Communication

Outcomes and Methods in Telemedicine Evaluation

NORIAKI AOKI, M.D., Ph.D., M.S.,1–4 KIM DUNN, M.D., Ph.D.,1,2
KATHY A. JOHNSON-THROOP, Ph.D.,3 and JAMES P. TURLEY, Ph.D., R.N.1

ABSTRACT

One hundred and four articles, published from 1966 to 2000, were reviewed to investigate telemedicine evaluation studies in terms of methods and outcomes. A total of 112 evaluations were reported in these 104 articles. Two types of evaluations were evaluated: clinical and non-clinical. Within the clinical evaluations, three were on clinical effectiveness, 26 on patient satisfaction, 49 on diagnostic accuracy, and nine on cost. In the non-clinical evaluations, 15 articles discussed technical issues relating to digital images, such as bandwidth, resolution, and color, and 10 articles assessed management issues concerning efficiency of care, such as avoiding unnecessary patient transfer, or saving time. Of the 112 evaluations, 72 were descriptive in nature. The main methods used in the remaining 40 articles used quantitative methods. Nineteen articles employed statistical techniques, such as receiver operating characteristics curve (three evaluations) and kappa values (seven evaluations). Only one article utilized a qualitative approach to describe a telemedicine system. Currently, there are a number of good reports on diagnostic accuracy, satisfaction, and technological evaluation. However, clinical effectiveness and cost-effectiveness are important parameters, and they have received limited attention. Since telemedicine evaluations tend to explore various outcomes, it may be appropriate to evaluate from a multidisciplinary perspective, and to utilize various methodologies.

INTRODUCTION

Telemedicine systems constitute integrated systems of health care delivery that employ telecommunications and computer technology as a substitute for face-to-face contact between provider and client and provider and provider. It has “the potential for ameliorating seemingly intractable problems in health care such as limited access to care among segments in the population—especially the geographically disadvantaged—uneven quality of care, and cost inflation.”

Since it can affect health outcomes and costs, it is important to identify appropriate evaluation methodologies for telemedicine. In this study, we examined the recent literature on telemedicine evaluation to identify the methods used. The aim was to assess the current status of evaluation methodology in this field. The

1School of Health Information Sciences, University of Texas Health Science Center–Houston, Houston, Texas.
2The Schull Institute, Houston, Texas.
3Information Research and Planning, Baylor College of Medicine, Houston, Texas.
4Center for Health Service, Outcomes Research and Development-Japan (CHORD-J), Tokyo, Japan.
present review summarizes the current state of knowledge, highlights where more information is needed, and shows the direction of future telemedicine evaluation research.

MATERIALS AND METHODS

A systematic search of the published literature (1966–2000) identified 104 articles related to the evaluation of telemedicine (Table 1). The literature was searched employing MEDLINE (National Library of Medicine) and a manual search of the major telemedicine journals using as key words “telemedicine,” “evaluation,” “assessment,” and “tele-.” All articles were classified into two categories: clinical and non-clinical evaluation. Subsequently, methods employed in each article were identified and results and conclusions extracted for this analysis.

RESULTS

Most articles focused on single clinical specialties (pathology,2–23 family and community medicine,24–39 ultrasound,40–55 dermatology,56–68 surgery,69–77 radiology,78–85 ophthalmology,86–91 otolaryngology,92,93 oncology,94,95 dental,96 and pediatrics97) or specific settings (emergency room,98–100 correctional setting,101–104 or offshore setting105). Only four articles evaluated clinical patients seen by several specialties in a telemedicine network. Eight of the articles employed two different evaluation methods. Since telemedicine research is relatively new, the rest of the studies—93% (95 out of 104)—were published between 1996 and 2000.

Outcomes for telemedicine evaluation

Table 2 presents data on the numbers of articles by clinical speciality. This list constitutes the universe of studies that were used in this analysis.

Clinical outcomes

Clinical outcomes were grouped into four categories: (i) clinical effectiveness, that is, reduction of mortality (death) and morbidity (disability), (ii) patient satisfaction, (iii) diagnostic accuracy, and (iv) cost.

Clinical effectiveness. Only three articles evaluated clinical effectiveness with mortality, morbidity and quality of life as main outcomes.39,55,72 All three articles reported either the same or improved clinical outcomes with telemedicine. For example, Whitlock et al.39 measured clinical parameters among diabetes mellitus patients to assess a home telemedicine consultation program. The telemedicine group showed significant reduction in body weight (average reduction from 214.3 to 206.7 lb) and in HbA1c (9.5% to 8.2%). Lambrecht et al.72 reported a decrease in unnecessary transportation without an increase in adverse clinical events with telemedicine.

Patient satisfaction. Twenty-six articles discussed patient satisfaction.6,7,17,23–26,29–32,38,44,63,64,75,83,89,93–95,97,98,100,104,106 These articles reported relatively high patient satisfaction, ranging from 61% to 100%. For example, Makhjian et al.104 investigated patient satisfaction in an Ohio maximum-security prison and reported that 91% of the patients were satisfied with the consultation. Huston and Burton31 reported that most patients were satisfied with their teleconsultation (on a seven-point Likert scale, the mean was 6.8). Most of these studies used questionnaires to obtain the data. Only a
few conducted in-depth interviews to uncover the underlying reasons for satisfaction or dissatisfaction.6,7

**Diagnostic accuracy.** Forty-nine studies investigated diagnostic accuracy of telemedicine versus non-telemedicine.2–8,9,11,13–23,27,28,36,37,40,41,43,51,53,54,56,58–62,65,66,69,70,73,76,77,79,80,84,85,88,90,91,96 Of the 49 articles, 41 discussed accuracy of digital images, mainly in radiology, dermatology, ultrasound, and pathology. Some studies focused on overall accuracy. For example, Pacht et al.,37 in a prospective, crossover study, found that two examiners showed substantial agreement (kappa statistics are 0.66 and 0.61, respectively) in auscultation of the lungs and diagnostic impression in a pulmonary medicine clinic.

**Cost.** We identified 9 articles that investigated the cost of telemedicine5,7,10,12,33,42,45,46,48–50,52,68,81,82 as compared to traditional face-to-face care. McCue et al.102 reported that their telemedicine program saved $14 per visit. Rendina et al.55 reported that the cost of a neonatal echocardiograms was reduced by $33 when compared to previous methods. Some articles mentioned the importance of patient volume for cost analysis. For example, Brunicardi et al.101 reported that their telemedicine system at the Corrections Medical Center experienced savings when 129 or more consults were performed during each quarter. Stoloff et al.105 analyzed the saving from the avoidance of unnecessary medical evacuations (MEDEVACs). They evaluated cost per MEDEVACs for various types of naval vessels: aircraft carriers (crews > 5,000), amphibious ships (500–2,000), small ships (<500), and submarines (<200), and concluded that telemedicine was cost-effective only on large ships (aircraft carriers and amphibious vessels).105

**Non-clinical evaluation**

Manuscripts were identified on two important non-clinical outcomes, namely, technical and management outcomes. Efficiency of patient management may be one of the most important outcomes in telemedicine implementation. Although these might not be directly related to clinical outcome, the enhancement of non-clinical outcomes give medical practitioners more time to provide care, thereby, gaining patient satisfaction and reducing resource consumption.

**Technical outcomes.** Fifteen articles discussed technical outcomes, such as bandwidth, resolution, and colors in digital images.5,7,10,12,33,42,45,46,48–50,52,68,81,82 These evaluations are crucial for telemedicine in terms of assuring quality of digital images, especially in pathology, radiology, ultrasound, dermatology, and ophthalmology. Some articles focused on the technical requirements for diagnostic quality. For example, Houston et al.48 concluded that for the echocardiographic assessment of the newborn, one (128 Kbps) or two ISDN2 channels (256 Kbps) will transmit images of satisfactory quality in many situations but three (384 Kbps) or more channels are necessary to ensure minimum degradation in image quality. Vidmar et al.68 evaluated the difference between low and high resolution images (720 × 500 pixel) versus (1490 × 1000 pixel) in dermatologic interpretation. Either resolution was found to be adequate for most store-and-forward teledermatology consultations.

**Management outcomes.** Ten articles addressed management issues, including time savings or avoidance of unnecessary patient transfers.9,34,35,47,71,72,74,75,78,87 All articles reported improvement in efficiency of care. For example, neurosurgical emergencies, Heautot et al.71 reported that 50% of unnecessary patient transfers were avoided using tele-consultation between a general hospital and a distant univer-

### Table 2. Outcomes Used for Telemedicine Evaluation

<table>
<thead>
<tr>
<th>Components</th>
<th>Number of evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Clinical effectiveness</td>
<td>3 (2.7%)</td>
</tr>
<tr>
<td>Patient satisfaction</td>
<td>26 (23.2%)</td>
</tr>
<tr>
<td>Diagnostic accuracy</td>
<td>49 (43.8%)</td>
</tr>
<tr>
<td>Cost</td>
<td>9 (8.0%)</td>
</tr>
<tr>
<td><strong>Nonclinical outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Technical evaluation</td>
<td>15 (13.4%)</td>
</tr>
<tr>
<td>Management evaluation</td>
<td>10 (8.9%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>112</td>
</tr>
</tbody>
</table>

OUTCOMES AND METHODS IN TELEMEDICINE EVALUATION 395
sity hospital, 100 km away. Lambrecht et al.\textsuperscript{72} also reported that 68 out of 100 trauma patients could remain in their rural community without any serious adverse effects. Roca et al.\textsuperscript{85} observed an exponential relationship between agreement with a gold standard (true diagnosis of the case based on pathology, surgery, and follow-up) and the time of training on screen diagnosis \((r = 0.97, p < 0.01)\), which increases \((r = 0.98, p < 0.004)\) with training.

**Summary of outcomes**

Almost all articles evaluating either clinical or non-clinical outcomes demonstrated that telemedicine is useful and asserted that it can play an important role in future health care. Patient satisfaction, diagnostic accuracy, and non-clinical outcomes have been evaluated, and, to a lesser extent, clinical effectiveness.

**Methods for telemedicine evaluation**

In reviewing the published literature, the three most common methodologies were statistical analysis, cost-analysis, and qualitative analysis, as shown in Table 3.

**Statistical analysis**

Appropriateness varies according to (a) the type of data, (b) underlying assumption of data (e.g., normality), and (c) the particular purpose of the evaluation. Two methods have been used extensively in telemedicine evaluation: kappa statistic and receiver operating characteristics curve (ROC).

**Statistical comparison.** Nineteen articles compared various differences between telemedicine and non-telemedicine encounters using statistical comparisons, such as the McNemar test, chi-square or Fisher’s exact test, and regression. Statistical analysis was used in two clinical effectiveness, six patient satisfaction, eight diagnostic accuracy, and three non-clinical evaluations. An example of statistical analysis is the paper described by Demartines et al.\textsuperscript{107} which evaluated 112 patients undergoing digestive or endocrine surgery to compare tele-transmission (telemedicine or teleradiology) and direct viewing of x-ray or computed tomography film (non-telemedicine). They found that the target organ was always visible and the structure and pathologic findings were analyzable in 98.2% of transmitted documents and 99.1% of live documents (difference not statistically significant). Details of the anatomic structures could be assessed in 89.3% of transmitted pictures and 95.5% of live pictures (difference not statistically significant).

**Agreement evaluation (kappa statistic)**\textsuperscript{108} Since overall agreement, such as concordance and discordance, includes some agreements arising by chance alone, the actual agreements beyond chance must be calculated in interpreting the agreement between two persons. Kappa is defined as the proportion of actual agreement beyond chance compared to the potential agreement beyond chance. In this review we found seven articles that employed the kappa statistic to eliminate agreement by chance. The kappa value indicates the degree of actual agreement: 0–0.2, slight agreement; 0.21–0.4, fair agreement; 0.41–0.6, moderate agreement; 0.61–0.8, substantial agreement; 0.81–1.0, perfect agreement. For example, Gilmour et al.\textsuperscript{58} compared traditional face-to-face care and tele-consultation in the diagnosis of skin lesions, and found almost complete agreement \((kappa = 0.96)\) between the two modalities. The kappa value is a useful means to measure the actual agreement among two or more persons, as long as the distribution is not highly skewed since kappa would be underestimated if the prevalence is skewed.\textsuperscript{109}

**Receiver operating characteristics (ROC) curve and area under the ROC curve (AUC).** The ROC method represents the trade-off between sensitivity and specificity, and involves plotting the true-positive rate (sensitivity) against the false-positive rate (specificity).\textsuperscript{110,111} The area under the ROC curve (AUC) has become a particularly important metric for evaluating diagnostic procedures because it is the average sensitivity over all possible specificities. It ranges from 0 to 1, where 0.5 is chance and 1 is perfect. O’Sullivan et al.\textsuperscript{83} conducted a study to address the diagnostic accuracy of an image to detect urinary calculi. They used the ROC method...
to compare the difference in accuracy between the observer with digital images and one with the original radiographs, but they did not describe the method and results well. Vidmar et al.\textsuperscript{68} used the ROC curve to evaluate 180 dermatologic cases in terms of degraded digital image resolution (as viewed on a monitor) on diagnostic accuracy. Physicians were blinded concerning the image resolution, and they were asked to record a diagnosis and level of confidence. The data were organized in a $2 \times 6$ matrix, which represented a summary of correct responses and the stated level of confidence, to generate the ROC curve. They did not find any consistent differences in digital image resolutions under the ROC curve. The ROC curve provides information on the overall performance of a diagnostic test, which is independent from disease prevalence and the decision threshold of observers.

Cost-analysis

Cost-analysis in telemedicine consists of four methods: cost-minimization, cost-effectiveness, cost-utility, and cost-benefit analysis.\textsuperscript{112,113} In this review, we identified 8 cost-minimization analyses and one cost-effectiveness analysis.\textsuperscript{55,57,67,92,99,101–103,105} Cost-minimization analysis is the simplest analysis of all; it merely compares cost between various strategies. For example, McCue et al.\textsuperscript{102} concluded that their telemedicine program saved $14 per visit. Rendina et al.\textsuperscript{55} reported that their telemedicine system for neonatal echocardiograms reduced the cost by $33 compared to previous methods. A cost-minimization analysis is valid only if the other factors, such as clinical effectiveness, can be assumed to be similar. If we do not know the total cost, we cannot ascertain cost savings.

Cost-utility analysis includes quality factors (e.g., quality of life) in addition to quantitative clinical effectiveness. Cost-effectiveness (cost-utility) analysis can provide more comprehensive evaluation, including economic, clinical and quality of life features, by using quality adjusted life years (QALYs).\textsuperscript{114} In QALYs, a quality factor, quality of life (QoL), can be combined with quantitative factors (e.g., life expectancy) as a utility score. Cost-utility analysis is appropriate for telemedicine, and includes both qualitative and quantitative factors.

Qualitative analysis

Whereas, qualitative methods may provide more insight, qualitative methods are more definitive in testing hypotheses.\textsuperscript{117–120} Qualitative analysis is gaining acceptance in medical research, but has long been the principal means employed by anthropologists to study the customs and behaviors of peoples in other cultures.\textsuperscript{117} However, this review identified only one article using qualitative analysis by Siden et al.\textsuperscript{38} for a needs assessment. They used the focus group method involving a small number of participants to generate data for further analysis.\textsuperscript{120} Focus groups revealed a number of important positive and negative attitudes regarding telemedicine and priorities for its implementation. “Uncertainty” and “trust” were two themes that emerged from all groups. Uncertainty referred to comments and concerns regarding unknown aspects of the technology. Trust comments were related to opinions regarding trust of professionals and technology.

CONCLUSION

Evaluation of some specific aspects of telemedicine, such as training and needs assessment, could provide much practical information to improve telemedicine projects. Therefore, future research can undoubtedly improve telemedicine programs and may encourage potential telemedicine providers to initiate such services. Finally, these results will be

<table>
<thead>
<tr>
<th>Methods</th>
<th>Number of evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical analysis</td>
<td></td>
</tr>
<tr>
<td>Statistical comparison</td>
<td>19 (17.0%)</td>
</tr>
<tr>
<td>Kappa</td>
<td>7 (6.3%)</td>
</tr>
<tr>
<td>ROC</td>
<td>4 (3.6%)</td>
</tr>
<tr>
<td>Cost analysis</td>
<td>9 (8.0%)</td>
</tr>
<tr>
<td>Qualitative analysis</td>
<td>1 (0.9%)</td>
</tr>
<tr>
<td>Ad hoc</td>
<td>72 (64.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>112</td>
</tr>
</tbody>
</table>
important evidence for demonstrating the effectiveness of telemedicine practice.

ACKNOWLEDGMENT

We appreciate the assistance of Dr. William Schull, professor emeritus at the University of Texas School of Public Health and president of the Schull Institute, in editing this paper.

REFERENCES


Address reprint requests to:
Noriaki Aoki, M.D.
School of Health Information Sciences
University of Texas Health Science Center-Houston
7000 Fannin, UCT-600
Houston, TX 77030

E-mail: Noriaki.Aoki@uth.tmc.edu
This article has been cited by:

