Three Principles for Determining the Relevancy of Store-and-Forward and Live Interactive Telemedicine: Reinterpreting Two Telemedicine Research Reviews and Other Research

Craig Locatis, PhD, and Michael Ackerman, PhD
National Library of Medicine, National Institutes of Health, Bethesda, Maryland.

Abstract

The Agency for Healthcare Research and Quality sponsored two telemedicine research reviews. The latest review concluded that telemedicine is most relevant to specialties, such as psychiatry and neurology, where high levels of patient interaction are crucial to assessment. Telemedicine research studies cited in the reviews having positive findings in the specialties of ophthalmology, otolaryngology, obstetrics and gynecology, gastroenterology, and cardiology and more recent research in these areas are reviewed to identify criteria other than degree of interaction for determining the appropriateness of telemedicine interventions. These criteria include congruity or the extent that procedures used in telemedicine are similar to those of in-person examination, fidelity or the degree to which the information used for assessment in remote examinations is of similar quality to that used in-person, and reliability or the consistency with which information can be gathered and transmitted.

Key words: telemedicine, technology, telecommunications

Introduction

The Agency for Healthcare Research and Quality (AHRQ) funded two comprehensive reviews of telemedicine research in 2001 and 2006. These reviews are significant for several reasons. One is their breadth, including research in all specialties impacting the Medicare population except pathology and radiology, whose providers typically do not see patients in-person, and studies related to women and children not covered by Medicare. Another is their strict inclusion criteria, focusing on rigorous, randomized control studies. Moreover, the reviews were performed by a research team, not just one or two individuals, that reached consensus on studies to include and findings to report. The reviews’ primary conclusions, especially those of the 2006 update, are summarized in this article. Its conclusion that telemedicine appears to work best in specialties, like psychiatry and neurology, where verbal interaction is a key assessment component.

AHRQ Research Reviews

The AHRQ reviews cover home-based monitoring, store-and-forward, and office/hospital-based telemedicine. Only the last two areas are relevant to this analysis, where diagnostic decisions are made either asynchronously (store-and-forward telemedicine) or synchronously by live-interactive videoconferencing (office/hospital-based telemedicine). The AHRQ update report’s findings generally mirror those of the original one.

The major findings concerning store-and-forward and office/hospital-based telemedicine are:

- Although telemedicine may be warranted when there is no alternative for providing care, it still is deployed in many areas where there is insufficient evidence justifying its use.
- There are insufficient rigorous studies, and many lack a “gold standard” for making comparisons. The update acknowledges that accuracy tests, such as comparing diagnoses with biopsy results, may be useful because clinicians can agree and still be wrong. Accuracy can be low, however, even when diagnoses are done face-to-face. Consequently, the most appropriate telemedicine standard is remote agreement with in-person exams performed by at least two concurring physicians. The standard is to prove telemedicine’s equivalency, not superiority.
- Store-and-forward and office/hospital-based telemedicine improve access to the extent healthcare was absent prior to telemedicine’s introduction. Store-and-forward interventions have only had modest impact on reducing the need for subsequent evaluations in-person, at least in certain domains such as dermatology.
- There may be less confidence in diagnostic accuracy in telemedicine because there is evidence that more tests are ordered and many telemedicine patients are still seen in-person.
- There is weak evidence that office/hospital-based telemedicine increases access and reduces travel in rural settings when assessments can be made by videoconferencing.
- The best evidence for telemedicine, especially live interactive office/hospital-based telemedicine, is in the psychiatry and neurology specialties, where verbal interaction is a key assessment component.
Telemedicine’s benefits in other specialties are more uneven. For example, tele-ophthalmology has high diagnostic agreement rates, but only for certain eye conditions, whereas tele-dermatology concordance rates are highly variable, and its research is problematic. Overall, telemedicine may best serve as an adjunct to care centered on the in-person visit.

These conclusions are plausible and consistent with the research reviewed and are generally reinforced in other research reviews and meta-analyses specific to the specialties the reports cover. Although one dermatology review took issue with the reports’ dermatology concordance findings, other dermatology reviews have documented variability or deplored the research rigor. A re-examination of the studies in specialties other than dermatology, neurology, and psychiatry having evidence of effectiveness cited in the AHRQ reports, as well as more recent research, suggests criteria other than direct human interaction for considering telemedicine’s appropriateness.

AHRQ Specialty Studies

Other specialties identified in the AHRQ reviews having evidence for telemedicine’s effectiveness include ophthalmology, otolaryngology, gastroenterology, and cardiology. Effectiveness studies cited in the original AHRQ report and update were re-examined and determined to have certain common features supporting broader telemedicine appropriateness criteria.

Seven ophthalmology studies cited in the reviews used slit lamps and video or still images, each focusing on assessing different eye conditions. Patients with trabeculectomized eyes were evaluated in-person using slit-lamp biomicroscopy and by telemedicine in one study. The telemedicine examinations used the same slit-lamp technology, but the images were transmitted by videoconference at a rate of 384 kilobits per second (Kbps), sufficient for full-screen, full-motion, standard-definition video. Agreement between ophthalmologists examining in-person and remotely was essentially the same. In another study involving a range of eye problems, in-person examinations using slit lamps were compared with remote video examinations at 384 Kbps using slit lamps or torchlight. The percentage of clinically important disagreement was the same for the slit lamp conditions (just 5% of cases), but doubled with torchlight. A third slit lamp strabismus assessment study found full agreement in 80% of the cases, but the 384 Kbps transmission rate was a problem in cases involving micro-eye movements. Technical problems necessitated using a 128 Kbps transmission rate in 4 cases that required repeated tests before a diagnosis could be made. Finally, three other studies compared photographs with digital images or digital images acquired with different cameras. Photographs were superior to digital images for diagnosing retinopathy in one study, but the digital images were only 640 × 480 pixels compressed to less than 1 megabyte. Only 56% of compressed images were graded acceptable. Other studies found higher-resolution digital cameras superior to a videocamera that captured still images at 768 × 576 pixels, and a nine-field photography system was better than a single-field system for retinopathy diagnosis.

Otolaryngology and speech pathology studies cited in the AHRQ update review involved real-time consultation. One study comparing in-person and remote endoscopic otolaryngology examinations found suboptimal and incorrect diagnoses for 8 of 42 patients, all of whom were in the first 20 examined. Diagnostic agreement for the last 22 patients was 100%, indicating the errors were due to the general practitioner’s initial inexperience operating the endoscope. Another study comparing a tele-audiometric system for hearing testing to an identical system used locally found the systems produced the same results.

Three obstetrics and gynecology studies in the original AHRQ review compared patients examined in-person and remotely through captured or real-time videos from colposcopes. Eighty-one patients examined in one study were classified the same (either normal or abnormal) except for 1 considered abnormal by telecolposcopic examination. Another study found distant and on-site experts had similar diagnostic accuracy but differed on certain treatment options. A third study assessed colposcopists’ judgments of confidence, comfort, care quality, and image resolution when images were transmitted in real time or recorded for later review. The two telecolposcopy methods were equivalent except that care quality was judged better for synchronous examinations even though image quality was judged superior for recorded ones. A quarter of the examinations had some visual irregularity, and the ability to scan, adjust focus, and magnification in real time may have compensated for lower resolution.

Two gastroenterology studies mentioned in the AHRQ reviews were a pilot study of 10 patients and a follow-up of 53 patients. There was 100% concordance in the first study for the five upper and five lower endoscopies conducted in-person and by videoconference. The follow-up study compared the major and minor findings of endoscopic examinations conducted by a gastroenterologist in-person with those of a gastroenterologist who observed the examination remotely by videoconference at a 512 Kbps transmission rate. Of the 47 major findings identified in-person, the remote observer disagreed in 11 instances, 10 of which were overdiagnoses of erosions and rings. In the 1 case where the remote diagnosis was less conservative than in-person, a repeat endoscopy was recommended after 4 weeks.

Two cardiology studies cited in the original AHRQ review involved faxing electrocardiograms (ECGs) or transmitting ECGs and echocardiogram videos. The first compared diagnoses of senior house officers still in training with those of remote cardiologists and found disagreements in 20 of 112 cases, 17 of which immediately benefited from the consultation and the remainder of which had non-diagnostic ECGs. There was diagnostic agreement on all but 1 of the 26 cases in the second study. A study cited in the update involving ambulances found an 86% successful ECG transmission rate, with 98% of these being diagnostically acceptable and with failures being due to geographically related transmission problems. Also cited were studies of telephone transmission of ECGs from 200 general practitioners to a consulting cardiologist and video transmission of ultrasound that found failure to detect problems in only 7 of 952 patients and concordance in treatment recommendations in 29 of 32.
patients, respectively. Differences for 2 patients in the latter study were minor (oral versus intravenous antibiotics), while the remote cardiologist noted symptoms of claudication in 1 patient who had no disease.

**Recent Studies**

Because the latest AHRQ report was published in 2006, PubMed searches were conducted for articles published since that date using the terms for each of the above specialties and telemedicine. Articles comparing effectiveness of telemedicine examination methodologies (remote with in-person, store-and-forward with live interactive) or use of different technologies affecting information quality (imaging methods or types of compression) were identified. No subsequent studies comparing examination methods or investigating information quality in gastroenterology were found, but studies were identified in ophthalmology, otolaryngology, obstetrics and gynecology, and cardiology. They generally complement findings in the AHRQ reviews.

A subsequent ophthalmology study compared digital photographs taken by a trained nurse following a defined protocol with in-person ophthalmoscopy to diagnose retinopathy and found agreement in 86% of the cases.29 Neither method showed a systematic tendency for over- or underdiagnosis, and most discrepancies concerned a particular disease in one eye region. Another comparison study of photographs and in-person ophthalmoscopy found 86.5% agreement and indications that in some cases of disagreement, the telemedicine images may have provided more accurate diagnosis.30 Finally, a study of eye images comparing film and uncompressed digital photographs at 2400×2000 pixels with those compressed by JPEG2000 compression found no significant diagnostic differences. The 37:1 compression ratio and the compression algorithm preserved clinically sufficient information.

A subsequent otolaryngology study was similar to an earlier one where data were transmitted from audiometric devices used for in-person examinations but used a video-otoscope instead. There was complete agreement between remote and subsequent in-person diagnosis in 67 of 68 cases and only minor disagreement for 1 case.32 Another otolaryngology study of normal and pathological voice recordings using lossless and lossy compression methods found identical waveforms for lossless recordings but jitter and shimmer artifacts in lossy ones.

Subsequent telemedicine studies on obstetrics and gynecology and cardiology overlap because both focus on heart conditions. Recent studies of echocardiogram and prenatal ultrasound transmission found high levels of agreement between examinations conducted remotely and in-person. Other studies have shown that ECG images, which are not bandwidth intensive, can be efficiently transferred by cell phone for cardiologists to review.34,35 In contrast, cell phone transmission of arthroscopic videos was error prone because the bit rate was only 55 Kbps and resolution was only 176×144 pixels. Although 95% of the pathologies could be diagnosed, the technology was inadequate for teleconsultation because of transmission latencies and interruptions and because correct anatomical orientation was identified only 60% of the time.

**Study Commonalities and Implications**

The effectiveness studies have several common features:

- Most comparison studies having agreement used certain optical technology (slit lamps, endoscopes, colposcopes, and cameras) to assess patients, no matter whether patients were examined face-to-face or remotely.
- If optical devices were not used to collect information, a common medium often was in both in-person or remote assessments. For example, ECG results were printouts that could be reviewed locally, faxed, or transmitted, printed out, and assessed remotely. Ultrasound and echocardiograms were displayed on local or remote videomonitors or stored on videotape.
- In studies where digital cameras and other instruments were substituted for those normally used, the quality and kind of the data obtained mattered. For example, the resolution and number of images acquired and following a protocol for data collection were pertinent in many ophthalmology camera studies. When video was used, torchlight substitution for slit lamps was insufficient.
- The quality with which data are transmitted or stored is a factor. Higher transmission rates providing full-motion, higher-resolution video work better overall than lower ones. Generally, lower compression is superior to higher compression, and the amount and kind of compression matter. The compression threshold may vary among specialties. The lossy audio compression methods in otolaryngology were inadequate. Early methods compressing ophthalmology images to 1 megabyte were also deficient, whereas later ones capturing images at very high resolution and reducing them 37:1 were satisfactory. Data rates were factors in ophthalmology and otolaryngology studies.
- Many studies reported user or technology problems, particularly at start up, affecting results. Transmission interruptions or the need to use lower bit rates affected the ability to collect data, as did the learning curve when new technologies or procedures were implemented.
- Although there was sometimes more variability in assessing certain signs and symptoms, in cases where disagreement mattered, remote assessments tended to be more conservative and precautionary.

**Telemedicine Application Principles**

The effectiveness studies suggest three principles for determining the appropriateness of using telemedicine in medical specialties or for assessing certain cases within specialties:

- The congruency principle: the extent to which the procedures used in telemedicine mimic those of examinations done in-person. Results were better in all ophthalmology studies where the telemedicine condition involved using the same slit lamps used for in-person examinations and worse in one study where they were not.11,12 Otolaryngology diagnoses were similar when outputs from local audiometric devices were transmitted.19 The only difference in the otolaryngology7 and gastroenterology22,23
studies involving endoscopes or otoscopes concerned transmitting local video for remote consultation. Diagnostic outcomes were the same given appropriate transmission rates. When remote otolaryngology consultations were poorer, they could be attributed to local operators initially learning an unfamiliar device. When competency was achieved, diagnostic outcomes were the same. Similarly, congruency is high in cardiology when ECGs, echocardiograms, and ultrasound are viewed locally or remotely. Physicians view printouts or video in both contexts. In obstetrics and gynecology, in-person and remote diagnostic outcomes in colposcopy are the same, whether looking through the colposcope lens directly or having its output transmitted or recorded by video, but real-time remote examinations are judged more favorably when magnification, focus, and view can be controlled as they can in-person.

- **The fidelity principle:** the degree to which the information being transmitted in telemedicine is like that observed in-person. The standard-definition video used to transmit patient data in the slit lamp studies had sufficient resolution and motion for remote ophthalmology to diagnose eye conditions except those involving micro-movement. In the specialties of obstetrics and gynecology, otolaryngology, and cardiology, standard-definition video provided adequate resolution for remote diagnosis with colposcopic endoscopic, echocardiology, and ultrasound examinations done remotely or in-person. The local and remote output of audiometric devices and otoscopes is also similar. Fidelity is affected by the amount and kind of audio, video, and still image compression. Fidelity may compensate for incongruity in cases when a telemedicine procedure differs from in-person examination but provides essentially equivalent clinically relevant information. Studies comparing different still pictures for diagnosing retinopathy indicate a resolution and compression threshold for image acceptability. Increasing the number of images may add information compensating for data that are missing or inadequately shown in smaller image sets.

- **The reliability principle:** the consistency with which information is provided. It is affected by the skills of individuals at examination sites, the establishment of protocols and procedures for acquiring appropriate information, and the quality of information transmission. Higher-quality information transmission may be required in some specialties, especially if video, high-resolution pictures, or large datasets must be transferred. Transmission may be interrupted, or, to accommodate bandwidth limitations, more compression may be applied, introducing artifacts or reducing quality. Trade-offs among transmission reliability, fidelity, and information quality are sometimes required for diagnosing certain problems, such as eye conditions with micro-movement, as was the case with the standard-definition video transmitted at 384 Kbps, the highest transmission rate commonly available at the time the study was done. Current cell phone technology, however, can accommodate the low bandwidth requirements of ECG transmission but not the higher ones for transmitting arthroscopic videos.

**Conclusions**

The principles of congruence, fidelity, and reliability do not contradict the AHRQ assessment that telemedicine may work best in psychiatric and neurology specialties involving high degrees of human interaction. Nor do they contradict the update’s finding of variable outcomes in teledermatology research. The videoconferencing technology that dominates telepsychiatry and telenurology applications easily accommodates the interview and motor assessment techniques typically used in-person; few changes are needed to provide these services at a distance. Moreover, the technology has sufficient fidelity to enable remote consultants to assess body language, motor behavior, and other verbal and nonverbal cues providing full-screen, full-motion video can be reliably sent. Similarly, congruence, fidelity, and reliability may be lower in dermatology, where there is often a need to palpate and where resolution requirements may vary depending upon specific skin conditions. Congruence, fidelity, and reliability may be more functional criteria for judging the appropriateness of telemedicine interventions than simply the degree of human interaction in diagnostic assessments.

**Acknowledgments**

This review was supported through the National Institutes of Health Intramural Research Program.

**Disclosure Statement**

No competing financial interests exist.

**REFERENCES**


DETERMINING THE RELEVANCY OF TELEMEDICINE


Address correspondence to:
Craig Locatis, PhD
Office of High Performance Computing and Communications
National Library of Medicine
National Institutes of Health
8600 Rockville Pike
Bethesda, MD 20894

E-mail: locatis@nlm.nih.gov

Received: March 8, 2012
Revised: April 19, 2012
Accepted: April 23, 2012