
Variability in audiometric recording

R. W. HOWELL and B. P. R. HARTLEY

British Steel Corporation, Head Office, Grosvenor Place, London SW1 and Stocksbridge Works, near Sheffield

Howell, R. W., and Hartley, B. P. R. (1972). *Brit. J. industr. Med.*, **29**, 432-435. **Variability in audiometric recording.** A study of initial and repeat audiograms by two operators on 143 young male new entrants to industry (free of previous occupational exposure) has shown that mean values (mean of both ears, readings at 3 and 4 kHz) differed significantly between operators, and that this difference was not attributable to transient medical conditions such as wax in the ears or temporary infection.

For nearly half the employees, the difference between the results obtained by the two operators amounted to 5 dB or more, with differences up to and including $21\frac{1}{2}$ dB. Of the two operators' lists of men in the lowest decile of hearing threshold levels, only half the names were common to both operators; there was clearly wide variation between the operators.

With such variability in audiometric recording, some of which may be attributable to variation in the patients' responses, it seems unlikely that small changes in recorded hearing levels will give confident early indication of deterioration in a susceptible ear.

It is suggested that further studies be carried out under industrial conditions to ascertain the advantages, if any, of two or more readings taken at one session (or within a short space of time) over single observations. Similar studies are needed using self-recording audiometry.

Consideration should be given to the basic training and potential of operators, to the need for monitoring their performance, and to periodic refresher training where appropriate.

The present investigation was undertaken to discover whether those subjects who had the poorest hearing at first examination would show the greatest deterioration at subsequent examination within a short period of time.

All figures in this paper are based on the mean pure tone audiometric readings for both ears at 3 and 4 kHz; the intention of the present investigation was also to see whether, in practice, small changes in measured hearing levels in the lowest decile were likely to be prognostic of ultimate occupationally related hearing damage, and to look at the variations in audiograms obtained by different operators.

Present investigation and results

A preliminary examination of the pre-employment records of male school-leavers at a number of

British Steel Corporation works showed that in practice the 10% of entrants with the greatest deterioration in recorded hearing levels had mean decreases, at all establishments, which were considerably greater than expected from theoretical calculations by Adam (1970, personal communication) on the basis of the work of Burns and Robinson (1970). The exclusion of those subjects with recognized medical impairment or noisy hobbies such as shooting or musical bands had little effect (and certainly no significant effect) on the mean values at any works.

Further examination of these audiograms showed that approximately two-thirds of the worst 10% on entry had a second audiogram which showed an increase of hearing levels, whereas in those with an initial reduced hearing level of not more than $8\frac{3}{4}$ dB (more normal levels) the second audiogram showed

that one-third of the total now had a small loss in hearing level. As the audiometer used recorded in 5 dB stages, and the mean of four readings was taken for each individual, mean progressions of 1½ dB were obtained (Table 1).

These figures suggested that there might be an element of chance in the composition of the lowest decile. For example, if subject error, observer error, and machine error all operate in the same direction in a proportion of initial audiograms, then some subjects may erroneously be placed in the lowest decile or, conversely, be wrongly omitted. It should be noted that those observations which vary from the mean by chance alone (as opposed to 'real' deviations) will tend on repetition of the test to revert, i.e., the low observations will tend to be higher and the high ones will tend to be lower. This is a consequence of the variability of observations (whether this is due to technique, operator or patient) and has no biological significance. It tends to obscure the effect of small significant variations in an overall series.

It was thought that some estimate of the extent of observer error might be obtained by repeating recent pre-employment audiograms at two works, at the same time exchanging operators. Eighty-seven dual recordings were obtained from one works and 56 from the other, so that in all cases one audiogram was taken in late 1970 and the other, using the other operator, in early 1971. Table 2 shows that one operator obtained means which were higher than those of the other operator at both works.

A learning effect or an apparent variation in measured hearing level associated with familiarization with the test is likely to affect both establishments. Similarly, both operators are likely to have been affected by familiarization because both

carried out repeat audiograms when the learning effect (if any, after a lapse of a few months) would operate. As one operator carried out slightly more second audiograms the effect may not be quite evenly balanced between operators.

Examination of subjects where there was evidence of wax or coryza suggested that these factors were not the cause of the variation between operators.

Figures 1 and 2 show the distribution of hearing levels at the two works as measured and recorded by the two operators.

Although the difference between the means obtained by the operators may not appear large when regard is paid to the fact that the audiometers in use are calibrated in 5 dB steps, the extent of the variation between operators can better be seen by comparing the results on each subject. The difference between the two operators' readings on each subject was therefore calculated; it was found that the difference was as high as 21½ dB on one subject. A Wilcoxon's signed rank test using the data from the 143 subjects showed that the differences between the operators was highly significant ($P < 0.02$). This test has the advantage that it takes into account the

TABLE 2
MEAN HEARING LOSS (dB) AT 3 AND 4 KHZ

Works	No. of men	Operator	
		X	Y
A	87	2.6 (1970)	3.7 (1971)
B	56	1.7 (1971)	6.7 (1970)
A and B		2.2	4.9

TABLE 1
COMPARISON OF FIRST AND SECOND AUDIOMETRIC READINGS
MALES, AGED 15 TO 18, WITHOUT INITIAL OCCUPATIONAL NOISE EXPOSURE

Interval between initial and second audiogram (yr)	Number of men					
	Worst 10% initially (smallest loss 13½ dB)			More normal levels (largest loss 8½ dB)		
	Second audiogram			Second audiogram		
	Better	Same	Worse	Better	Same	Worse
1	3	0	1	0	0	1
2	1	0	0	1	0	4
3	7	1	1	5	0	3
4	11	1	6	7	2	11
5	2	0	4	7	3	23
Total (1 to 5 years)	24	2	12	20	5	42
(3 to 5 years)	20	2	11	19	5	37

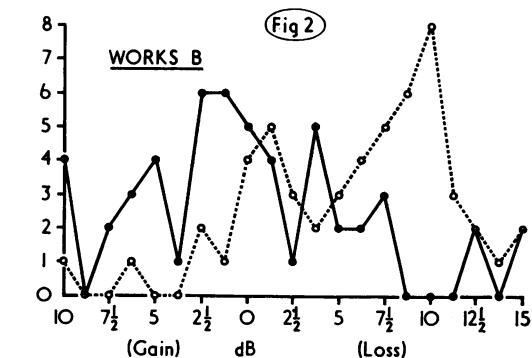
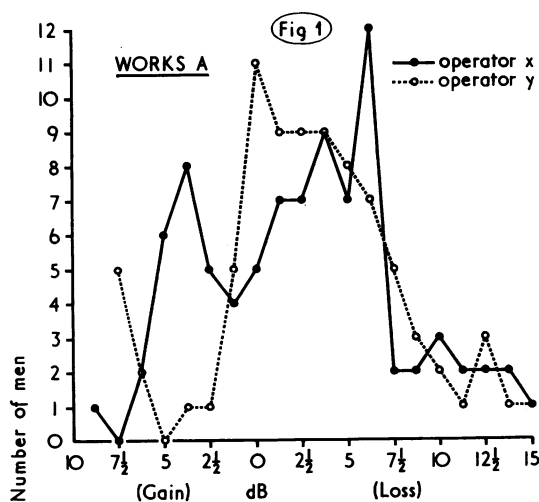


FIG. 1. Mean dB hearing levels at 3 and 4 kHz (both ears). Initial and second audiograms for males aged 15 to 18. Excludes four males with mean losses greater than 15 dB (1, operator x; 3, operator y).

FIG. 2. Mean dB hearing levels at 3 and 4 kHz (both ears). Initial and second audiograms for males aged 15 to 18. Excludes 7 males with mean losses greater than 15 dB (4, operator x; 3, operator y).

positive and negative differences between the operators and the size of the difference. Each subject has been classified according to operator difference (usually in $1\frac{1}{2}$ dB stages) and the numbers of subjects in each category are shown in Table 3. For example, if one operator had a mean level of 5 dB for a subject for whom the other operator recorded $12\frac{1}{2}$ dB then this subject would be counted as a difference of $7\frac{1}{2}$ dB in Table 3.

Discussion

Although Tables 2 and 3 show considerable operator variability, it must also be borne in mind that subject variability may contribute to the differences observed. There may be some merit in considering the use of self-recording audiometry in an endeavour to reduce variability, but Gosztonyi, Vassallo, and Sataloff (1971) found 'the prevalence of unreliability was far higher in audiograms obtained by self-recording audiometry, than by manual audiometry'. They found the complicating factor 'could be called malingering'. It may be that such complications are absent in this country but it would be useful to have some confirmation from industrial manual populations.

The composition of the lowest decile is clearly important in this concept of detecting relatively small changes in a subject in a short period. When the worst 10% on entry according to each operator's results were compared, it was found that only just over half the subjects were common to both operators. This finding does not contradict the suggestion from Table 1 that some element of variability arises in the composition of these worst 10% groupings.

Table 3 shows that in 11% of subjects there was agreement between operators' mean readings, but about half of the subjects showed differences of 5 dB or more. For one subject in eight the operators differed by 10 or more dB; the highest recorded difference between the operators was $21\frac{1}{2}$ dB, followed by 20 and $18\frac{1}{2}$ dB.

There is no evidence that these two operators represent the extremes of operator variability. Given a larger series of operators no doubt some would give results with less average variation but

TABLE 3
DIFFERENCE BETWEEN OPERATORS' READINGS ON SAME SUBJECT (MEAN 3 AND 4 KHZ BOTH EARS)

	Operator difference (dB)								
	0	$1\frac{1}{2}$	$2\frac{1}{2}$	$3\frac{1}{2}$	5	$6\frac{1}{2}$	$7\frac{1}{2}$	$8\frac{1}{2}$	10+
No. of men	16	23	21	17	19	10	11	8	18
% of total	11.2	16.1	14.7	11.8	13.3	7.0	7.7	5.6	12.6

undoubtedly some would provide results with larger differences than in this series. Similar considerations probably apply to subject variability.

It seems unlikely, with the variabilities in audiometric recording shown in this survey, that a mean loss of 5 dB in one year for a man with a pre-exposure mean of 8 dB (i.e., a measured hearing level change of from 8 to 13 dB) will indicate with any degree of certainty deterioration in a so-called sensitive ear. It may be that single observations are not reliable enough and that the mean of two or three readings, taken within a short interval of time, should be used (Burns, 1968). On the other hand, in the practice of routine audiometry in industry as opposed to special or laboratory studies, the chances of getting a man released on three separate occasions within a short space of time may not be realistic, and the repeats may have to be done at the initial attendance. It seems desirable that studies should be carried out in industrial settings to establish the practical value of such a procedure. Already there are difficulties in avoiding temporary threshold shift when testing industrial workers. In consequence the practical problem is not only of having men released from work but having them released at particular times (e.g., after the weekend) and for quite long periods of examination time. Trevethick (1971, personal communication) has suggested that effort should be concentrated on the worst 20% on

entry rather than on the worst 10%, although such levels are likely to be arbitrary until large series establish reasonable thresholds of entry for any given age. Further experimental study is probably needed.

Given the variations apparently possible between identically trained operators (and probably larger variations between those trained by different methods), consideration should also be given to the basic training requirements for operators, the need for re-testing them at intervals, and the need for refresher courses.

We are grateful to Dr. R. A. Trevethick, who so kindly agreed to this survey and provided facilities at one of the works, to Mr. J. Adam for useful discussion, and most importantly, to the two anonymous operators.

References

- Adam, J. A. (1970). Personal communication.
 Burns, W. (1968). *Noise and Man*, 1st ed., pp. 302-304. John Murray, London.
 —, and Robinson, D. W. (1970). *Hearing and Noise in Industry*, 1st ed. H.M.S.O., London.
 Gosztonyi, R. E., Vassallo, L. A., and Sataloff, J. (1971). Audiometric reliability in industry. *Arch. environm. Hlth*, 22, 113.
 Trevethick, R. A. (1971). Personal communication.

Received for publication December 15, 1971.