J Am Acad Audiol 7: 251-259 (1996)

Interexaminer Reliability of Otoscopic Signs and Tympanometric Measures for Older Adults

David M. Nondahl*
Karen J. Cruickshanks*
Terry L. Wiley†
Ted S. Tweed†
Barbara E. K. Klein*
Ronald Klein*

Abstract

To accurately classify hearing loss and otic disorders among older adults, examiners must be able to consistently assess otoscopic signs and perform middle-ear screening tympanograms. As part of a population-based study of hearing loss in Beaver Dam, Wisconsin, the interexaminer reliability of otoscopic examinations and screening tympanograms was evaluated using 45 replicate examinations. Data from 1941 participants 48 to 91 years of age were used to compare otoscopic and tympanometric results. Overall agreement for nine otoscopic signs ranged from 73 percent (vascularity) to 100 percent (drainage). There were small examiner differences in tympanometric measures of equivalent ear-canal volume (V_{ea}) and tympanogram width. No significant differences were observed for peak compensated static acoustic admittance and tympanogram peak pressure. Our findings suggest that examiners can be trained to consistently and accurately assess otoscopic signs and obtain reliable tympanometric results.

Key Words: Middle ear, observer variability, otoscopy, tympanometry

Hearing loss is one of the most common chronic conditions affecting older adults (NCHS, 1990). Thirty-six percent of those in the United States at least 75 years of age report a disabling hearing loss (NCHS, 1990). The presence of middle-ear disease, while most common in children (Meyrick, 1951; Moran et al, 1979; Pedersen and Zachau-Christiansen, 1988), can affect older adults as well (Moran et al, 1979; Rudin et al, 1983; Browning and Gatehouse, 1992). Moscicki et al (1985) suggest that the incidence of mixed hearing loss and associated middle-ear disorders may be as high as 24 percent for older subjects. Current American Speech-Language-Hearing Association (ASHA) guidelines state that both otoscopy and tympanometry are essential elements of a middle-ear screening program that will minimize over-referral rates (ASHA, 1990). Therefore, audiologic evaluations must include examinations of the outer and middle ear to identify otic conditions that may affect hearing or require a medical referral. Good otoscopic and tympanometric interexaminer reliability are fundamental to the contribution that these assessments make to the differential diagnosis of middle-ear abnormalities and hearing loss.

Accordingly, examiners involved in field studies, screening programs, or clinical evaluation must be able to consistently assess otoscopic signs and perform middle-ear screening tympanometry. However, we know of no study that reports on interexaminer reliability of tympanometry or otoscopy with adult subjects. Furthermore, most related studies of children focus on comparing examiners' or instruments' diagnostic assessments with a known, often postsurgical, diagnosis (Gates, 1986; Mills, 1987;

*Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, Madison, Wisconsin; †Department of Communicative Disorders, University of Wisconsin, Madison, Madison, Wisconsin

Reprint requests: David M. Nondahl, 460 WARF Office Building, 610 N. Walnut Street, Madison, WI 53705-2397

Wazen et al, 1988; Toner and Mains, 1990; Finitzo et al, 1992; Kaleida and Stool, 1992; Vaughan-Jones and Mills, 1992; Fields et al, 1993; Ovesen et al, 1993). These same studies do not assess interexaminer reliability. Among the few studies of interexaminer reliability with children are reports on interexaminer reliability of the otoscopic assessment of middle ear effusion (Mains and Toner, 1989), assessment of several otoscopic signs by two pediatricians (Margolis et al, 1979), otoscopic and tympanometric assessment of otitis media (Bluestone and Cantekin, 1979), and classification of tympanogram shape (de Melker, 1992). Parallel studies have not been conducted in adult populations. Accordingly, as part of the population-based Epidemiology of Hearing Loss Study (EHLS), we addressed two research questions: (1) Can examiners consistently assess otoscopic signs and obtain reliable tympanometric results and (2) Are the otoscopic assessments consistent with tympanometric results?

METHOD

The EHLS is an ongoing population-based study of hearing loss in adults 48 to 91 years of age residing in the city or township of Beaver Dam, WI. All participants in the Beaver Dam Eye Study (Klein et al, 1991; Linton et al, 1991) who were alive on March 1, 1993 were eligible to participate in the EHLS (n = 4541). Two thousand forty-one participants, including 56 nursing home participants, were seen by the EHLS between March 19, 1993 and March 18, 1994. Baseline data collection was completed in the summer of 1995.

The EHLS examination includes a questionnaire administered by one of three trained examiners that covers medical and family history, history of noise exposure (including occupational history, hobbies, and military service) (Talbott et al, 1990), and quality of life. The physical examination includes otoscopy, screening tympanometry (ASHA, 1990), pure-tone air- and bone-conduction audiometry (ASHA, 1978), and tests of word recognition. The tympanometers and audiometers were calibrated according to ANSI standards (ANSI, 1987, 1989).

Three EHLS examiners (one who had recently received her M.S. in audiology and two with no previous audiologic training) were trained to perform tympanometry and otoscopy by two ASHA certified audiologists (TLW, TST). Training consisted of a tutorial with slides, handouts, and group practice sessions.

Sample Selection

A hermetic seal could not be obtained during tympanometry in either ear for 76 participants, primarily due to the presence of excessive hair in the external auditory meatus. These 76 participants were excluded from the analysis.

Each tympanogram was graded on five characteristics: the visibility of the peak, the appearance of a baseline for both negative and positive pressures, the presence of any sharp drop-off or notch, and any other unusual abnormalities (e.g., bumps). On a separate day, the curve was evaluated by a certified audiologist (TST) to determine if it was acceptable for analysis. Those judged to be unacceptable displayed strong evidence of one or more of the following test problems: (1) a participant swallowing, moving, or clearing his throat during tympanometry; (2) movement of the tympanometer by the examiner during tympanometry; or (3) sudden loss of seal. This screening process resulted in 24 additional participants being excluded from the analysis, for a final sample size of 1941 participants and 3628 ears.

Equipment and Measures

Otoscopy was performed using the #25020 Welch Allyn Otoscope/Throat Illuminator. Nine otoscopic signs were assessed (Table 1). Among the 30 of 1941 participants judged to have one or both ear canals impacted, 17 visited their doctor for cleaning and returned later to complete their EHLS exam. Otoscopic and tympanometric data from the first visit were used in these cases.

The study used two Lucas Grason Stadler, Inc. GSI 37 Auto Tymps. Both were calibrated to ANSI standards (ANSI, 1987) before examinations began and were rechecked at the beginning of each examination day. This unit employs a 226-Hz probe frequency, a pump speed of 200 daPa/sec and a positive-to-negative direction of pressure change. Tympanometry was repeated if a tympanogram showed evidence of any technical problems. The Auto Tymp records four measures (ANSI, 1987). Peak compensated static acoustic admittance (peak Y_{tm}) is represented in cubic centimeters (cc) on the tympanogram from the Auto Tymp. This peak Y_{tm} measure is more properly expressed in acoustic mmhos. A low peak Y_{tm} could indicate a stiffening of the ossicular chain due to otosclerotic bone formation/adhesions or the presence of middle-ear effusion, while a high peak Y_{tm} could indicate

Table 1 Tabulation of Otoscopic Signs

	1842 EHLS Right Ears*		*45 Quality Control Right Ears†*	
Otoscopic Sign	*Number*	*%*	*Number*	*%*
Drainage present				
No	1837	99.7	45	100.0
Yes	5	0.3	0	0.0
Tympanic membrane color				
Gray/White	1464	79.5	35	77.8
Yellow/Amber	87	4.7	0	0.0
Blue/Black	28	1.5	2	4.4
Unknown	263	14.3	8	17.8
Tympanic membrane position				
Normal	1580	85.8	37	82.2
Bulging	1	0.1	0	0.0
Retracted	2	0.1	0	0.0
Unknown	259	14.0	8	17.8
Liquid present				
None	1571	85.3	37	82.2
Bubbles	4	0.2	0	0.0
Visible air-liquid line	0	0.0	0	0.0
Unknown	267	14.4	8	17.8
Perforation present				
No	1573	85.4	37	82.2
Yes	3	0.2	0	0.0
Unknown	266	14.4	8	17.8
Ear-canal collapse				
None	1768	96.0	44	97.8
Partial	36	2.0	1	2.2
Complete	2	0.1	0	0.0
Unknown	36	2.0	0	0.0
Debris in canal				
None	457	24.8	9	20.0
Some	1111	60.4	28	62.2
A lot	253	13.8	7	15.6
Impacted	19	1.0	1	2.2
Unknown	2	0.0	0	0.0
General appearance of the tympanic membrane				
Normal	1053	57.2	19	42.2
Dull	524	28.5	18	40.0
Unknown	265	14.3	8	17.8
Vascularity				
None	685	37.2	12	26.7
Mild	813	44.1	24	53.3
Considerable	76	4.1	1	2.2
Unknown	268	14.5	8	17.8

Since only the right ear was used during the quality control replications (n = 45), right ear data is shown for the larger sample as well (n = 1842 right ears). Left ear data (n = 1786) for the larger sample was similar.

*EHLS: Epidemiology of Hearing Loss Study.

†Results from first assessment are shown.

ossicular discontinuity (Shanks and Shelton, 1991).

Equivalent ear-canal volume (V_{ea}) also was measured. A low V_{ea} can indicate an ear canal impacted with cerumen or improper probe placement, whereas a high V_{ea} accompanied by a flat tympanogram can indicate possible perforation of the tympanic membrane (TM) (Margolis and Heller, 1987; Shanks and Shelton, 1991; Shanks et al, 1992).

Tympanogram width (TW) represents the pressure interval (in daPa) at the 50 percent point between the peak and the tails of the tympanogram. Abnormally large TWs can be a criterion for medical referral (ASHA, 1990), although the clinical usefulness of this measure continues to be studied.

Tympanogram peak pressure (TPP), also measured in daPa, is the ear-canal pressure at which the peak Y_{tm} occurs. TPP correlates highly with actual middle-ear pressure (Van Camp et al, 1986). Abnormally positive or negative TPP in the absence of other TM abnormalities is not usually cause for medical referral (Margolis and Heller, 1987; ASHA, 1990).

Quality control replications were done on the right ear every 4 months (June/July, November, March) for both tympanometry and otoscopy. During these periods, for selected participants, a second examiner replicated the tests conducted by the first examiner. Each replication was done during one examination on a single day. Participants were chosen for replication based on examiners' time and schedule constraints, not a structured randomization plan. During each replication period, there were 5 replications for each possible pairing of the three examiners, or 15 replications. Therefore, a participant whose examination was being replicated was seen by two of the three examiners. (The number of replications was chosen based on what the examiners could reasonably handle in a 30-day period over and above their regular examination load.) Thus, 45 replications were done during the first year of the study.

Statistical analyses, including t-tests, paired t-tests, Fisher's Exact Test, contingency tables, and tests for significant regression slopes were performed using SAS (SAS Institute Inc., 1988); Kappa statistics (Fleiss, 1981) were calculated from the contingency tables.

RESULTS

Interexaminer Reliability: Otoscopy

The tabulation of the otoscopic signs for the first assessment of the 45 quality control participants is shown in Table 1. Overall interexaminer agreement ranged from 73 percent (vascularity) to 100 percent (drainage) for the nine otoscopic signs (Table 2). The overall Kappa statistics ranged from .49 to .94 and were highly significant ($p < .001$). Since the ability of examiners to visualize the relevant components of the middle ear affects the assessments, the "unknown" category (meaning unable to visualize) was included in evaluation of interexaminer agreement.

Twelve of the 45 replications resulted in differing assessments of vascularity. Nine of the 12 cases were disagreements between "none" and "mild," 2 were disagreements between "mild" and "considerable," and 1 was between "none" and "unknown." Seven of 45 replications resulted in differing assessments of the general appearance of the TM. Six of these cases were disagreements between "normal" and "dull" and one was between "dull" and "unknown."

Examiner-specific comparisons of otoscopic assessments are also shown in Table 2. Agreement ranged from 67 percent (general appearance: examiner 1 vs examiner 2) to 100 percent. All five disagreements between examiner 1 and examiner 2 regarding general appearance of the TM were disagreements between "normal" and "dull," with two of the disagreements in one direction and three in the other. The Kappa statistics ranged from .45 to 1.00 and were all significant ($p < .02$).

Interexaminer Reliability: Tympanometry

Pairwise comparisons (Table 3) for examiners show statistically significant mean differences in V_{ea}. Specifically, examiner 3 tended to obtain the largest V_{ea} measures and examiner 2 tended to obtain the smallest.

One examiner-specific comparison of TW resulted in a statistically significant mean difference of 9.33 daPa ($p = .01$; Table 3), with examiner 2 having higher values on average than examiner 1. The other two examiner-specific pairwise comparisons for TW were not significant. None of the examiner-specific pairwise comparisons for peak Y_{tm} or TPP resulted in statistically significant differences. Adjusting for multiple comparisons made no difference in the test results in terms of statistical significance; results shown in Table 3 are unadjusted.

Tympanometry and Otoscopy Compared

When comparing the consistency of tympanometric and otoscopic findings, we used the tympanometric results as the reference standard and compared the otoscopic assessments against that standard to provide evidence for or against the reasonableness of the otoscopic assessments.

Table 2 Interexaminer Agreement for Otoscopy

Otoscopic Sign	*Examiner Pair*	*n*	*% Agreement*	*Kappa*
Drainage present	1 vs 2	15	100	NA*
	1 vs 3	15	100	NA
	2 vs 3	15	100	NA
	Overall	45	100	NA
Tympanic membrane color	1 vs 2	15	100	1.00
	1 vs 3	15	93	0.87
	2 vs 3	15	100	1.00
	Overall	45	98	0.94
Tympanic membrane position	1 vs 2	15	100	1.00
	1 vs 3	15	93	0.82
	2 vs 3	15	100	1.00
	Overall	45	98	0.93
Liquid present	1 vs 2	15	100	1.00
	1 vs 3	15	93	0.82
	2 vs 3	15	100	1.00
	Overall	45	98	0.93
Perforation present	1 vs 2	15	100	1.00
	1 vs 3	15	93	0.82
	2 vs 3	15	100	1.00
	Overall	45	98	0.93
Ear-canal collapse	1 vs 2	15	93	0.64
	1 vs 3	15	100	NA
	2 vs 3	15	93	NA
	Overall	45	96	0.49
Debris in canal	1 vs 2	15	93	0.83
	1 vs 3	15	87	0.78
	2 vs 3	15	100	1.00
	Overall	45	93	0.88
General appearance	1 vs 2	15	67	0.45†
	1 vs 3	15	93	0.90
	2 vs 3	15	93	0.90
	Overall	45	84	0.76
Vascularity	1 vs 2	15	73	0.58
	1 vs 3	15	73	0.62
	2 vs 3	15	73	0.56
	Overall	45	73	0.61

n = 45.
*Contingency table had fewer than two rows or two columns so Kappa could not be calculated.
†p = .011. For all other Kappas, p < .001.

As V_{ea} decreased, the estimated amount of debris, based on otoscopy, increased (Fig. 1). Specifically, the mean values for V_{ea} for 1840 right ears were 1.45, 1.37, 1.26, and 0.94 cc for the otoscopic categories none, some, a lot, and impacted, respectively. For 1785 left ears, the means were 1.47, 1.39, 1.22, and 0.93 cc for the same respective otoscopic categories. The linear trend was highly significant for both ears (regression slope: $p < .001$).

Table 3 Paired T-test Results for Tympanometry: Comparison of Individual Examiners

Measure	*Examiner Pair*	*n*	*Mean Difference*	*T (df)*	*p Value (2-tailed)*
Peak Y_{tm} (mmhos)*	1 – 2	15	0.00	0.00 (14)	1.00
	1 – 3	15	0.02	0.59 (14)	0.57
	2 – 3	15	–0.03	–0.50 (14)	0.62
V_{ea} (cc)†	1 – 2	15	0.17	2.71 (14)	0.02
	1 – 3	15	–0.22	–3.06 (14)	0.01
	2 – 3	15	–0.26	–5.46 (14)	< 0.01
TW (daPa)‡	1 – 2	15	–9.33	–3.29 (14)	0.01
	1 – 3	15	–4.00	–1.70 (14)	0.11
	2 – 3	15	2.33	–0.44 (14)	0.67
TPP (daPa)§	1 – 2	15	1.00	0.36 (14)	0.72
	1 – 3	15	–1.00	–0.37 (14)	0.72
	2 – 3	15	2.00	0.81 (14)	0.43

*Peak compensated static acoustic admittance, in acoustic mmhos.
†Equivalent ear-canal volume, in cc.
‡Tympanogram width, in daPa/sec.
§Tympanogram peak pressure, in daPa/sec.

For the right ear, mean V_{ea} was significantly higher for the three ears judged to have a perforated TM than for the 1573 ears judged not to have TM perforations (3.97 vs 1.39 cc; t-test: $p < .001$; Fig. 2). Similar results were obtained for the left ear (4.17 vs 1.41 cc; t-test: $p = .099$), although the difference was not statistically significant. Five of the six ears judged to have TM perforations either had an air-bone gap of greater than 10 dB (three ears) or the air-bone gap could not be determined (two ears) due to output limitations for bone-conduction audiometry.

Three of 22 (13.6%) right ears with a high V_{ea} (> 2.5 cc; Shanks and Shelton, 1991) were judged to have TM perforations, compared to none of the 1554 (0.0%) right ears without a high V_{ea}. Three of 28 (10.7%) left ears with a high V_{ea} were judged to have TM perforations, compared to none of the 1523 (0.0%) left ears without a high V_{ea}. These differences were highly significant for both ears (Fisher's Exact Test: $p < .001$).

Among subjects with a negative history of ear surgery whose TM could be visualized, 15 of 1482 (1.0%) right ears and 20 of 1462 (1.4%) left ears without otoscopically detected perforation had V_{ea} greater than 2.5 cc. Maximum values were 3.7 and 4.9 cc for the right and left ears, respectively.

For the right ear, mean peak Y_{tm} was significantly lower for the 19 ears judged to have impacted ear canals than for the 1821 ears judged not to be impacted (0.25 vs 0.65 mmhos; t-test: $p = .003$; Fig. 3). Similar results were obtained for the left ear (0.40 vs 0.65 mmhos; t-test: $p = .107$), although the difference was not statistically significant.

Fourteen of 260 (5.4%) right ears with low peak Y_{tm} (< 0.3 mmhos; ASHA, 1990) were judged to be impacted, compared to 5 of 1580 (0.3%)

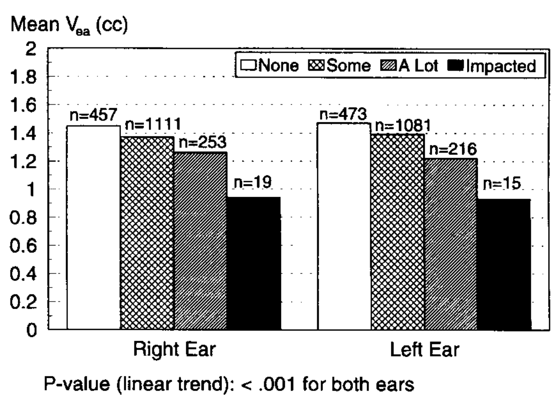

Figure 1 Mean equivalent ear-canal volume (V_{ea}) and otoscopic signs of ear-canal debris for 1840 right ears and 1785 left ears. V_{ea} is measured in cc.

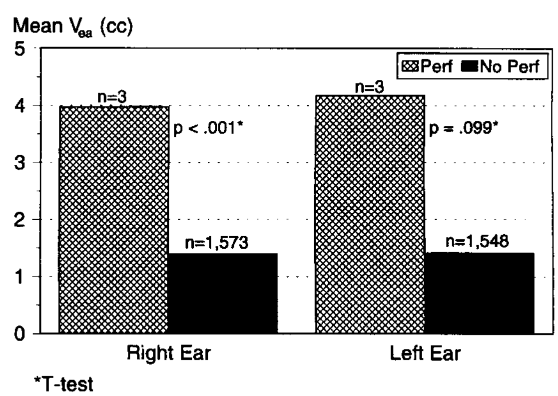

Figure 2 Mean equivalent ear-canal volume (V_{ea}) and otoscopic signs of tympanic membrane perforation for 1576 right ears and 1551 left ears. V_{ea} is measured in cc. Ears for which the tympanic membrane could not be visualized due to excessive cerumen or ear canal collapse are excluded.

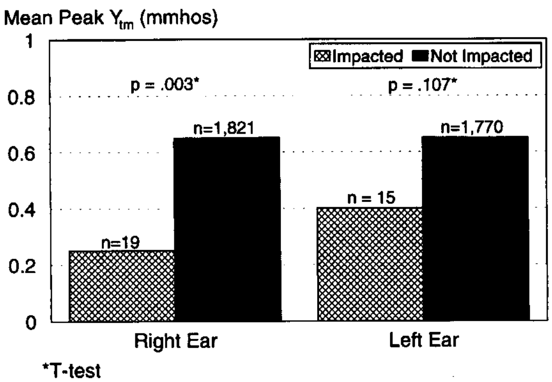

Figure 3 Mean peak compensated static acoustic admittance (peak Y_{tm}) and otoscopic signs of ear-canal impaction for 1840 right ears and 1785 left ears. Peak Y_{tm} is measured in acoustic mmhos.

right ears without low peak Y_{tm}. Eight of 261 (3.1%) left ears with low peak Y_{tm} were judged to be impacted, compared to 7 of 1524 (0.5%) left ears without low peak Y_{tm}. These differences were highly significant for both ears (Fisher's Exact Test: $p < .001$). Twelve of 19 right ears judged to be impacted had a peak Y_{tm} of zero, as did 7 of 15 left ears judged to be impacted.

DISCUSSION

Interexaminer Reliability: Otoscopy

After careful training, the nonphysician examiners were consistent graders of otoscopic signs. The two signs with less than 90 percent overall agreement, general appearance of the TM, and vascularity are among the least important for interpreting other audiologic findings.

For each sign scored, the level of overall agreement was considered to be moderate to high based on the Kappa statistics for first versus second examiner. Ear-canal collapse had the lowest overall Kappa (0.49), a value that nevertheless represents a moderate level of agreement above chance. Only 2 of 45 ear-canal collapse assessments were not in agreement. The small number of categories used for this sign resulted in a high level of expected agreement, making levels of agreement above chance more difficult to attain (Thompson and Walter, 1988). Examiner-specific Kappas were also moderate (0.45) to high (1.00).

To our knowledge, there are no published reports on interexaminer reliability of otoscopic assessments among adults. However, our results are consistent with those of Mains and Toner (1989), who report a concordance rate of 91 percent for two experienced examiners assessing the presence or absence of TM mobility in 209 ears of children who were admitted to the Belfast City Hospital for myringotomy.

Bluestone and Cantekin (1979) studied interexaminer reliability with 425 ears of children. Using three categories to otoscopically assess otitis media, two otolaryngologists agreed in their diagnosis in 82 percent of the ears.

Possibly the most comparable study to ours was carried out by Margolis et al (1979). Interexaminer reliability was assessed in the evaluation of otoscopic signs in 350 ears of young children by two pediatricians. The categories used were color (gray/pink, injected, red), light reflex (clear, blurred, absent), landmarks (distinct, blurred, absent), and air-fluid level (present, absent). The two examiners' diagnoses were also compared (normal, otitis media, serous otitis media). Overall percent agreement ranged from 50 percent for landmarks to 85 percent for air-fluid level. Kappa statistics ranged from –0.02 for air-fluid level to 0.54 for color.

In contrast to our results, Holmberg et al (1985) report on the correlation between otoscopy and otomicroscopy in 50 children who had suffered from acute otitis media about four weeks earlier. Otoscopic signs assessed by four otolaryngologists included TM color, position, mobility, and appearance. While the authors do not show specific interexaminer agreement results for otoscopy, they do note that, in cases where the TMs were judged to be abnormal, interexaminer agreement variation was great for both otoscopy and otomicroscopy. In addition, the otoscopic diagnoses of the abnormal TMs were often changed at otomicroscopy.

The present population-based study included few ears with drainage, bulging, or retracted TMs, abnormal TM color, fluid, or TM perforation. Accordingly, this was true of the quality control replications as well (see Table 1). If the finding (that interexaminer agreement is worse in the presence of abnormal TMs) of Holmberg et al (1985) can be generalized, it is likely that interexaminer agreement percentages in the present study may be higher than they would have been if more abnormal signs had been found. Kappa statistics (see Table 2) adjust for the higher expected agreement obtained in such instances.

Interexaminer Reliability: Tympanometry

The only significant interexaminer differences for tympanometry observed in the present study were for V_{ea} and TW. Examiner 3 tended to obtain the largest V_{ea} measures and examiner 2 the smallest. V_{ea} will vary somewhat depending on how deeply the probe tip is inserted into the ear canal. It is possible that examiner 3 inserted the probe tip the least distance into the ear canal and examiner 2 the farthest. A difference of 0.26 cc in absolute value represents only 29 percent of the normal V_{ea} range for adults (ASHA, 1990) and would be unlikely to affect the clinical interpretation of the results.

Similarly, the mean difference in TW of 9.33 daPa between examiner 1 and examiner 2 represents only 16 percent of the normal TW range (ASHA, 1990) and would be unlikely to affect the clinical interpretation. Other than the V_{ea} differences discussed above, this was the only case of a significant mean difference in a tympanometric measure between the examiner with the M.S. in audiology (examiner 1) and one of the other examiners. It is noteworthy that there were no significant differences between examiners for peak Y_{tm}, a measure that can be used to enhance the reliability of the classification of tympanogram shape (Margolis and Heller, 1987).

The interexaminer reliability assessed in this study is more technically a combination of interexaminer and instrument reliability. The same tympanometer was used in 38 of 45 (84%) replications, whereas the remaining 7 replications each used two tympanometers. Since the calibration of both units was checked at the beginning of each examination day, substantial tympanometer differences can be ruled out. Thus, there remains an unknown degree of test–retest reliability contributing to the interexaminer reliability we report (as would be the case with any evaluation of interexaminer reliability that uses examination instruments). Unfortunately, we know of no study of test–retest reliability of tympanometers using adult subjects.

Furthermore, we know of no reports on interexaminer reliability of tympanometry for adults. Bluestone and Cantekin (1979) had two audiologists examine 187 ears of children. The audiologists agreed on a five-step classification of tympanogram shapes in 87 percent of the ears and on a three-category classification of otitis media based on tympanometry in 97 percent of the ears. De Melker (1992) reports a Kappa statistic of .95 for the interexaminer reliability of a four-step classification of tympanogram shapes by a general practitioner and a nurse for 215 ears of schoolchildren.

A potential limitation of the current study is that only one of the three examiners had previous audiologic training. Otoscopic evaluations, whether by ASHA-certified audiologists, experienced otolaryngologists, or trained examiners, are all subject to observer variation. The lack of an accepted "gold standard" for otoscopy makes it difficult to establish the accuracy of their judgments. While the limited time frame of the present study and repetitive nature of the work made it impractical to use three experienced audiologists as examiners, all were trained to follow standardized study protocols and were monitored closely by two ASHA-certified audiologists (TLW, TST). Furthermore, examiner-specific comparisons demonstrate that the two examiners with no previous audiologic training (examiners 2 and 3) had good agreement with the examiner with the M.S. in audiology (examiner 1), who has since become ASHA certified. While the use of previously inexperienced examiners in the current study is in no way intended to advocate this approach in routine clinical practice, our results show that, with careful training and supervision, it is possible to use such examiners in a research setting.

As in the assessment of accuracy, the lack of a gold standard against which to compare the three examiners' otoscopic and tympanometric evaluations limits the assessment of validity. Good agreement between the three examiners does not necessarily prove that their otoscopic assessments and tympanometric results were valid. The examiners' thorough training and continued monitoring by two certified audiologists, however, provided some support for the expectation of reasonable validity. In addition, good agreement between the assessment of otoscopic signs and related tympanometric measures, as discussed below, provided some evidence of validity of the otoscopic assessments.

Tympanometry and Otoscopy Compared

Our study demonstrated good agreement between the assessment of otoscopic signs and related tympanometric measures. Whereas the interexaminer comparisons provided evidence that both otoscopic assessments and tympanometric measures were reliable, good agreement between the two provided some internal evidence that the otoscopic assessments were valid, using tympanometric results as the basis for validity.

The large sample from this population-based study compared to previous studies (Bluestone and Cantekin, 1979; Margolis et al, 1979; Mains and Toner, 1989; de Melker, 1992) provided added precision for summary measures and statistical tests. In addition, it allowed for comparisons of results from the examination components that involved relatively infrequent conditions, such as impacted ear canals or perforated TMs, while maintaining representation of the population under study. It may be the largest sample of adult ears ever used in research on otoscopic and tympanometric comparisons.

There was an expected relation between otoscopically observed TM perforations and measures of V_{ea}, although the number of observed perforations was small. While mean V_{ea} was higher for ears judged to have TM perforations than for ears without TM perforations, the relation failed to reach statistical significance for the left ear ($p = .099$). This was due to the larger standard deviation in the data from the three left ears judged to have TM perforations (SD = 1.63) compared to the three right ears (SD = 0.11). The three left ear values ranged from 2.9 to 6.0 cc, compared to 3.9 to 4.1 cc for the right ear.

Shanks and Shelton (1991) indicate that, in adults, all TM perforations will be accompanied by a V_{ea} greater than 2.5 cc. (This presumably refers to adults without active middle ear disease.) All six ears judged in our study to have TM perforations had V_{ea} greater than 2.5. The converse, however, is not true. All ears with V_{ea} greater than 2.5 do not necessarily have TM perforations: such participants may simply have large ear canals. And some cases of TM perforation may go unnoticed in routine otoscopy. The perforation may be very small or may be on the outer edge of the TM, where it is difficult to visualize. Finally, it is possible that selected subjects presented with an unknown history of otic surgery or otic disease that accounted for the large V_{ea} values.

Similarly, the relation between otoscopically assessed impaction and peak Y_{tm} values was reasonable. Again, perfect agreement would not be expected. A low peak Y_{tm} value could occur for other reasons besides ear-canal impaction. It could be the result of a normal but relatively stiff middle ear system, or there may exist some subclinical pathology. In addition, an ear canal could appear to be impacted even when tiny openings are present in the cerumen.

Our findings indicate that examiners can be trained to consistently and reliably assess otoscopic signs and obtain reliable tympanometric results. These results have important implications for screening programs and epidemiologic studies of hearing loss and otic diseases.

Acknowledgment. This research was supported by National Institutes of Health grant AG11099. These data were presented in part at the Society for Epidemiologic Research 27th Annual Meeting, June 1994, Miami, FL.

REFERENCES

American National Standards Institute. (1987). *Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance).* (ANSI S3.39-1987). New York: Acoustical Society of America.

American National Standards Institute. (1989). *Specification for Audiometers.* (ANSI S3.6-1989). New York: Acoustical Society of America.

American Speech-Language-Hearing Association. (1978). Guidelines for manual pure tone threshold audiometry. *ASHA* 20:297–301.

American Speech-Language-Hearing Association. (1990). Guidelines for screening for hearing impairment and middle-ear disorders. *ASHA* 32 (Suppl 2):17–24.

Bluestone CD, Cantekin EI. (1979). Design factors in the characterization and identification of otitis media and certain related conditions. *Ann Otol Rhinol Laryngol Suppl* 88(Suppl 60):13–28.

Browning GG, Gatehouse S. (1992). The prevalence of middle ear disease in the adult British population. *Clin Otolaryngol* 17:317–321.

de Melker R. (1992). Diagnostic value of microtympanometry in primary care. *Br J Gen Pract* 304:96–98.

Fields MJ, Allison RS, Corwin P, White PS, Doherty J. (1993). Microtympanometry, microscopy and tympanometry in evaluating middle ear effusion prior to myringotomy. *N Z Med J* 106:386–387.

Finitzo T, Friel-Patti S, Chinn K, Brown O. (1992). Tympanometry and otoscopy prior to myringotomy: issues in diagnosis of otitis media. *Int J Pediatr Otorhinolaryngol* 24:101–110.

Fleiss JL. (1981). Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons.

Gates A. (1986). Differential otomanometry. *Am J Otolaryngol* 7:147–150.

Holmberg K, Axelsson A, Hansson P, Renvall U. (1985). The correlation between otoscopy and otomicroscopy in acute otitis media during healing. *Scand Audiol* 14: 191–199.

Kaleida PH, Stool SE. (1992). Assessment of otoscopists' accuracy regarding middle–ear effusion. Otoscopic validation. *Am J Dis Child* 146:433–435.

Klein R, Klein BEK, Linton KLP, De Mets DL. (1991). The Beaver Dam Eye Study: visual acuity. *Ophthalmology* 98:1310–1315.

Linton KLP, Klein BEK, Klein R. (1991). The validity of self-reported and surrogate reported cataract and age-related macular degeneration in the Beaver Dam Eye Study. *Am J Epidemiol* 134:1438–1446.

Mains BT, Toner JG. (1989). Pneumatic otoscopy: study of inter-observer variability. *J Laryngol Otol* 103: 1134–1135.

Margolis CZ, Porter B, Barnoon S, Pilpel D. (1979). Reliability of the middle ear examination. *Isr J Med Sci* 15:23–28.

Margolis RH, Heller JW. (1987). Screening tympanometry: criteria for medical referral. *Audiology* 26:197–208.

Meyrick PS. (1951). The incidence of diseases of the ear, nose and throat. A survey of a remote native reserve. *S Afr Med J* 25:701–704.

Mills RP. (1987). Tympanometry and otoscopy in middle ear effusion. *Clin Otolaryngol* 12:97–101.

Moran DJ, Waterford JE, Hollows F, Jones DL. (1979). Ear disease in rural Australia. *Med J Aus* 2:210–212.

Moscicki EK, Elkins EF, Baum HF, McNamara PM. (1985). Hearing loss in the elderly: an epidemiologic study of the Framingham Heart Study Cohort. *Ear Hear* 6: 184–190.

National Center for Health Statistics. (1990). *Current Estimates from the National Health Interview Survey, 1989.* Vital and health statistics, Series 10: no. 176. DHHS publication no. (PHS) 90–1504. Washington, DC: Government Printing Office.

Ovesen T, Paaske PB, Elbrond O. (1993). Accuracy of an automatic impedance apparatus in a population with secretory otitis media: principles in the evaluation of tympanometrical findings. *Am J Otolaryngol* 14:100–104.

Pedersen CB, Zachau-Christiansen B. (1988). Chronic otitis media and sequelae in the population of Greenland. *Scand J Soc Med* 16(1):15–19.

Rudin R, Svärdsudd K, Tibblin G, Hallén O. (1983). Middle ear disease in samples from the general population. Prevalence and incidence of otitis media and its sequelae. The study of men born in 1913–23. *Acta Otolaryngol* 96:237–246.

SAS Institute Inc. (1988). *SAS/STAT User's Guide, Release 6.03 Edition.* Cary, NC: SAS Institute Inc.

Shanks J, Shelton C. (1991). Basic principles and clinical applications of tympanometry. *Otolaryngol Clin North Am* 24:299–328.

Shanks JE, Stelmachowicz PG, Beauchaine KL, Schulte L. (1992). Equivalent ear canal volumes in children pre- and post-tympanostomy tube insertion. *J Speech Hear Res* 35:936–941.

Talbott EO, Findlay RC, Kuller LH, Lenkner LA, Matthews KA, Day RD, Ishii EK. (1990). Noise-induced hearing loss: a possible marker for high blood pressure in older noise-exposed populations. *J Occup Med* 32:690–697.

Thompson WG, Walter DW. (1988). A reappraisal of the kappa coefficient. *J Clin Epidemiol* 41:949–958.

Toner JG, Mains B. (1990). Pneumatic otoscopy and tympanometry in the detection of middle ear effusion. *Clin Otolaryngol* 15:121–123.

Van Camp KJ, Margolis RH, Wilson RH, Creten WL, Shanks JE. (1986). *Principles of Tympanometry.* ASHA Monogr No. 24. Washington, DC: ASHA.

Vaughan-Jones R, Mills RP. (1992). The Welch Allyn Audioscope and Microtymp: their accuracy and that of pneumatic otoscopy, tympanometry and pure tone audiometry as predictors of otitis media with effusion. *J Laryngol Otol* 106:600–602.

Wazen JJ, Ferraro JA, Hughes R. (1988). Clinical evaluation of a portable, cordless, hand-held middle ear analyzer. *Otolaryngol Head Neck Surg* 99:348–350.