about the criterion used makes it very difficult to compare prevalence figures obtained in different studies, as long as a standardised decision criterion is not generally accepted.

It is not possible for all private practitioners or small screening teams to gain experience in "visual

interpretation" of OAE curves. For these users, who will become a fast growing group as price of equipment goes down, a clear and well defined numerical criterion is needed. If one wants to come to automated decision making (as implemented in some of the newest commercial apparatuses), or if one wants the screening to be done by non-specialised personnel, numerical scores will have to be used rather than intuitive decision making (based on visual inspection of curves) which can only be obtained after extensive training. As pointed out by Kemp & Ryan (18), good automatic scoring should be based on multiparameter analysis.

In the present paper we will address the question whether at this moment a consensus exists amongst screening teams concerning the numerical decision criterion to be used in the scoring of OAE screening results. We will give an overview of decision criteria currently used by a number of screening teams, and we will show the effect on prevalence figures when these different criteria are applied to a given set of OAE measurements obtained in daily clinical practice.

MATERIAL AND METHODS

In November 95, a questionnaire was sent to 95 groups who appeared on the mailing list of the European Concerted Action on Otoacoustic Emissions. The questionnaire was accompanied by a letter, explaining that we wanted to obtain an overview of the criteria which are currently in use, and proposing that all participants in the enquiry would obtain an overview of the results. The questionnaire contained the following questions:

- How many patients do you test annually for OAE's (adults, babies)?
- Which OAE system do you use?
- Which number of averagings is applied?
- Which kind of clicks do you use (linear, non-linear, both)
- Does the baby pass only if both linear and non-linear clicks give an OAE response, or it is sufficient that the linear clicks give an OAE?
- Do you use visual interpretation of the response curves? If so, how is this done?
- Which criteria do you use for the numerical scores 'global reproducibility, reproducibility in several bands, global S/N ratio, S/N ratio in several bands, correlation'?
- Specify any other methods used than those mentioned?

The questions were ordered in a table on one A4 page, using boxes to indicate the answers, and using multiple choice indicating boxes where possible. The page contained return address and fax number. We

think it is a fair guess to say that filling out the form took no more than 10 min.

Of the 95 questionnaires sent, 3 were returned to sender because the address in the mailing list was wrong. Of the other 92, 40 groups sent the answering page back by mail or fax. We waited 4 months to obtain the last answer (up till now).

As the letters were directed towards groups which are active in the field of OAE screening, we did not explain the concepts such as "reproducibility" or "linear clicks", as this is common knowledge for those involved in OAE work. For the not specialised reader, we will shortly explain the terms used.

To eliminate stimulus and middle ear artefact from the ear canal response, two responses following clicks of the same polarity and amplitude are added to the response following a click with reversed polarity and double amplitude. All linear components in the response will disappear in this sum, as they will be twice as large and of inverse sign for the third click. Due to the saturating nature of the OAE response, the non-linear residue will remain. This process is referred to as non-linear click OAE testing. It has the advantage that linear artefacts in the response cannot be mistaken for OAE's. The subtraction process, however, also eliminates part of the OAE response itself, leading to a smaller signal to noise ratio. Therefore, if a non-linear OAE response is found, a linear response is also due to be present.

To improve the signal to noise ratio in the final response, a number of responses is averaged (usually between 25 and 1,000). A standard procedure to differentiate between OAE response and noise uses the acquisition of two sub averages. In one buffer, the average is made of the responses following the even numbered clicks, and in a second buffer the average is made of the responses following the odd numbered clicks. (In the non-linear measurement mode, these responses are in fact the result of three clicks. The sum of the three measurements is regarded as one response, which is added to one of the average buffers.) At the end of the measurement the content of the two buffers is displayed on the same graph. If an OAE signal is present, the two curves will overlap well. If the signal only contains noise there will be no systematic overlap. Interpreting this quality of overlapping is referred to as visual interpretation of the OAE response.

To eliminate noise in the measurement, the user will set a rejection level. Responses with a higher amplitude than this level are rejected, and will not be added to the averaged response, as they are regarded as artefacts. A lower rejection level will decrease the noise in the measurement, but will increase measurement time.
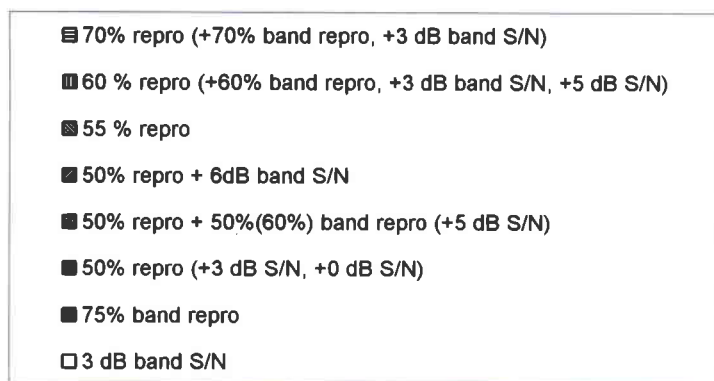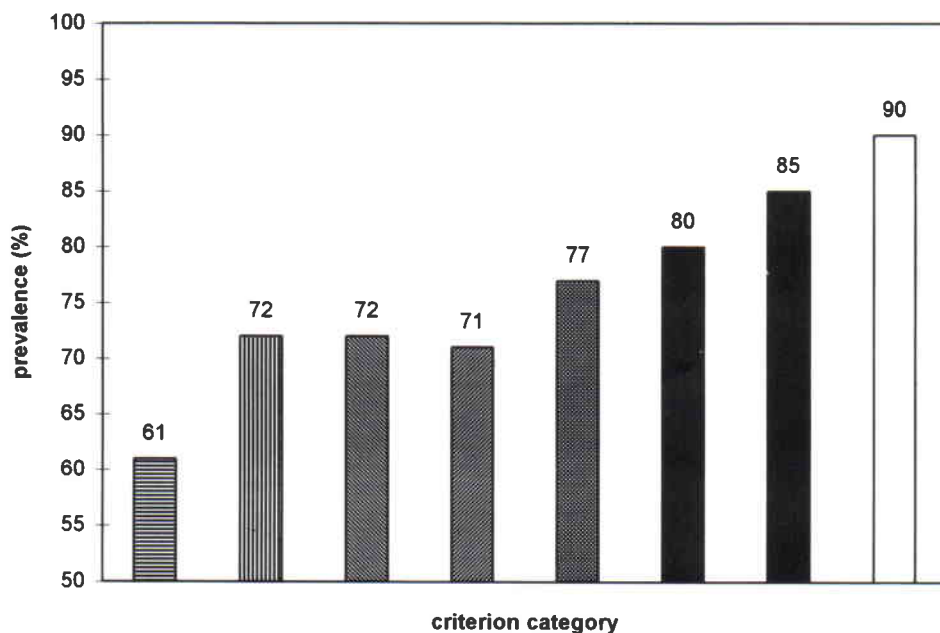
Fig. 1. prevalence rates obtained by applying different categories of numerical decision criteria to a given set of OAE results. The composition of the categories, indicated shortly in the legend, is explained further in paragraph four.

for 70% reproducibility, and then fail for the band criteria which are added by some groups.

The criterion of 60% global reproducibility, be it or not in conjunction with other global or frequency band scores, delivers a prevalence figure of 72%. This scoring is used by 23.8% of the groups (8.5% babies). The criterion of 55% reproducibility leads to the same prevalence figure, and is used by one group or 4.8% (0.5% babies). In total, 28.6% of the groups (9% babies) use 60% or 55% global reproducibility.

When using 50% global reproducibility, together with a S/N ratio of 6 dB in at least 3 frequency bands, we obtained a prevalence figure of 71%. One group, or 4.8%, (1.5% babies) uses this decision score.

Fifty percent global reproducibility, together with 50% or 60% reproducibility in at least 3 frequency bands, leads to 77% prevalence. This procedure is used by two groups. One more group adds global

S/N ratio of 5 dB and S/N ratio of 3 dB in at least 3 frequency bands. This criterion leads to the same prevalence figure. Taken as a whole, we see that 50% global reproducibility, together with 60% (or 50%) reproducibility in three bands, gives a prevalence of 77% and is used by 14.3% of the groups (7.4% babies).

Using 50% global reproducibility on its own leads to a prevalence figure of 80%. As stated above, 5 groups share this way of working. Combining 50% global reproducibility with 3 dB global S/N ratio or with 0 dB S/N ratio, each used by one group, does not alter the obtained prevalence figure. In total, therefore, 28.5% of the groups (13.2% babies) base their decision on the 50% global reproducibility criterion.

Finally, there were two groups which did not apply global criteria. One group, the one which screens nearly half of the babies involved in this study (50.7%

effect of scoring procedure on sensitivity will only become clear when for each scoring procedure figures become available of follow up of large numbers of babies who passed the OAE screen.

If OAE testing is to be performed by untrained personnel, a clear and simple, numerically based assessment of the measurement result is needed. At this moment, a consensus on the criterion to use does not exist. With the proliferation of the OAE test, clear guidelines on the scoring procedure to be used is urgently needed. This consensus has to be based on figures of specificity and sensitivity for each procedure currently in use.

ACKNOWLEDGEMENT

REFERENCES

1. Kemp DT. Stimulated acoustic emissions from within the human auditory system. J Acoust Soc Am 1978; 64: 1386–91.
2. Wit HP, Ritsma RJ. Stimulated acoustic emissions from the human ear. J Acoust Soc Am 1979; 66: 911–3.
3. Johnsen NJ, Elberling C. Evoked acoustic emissions from the human ear I. Equipment and response parameters. Scand Audiol 1982; 11: 3–12.
4. Probst R, Coats AC, Martin GK, Lonsbury-Martin BL. Spontaneous, click- and toneburst-evoked otoacoustic emissions from normal ears. Hear Res 1986; 21: 261–75.
5. Bonfils P, Bertrand Y, Uziel A. Evoked otoacoustic emissions: normative data and presbycusis. Audiology 1988; 27: 27–35.
6. Stevens JC, Ip CB. Click-evoked otoacoustic emissions in normal and hearing impaired adults. Br J Audiol 1988; 22: 45–9.
7. Rutten WLC. Evoked acoustic emissions from within the normal and abnormal human ears: comparison with audiometric and electrocochleographic findings. Hear Res 1980; 2: 263–71.
8. Grandori F. Evoked otoacoustic emissions stimulus-response relationship. Revue Laryngol (Bordeaux) 1983; 104: 153–5.
9. Zwicker E. Delayed evoked otoacoustic emissions and their suppression by Gaussian-shaped pressure impulses. Hear Res 1983; 2: 359–71.
10. Dijk P Van, Wit HP The occurrence of click-evoked otoacoustic emissions in normal-hearing ears. Scand Audiol 1987; 16: 62–4.
11. Elberling C, Parbo J, Johnsen NJ. Evoked acoustic emission: clinical application. Acta Otolaryngol (Stockh) 1985; Suppl 421: 77–85.
12. Johnsen NJ, Bagi P, Parbo J, Elberling C. Evoked acoustic emissions from the human ear IV. Final results in 100 neonates. Scand Audiol 1988; 17: 27–34.
13. Bonfils P, Dumont A, Marie P, Francois M, Narcy P. Evoked otoacoustic emissions in newborn hearing screening. Laryngoscope 1990; 100: 186–9.
14. Kok MR, Van Zanten GA, Brocaar MP, Wallenburg HCS. Click-evoked oto-acoustic emissions in 1036 ears of heathy newborns. Audiology 1993; 32: 213–24.
15. Stevens JC, Webb HD, Hutchinson JT, Buffin JT, Smith MF. Click evoked otoacoustic emissions in neonatal screening. A preliminary analysis of an 8 year study. Advances in Otoacoustic Emissions 1994; 1: 124–30.
16. Meredith R, Stephens D, Hogan S, Cartlidge P, Drayton M. Screening for hearing loss in an at-risk neonatal poulation using evoked otoacoustic emissions. Scand Audiol 1994; 187–93.
17. Salomon G, Anthonissen B, Groth J, et al. Pilot study preceding a systematic screening of newborns based on evoked otoacoustic emissions. Advances in Otoacoustic Emissions 1994; 1: 131–40.
18. Kemp D, Ryan S. Otoacoustic emission tests in neonatal screening programmes. Acta Otolaryngol (Stockh); Supp 482: 73–84.
19. Don M, Elberling C, Warring M. Objective detection of average auditory brainstem responses. Scand Audiol 1984; 13: 219–28.
20. Kemp D, Siobhan R, Bray P. A guide to the effective use of otoacoustic emissions. Ear Hear 1990; 11: 2, 93–105.

Address for correspondence:
J. Dirckx, PhD
University Department of Otorhinolaryngology
St. Augustinus Hospital
Oosterveldlaan 24
B-2610 Antwerp-Wilrijk
Belgium
Fax: +32 3 2180318