

## A Comparison of In-Person Examination and Video Otoscope Imaging for Tympanostomy Tube Follow-Up

CHRIS PATRICOSKI, M.D., JOHN KOKESH, M.D., A. STEWART FERGUSON, Ph.D.,  
KATHRYN KOLLER, R.N., M.S.N., GREG ZWACK, M.D., ELLEN PROVOST, D.O., M.P.H.,  
and PETER HOLCK, Ph.D.

### ABSTRACT

The objective of this study was to determine if video otoscope still images (640 × 480 pixel resolution) of the tympanic membrane following surgical placement of tympanostomy tubes are comparable to an in-person microscopic examination. Forty patients having undergone tympanostomy tube placement in both ears were independently examined in-person by two otolaryngologists and imaged using a video otoscope and telemedicine software package. The two physicians later reviewed images at 6 and 12 weeks. Physical examination findings and diagnosis were documented and compared for their concordance using kappa statistics. For both physicians, the intraprovider concordance between the in-person examination and the corresponding image review was high for each of the physical examination findings: Tube In 93–94% (K 0.85–0.87), Tube Patent 86–93% (K 0.74–0.85), Drainage 94–98% (K 0.42–0.66), Perforation 85–98% (K 0.40–0.84), Granulation 95–99% (K –0.01 to 0.00), Middle Ear Fluid 89–91% (K –0.03 to 0.50), and Retracted 89–94% (K 0.13–0.43). These agreement rates are similar to the normal interprovider concordance observed when two physicians independently examined the same patient in-person for physical exam findings: Tube In 96% (K 0.93), Tube Patent 94% (K 0.88), Drainage 96% (K 0.56), Perforation 90% (K 0.60), Granulation 96% (K 0.39), Middle Ear Fluid 88% (K 0.14), and Retracted 91% (K 0.43). For both physicians, the intraprovider diagnostic concordance between the in-person examination and the corresponding image review was high 79–85% (K 0.67–0.76). The interprovider diagnostic concordance for the in-person exam was 88% (K 0.81). The interprovider diagnostic concordance when two physicians independently reviewed all images was 84% (K 0.74), and 89% (K 0.80) when poor images were excluded. This study demonstrates that physician review of video otoscope images is comparable to an in-person microscopic examination. Store-and-forward video otoscopy may be an acceptable method of following patients post-tympanostomy tube placement.

## INTRODUCTION

**V**ISUALIZATION of the tympanic membrane is an important aspect of primary care and otolaryngology. Otitis media (OM) accounts for 17–18% of all office visits under the age of 5.<sup>1</sup> The diagnosis and treatment for OM is estimated at \$5.3 billion<sup>2</sup> and the expenses multiply when one factors in the loss of parental work, transportation, and other indirect costs.<sup>3</sup> The most common indications for tympanostomy tube surgery in children are recurrent acute otitis media (three or more episodes of OM in a 6 month period or 4 episodes during a 1 year period) and chronic otitis media with effusion (bilateral OME which has been unresponsive to non-surgical therapy for three months or more), with a documented associated hearing loss.<sup>4,5</sup>

Myringotomy with tube insertion is the leading procedure for children under 15 years old, with 512,000 surgeries being performed in ambulatory and inpatient settings in 1996.<sup>6</sup> Post-surgical follow-up of these children typically involves examination at 1 month (or earlier) and then at intervals no longer than 6 months.<sup>7</sup> Otolaryngologists, however, have individual preferences; most follow-up within 1 month (97%), and the second and subsequent visits occurs at 3 months (29%), 4 months (25%) or 6 months (37%).<sup>8</sup> Patient history and otoscope or microscope examination are used in determining tube patency and function. With the large number of tympanostomy tubes (TT) (or pressure equalization tubes) placed in children throughout rural U.S., the follow-up is challenging. It is possible that telemedicine could play an important role in postsurgical follow-up of TT placement, especially in settings where access to quality follow-up care is difficult or limited.

Technology-assisted direct visual inspection using the otoscope is one of the proposed areas of telemedicine evaluation.<sup>9</sup> If telemedicine is to be incorporated into a protocol for follow up of pressure equalization tube surgery, then it needs to be established that video otoscope images are an acceptable replacement for the in-person microscope examination. Some studies have adequately addressed image quality and diagnostic accuracy using inter-provider

and intra-provider comparisons in the fields of teleophthalmology<sup>10–15</sup> and teledermatology.<sup>16–29</sup> Few studies have addressed visualization and image capture in the other telemedicine specialties. Telemedicine concordance studies in otolaryngology are limited to video endoscopy. The vocal cords and the nasopharyngeal anatomy have been analyzed in two studies.<sup>30,31</sup> There have been no comparison, evaluation, or concordance studies on digital image capture of the tympanic membranes for post-surgical follow-up of TT. In this study, a store-and-forward software product is utilized with a video otoscope to capture images of tympanic membranes. The purpose of this study is to determine whether video otoscope images of tympanic membranes following TT placement are comparable to an in-person microscope examination. If concordance is acceptable to otolaryngologists, then these still images may be used for routine follow-up after TT placement.

## MATERIALS AND METHODS

The protocol was reviewed and approved by the Indian Health Service Alaska Area Institutional Review Board, the Alaska Native Medical Center Joint Operating Board, Southcentral Foundation Board, Alaska Native Tribal Health Consortium Board, Manililaq Association and the Indian Health Service National Institutional Review Board. Informed consent was obtained for each participant. Patients having had TT placement since April 1, 2000 were identified from the Alaska Native Medical Center (ANMC) operating room database in Anchorage, Alaska. ANMC provides otolaryngology specialty care for all Alaskan Natives within the state of Alaska. From this list, patients from the Anchorage ( $N = 20$ ) and Kotzebue ( $N = 20$ ) Service Units were recruited to voluntarily participate in the study. Minors were accompanied by parent or guardian and were interviewed to verify demographics, dates of tube insertion, dates of follow-up, and to inquire about significant post-surgical events: drainage, ear infection, hearing problems, pain, and if the tubes had fallen out. Then, the encounter proceeded similar to the routine normally used in the ENT

specialty clinic with the addition of a second in-person exam by a physician and a video otoscope imaging exam by a non-physician.

Two board certified otolaryngologists using a Zeiss Otology examination microscope examined participants independently. During this in-person examination, the physician documented the Physical Exam, Assessment, and Plan using a standard follow-up form. The Physical Exam portion of the form included right and left check lists noting: Tube in the TM, Tube Patent, Drainage, Perforation, Granulation Tissue, Fluid in Middle Ear, Retracted TM, and Other. It was also noted if Cerumen Removal and/or Pneumato-otoscopy were required for assessment. The Assessment portion of the form included a checklist of seven possible diagnosis: Intact and Functional Tube, Nonfunctional Tube, Tube Extruded/Normal TM, Perforation and Dry, Perforation and Draining, Otitis Media, and Retracted TM. In Kotzebue, the physicians first saw the patient, removed any cerumen obstructing the view, documented findings, then sent the patient for imaging; in Anchorage, cerumen was not removed prior to imaging. Patient forms and images were assigned a random investigation code. Patients completed a satisfaction survey at the end of the visit.

The participant was then seen by a community health aide, nurse, or nurse practitioner who had received training in video otoscopy. Training included hand position, focusing, cleaning the tip, setting white balance, using image enhancement, lighting, maneuvering the tip, and capturing images. The staff member used the video otoscope to illuminate and capture a still image of the tympanic membrane of each ear. For younger children, it was necessary to have a nurse assist in holding the child. Staff was allowed to discard a poor image and repeat the procedure to obtain a new image. Regardless of wax or an uncooperative patient, images were required for all ears. Images were saved as 24-bit color JPEG images with  $640 \times 480$  pixel resolution, and typically compressed at a 13:1 ratio. Equipment included the AMD/Welch Allyn 300S Imaging and Illumination Platform with Solarc lamp, corresponding fiber optic bundle, camera cable, camera head, c-mount 45 mm optical coupler (AMD

2450), and otoscope ENT probe (AMD 2015). The image was transferred using S-Video output to a workstation with an Integral Technologies FlashBus MV Pro video capture board. Images were captured using a custom web-based telehealth software package developed by the Alaska Federal Health Care Access Network–AFHCAN Project, and stored on a server for later viewing.

Each ear was imaged on the same day as the in-person examination. The otolaryngologists reviewed the video otoscope images independently at 6 weeks and 12 weeks following the in-person examination. The face-to-face in-person examination is hereafter referred to as Exam<sub>0</sub>. The first review of images at 6 weeks is Review<sub>1</sub>, and the second review of images at 12 weeks is Review<sub>2</sub>. The order of cases was randomized and differed between the various times Exam<sub>0</sub>, Review<sub>1</sub>, and Review<sub>2</sub>. The otolaryngologists used a review form during the image reviews (Review<sub>1</sub> and Review<sub>2</sub>) to document the Physical Exam and Assessment based on their interpretations of the digital images. Each physician separately reviewed the images on a workstation with a Viewsonic VP150m LCD Viewpanel and utilized the AFHCAN Telemedicine Software to pull up and display the images. Basic historical information such as the dates of tube placement and significant post-surgical events were made available at the time of image review. During Review<sub>1</sub> and Review<sub>2</sub>, the physicians also rated various parameters using a Likert five-point scale (such as overall image quality and confidence in diagnosis).

Data analysis included examination of inter-provider and intraprovider concordance. Interprovider concordance is a comparison of the results of one physician with another. This was calculated comparing the two physicians' results for the face-to-face in-person examination (Exam<sub>0</sub>) and also for their interpretation of images (Review<sub>1</sub> and Review<sub>2</sub>). The interprovider concordance during Exam<sub>0</sub> is a critical measurement and is often used throughout this paper as a benchmark or gold standard to compare with other agreement rates obtained. For this reason, the notation IE<sub>0</sub>C is used to refer to the "Interprovider Exam<sub>0</sub> Concordance".

In some instances, the data from the two im-

age review sessions (Review<sub>1</sub> and Review<sub>2</sub>) were aggregated and reported as Review<sub>1,2</sub>. Intraprovider concordance is a comparison of a single physician's results from two different points in time. For example, intraprovider concordance was calculated for Exam<sub>0</sub> versus Review<sub>1</sub>, Exam<sub>0</sub> versus Review<sub>2</sub>, and Review<sub>1</sub> versus Review<sub>2</sub>. Kappa statistics (K), confidence intervals, and other statistical analysis were determined where appropriate. It should be noted that the kappa statistic, as a measure of agreement, could loosely be defined to indicate "very good" agreement (K > 0.8), "good" agreement (K = 0.61–0.80), "moderate" agreement (K = 0.41–0.60), "fair" agreement (K = 0.21–0.40), and "poor" agreement (K = 0.01–0.20).

## RESULTS

Forty Alaska Natives, 1–21 years of age, participated in the study. The first 20 participants were recruited from the Kotzebue region and seen at the Manililaq Health Center in Kotzebue, Alaska; the second 20 participants were recruited from the Anchorage region and seen at the Alaska Native Medical Center in Anchorage, Alaska. The patients had a history of TT placement within 1.5 years prior to Exam<sub>0</sub>. Exam<sub>0</sub> resulted in 160 evaluations because there were 40 participants, or 80 ears, examined by two otolaryngologists. Review<sub>1</sub> resulted in 160 evaluations as images from 80 ears were reviewed by two otolaryngologists. Review<sub>2</sub> similarly resulted in 160 evaluations. The combined review process, Review<sub>1,2</sub>, represents 320 image reviews.

### Overall image quality

Physicians were asked to rate the overall image quality for each ear during Review<sub>1</sub> and Review<sub>2</sub> (Table 1). Ratings for the 80 video otoscope ear images resulted in a total of 160 responses from each physician. Overall, 261 of the 320 total responses (82%) rated the images as "Adequate" or better. Of the 59 images rated as "poor" or "very poor," 49 occurred in children younger than 4 years old (Table 2). This accounts for 83% of the bad images, although children younger than 4 years old accounted for 43% (18 of 40) of the total of number of subjects. One-year-old patients had the highest percentage of poor images (66%).

### Physical examination

Descriptions in the physical examination are the most useful parameter for evaluating physician concordance of the tympanic membrane. The physical exam utilized seven (yes/no) descriptors: "Tube In," "Tube Patent," "Drainage," "Perforation," "Granulation," "Middle Ear Fluid," and "Retracted." During Exam<sub>0</sub>, the IE<sub>0</sub>C was 88% or higher for all descriptors and this was closely paralleled by intraprovider concordance (Exam<sub>0</sub> versus Review<sub>1</sub> and Exam<sub>0</sub> versus Review<sub>2</sub>) which was 85% or higher. The ranges of intraprovider concordance (maximum and minimum) are shown in Figure 1, and are generally seen to lie close to, or overlap, the IE<sub>0</sub>C. Excluding the ears for which the image quality was rated as very poor (1) or poor (2) generally resulted in a slightly higher concordance between Exam<sub>0</sub> and Review<sub>1,2</sub>. This was especially true for the de-

TABLE 1. OVERALL IMAGE RATING

	Blank	1 = Very poor	2 = Poor	3 = Adequate	4 = Good	5 = Excellent
Provider AA	0	11	26	40	43	40
Provider BB	1	9	12	28	46	64
Total	1 (0%)	20 (6%)	38 (12%)	68 (21%)	89 (28%)	104 (33%)
Aggregates	1 (0%)	58 (18%)			261 (82%)	

Results for each provider, aggregate counts, and percentages are shown for the physician ratings of overall image quality during Review<sub>1</sub> and Review<sub>2</sub>. The two physicians are referred to as "AA" and "BB." A total of 160 images (or ears) were reviewed by each physician (2 ears per patient, 40 patients, 2 image review sessions). The final rows in the table indicate the aggregate numbers and percentages when image quality is broken into two categories—"poor" or worse, and "adequate" or better.

TABLE 2. PATIENT AGE DISTRIBUTION AND OVERALL IMAGE QUALITY

Patient age (years)	Number of patients	Images rated "poor" or worse	Images rated "adequate" or better
1	4 (10%)	21 (66%)	11 (34%)
2	9 (23%)	20 (28%)	52 (72%)
3	5 (13%)	8 (20%)	32 (80%)
4-5	7 (17%)	0 (0%)	55 (98%)
6-8	8 (20%)	2 (3%)	62 (97%)
9-12	4 (10%)	0 (0%)	32 (100%)
13-16	2 (5%)	7 (44%)	9 (56%)
14-21	1 (2%)	0 (0%)	8 (100%)

The distribution of patient ages at the time of Exam<sub>0</sub> is shown, with the percentage of the overall patient population accounted for in this age range. The number of "poor" or worse images, and also "adequate" or better images, is shown for each patient age category. The percentage of images is the portion of images for that age category that were "poor" or worse, or "adequate" or better. Note that the single blank image rating occurred in the 4-5-year-old category, hence the percentages of images do not add to 100%.

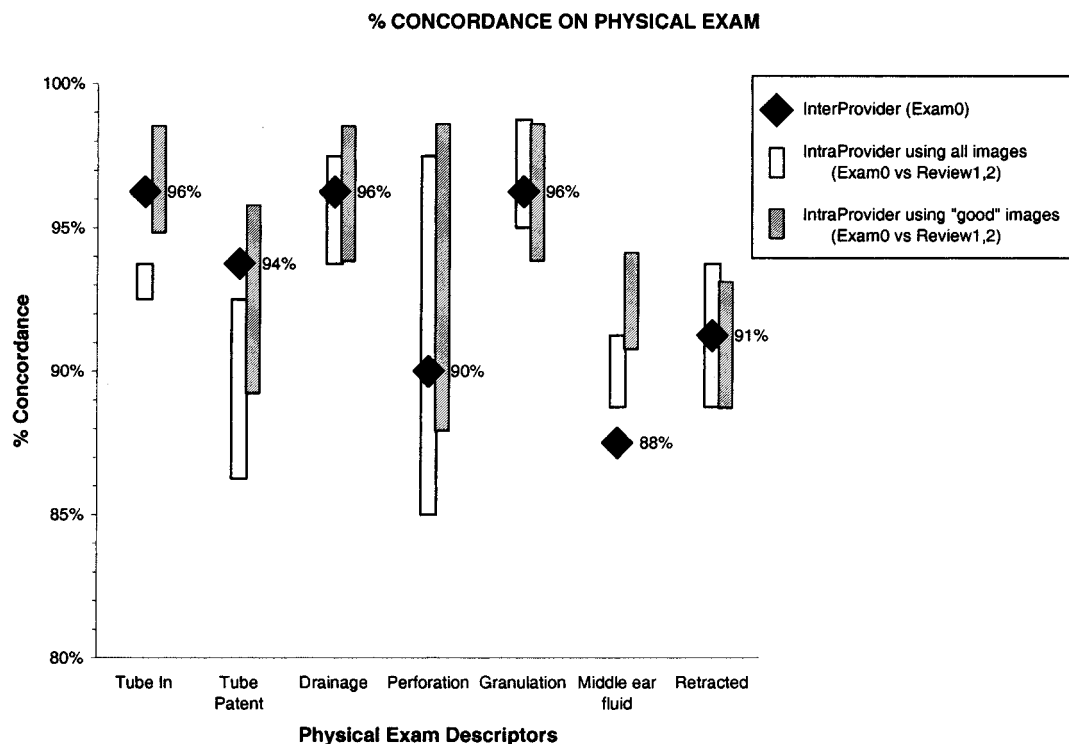


FIG. 1. Physical exam agreement. Percent concordance on physical exam descriptors. The black diamonds indicate the percent concordance between providers during the in-person exam (Exam<sub>0</sub>). The white bar indicates the range of concordance for individual physicians (intraprovider) between their descriptors at Exam<sub>0</sub> and their descriptors during the image review (Review<sub>1</sub> and Review<sub>2</sub>). The range bars are therefore the minimum and maximum of four values (two physicians each with Exam<sub>0</sub> versus Review<sub>1</sub> and Exam<sub>0</sub> versus Review<sub>2</sub>). The shaded bars represent a similar min/max range for intraprovider concordance when ears with Poor or Very Poor image quality are removed from consideration.

scriptors "Tube In," "Tube Patent," and "Middle Ear Fluid."

Concordance values for the physical exam are shown in Table 3. The IE<sub>0</sub>C value was determined for each physical exam finding: Tube In 96% (K 0.93), Tube Patent 94% (K 0.88), Drainage 96% (K 0.56), Perforation 90% (K 0.60), Granulation 96% (K 0.39), Middle Ear Fluid 88% (K 0.14), and Retracted 91% (K 0.43). These values represent the interprovider concordance observed when two physicians saw the same patient in-person. Interprovider concordance when physicians used images was also ≥ 90% for all physical exam findings in Review<sub>1</sub>, Review<sub>2</sub>, and Review<sub>1,2</sub>.

The intraprovider concordance values were 85–100%, and most physical exam findings were 91% or higher. For both physicians, the intraprovider concordance between their in-person examination and their image reviews was high for each of the physical examination

findings: Tube In 93–94% (K 0.85–0.87), Tube Patent 86–93% (K 0.74–0.85), Drainage 94–98% (K 0.42–0.66), Perforation 85–98% (K 0.40–0.84), Granulation 95–99% (K -0.01 to 0.00), Middle Ear Fluid 89–91% (K -0.03 to 0.50), and Retracted 89–94% (K 0.13–0.43).

The kappa statistics associated with the concordance results are also listed in Table 3. The kappa values generally indicated agreement for most categories of interprovider and intraprovider concordance. However, the kappa statistic varied greatly depending on the descriptor. For example, the kappa values indicate there was very good to good agreement between providers and also within each provider's physical exam for the descriptors Tube In and Tube Patent. There was moderate to fair agreement when Drainage or Perforation was described, and the least agreement when Retracted or Granulation or Middle Ear Fluid was described. A Chi Square test for equality

TABLE 3. KAPPA VALUES AND PERCENT CONCORDANCE FOR PHYSICAL EXAMINATION FINDINGS

	<i>Tube in</i>	<i>Tube patent</i>	<i>Drainage</i>	<i>Perforation</i>	<i>Granulation</i>	<i>Middle ear fluid</i>	<i>Retracted</i>
<b>Interprovider</b>							
AA vs. BB							
Exam <sub>0</sub>	0.93 (96%)	0.88 (94%)	0.56 (96%)	0.60 (90%)	0.39 (96%)	0.14 (88%)	0.43 (91%)
Review <sub>1</sub>	0.93 (96%)	0.90 (95%)	0.75 (99%)	0.66 (94%)	0.00 (99%)	0.23 (90%)	0.27 (94%)
Review <sub>2</sub>	0.95 (98%)	0.88 (94%)	0.65 (96%)	0.72 (95%)	1.00 (100%)	0.49 (91%)	-0.03 (91%)
Review <sub>1,2</sub>	0.94 (97%)	0.89 (94%)	0.68 (98%)	0.69 (94%)	0.00 (99%)	0.38 (91%)	0.12 (93%)
<b>Intraprovider for provider AA</b>							
Exam <sub>0</sub> vs. review <sub>1</sub>	0.85 (93%)	0.79 (89%)	0.66 (98%)	0.40 (85%)	0.00 (95%)	0.50 (91%)	0.43 (94%)
Exam <sub>0</sub> vs. review <sub>2</sub>	0.85 (93%)	0.74 (86%)	0.42 (94%)	0.49 (86%)	0.00 (95%)	0.42 (89%)	0.20 (91%)
Review <sub>1</sub> vs. review <sub>2</sub>	0.95 (98%)	0.95 (98%)	0.56 (96%)	0.86 (98%)	1.00 (100%)	0.82 (98%)	0.49 (98%)
<b>Intraprovider for provider BB</b>							
Exam <sub>0</sub> vs. review <sub>1</sub>	0.87 (94%)	0.83 (91%)	0.59 (98%)	0.73 (95%)	-0.01 (98%)	-0.03 (90%)	0.14 (89%)
Exam <sub>0</sub> vs. review <sub>2</sub>	0.87 (94%)	0.85 (93%)	0.55 (96%)	0.84 (98%)	0.00 (99%)	0.17 (90%)	0.13 (89%)
Review <sub>1</sub> vs. review <sub>2</sub>	0.93 (96%)	0.81 (90%)	0.49 (96%)	0.79 (96%)	0.00 (99%)	0.46 (91%)	0.37 (93%)

Kappa values are shown (with percent concordance in parenthesis) for both interprovider and intraprovider comparisons.

of kappa values of intraprovider agreement and interprovider agreement on each physical examination descriptor did not find statistically significant differences ( $p > 0.05$ ). This indicated that agreement between a physician's findings from a physical exam in-person and using an image review is similar to the agreement observed between two providers conducting in-person exams.

*Diagnosis*

Interprovider diagnostic concordance at Exam<sub>0</sub> (IE<sub>0</sub>C) was 88% (K 0.81) as shown in Table 4. The interprovider diagnostic concordance when two physicians viewed images was 84% (K 0.74), and 89% (K 0.80) when poor images were excluded. For both physicians, the

intraprovider diagnostic concordance between their in-person examination and their image reviews was high 79–85% (K 0.67–0.76), and 77–88% (K 0.65–0.78) when poor images were excluded.

Comparing the concordance for each provider between Review<sub>1</sub> and Review<sub>2</sub> provided a measure of the reproducibility of the results between the different image review sessions. The intraprovider comparison for Review<sub>1</sub> versus Review<sub>2</sub> ranged from 84% to 90% with a kappa value indicating “good” to “very good” agreement of the results between image review sessions. Restricting the image review to ears with adequate or better images improved the concordance slightly to a range of 86% to 93%.

The methodology used to measure diagnosis required physicians to select a single descriptor. The above results indicate good agreement or better using this multiple choice selection process. It is instructive to look at the outliers—the combinations of descriptors—which were not the same. These are summarized in Tables 5A–C. During Exam<sub>0</sub>, for example, the physicians agreed with one another's diagnosis 69 times and there were 11 discrepancies (14%). This can be seen in Table 5A by summing the elements on the diagonal (69) for the number of concordant combinations out of the total number of possibilities in the lower right corner (80).

The outliers were scattered among categories. With respect to the image reviews, physicians agreed with one another's image diagnosis 136 times, and there were 24 discrepancies (15%) (Table 5B). Of these outliers, 15 arose when the TT was extruded and the providers disagreed as to the presence of otitis media, perforation, or retraction. Three times there was disagreement as to whether TT was extruded or intact. Restricting the image reviews to Adequate images or better (121 total), the providers agreed with one another 110 times and there were 11 discrepancies (9%) (Table 5C). Eight of these outliers arose when the TT was extruded and the providers disagreed as to the presence of otitis media, perforation, or retraction. One time there was disagreement as to whether the tube was extruded

TABLE 4. KAPPA VALUES AND PERCENT CONCORDANCE FOR DIAGNOSIS (PURPOSE OF VISIT)

	<i>All images</i>	<i>“Good” images</i>
Interprovider Exam	0.81 (88%)	
Review <sub>1</sub>	0.78 (86%)	0.84 (91%)
Review <sub>2</sub>	0.71 (81%)	0.77 (86%)
Review <sub>1,2</sub>	0.74 (84%)	0.80 (89%)
Intraprovider for provider AA		
Exam <sub>0</sub> vs. review <sub>1</sub>	0.76 (85%)	0.78 (88%)
Exam <sub>0</sub> vs. review <sub>2</sub>	0.69 (80%)	0.69 (82%)
Review <sub>1</sub> vs. review <sub>2</sub>	0.82 (89%)	0.84 (91%)
Intraprovider for provider BB		
Exam <sub>0</sub> vs. review <sub>1</sub>	0.67 (79%)	0.73 (84%)
Exam <sub>0</sub> vs. review <sub>2</sub>	0.67 (79%)	0.65 (77%)
Review <sub>1</sub> vs. review <sub>2</sub>	0.78 (86%)	0.78 (87%)

Kappa values are shown (with percent concordance in parenthesis) for both interprovider and intraprovider comparisons. The column labeled “All images” is the result when all images are included. The column labeled “Good images” is the result when only images rated as “adequate” or “better” are included. Note: Images are not involved in the Interprovider Exam<sub>0</sub> (in-person examination); therefore, the results are shown as a separate column.

TABLE 5. AGREEMENT AND OUTLIERS FOR DIAGNOSIS (PURPOSE OF VISIT)

Table 5A. Exam <sub>0</sub>								Table 5B. Review <sub>1,2</sub>									
AA assessment								AA assessment									
BB assessment								BB assessment									
	Intact & functional tube	Nonfunctional tube	Otitis media	Perforation, draining	Perforation, dry	Retracted TM	Tube extruded/normal TM		Intact & functional tube	Nonfunctional tube	Otitis media	Perforation, draining	Perforation, dry	Retracted TM	Tube extruded/normal TM		
Intact and functional tube	40	1	0	0	0	0	1	42	Blank	0	0	0	1	0	0	1	2
Nonfunctional tube	1	1	0	0	0	0	1	3	Intact and functional tube	84	0	1	0	0	0	0	85
Otitis media	1	0	1	0	0	0	0	2	Nonfunctional tube	1	1	0	0	0	0	0	2
Perforation, draining	0	0	1	0	0	0	0	1	Otitis media	0	1	5	0	0	0	4	10
Perforation, dry	0	0	0	0	7	0	0	7	Perforation, draining	0	0	0	1	0	0	0	1
Retracted TM	0	0	1	0	0	2	2	5	Perforation, dry	0	0	0	1	9	0	2	12
Tube extruded/normal TM	0	0	2	0	0	0	18	20	Retracted TM	0	0	0	0	0	1	3	4
	42	2	5	0	7	2	22	80	Tube extruded/normal TM	2	1	3	1	1	1	35	44
										87	3	9	4	10	2	45	160

Table 5C. Review<sub>1,2</sub> (with Adequate or Better Images Only)

AA assessment							
BB assessment							
	Intact & functional tube	Nonfunctional tube	Otitis media	Perforation, draining	Perforation, dry	Retracted TM	Tube extruded/normal TM
Intact and functional tube	72	0	0	0	0	0	72
Nonfunctional tube	1	0	0	0	0	0	1
Otitis media	0	0	3	0	0	0	4
Perforation, draining	0	0	0	0	0	0	0
Perforation, dry	0	0	0	1	6	0	9
Retracted TM	0	0	0	0	0	1	3
Tube extruded/normal TM	1	0	2	0	0	1	28
	74	0	5	1	6	2	33
							121

Two physicians examining the same ear during an inpatient exam, or the same images during an image review show the distribution of results in a matrix for all possible combinations of "Purpose of Visit." Values on the diagonal indicate the number of ears with concordance between the two physicians. Table 5A indicates the results for the in-person exam only (Exam<sub>0</sub>). Table 5B indicates the results for all image reviews (Review<sub>1,2</sub>). Table 5C restricts the image reviews to only ears in which the images were rated adequate or better.

or intact and functional in the tympanic membrane.

Evaluation questions

Physicians were asked to respond to several evaluation questions related to image quality,

confidence in diagnosis, and applicability. Responses were rated on a Likert scale of one through five. Confidence ratings are shown in Table 6. Seventy-three percent of the time physicians were Confident to Very Confident in their diagnosis using the video-otoscope im-

TABLE 6. EVALUATION QUESTION, CONFIDENCE IN DIAGNOSIS

Table 6A. "Please rate how confident you are in the diagnosis/assessment using telemedicine in this case"

	Blank	1 = Not confident at all	2	3 = Somewhat confident	4	5 = Very confident
Provider AA	0	5	7	14	33	21
Provider BB	1	2	4	10	23	40
Total	1 (1%)	7 (4%)	11 (7%)	24 (15%)	56 (35%)	61 (38%)

Table 6B. "Please rate how confident you are in the diagnosis/assessment using telemedicine in this case"

	Blank	1 = Not confident at all	2	3 = Somewhat confident	4	5 = Very confident
Provider AA	0	1	4	12	33	21
Provider BB	2	0	2	5	22	40
Total	2 (1%)	1 (1%)	6 (4%)	17 (12%)	55 (39%)	61 (43%)

Table 6A shows the responses to the Likert scale evaluation questions. Responses were provided during Review<sub>1</sub> and Review<sub>2</sub> with one response per physician per patient at T1 and again at T2 (total of 80 responses per provider). Table 6B only includes images with adequate or better image quality.

ages (Table 6A). There were some differences in the distribution of answers between the two providers, with provider AA most often responding Confident and BB responding Very Confident. Provider AA had a mean confidence score of 3.7, whereas Provider BB had a mean score of 4.2. Confidence improved slightly when poor and very poor images were removed (Table 6b). When considering only those answers for which the physician rated the image quality as Adequate or higher, physicians were Confident or higher in their diagnosis 82% of the time. The overall confidence

rating for the physicians rose from 3.96 (Confident) with all images to 4.2 when poor images were excluded.

Overall image quality ratings are shown in Table 1. The close correlation between confidence and image quality is further shown in Table 7. Note that the higher confidence levels are bunched with the higher image quality values. It is interesting that 42% of the responses from the physicians had confidence exactly matching image quality (the sum of the values along the diagonal), and 85% of the responses had confidence and image quality differing by

TABLE 7. IMAGE QUALITY VERSUS CONFIDENCE IN DIAGNOSIS

	Image quality <sup>a</sup>				
	1 = Very poor	2 = Poor	3 = Adequate	4 = Good	5 = Excellent
Confidence in diagnosis <sup>b</sup>					
1 = Not confident at all	4%	1%	0%	1%	0%
2	1%	3%	1%	1%	1%
3 = Somewhat confident	1%	6%	5%	2%	0%
4	0%	2%	8%	10%	11%
5 = Very confident	0%	0%	7%	14%	20%

<sup>a</sup>"Please rate the image quality overall (Overall Image Rating)."

<sup>b</sup>"Please rate how confident you are in the diagnosis/assessment using telemedicine in this case."

Percentage of occurrence of a specific rating for confidence versus a rating for image quality (n = 317). Each cell in the table provides the percentage of times a physician provided that combination of image quality and confidence rating. (This table summarizes the 317 occurrences out of 320 possible. In two instances the Diagnosis rating was left blank, and in one instance the Image Quality rating was blank.)

TABLE 8. EVALUATION QUESTION, APPLICABILITY: "TO WHAT EXTENT DOES THIS TELEMEDICAL CONSULT SEEM TO PROVIDE THE SAME INFORMATION AS AN IN-PERSON CONSULT?"

		1 = <i>Insufficient</i>	2 = <i>Poor</i>	3 = <i>Fair</i>	4 = <i>As good as</i>	5 = <i>Better than in-person</i>
Provider AA	0	4	10	24	34	8
Provider BB	0	2	6	23	48	0
Total	0	6	16	47	82	8
	(0%)	(4%)	(10%)	(29%)	(52%)	(5%)

no more than 1 (lying within one square of the diagonal). Only 3% of the responses had a disparity between image quality and confidence of 3 or more.

Applicability ratings are shown in Table 8. The two physicians believed that a telemedicine consultation was Fair, As Good As, or Better Than an in-person consultation 86% of the time. The other 14% of the time, the physicians considered the telemedicine case to be Poor or Insufficient compared to an in-person exam. When poor images were excluded, physicians believed the consultation was Fair or better 96% of the time.

## DISCUSSION

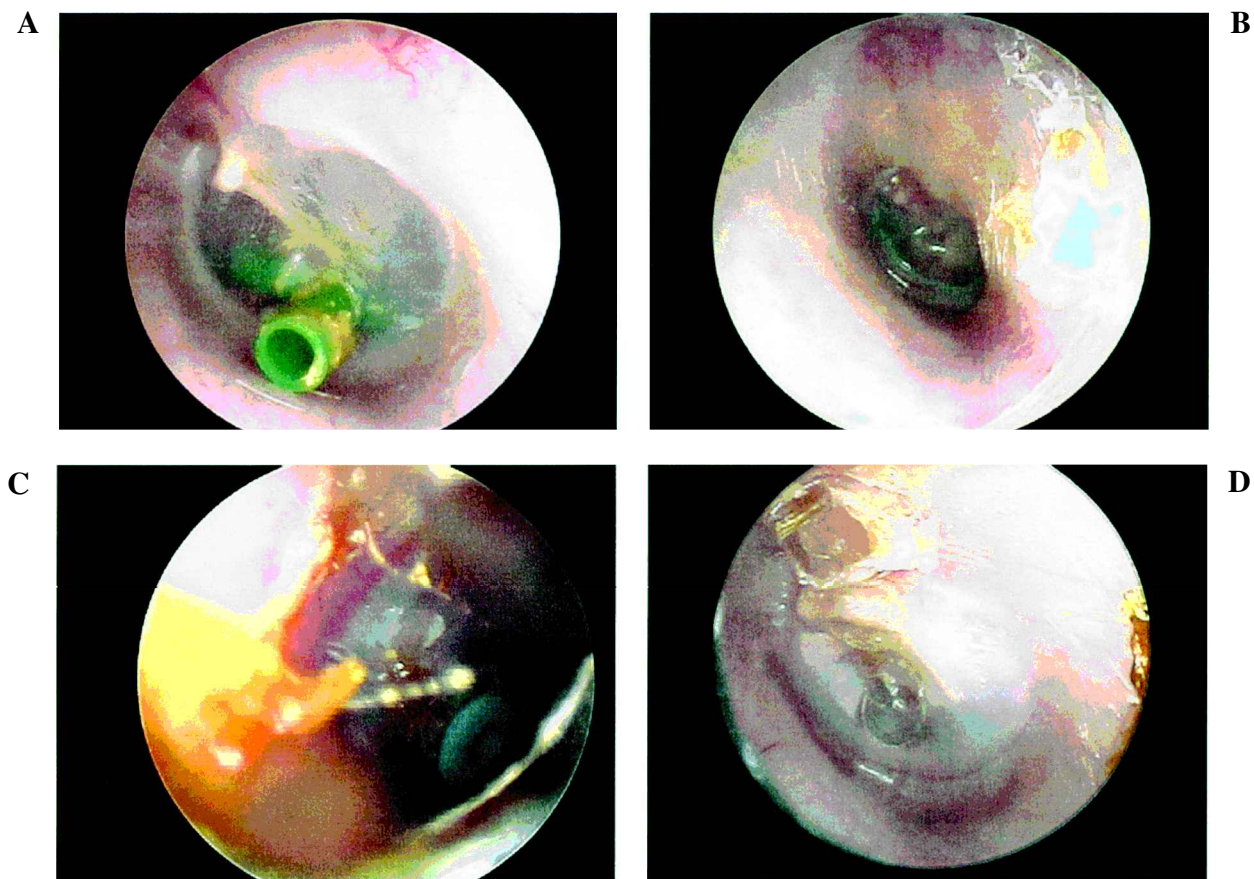
While there is a long history in the medical field of tympanic membrane photography,<sup>32-34</sup> telescopic video otoscopy is a relatively new method of ear imaging,<sup>35,36</sup> which continues to evolve.<sup>37</sup> Clinicians are finding increasing value in using video otoscopy for teaching students and instructing patients.<sup>38-40</sup> It is now finding its way into daily ENT and audiology practice<sup>39-41</sup> and being used more often in telemedicine applications.<sup>31,40,42-49</sup> The video otoscope has even been used for quantitative analysis of the tympanic membrane to determine and monitor disease progression.<sup>39</sup> While the usage is expanding, little has been published regarding the reliability of video otoscope imaging compared to an in-person examination. Imaging of the eardrum has not been studied for diagnostic reliability in telemedicine applications.

In this study, using video otoscope still images with 640 × 480 pixel resolution, the intraprovider concordance rates were 85-99% for physical exam findings and 79-85% for diag-

nosis. Kappa statistics generally indicated good agreement, not likely due to chance. Further statistical analysis indicated that agreement between a physician's findings from a physical exam in-person and using an image review is similar to the agreement observed between two providers conducting in-person exams.

However, kappa values did vary. It may seem surprising that the concordance was high for some physical exam descriptors yet the corresponding kappa value was low. This is explained by the descriptors being rarely selected. The high uniformity of No responses resulted in high concordance; however, the kappa statistic is lowered by the uniformity of the response. The descriptors Tube In and Tube Patent had both physicians saying Yes on 55% and 53% of the cases, respectively—and these exhibited the highest kappa values. Perforation was the third most common descriptor, but only had both physicians saying Yes 9% of the time and had a lower kappa. None of the other descriptors exceed a 4% rate of agreement on a Yes answer, and consequently had very low kappa values despite high levels of concordance.

The significance of the disagreements was informally assessed with a case-by-case physician evaluation and discussion. Most disagreements between providers for the in-person exams and image analysis were related to differences in terminology describing similar conditions. For example, an ear with otitis media may also be retracted. Disagreements occurred when each provider chose a different descriptor despite agreeing on the overall assessment. A few of the outliers were clinically relevant. The remaining outliers were plausible disagreements in interpretations of the image. For example, eight of the 11 outliers arose when the TT was extruded and the providers disagreed



**FIG. 2.** Video otoscope images. **A.** Physicians agreed this left TM had an intact, patent, functioning tympanostomy tube. They were confident to very confident in their diagnosis. **B.** Physicians disagreed as to whether this left TM was normal, retracted, or demonstrated otitis media (post tube extraction). They were somewhat to very confident in their diagnosis. **C.** Physicians disagreed as to whether the tympanostomy tube in this right TM was extruded or intact. They were not confident in their diagnosis. **D.** Physicians disagreed if this was a normal left TM with tube extruded or a dry perforation. It is difficult to determine if there is a thin membrane covering the previous location of tympanostomy tube. Physicians were confident in their diagnosis.

as to whether the intact drum and middle ear were showing otitis media, perforation, or retraction. Figure 2A demonstrates an image for which the providers readily agreed and confidence was high. Physicians noted that the degree of tympanic membrane retraction could be difficult to distinguish with still images. Figure 2B demonstrates disagreement as to whether the tympanic membrane is normal, retracted, or has (serous) otitis media. Clinicians soon recognized the inability to assess eardrum mobility using a store-and-forward system of still images. They felt that the inability to perform tympanic insufflation or tympanometry hindered their capabilities of diagnostic accuracy. Insufflation and tympanometry help to determine tympanic membrane mobility, retraction, perforation and tube patency. Figure 2C dem-

onstrates the one situation where there was disagreement as to whether the tube was extruded or intact. Figure 2D is an example where it was difficult to assess whether the tympanic membrane had a perforation or was intact. This may be explained by difficulty in distinguishing between small perforations and perforations (which heal) with monomeric membranes.

Overall, the biggest source of problems seemed to be those cases where an intact tympanic membrane looked relatively normal but the presence of fluid or negative pressure was in question. For in-person exams, this is not generally a determination that can be made by visual assessment alone. Or, if it can be made on visual assessment, it takes the kind of depth perception afforded by live interactive binocular vision. In clinical settings, these cases are

the kinds where the clinician would insufflate to assess mobility or obtain a tympanogram to assess compliance and pressure.

Of course, image quality affected most aspects of the study. When Poor and Very Poor images were not included in the results, clinical outliers were reduced from 24 to 11, and in almost all cases, concordance and kappa values increased slightly. In fact, the otolaryngologists noted that they would refuse consultation on those cases with poor quality images, so the concordance should be higher with real clinical consultations. Physicians were confident in their diagnosis of the tympanic membrane using the images. Their confidence improved with image quality, and there was a tight relationship between image quality ratings and confidence ratings on each case. In general, the youngest patients were the most difficult to image. This limited the overall good images from one and two year olds and brings into question the utility of using video otoscopy on this age group for follow up. One wonders if physicians would end up being less confident for the young pediatric population due to image quality. Regarding applicability, it was the providers' opinion that most of the time the image provided the same information as an in-person exam. This could prove important as physicians ponder acceptance of a new technology in the clinical setting.

There were limitations to this study. The study was performed in two small hospitals with trained imagers. It is not known how this application will fair with less experienced clinicians in the field. Additionally, some patients had earwax removed microscopically prior to imaging; not likely to occur at a site originating a telemedicine case. Finally, the sample size was large enough to determine overall concordance, but too small to assess if specific outliers would delineate any consistent clusters. The term "concordance" is used because it is clinically relevant and common in the literature; the term "inter-rater agreement" may be more descriptive. Other comparative terms in the literature include interobserver agreement, reliability, and accuracy. Physicians did not have serous otitis media as a choice for diagnosis and occasionally wrote it in (and then selected otitis media on the form). Furthermore, it is un-

certain if these findings could be extrapolated to other tympanic membrane disorders or post-surgical states, such as cholesteatoma or tympanoplasty.

## CONCLUSION

Most telemedicine applications using the video otoscope are real time videoconferencing interactions. In Alaska, store-and-forward telemedicine is being used to allow the consulting physician to review images asynchronously. Store-and-forward telemedicine also requires less transmission bandwidth. This study demonstrates the potential of store-and-forward video otoscopy for the purposes of post surgical follow-up. It appears that video otoscope images of tympanic membranes after TT placement are comparable to an in-person microscopic examination. This study suggests that store-and-forward video otoscopy may be an appropriate method of following TT after placement in a majority of cases. In time, utilization of this method of follow-up may lead to improved access to quality care, while saving time and travel costs when applied to rural populations.

## ACKNOWLEDGMENTS

Special thanks are extended to Andrew Tooyak CHA/P, Darlene Hoff R.N., Rena Spieker R.N., Denise Statz PA, Sharon Ree, Joanna Oudal, RN, and Darren Coolidge, for patient imaging, organization of study, and software technical assistance. This research was funded in part by a grant from the DHHS Office for Advancement of Telehealth. The Alaska Federal Health Care Access Network is an initiative of the Alaska Federal Health Care Partnership, which receives congressional funding from Indian Health Service, Department of Defense, and Department of Veterans Affairs.

## REFERENCES

1. Schappert SM. Office Visits for Otitis Media: US, 1975-1990. Publication no. 214. Advance Data from

- Vital and Health Statistics of the Centers for Disease Control/National Center for Health Statistics. Hyattsville, MD, 1992.
2. Bondy J, Berman S, Glanzer J, et al. Direct expenditures related to otitis media diagnosis: extrapolations from a pediatric Medicaid cohort. *Pediatrics* **2000**;105:E72.
  3. Alsarraf R, Gates GA. The real cost of otitis media. *Acta Paediatr* **1999**;88:487-488.
  4. Clinical Indicators for Otolaryngic—Head and Neck Surgery. American Academy of Otolaryngology—Head and Neck Surgery, Inc., 1995.
  5. Clinical Practice Guideline, Otitis Media with Effusion in Young Children (AHCPR publication no. 94-0622) and the Quick Reference Guide for Clinicians, Managing Otitis Media with Effusion in Young Children (AHCPR Publication no. 94-0623), 1994.
  6. Owings MF, Kozak LJ. Ambulatory and Inpatient Procedures in the United States, 1996. National Center for Health Statistics. *Vital Health Stat Series* **13** **1998**;139:1-13.
  7. Cunningham MJ, Darrow DH, Goldstein MN, et al. Follow-up management of children with tympanostomy tubes. *Pediatrics* **2002**;109:328-329.
  8. Derkay CS, Carron JD, Wiatrak BJ, et al. Postsurgical follow-up of children with tympanostomy tubes: results of the American Academy of Otolaryngology—Head and Neck Surgery Pediatric Otolaryngology Committee National Survey. *Otolaryngol Head Neck Surg* **2000**;122:313-318.
  9. ZA. Yawn BP. Telemedicine: a new framework for evaluation. *Telemed J* **2000**;6:55-61.
  10. Laird M, Deen M, Brooks S, et al. Telemedicine diagnosis of diabetic retinopathy, cataract and glaucoma by direct ophthalmoscopy. *Invest Ophthalmol Vis Sci* **1996**;37:104.
  11. Nitzkin JL, Zhu N, Marier RL. Reliability of telemedicine examination. *Telemed J* **1997**;3:141-157.
  12. Li HK, Tang RA, Oschner K, et al. Telemedicine screening of glaucoma. *Telemed J* **1999**;5:283-290.
  13. Lattimore MR. A store-forward ophthalmic telemedicine case report from deployed U.S. Army Forces in Kuwait. *Telemed J* **1999**;5:309-313.
  14. Yogesan K, Constable IJ, Barry CJ, et al. Telemedicine screening of diabetic retinopathy using a hand-held fundus camera. *Telemed J* **2000**;5:219-223.
  15. Rayner S, Beaconsfield M, Kennedy C, et al. Subspecialty adnexal ophthalmological examination using telemedicine. *J Telemed Telecare* **2001**;7:29-31.
  16. Lefell DJ, Chen YT, Berwick M. Interobserver agreement in a community skin cancer screening setting. *J Am Acad Dermatol* **1993**;28:1003-1005.
  17. Lyon CC, Harrison PV. A portable digital imaging system in dermatology: diagnostic and educational applications. *J Telemed Telecare* **1997**;3:81-83.
  18. Kvedar JC, Edwards RA, Menn ER, et al. The substitution of digital images for dermatologic physical examination. *Arch Dermatol* **1997**;133:161-167.
  19. Phillips CM, Burke WA, Shechter A, et al. Reliability of dermatology teleconsultations with the use of teleconferencing technology. *J Am Acad Dermatol* **1997**;37:398-402.
  20. Phillips CM, Burke WA, Allen MH, et al. Reliability of telemedicine in evaluating skin tumors. *Telemed J* **1998**;4:5-9.
  21. Krupinski EA, LeSueur B, Ellsworth L, et al. Diagnostic accuracy and image quality using a digital camera for teledermatology. *Telemed J* **1999**;5:257-263.
  22. Vidmar DA, Cruess D, Hsieh P, et al. The effect of decreasing digital image resolution on teledermatology diagnosis. *Telemed J* **1999**;5:375-383.
  23. Whited JD, Hall RP, Simel DL, et al. Reliability and accuracy of dermatologist's clinic-based and digital image consultations. *J Am Acad Dermatol* **1999**;41:693-702.
  24. Barnard CM, Goldyne ME. Evaluation of an asynchronous teleconsultation system for diagnosis of skin cancer and other diseases. *Telemed J* **2000**;6:379-384.
  25. High WA, Houston MS, Calobrisi SD, et al. Assessment of the accuracy of low-cost store-and-forward teledermatology consultation. *J Am Acad Dermatol* **2000**;42:776-783.
  26. Loane MA, Bloomer SE, Corbett R, et al. A comparison of real-time and store-and-forward teledermatology: a cost-benefit study. *Br J Dermatol* **2000**;143:1241-1247.
  27. Loane MA, Bloomer SE, Corbett R, et al. A randomized controlled trial assessing the health economics of realtime teledermatology compared with conventional care: an urban versus rural perspective. *J Telemed Telecare* **2001**;7:108-118.
  28. Van den Akker TW, Reker CHM, Knol A, et al. Teledermatology as a tool for communication between general practitioners and dermatologists. *J Telemed Telecare* **2001**;7:193-198.
  29. Nordal EJ, Moseng D, Kvammen B, et al. A comparative study of teleconsultations versus face-to-face consultations. *J Telemed Telecare* **2001**;7:257-265.
  30. Furukawa M, Furukawa MK, Mizojiri G, et al. Telemedicine in laryngology. *Telemed J* **1998**;4:329-333.
  31. Sclafani AP, Heneghan C, Ginsburg J, et al. Teleconsultation in otolaryngology: live versus store and forward consultations. *Otolaryngol Head Neck Surg* **1999**;120:62-72.
  32. Stein ST. Apparat zur photographischen aunahme des trommelfells. *Arch Ohren Nasen Kehlkopfheilkd* **1873**;7:56-58.
  33. Buckingham RA. Photography of the ear. In: English GM, ed. *Otolaryngology*. Vol. 1. Hagerstown, MD: Harper and Row, **1981**:301-333.
  34. Pensak MI, Yanagisawa E. Tympanic membrane photography: historical perspective. *Am J Otol* **1984**;5:324-332.
  35. Yanagisawa E, Carlson RD. Telescopic video-otoscopy using a compact home video color camera. *Laryngoscope* **1987**;97:1350-1355.
  36. Konrad HR, Berci G, Ward P. Pediatric otoscopy and photography of the tympanic membrane. *Arch Otolaryngol* **1979**;105:431-433.

37. Crump WJ, Kumar R, Orsak G, et al. A field trial of two telemedicine camera systems in a family practice. *Arch Fam Med* **1998**;7:174-176.
38. Yanagisawa E. The use of video in ENT endoscopy: its value in teaching. *ENT J* **1994**;73:754-763.
39. Jaisinghani VJ, Hunter LL, Yaoli L, et al. Quantitative analysis of tympanic membrane disease using video-otoscopy. *Laryngoscope* **2000**;110:1726-1730.
40. Burgess LPA, Holtel MR, Syms MJ, et al. Overview of telemedicine applications for otolaryngology. *Laryngoscope* **1999**;109:1433-1437.
41. Sullivan RF. Video-otoscopy in audiologic practice. *J Am Acad Audiol* **1997**;8:447-467.
42. Rinde E, Nordrum I, Nymo BJ. Telemedicine in rural Norway. *World Health Foun* **1993**;14:71-77.
43. Pedersen S, Hartviksen G, Haga D. Teleconsultation of patients with otorhinolaryngologic conditions. *Arch Otolaryngol Head Neck Surg* **1994**;120:133-136.
44. Crump WJ, Driscoll B. An application of telemedicine technology for otorhinolaryngology diagnosis. *Laryngoscope* **1996**;106:595-598.
45. Blakeslee DB, Grist WJ, Stachura ME, et al. Practice of otolaryngology via telemedicine. *Laryngoscope* **1998**;108:1-7.
46. Stern J, Heneghan C, Sclafani AP, et al. Telemedicine applications in otolaryngology. *J Telemed Telecare* **1998**;4:74-75.
47. Heneghan C, Sclafani AP, Stern J, et al. Telemedicine applications in otolaryngology. *IEEE Eng Med Biol Mag* **1999**;18:53-62.
48. Whitten PS, Cook DJ. School-based telemedicine: using technology to bring health care to inner-city children. *J Telemed Telecare* **1999**;5:S1:23-25.
49. Made C, Carle L, Soderberg O, et al. Tele-otolaryngology consultations between two rural primary care centres in Southern Lapland and the University Hospital of Umea. *J Telemed Telecare* **1999**;5:S1:93-94.

Address reprint requests to:

Chris Patricoski, M.D.

Alaska Federal Health Care Access Network

Suite 310

4201 Tudor Centre Drive

Anchorage, AK 99508

E-mail: [cpatricoski@afhcan.org](mailto:cpatricoski@afhcan.org)