ORIGINAL ARTICLE

Agreement between hearing thresholds measured in non-soundproof work environments and a soundproof booth

T W Wong, T S Yu, W Q Chen, Y L Chiu, C N Wong, A H S Wong

Occup Environ Med 2003;**60**:667–671

See end of article for authors' affiliations

Correspondence to:
Professor Tze Wai Wong,
Department of Community
and Family Medicine, The
Chinese University of Hong
Kong, 4/F, School of
Public Health, Prince of
Wales Hospital, Shatin,
NT, Hong Kong;
tw Wong@cuhk.edu.hk

Accepted 12 September
2002

Aims: To study the agreement between audiometric test results measured in non-soundproof environments at the worksite, and in a soundproof booth.

Methods: In a cross sectional prevalence study on noise induced hearing loss, 885 transport workers whose hearing thresholds were measured by a standard audiometric test method in non-soundproof environments at the worksite were identified to have some hearing loss (>25 dB), and were retested in a soundproof booth.

Results: At 4–8 KHz, the mean of the absolute differences in hearing threshold obtained by these two methods was 2 dB or less. When the proportions of hearing loss (≥ 30 dB for any frequencies at 3–8 KHz, or ≥ 90 dB for three low frequencies at 0.5–2 KHz, or ≥ 90 dB for three high frequencies at 3–6 KHz) were compared, considerable differences existed. A much better agreement was obtained when the criteria for hearing loss as measured in the field test under non-soundproof conditions were relaxed by 5 dB. At 4 KHz, the difference between the proportion of subjects with hearing loss as measured in the field and that as measured in the booth was the smallest. The kappa statistic was highest at 3 and 4 KHz.

Conclusions: Audiometric test results conducted in non-soundproof environments in the field are comparable to those obtained in a soundproof environment among transport workers with a hearing loss of >25 dB. The hearing threshold at 4 KHz appears suitable for the estimation of the prevalence of hearing loss when appropriate adjustments are made in the diagnostic criteria.

Sound is a form of energy generated by vibration. It can be characterised by its frequency, measured in Hertz (Hz), which represents the number of vibration cycles per second, and its intensity, a measure of energy level, expressed in watts per square metre (W/m^2). Sound intensity level is measured by a logarithmic scale, in decibels (dB), because the human ear can detect a wide range of intensities. A reference intensity of 10^{-12} W/m^2 , corresponding to the human hearing threshold, is arbitrarily set as 0 dB. Noise, defined as unwanted sound, is one of the most common occupational and environmental hazards. Prolonged exposure to excessive noise causes a sensorineural hearing deficit that begins at the higher frequencies (3–6 KHz). This deficit is commonly described as “noise induced hearing loss” (NIHL).

It has been shown that once exposure to damaging noise levels is discontinued, further significant progression of hearing loss will stop.¹ This implies that the early detection of NIHL through audiometry among high risk workers is useful in the prevention of further hearing losses. Periodic screening for hearing impairment among the workers exposed to excessive sound levels is therefore an important component of hearing conservation. Ideally, audiometric measurements are made in a soundproof environment at different frequencies and intensities to detect the hearing threshold of the subject at the respective frequencies. However, this environment is not always available in field surveys, especially when conducted at the worksite. Mobile soundproof booths are expensive and there may be problems of access. Hence, when audiometric screening is conducted at the worksite, it is often performed in a quiet (but non-soundproof) room where the sound level might be above the maximum limit recommended by the American National Standards Institute (ANSI).² The background noise level in the test environment affects the worker's hearing threshold. However, differences in the hearing

threshold obtained in field screening and in a soundproof booth have not been addressed in previous studies. To assess the agreement of results obtained from the two test environments, we compared the hearing thresholds of 885 workers who were detected to have some hearing loss (>25 dB in any of the frequencies tested) in an audiometric screening test and who underwent a subsequent test in a soundproof booth, as part of a prevalence study of NIHL in the transport industry in Hong Kong. Information on the mean absolute differences between the hearing thresholds measured under the two conditions is of clinical relevance. In addition, we also compared the proportions of hearing loss at different frequencies among the test subjects measured in the field with those obtained in the soundproof booth. This information is useful in epidemiological studies of the prevalence of NIHL.

SUBJECTS AND METHODS

Subjects

In a prevalence study of NIHL, 5590 transport workers were recruited from all major transport companies in Hong Kong.³ These included railway workers, underground railway workers, bus drivers, crew members of a ferry company, air cargo workers, and other staff from the airport. The response rates of the prevalence study varied between companies, ranging from 70% to 90%. After a screening audiometric test conducted in the worksites, 2558 workers (45.8%), who had a hearing threshold higher than 25 dB in any frequencies tested in either ear, were invited to undergo a diagnostic audiometric test

Abbreviations: ANSI, American National Standards Institute; NIHL, noise induced hearing loss

inside a standard soundproof audiometric booth. Of these, 885 responded (response rate 35%) and their test results in both settings were compared.

Audiometric tests

Two audiometric examinations were conducted successively in this study—a screening test, and a diagnostic test conducted one day to several weeks later. All subjects were asked to avoid exposure to noise for 12 hours or longer prior to the examinations. The tests were conducted by trained technicians; pure tone air conduction hearing thresholds were measured following a standard procedure in accordance with British Standards Institution (BSI) standards.⁴

Screening test

As the first part of the prevalence study on NIHL, workers underwent audiometric screening tests for hearing loss at the worksites, in quiet rooms provided by the companies concerned. The background noise level was monitored using a precision sound level meter (Rion NL14, Japan) with a type NX-05 octave band filter unit. For all subjects, pure tone air conduction hearing thresholds at different frequencies were measured manually by the abridged ascending method following the procedure described above. The hearing thresholds were measured at the octave frequencies 0.5, 1, 2, 3, 4, 6, and 8 KHz for both ears. A portable audiometer (Interacoustics AD 25 or AD 27, Denmark) with earphone (TDH-39P, Denmark) was used to perform the screening audiometric test. The earphone was equipped with an audiocup (Amplivox, UK) to enable testing in non-soundproof environments. Those who were found to have a hearing loss of more than 25 dB in either ear at any of the frequencies tested were requested to attend a diagnostic pure tone audiometric test in a soundproof booth. This level was arbitrarily chosen as a clinically significant level of hearing loss. Workers with “normal hearing” (hearing threshold at 25 dB or less at any one of the frequencies tested in any ear) were not called back for further tests.

Diagnostic audiometric test

The diagnostic audiometric test included the measurements of pure tone air conduction hearing thresholds at the octave frequencies 0.5, 1, 2, 3, 4, 6, and 8 KHz in both ears. It was performed inside a soundproof audiometric booth (NAP Acoustic Silentflo Room, Australia) using a Madsen audiometer (Orbiter 922, Denmark). The background sound level inside the booth was also measured and conformed with the BSI standard.⁴

Criteria for the diagnosis of hearing loss

We evaluated hearing loss at each of the high frequencies (3, 4, 6, and 8 KHz) and the sum of hearing losses (in dB) at three low (0.5, 1, and 2 KHz) and three high (3, 4, and 6 KHz) frequencies. In terms of hearing threshold, there are no universally accepted criteria for NIHL, other than a positive history of noise exposure and hearing loss at high frequencies (typically with a “4 KHz dip” in the audiogram). Although low frequency hearing loss is not a characteristic of early NIHL, low frequency background noise in the environments where the screening was performed often exceeded the recommended maximum levels by ANSI² and might have affected the hearing thresholds at lower frequencies more than at higher ones.

Criteria of hearing loss used in the diagnostic test

- (1) Low frequency hearing loss: When the sum of the hearing threshold at three low frequencies (0.5, 1, and 2 KHz) was ≥ 90 dB in any ear.
- (2) High frequency hearing loss: When the sum of the hearing threshold at three high frequencies (3, 4, and 6 KHz) was ≥ 90 dB in any ear.

- (3) Hearing loss at an individual high frequency (3, 4, 6, and 8 KHz): When the hearing threshold was ≥ 30 dB at the respective frequency in any ear.

Criteria of hearing loss used in the screening test

We arbitrarily used three different sets of criteria for the definition of hearing loss at various frequencies in the screening test. In criterion 1, we used the same criteria as in the diagnostic examination (≥ 30 dB for individual high frequencies and ≥ 90 dB for the sum of the hearing threshold at the three low/high frequencies, as described above). Criteria 2 and 3 were obtained by raising the hearing threshold in the screening test in two steps. Each step increased the threshold by 5 dB for the individual frequencies and 15 dB for the sum of the hearing threshold at the three low and three high frequencies, to compensate for the effects of background noise on hearing threshold. In criterion 2, we used ≥ 35 dB for the individual high frequencies and ≥ 105 dB for the sum of the hearing thresholds at the three low/high frequencies as the criteria for hearing loss. For criterion 3, we used ≥ 40 dB for the individual high frequencies and ≥ 120 dB for the sum of the hearing thresholds at the low/high frequencies. The proportions of hearing loss obtained from the three criteria used in the screening tests were then compared with those based on the criteria used in the diagnostic test. The purpose of these comparisons was to identify the screening test criteria which produced results that agreed best with those of the diagnostic test.

Data analysis

The absolute difference (in dB) in the hearing thresholds for each frequency tested, obtained from the screening and diagnostic tests, was calculated for every subject. The means and standard deviations of these differences were then computed for all frequencies. The proportions of hearing loss among the subjects were calculated using the criteria described earlier. The differences in these proportions between screening and diagnostic examinations were then compared using the three sets of criteria for the screening test results. The agreement on the screening and diagnostic test was assessed with Cohen's κ .⁵ The latter was computed as outlined by Fleiss.⁶ The interpretation of κ is as follows:

$\kappa = 1$ if there is complete agreement.

$\kappa \geq 0$ if the observed agreement is greater than or equal to chance agreement.

$\kappa \leq 0$ if the observed agreement is less than or equal to chance agreement.

As a general rule, when $\kappa > 0.8$, agreement is considered to be high. Agreement is “good” when $\kappa = 0.6$ –0.8, “fair” when $\kappa = 0.4$ –0.6, and “poor” when $\kappa < 0.4$.

RESULTS

Comparison with non-respondents

A total of 885 (35%) of 2558 subjects responded to our request for a diagnostic test in the booth. To assess the extent of selection bias among the respondents, we compared the screening test results of those who responded with those who did not. There was little difference in the gender, marital status, and education level between the two groups. The mean age among the respondents was slightly (and significantly) higher than in non-respondents (45.3 years *v* 43.4 years), as was the mean duration of work at the current job (14.7 years *v* 12.0 years). The mean hearing thresholds among the 1673 non-respondents were slightly lower (by less than 2 dB at 0.5, 1, 2, 3, and 4 KHz, and less than 4 dB at 6 and 8 KHz) than among the 885 respondents.

Table 1 Sociodemographic characteristics of the 885 subjects

Sociodemographic characteristics	No. of subjects (%)
Age (y)	
<36	124 (14.0)
36–45	340 (38.4)
46–55	293 (33.1)
>55	128 (14.5)
Sex	
Male	849 (95.9)
Female	36 (4.1)
Marital status	
Single	80 (9.0)
Married	798 (90.2)
Divorced	10 (0.8)
Education level	
Not completed primary	94 (10.6)
Completed primary	202 (22.8)
Not completed secondary	259 (25.3)
Completed secondary	274 (31.0)
Post-secondary or above	56 (6.3)

Sociodemographic characteristics of subjects and noise exposure

Table 1 presents the sociodemographic characteristics of the subjects. The mean age was 45.3 years (SD 8.7 years), with a range from 20 to 66 years. The majority of the workers were male (95.9%) and married (90.2%).

Background sound pressure levels in the testing rooms in the field

Sound pressure levels in these testing rooms varied considerably, ranging from 11.2 to 68.7 dB (table 2). Sound level measurements at 0.5 KHz in all the testing rooms in the field were higher than the maximum level recommended by ANSI,² while the mean sound levels for 1 KHz and 2 KHz were also above the recommended levels.

Time interval between screening test and diagnostic test

The time interval between the screening test and the diagnostic test ranged from one day to 12 weeks. Seven per cent were diagnostically tested one day after screening, and 2.2%, two days after. In 92.8% of subjects, the time interval between the two tests was three days or more, and in 83.4% of subjects, seven days or more.

Absolute differences between hearing thresholds obtained from screening and diagnostic tests at all testing frequencies

Table 3 shows the means and standard deviations of the “absolute value of the difference” between the hearing thresholds obtained from the screening and diagnostic tests, for all the frequencies tested. The mean absolute difference was largest at 0.5 KHz and progressively decreased at higher

Table 2 Background sound levels in testing areas

Octave band (Hz)	Background sound levels measured			Maximum dB allowed
	Minimum	Maximum	Mean (SD)	
500	20.9	68.7	44.8 (11.2)	19.5
1000	12.3	61.3	41.4 (10.7)	26.5
2000	11.2	58.1	38.1 (10.5)	28.0
4000	11.3	55.1	32.7 (10.6)	34.5
8000	12.1	51.1	28.1 (9.6)	43.5

Table 3 Mean absolute difference* in hearing threshold between screening and diagnostic tests by frequency

Frequency	Ears	Absolute difference (dB)	
		Mean*	SD
500 Hz	Left	8.78	11.33
	Right	8.14	10.42
1000 Hz	Left	5.48	8.05
	Right	4.03	8.04
2000 Hz	Left	2.89	7.69
	Right	2.59	8.00
3000 Hz	Left	3.74	8.19
	Right	3.65	8.07
4000 Hz	Left	1.94	9.28
	Right	1.79	8.59
6000 Hz	Left	2.02	11.56
	Right	1.78	11.63
8000 Hz	Left	1.47	10.31
	Right	1.74	10.87

*The mean absolute differences in hearing threshold between the two tests were statistically significant ($p < 0.001$) by the Wilcoxon signed rank test at all frequencies, for both ears.

frequencies, at around 2 dB or less at 4, 6, and 8 KHz in either ear. With our large sample size, even the small absolute differences at the higher frequencies from 4 to 8 KHz were statistically significant. The standard deviations of the differences were about 8 dB at 1–3 KHz and 9 dB at 4 KHz.

Proportions of hearing loss at different frequencies and the difference in proportions between screening testing and diagnostic testing

We calculated the proportions of subjects with hearing loss according to the criteria adopted for the diagnostic test and compared the results with those according to the three criteria of hearing loss used in the screening test (table 4).

When we used the same criteria (≥ 30 dB for any of the high frequencies, and ≥ 90 dB for the sum of three low/high frequencies) in both screening and diagnostic tests, the proportions of hearing loss from the screening test results were higher than those from the diagnostic test at all frequencies. The difference was greatest (at 26.5%) with the criteria for low frequencies (≥ 90 dB at 0.5, 1, and 2 KHz), and smallest (at 6.5%) at 8 KHz, followed by 4 and 6 KHz (at 8% and 8.8% respectively).

When we relaxed the criteria in the screening test by 5 dB to ≥ 35 dB for any of the high frequencies and by 15 dB to ≥ 105 dB for the sum of the three low/high frequencies (criterion 2), and compared the results with those from the diagnostic tests (that still used the ≥ 30 dB and ≥ 90 dB criteria), the differences in the proportions were much smaller and statistically insignificant, being 1.8% at 4 KHz, 2% at 6 KHz, and 2.4% at 3 KHz. At these three frequencies, the prevalence of hearing loss obtained from the screening tests was still higher than that from the diagnostic tests. This difference was, however, reversed in the three low frequencies, three high frequencies, and at 8 KHz.

When we further relaxed the criteria in the screening tests to ≥ 40 dB for any high frequencies and to ≥ 120 dB for the sum of the three low/high frequencies (criterion 3), the prevalence of hearing loss based on screening test results was much lower than that based on diagnostic test results, and their differences were too great for the criteria to be meaningful.

The agreement between the prevalence of hearing loss obtained from screening tests and diagnostic tests were assessed with the kappa (κ) statistic. Using the ≥ 30 dB/ ≥ 90 dB criterion (criterion 1) for both tests, the κ values generally indicated good agreement, at 0.6 or above for most frequencies except 0.5 KHz (at 0.49) and 6 KHz (at 0.40). κ was highest at

Table 4 Difference in proportions of hearing loss* between screening test† (using three different sets of criteria) and diagnostic test results, by frequency

Frequency	Diagnostic test	Screening test (criterion 1)			Screening test (criterion 2)			Screening test (criterion 3)		
	%	%	Difference‡	Kappa	%	Difference‡	Kappa	%	Difference‡	Kappa
3 low frequencies	36.9	63.4	26.5**	0.491	26.3	-10.6**	0.629	3.4	-33.5**	0.120
3 high frequencies	70.3	81.1	9.8**	0.614	65.5	-4.8**	0.654	5.8	-64.5**	0.093
3 KHz	52.0	65.6	13.6**	0.641	54.4	2.4	0.712	12.4	-39.6**	0.153
4 KHz	73.1	81.1	8.0**	0.652	74.9	1.8	0.669	15.9	-57.2***	0.083
6 KHz	81.8	90.6	8.8**	0.401	83.8	2.0	0.468	20.3	-61.5**	0.019
8 KHz	65.6	72.1	6.5**	0.607	62.6	-3.0	0.659	13.6	-52.0**	0.037

*Hearing loss was defined according to the criteria described in "Subjects and methods" section—that is, for the diagnostic test, the sum of low frequencies (0.5, 1, and 2 KHz), ≥ 90 dB; the sum of high frequencies (3, 4, and 6 KHz), ≥ 90 dB; and for the individual frequencies (3, 4, 6, and 8 KHz), ≥ 30 dB.

†Criterion 1: same criteria as in diagnostic test. Criterion 2: sum of three low frequencies ≥ 105 dB; sum of three high frequencies ≥ 105 dB; individual frequencies (3, 4, 6, and 8 KHz) ≥ 35 dB. Criterion 3: sum of three low frequencies ≥ 120 dB; sum of three high frequencies ≥ 120 dB; individual frequencies (3, 4, 6, and 8 KHz) ≥ 40 dB.

‡Difference = % in screening test – % in diagnostic test.

** $p < 0.01$, *** $p < 0.001$.

0.65 for 4 KHz. With criterion 2 (≥ 35 dB/ ≥ 105 dB), the agreement improved as κ rose to 0.71 for 3 KHz and 0.67 for 4 KHz. With criterion 3, where the criteria for the screening test were further relaxed to ≥ 40 dB/ ≥ 120 dB, κ fell to 0.15 and below, indicating very poor agreement.

DISCUSSION

Field surveys on hearing loss among workers are often conducted in an environment with moderate to substantial levels of background noise. Such testing environments might produce results that are biased towards an overestimation of the shift in hearing threshold and consequently, the prevalence of hearing loss, when measured against results from a standard audiometric testing environment. Differences in the hearing threshold measured in the two settings, and the consequent variations in the estimate of the prevalence of hearing loss have not been previously reported in the literature. In this study, we compared the hearing test results in a group of transport workers conducted in an environment with substantial levels of low frequency background noise with those obtained in a soundproof booth environment.

The procedures used in field screening and diagnostic testing were standardised. However, inter-observer variations (between the technicians conducting the tests), the subjects' exposure to noise prior to testing (that might have resulted in temporary threshold shifts), and random variations in the subjects' response to audiometric tests might have affected the validity of our results. The sequence of the two tests (screening preceding diagnostic test) might produce a "learning effect", but the time interval between the two tests was three days or more for 92.8% of subjects, and seven days or more for 83.4% of subjects. Hence we concluded that any "learning effect" would be insignificant. We also plotted the differences versus the mean values of the hearing loss in both tests to look for systematic error. The slopes of the regression lines were almost parallel to the x-axis, and the regression coefficients (β) ranged from about -0.1 to -0.07 at different frequencies. The intercept, which represented the systematic difference between the two test results, was smallest (4.6 dB) at 4 KHz, and greatest at the low frequencies of 0.5 and 1 KHz (at 11.6 dB and 8.8 dB respectively). This suggested that the presence of low frequency background noise in the screening test environment increased the discrepancies in low frequency hearing thresholds obtained by the two test methods (screening and diagnostic), while higher frequencies were less affected. Systematic error because of learning effects should theoretically be uniform across all frequencies.

An important limitation of our study was that the subjects belonged to a selected group who were occupationally exposed to noise, detected to have some form of hearing impairment

(>25 dB in any frequencies tested) at the screening test, and volunteered to be tested in a soundproof booth environment. Ideally, testing subjects with hearing loss ≤ 25 dB would enable the evaluation of misclassification, and of the sensitivity and specificity of the screening test against the "gold standard", that of the diagnostic test. In practice, this ideal can rarely be achieved. Our study made use of data from a prevalence survey on NIHL, and a follow up diagnostic test could only be justified when some hearing loss was detected in the screening tests. Even among those with >25 dB hearing loss, the response to the request for diagnostic testing was low (at 34.5%). One possible bias was that our subjects were more concerned about their hearing and possibly represented those with greater hearing problems. A comparison between the mean hearing thresholds of the respondents and the non-respondents (who refused to come for the diagnostic test) showed indeed that the respondents had slightly worse hearing, but the mean hearing thresholds obtained from the screening test differed little among the two groups (<4 dB in 6 and 8 KHz, and <2 dB in all the other tested frequencies). Although our subjects were not representative of all transport workers, this study of the extent of agreement between the two test methods among those with some degree of hearing loss contributes towards epidemiological research and clinical applications of audiometry in field surveys. We had insufficient data to elucidate the relation between the level of background noise and the subjects' hearing threshold at various frequencies. While background noise measurements were made at all test sites, they were not performed on all test days. Moreover, the background noise levels fluctuated during the screening tests. Therefore, we could not further evaluate its influence on the hearing test results of individual subjects. In our comparison between the two test results, multiple comparisons were made and one must caution against the chance findings of "statistically significant" results. However, the mean absolute differences in hearing thresholds (table 3) and differences in proportions of hearing losses (table 4) were highly significant, at $p < 0.001$ and $p < 0.01$ respectively. Hence we believe our findings were unlikely to be caused by chance.

In pure tone audiometry, the absolute consistency of hearing threshold measurements is considered to be more important than the relative consistency. The former was defined by Jerger⁷ as the "absolute variability in performance from test to test" and reflected the precision with which a test instrument predicts the criterion measure across repeated applications.⁸ "Relative consistency, as expressed in the coefficient of correlation between test-retest threshold measurements, is not relevant if the difference is inordinately large and variable".⁸ In

this study, we used the means of the “absolute value of differences” in the hearing thresholds measured by the two methods to reflect the absolute consistency. The means of the absolute differences between two measurements ranged from 1.5 to 8.8 dB. At 0.5 KHz (the lower end of the frequencies tested), where the background noise level in the field environment was high, considerable differences (8–9 dB) existed in the subjects’ mean hearing thresholds by the two testing methods. The differences steadily diminished with progressively higher frequencies, to about 2 dB at 4–8 KHz. Even though the differences from the two tests were statistically significant at all frequencies, differences of less than 5 dB are of no practical relevance because a 5 dB interval was used in measuring the hearing threshold, a standard procedure in audiometry. The higher background noise levels at 0.5 KHz might have masked the subjects’ response to sound and accounted for the larger discrepancies in the hearing thresholds by the two tests. The standard deviations were largest (at 10–12 dB) at both ends of the frequencies, decreased to about 9 dB at 4 KHz, and to about 8 dB at 1–3 KHz. McBride and Williams reported that a dip in the audiogram at 4 KHz, but not at 6 KHz, was associated with noise exposure in a group of electrical transmission workers.⁹ Impairment at 6 KHz was considered transient and reversible.⁹ This agrees with our findings of a larger standard deviation in hearing threshold at 6 KHz than at 4 KHz. As a hearing conservation programme, our two stage approach to test noise exposed workers is appropriate and of clinical usefulness, but the sensitivity and specificity of the screening audiometric test require further evaluation.

We compared the proportions of hearing loss obtained from the two test methods, at the frequencies 3, 4, 6, and 8 KHz (≥ 30 dB in each frequency), at three low frequencies (a total of ≥ 90 dB at 0.5, 1, and 2 KHz), and three high frequencies (a total of ≥ 90 dB at 3, 4, and 6 KHz). The proportions differed the least at 4 KHz, by 8% when the same criteria of hearing loss were used for both tests, and by 1.8% when the screening test criteria were relaxed to ≥ 35 dB (single frequency) or ≥ 105 dB (three frequencies combined). Likewise, the highest κ value was found at 4 KHz ($\kappa = 0.65$), when the agreements between the two test results on the proportion of hearing loss were evaluated using the same criteria (≥ 30 dB) for both tests. When the screening test criteria were relaxed (≥ 35 dB), however, the best agreement was seen at 3 KHz ($\kappa = 0.71$).

A standard otolaryngological text advised that: “the very earliest changes in young subjects exposed to broad band noise for 1–2 years occur around 6 kHz. With the duration of exposure to noise of 2–5 years, noise induced permanent threshold shift slides into the 4 kHz region”.¹⁰ The widely accepted early indicator of NIHL is the 4KHz “dip” or “notch”, spreading to lower frequencies if the individual is continuously exposed to an excessively noisy environment.¹

After evaluating the mean absolute differences in the hearing thresholds and their standard deviations, and the proportions of hearing loss by both methods, we found that at 4 KHz, the mean absolute difference and its standard deviation are the smallest of all the frequencies tested. In addition, the proportion of hearing loss at 4 KHz showed the best agreement (in terms of the difference in proportions and the κ value) among subjects who had more than 25 dB hearing loss in the screening test. Hearing loss at 4 KHz is a classical indicator for NIHL, and audiometric screening tests conducted in the field may be used epidemiologically to estimate its true prevalence, after adjusting for the effects of background noise. After the screening test, a representative sample of the subjects should be retested in a soundproof booth and the agreement between

Main messages

- The mean hearing thresholds at 4 KHz among noise exposed workers, measured under non-soundproof environments, were similar to those measured in a soundproof booth.
- The prevalence of NIHL can be estimated from results of hearing threshold at 4 KHz obtained from field surveys under non-soundproof conditions.
- The criteria for the diagnosis of NIHL used in the field survey should be adjusted for the effects of background noise by comparison with hearing thresholds obtained from a soundproof booth.

Policy implications

- Our study confirms the usefulness of screening audiometric tests in the field in identifying workers with hearing loss and estimating its prevalence.
- Hearing conservation programmes may take the form of field screening tests in the worksite, followed by confirmatory tests in a soundproof booth.

the two test results compared. The true prevalence of NIHL can then be estimated by adjusting the criteria for the definition of hearing loss used in the screening test. In our experience, an impairment of ≥ 35 dB at 4 KHz measured in non-soundproof environments appears to be a suitable criterion for estimating the prevalence of NIHL defined as a 30 dB loss at 4 KHz among noise exposed transport workers.

ACKNOWLEDGEMENTS

This work was supported by the Hong Kong Occupational Safety and Health Council. We thank the transport industries and all the workers for their participation, all the field researchers for their efforts, and Prof. J L Tang for his useful comments.

Authors’ affiliations

T W Wong, T S Yu, W Q Chen, Y L Chiu, C N Wong, A H S Wong, Department of Community and Family Medicine, The Chinese University of Hong Kong

REFERENCES

- 1 Rabinowitz PM. Noise-induced hearing loss. *Am Fam Physician* 2000;**61**:2749–56.
- 2 American National Standards Institute. *Maximum permissible ambient noise levels for audiometric test rooms*, ANSI S3.1-1991. New York: ANSI, 1991.
- 3 Wong TW, Yu TS. *A survey on noise exposure and hearing impairment among transport workers in Hong Kong (Research Report)*. Hong Kong: Department of Community and Family Medicine, The Chinese University of Hong Kong, 1999.
- 4 British Standards Institution. *British standard specification for pure tone air conduction threshold audiometry for hearing conservation purposes*, BS 6655:1986. Milton Keynes: BSI, 1986.
- 5 Rigby AS. Statistical methods in epidemiology. v. Towards an understanding of the kappa coefficient. *Disabil Rehabil* 2000;**22**:339–44.
- 6 Fleiss JL. *Statistical methods for rates and proportions*, 2nd edn. New York: John Wiley, 1981:212.
- 7 Jerger JF. Comparative evaluation of some auditory measures. *J Speech Hear Res* 1962;**5**:3–17.
- 8 McDermott JC, Fausti S, Henry JA, et al. Masked high-frequency bone-conduction audiometry: test reliability. *J Am Acad Audiol* 1991;**2**:99–104.
- 9 McBride DI, Williams S. Audiometric notch as a sign of noise-induced hearing loss. *Occup Environ Med* 2001;**58**:46–51.
- 10 Maran A, Stell P. *Clinical otolaryngology*. Oxford: Blackwell, 1979:102.