An Evaluation of Telemedicine in Surgery

Telediagnosis Compared With Direct Diagnosis

Nicolas Demartines, MD; Ulrike Otto, MD; Didier Mutter, MD, PhD; Ludwig Labler, MD; Alex von Weymarn, MD; Michel Vix, MD; Felix Harder, MD

Hypothesis: Telemedicine for real-time transmission of clinical documents and interactive remote telediagnosis allows accurate clinical application in surgery.

Design: Prospective cohort study in which 2 hospitals, 120 miles apart, were connected via integrated services digital network (ISDN) teleconferencing units, and each evaluated clinical cases in real time.

Setting: A tertiary care university hospital and primary care county hospital.

Participants: Between May 1, 1998, and June 30, 1998, 112 patients undergoing digestive or endocrine surgery were evaluated by teletransmission (study group) and direct vision (control group). Diagnosis had to be known by the viewer, and either conventional magnetic resonance imaging or computed tomographic scans were available.

Main Outcome Measures: Picture quality, organ structure, and pathologic finding viewed on telemedicine documents were evaluated by radiologists and surgeons blind to diagnosis. Accuracy of remote 128–kilobit (kb)/s transmission-rate diagnoses and results were compared with those obtained directly.

Results: Picture quality was “good” or “excellent” in 92.9% of transmitted documents and 95.5% of live images (P<.4). The target organ was always recognized, structure and pathologic finding were analyzable in 98.2% of transmitted documents and 99.1% of live documents, and fine structures were assessable in 89.3% of transmitted pictures and 95.5% of live pictures (P=.05). Diagnosis was made in 84.8% of transmitted cases and 93.8% of live cases (P=.02).

Conclusions: Low bandwidth (128 kb/s) telemedicine application in surgery is reliable in evaluating remote cases. Loss of image quality through teletransmission occurred in 2.7% of cases, and diagnosis was not possible in 15.2% of transmitted vs 6.2% of live cases, suggesting factors other than technical quality (choice of radiological studies, additional clinical information required, etc). This underscores the importance of real-time interactive discussion during surgical teleconferences.


ELEMEDICINE is the utilization of new computer-based communication technology for medical purposes. Though still a relatively new area of interest, it offers enormous potential for applications in medical and surgical practice as well as education. Telemedicine is believed to favor and simplify the exchange and diffusion of information, knowledge, and surgical education by permitting broader access to expertise and second opinions without travel. Telemedicine applications have the potential to readily bring specialists and subspecialists to several different locations, thus simplifying and increasing the availability of expertise in the medical community. This may be of great interest in surgical specialties, but surprisingly, until recently, telemedicine in surgery was scarcely investigated. Analysis of telemedicine potential and quality control are important and the implementation of poor telemedicine services should be avoided. Before broad clinical applications of remote diagnosis and therapy may be discussed, the value of telediagnosis and the accuracy of transmission of clinical documents still require proper assessment.

Having been involved in an international academic network in the field of digestive and endocrine surgery since 1996, we have experience in the surgical application of telemedicine. To develop a regional surgical network linked to the University Hospital of Basel, Basel, Switzerland, we first had to evaluate the accuracy of remote diagnoses for clinical
PATIENTS AND METHODS

PATIENTS

Between May 1, 1998, and June 30, 1998, we performed a simple selection of ongoing cases hospitalized in the Surgical Clinic of the University Hospital and in the Surgical Clinic of Hospital of Frauenfeld, Frauenfeld, Switzerland, a county hospital some 120 miles away. The criteria of selection included cases of digestive or endocrine surgery for which the final diagnosis was known and the original conventional x-ray film or computed tomographic (CT) scan was available. All cases seen during the study period that fulfilled the above criteria were included in this study. No cases were excluded because of complexity or because the findings were obvious (Table). The patients themselves were not present during the teleconference, and the documents were anonymized according to the 97/66/EC directive from the European parliament and from the council held on December 15, 1997, regarding the processing of personal data and the protection of privacy in the telecommunication sector.11

TRANSMISSION TECHNOLOGY

The teleconference unit used in Basel was the videoconferencing system PictureTel Venue 2000 (PictureTel Corp, Danvers, Mass) configured for a 128 kilobit (kb)/s transmission rate, instead of the standard configuration of 384 kb/s, to be compatible with the system at the remote site. In Frauenfeld, the system used was a compact videoconferencing system (PictureTel Swift Site; PictureTel Corp) in a standard configuration for 128 kb/s. The systems are composed of a video coding/decoding system (CODEC) with different components connected to them: transmitting devices, such as the microphone and the 3-chip camera, and reception devices, such as the monitor and speakers. Transmission was achieved using 2 integrated services digital network (ISDN) lines (64 kb/s each) at the worldwide standards of H320 and H323 at a rate of 128 kb/s and 15 frames per second. The television monitor was a Sony Trinitron 70 cm (Sony Electronics Corp, Park Ridge, NJ), and live pictures were seen on standard negatoscopes.

ORGANIZATION OF THE TELECONFERENCES

Prior to the study, the panel of surgeons and radiologists knew each other personally and had worked together for at least 2 years. Furthermore, a minimum of 2 members from each panel had been trained in the use of the telemedicine unit and had performed teleconferences prior to the study.

On the conference day, each team alternatively evaluated their own pictures directly and then the pictures from the other team. The panel was blind to diagnoses of all cases. Each document was evaluated with a systematic multiple-choice questionnaire filled out immediately after the presentation of each case. Diagnosis at each site was reached by consensus.

The types of documents to be transmitted to the remote site were chosen by the panel at each site. During the meeting, clinical and radiological document presentation was performed together in real time; pictures were seen on the television monitor accompanied by a commentary with the presentation. At the same time, both panels could speak to each other and had the opportunity to ask and answer questions, to request that the picture be moved, to focus on a detail, or to ask for another document. To avoid fatigue, the videoconferences were scheduled for 90 minutes.

DATA ANALYSIS

We focused on picture evaluation, case evaluation, and diagnostic accuracy, and compared the results from teletransmission (study group) with the results of direct vision (control group). For each case, both direct and transmitted, the panel was asked to judge the following points:

- Was the technical quality of pictures excellent, good, fair, or insufficient? Excellent and good were defined as a picture that could be evaluated without any technical problem or artifact; fair, a picture with some technical imperfection; and insufficient, a picture that could not reasonably be evaluated.
- Was the target organ recognizable (yes or no)? The target organ was defined as the organ presently under evaluation (eg, the head of the pancreas on a CT scan).
- Were the organ structures recognizable (yes or no)? Organ structures were defined as the recognition of anatomical structures that could be delimited on the document under evaluation.
- Was it possible to analyze fine organ structures and the pathologic finding(s) (yes or no)? The fine structures were defined as the finest anatomical structure analyzable on such clinical documents, and the pathologic finding was a finding of something not normally belonging to the studied anatomical structures.
- Was a diagnosis accurate, probable, possible but uncertain, or impossible at the end of the case presentation?
- Did you need more images of the same imaging technique to assess your diagnosis (yes or no)?
- Did you need other types of imaging techniques to reach a diagnosis (no, may be helpful, or mandatory), for example, a magnetic resonance image instead of a CT scan?
- How did you characterize this case? This was a subjective evaluation by the panel relating to their own clinical experience about the presented cases. Was the case routine (unusual but well known and/or obvious) or challenging (difficult because of the complexity and combination of clinical problems)?
- Interactivity was defined as the live, real-time exchange of questions and answers as well as discussion and debate among teleconference participants at the different remote sites with the goal of eliciting new evidence or insight contributing to case resolution. To judge whether the interactive discussion allowed one to receive supplementary data contributing to case resolution, the possibility of diagnosis was assessed at 2 different times: at the end of the initial clinical presentation and after the interactive discussion. Differences in judgment before and after interactive discussion were rated and compared to assess the relevance of the interaction for case resolution. The results for direct evaluation were compared with the results for evaluation by teletransmission.

Questionnaire results were entered into a personal computer and processed with Microsoft Excel 97 (Microsoft Corp, Redmond, Wash). Results are expressed as absolute values and percentages of the answers. The differences of evaluation between direct vision and teletransmission were compared using χ² analyses or the t test when appropriate with significance defined at P<.05.
applications with low transmission rates. We performed a comparative analysis on the evaluation of clinical documents viewed directly (live) with similar documents transmitted in real time per teleconference and evaluated at a remote site. The aim of the study was to assess the method of comparison between direct and remote document evaluation, to quantify the potential loss of information introduced by the telemedicine transmission in a clinical situation, and to compare the diagnosis from direct vision with diagnostic possibilities using teleconferencing for surgical patients.

RESULTS

ORIGINAL IMAGES

Pictures analyzed during this study included 62 CT scans and 25 conventional x-ray films with or without a contrasting agent. The evaluation of clinical documents took an average of 2.4 minutes (range, 1.0-6.0 minutes) for live documents and 2.1 minutes (range, 0.5-8.0 minutes) for transmitted documents ($P$ = .5).

The original picture definition in both hospitals was similar, ranging from 512 to 2048 pixels (conventional range for x-ray films and CT scans). The quality of the image itself was judged equally at both sites: 98.2% of the original images were rated as excellent or good in one hospital and 94.6% in the other ($x^2_1$ = 1.88, $P$ > .1). Considering these facts, the comparison of transmitted vs live (direct vision) images relates the following aspects.

COMPARISON OF TRANSMITTED VS DIRECT VISION IMAGES

Picture Quality and Target Organ Recognition

The transmitted picture quality was judged as being good or excellent in 104 cases (92.9%) compared with 107 live cases (95.5%) ($x^2_1$ = 0.73, $P$ > .4). The images were fair or insufficient in 8 transmitted cases (7.1%) and in 5 live cases (4.5%). A loss of picture quality was therefore observed in 2.6% of cases. One hundred percent of the target organs were recognized in the live as well as the transmitted documents.

Organ Structure, Fine Organ Structure, and Pathologic Findings

It was possible to judge the organ structure in 111 of 112 live pictures, while for transmitted documents, this was possible in 110 of 112 cases (98.2% vs 99.1%, $x^2_1$ = 0.34, $P$ > .5). It was possible to analyze the fine organ structure and/or the pathologic findings in 107 of 112 live images and in 100 of 112 transmitted images (95.5% vs 89.3%, $x^2_1$ = 3.11, $P$ > .05).

DIAGNOSIS

On live images, diagnosis was accurate in 79 (70.5%) of 112 cases, possible in 26 (23.2%), and not possible in 7 (6.3%). With transmitted images, diagnosis was accurate in 69 (61.6%) of 112 cases, possible in 26 (23.2%), and not possible in 17 (15.2%). Accurate or possible diagnosis using live vs transmitted documents was 93.8% vs 84.8% ($x^2_1$ = 4.67, $P$ = .02).

The need to obtain more pictures by the same imaging technique, respective of the same radiological study, was judged not necessary for 101 live cases (90.2%) and for 99 transmitted cases (88.4%) ($x^2_1$ = 0.18, $P$ > .6).

In live cases, the panel judged that other imaging technique were mandatory in 6 cases (5.4%), may be helpful in 16 (14.3%), and were not necessary in 90 (80.4%). The results for transmitted cases were that other imaging techniques were mandatory in 5 cases (5.5%), may be helpful in 21 (18.8%), and were not necessary in 86 (76.3%) ($x^2_1$ = 0.42, $P$ > .4).

The cases were graded as routine in 95 live cases (84.8%) and in 94 transmitted cases (83.9%). Seventeen live cases (15.2%) were judged to be challenging as were 18 transmitted cases (16.1%) ($x^2_1$ = 0.03, $P$ > .7).

More clinical information in connection with the cases was desired by the panel in 21 live cases (18.8%) and in 45 transmitted cases (40.2%) ($x^2_1$ = 12.4, $P$ < .001). Thus, the initial clinical information was rated sufficient in 91 live cases (81.2%) and in 67 transmitted cases (59.8%). After the interactive discussion, the panel rated the clinical information provided with the radiological images as sufficient in 100% of live cases and in 99.1% of transmitted cases. Thereafter, the interactive discussion significantly increased the level of information from an initial 91% to 100% of live cases ($x^2_1$ = 21.3, $P$ < .001), and from 81.2% to 99.1% in transmitted cases ($x^2_1$ = 52.7, $P$ < .001).

COMMENT

Telemedicine is a relatively new concept based on the application of modern telecommunication infrastructures. Telemedicine will increase and simplify the exchange and diffusion of medical knowledge, education, and information. However, these applications are not yet widely employed, surgical application remains lim-
limited, and the value of this new technology is still not generally as recognized as it could be. The main reason for this is probably a lack of information about its potential and a need for further investigation of telemedicine, especially in the field of surgery.

Telemedicine includes all the telecommunication technologies such as videoconferencing and digital image transmission via telephone, fax, mail, paging systems, and satellites. Currently, the use of videoconferencing systems is based on the new ISDN telephone line. The ISDN seems to offer the best quality-price ratio for real-time interactive exchange of information with sound and images in motion. Based on our 2 years of prior experience with international academic teleconferencing in surgery, we initiated a low-cost ISDN based on low-bandwidth technology at 128 kbps. The choice of such a basic system was believed to ease the introduction and the utilization of telemedicine for surgical applications, provided this technology allowed a qualitatively good level of remote diagnosis and advice.

Our analysis of the clinical documents of surgical patients shows that there was not a significant loss of technical quality of the images sent via teleconferencing; the technical quality from 7.2% of transmitted images was rated negatively, compared with only 4.5% of live images that received this rating ($P = .4$), resulting in a loss of technical quality based on telemedicine transmission in only 2.7% of cases. Despite this slight loss of quality, 100% of the target organs were recognizable via transmission, and the fine organ structure was analyzable in 89.9% of transmitted vs 95.5% of live images ($P = .03$). However, diagnosis based primarily on images was not possible in 15.2% of transmitted vs 6.3% of live cases ($P < .05$), suggesting the influence of factors other than just technical picture quality in the accuracy of remote diagnosis.

The clinical relevance of our results has to be assessed by a critical review of the methods employed. Basically, research in telemedicine involves 2 different aspects: the evaluation of the telemedicine system itself, thus assessing the value of the technology, and the evaluation of the telemedicine service, analyzing the effectiveness of the use of the technology. Most studies judge the technical image quality and do not consider the simultaneous evaluation of the clinical information provided. In the present study, both the quality of images transmitted via teleconferencing and the clinical relevance of real-time interactive discussion is analyzed to determine the benefits, if any, to the diagnostic accuracy in the field of surgery.

Our case selection was based on explicit criteria to eliminate the risk of later exclusion; thus, no cases had to be excluded, while in other studies, high retrospective exclusion rates have been reported. Our cases came from a prospective sampling of ongoing cases and therefore accurately reflect clinical reality. However, in other studies, the cases may come from a compilation of archives, selected, or unselected. This compilation enables one to answer more specific questions with a smaller set of images, but the study conditions are more artificial. Ideally, the same observers and the same cases should be used in both the study and the control conditions. However, the learning effect introduces an important bias in this type of analysis. Even if a random order is applied, the panel sees the same images in direct vision as well as via the teleconferencing unit, resulting in a bias in favor of the pictures seen the second time. We therefore used 2 different panels of surgeons and radiologists, blind to the diagnosis, who alternatively evaluated their own pictures in direct vision and then the pictures from the other team. This method also introduces a bias: 50% of transmitted images are judged by one panel, while 50% are judged by the other. At the end of the study, 100% of pictures, live or transmitted, were judged by 2 different panels, but each panel judged each picture only once, thus avoiding the learning effect. This bias may be somewhat offset by a good concordance in the rating of images by the different panels. Both panels rated their own images as good or excellent for 98.2% of cases and rated the images from the other hospital as good or excellent for 94.6% of cases ($P > .1$), thus achieving a similar assessment and validating their rating of transmitted images. Moreover, the members of the panels should be accustomed to using the teleconferencing unit (as was the case in our study) to avoid problems with the ergonomics of the system that could negatively influence clinical evaluation and to minimize the bias produced by the enthusiasm of new users.

The determination of a criterion standard or of a reference level is an important step in the analysis. The criterion standard in the present study was the diagnosis set by the control group, and the main outcome measure was the agreement between the remote and the live diagnosis. This is a recognized method but may introduce a bias in favor of the control group even if diagnosis is reached by consensus. Our panels reached a consensus in each live and remote case, and no case had to be excluded. This is not always possible. In other studies, failure to reach a consensus is described in up to 10% of cases and leads to exclusion. One reason for the high level of consensus in our study was that the panel members knew each other and had worked together, which may be regarded as an advantage, but may also introduce a bias in favor of the study group.

The question addressed by our findings is the clinical significance of the loss of performance through teletransmission: only 2.7% of our cases lost image quality. However, a significant loss of diagnostic accuracy in 8% of cases ($P = .03$) was observed. This suggests the influence of other factors, such as the choice of images studied (additional radiological studies were desired by the panel in up to 5% of cases), the choice of images themselves (more pictures were desired in up to 12% of cases), and the clinical comments provided, which had to be completed in up to 19% of cases. The need for additional imaging studies during the course of teleconferences is a recognized problem, but thanks to interactivity, such supplemental data may be requested and eventually immediately provided during real-time teleconferences. As confirmed by others, real-time interactive discussion permits a significant increase in the level of clinical
information, eg, in our study from 81.2% to 99.1% satisfaction for sufficient clinical information of the transmitted cases after discussion (P<.001). Thus, a diagnostic accuracy of 85% is similar to the 88% accuracy in interpreting digitized images on teleradiology systems, as described in a recent study in urology.23

The remote diagnosis was positively evaluated by a few surgical teams,26,27 demonstrating that telemedicine evaluation was similar to the evaluation by direct clinical examination with a diagnostic sensitivity between 78% and 98%.27 These results emphasize the importance of the interactive approach at both the remote and local site, thus favoring a real-time discussion, and suggesting that interactivity is a key factor in telemedicine.28 However, we have no data to support the idea that interactivity alone allows one to counterbalance poor image quality.

LIMITATIONS

One general limitation to telemedicine is the availability of advanced technology. The basic teleconference unit used for this trial cost $10000. It is believed to transmit pictures that are less clear than would a more sophisticated system, transmitting at up to 384 kb/s,28 such as we used in our previous study (cost, $40000). The benefit of this small unit is that it is not difficult to manage and is cost-effective and therefore easy to introduce and utilize. Since a relatively small loss of quality was measured for most routine cases (85%), it suggests that this basic unit is valuable for surgical remote consultation.

Medicolegal aspects were not addressed in this study.29 However, it is important to establish medical responsibility at the beginning of a telesurgery program and internal standards prior to submitting cases for telediagnosis.30,31 Usually, it is recommended that one assigns responsibility to the team in charge of the patient.10

The results of this study reveal that the use of low-bandwidth teleconferencing is possible and even reliable for clinical application in digestive and endocrine surgery. If the results of this study are confirmed by others, it will further demonstrate that telemedicine offers a new valuable avenue for providing broad access to expertise without travel and is therefore time-saving for experts and patients. Thus, remote diagnostic and therapeutic advice may become part of clinical routine practice.10 The next application of this technology, however, with the utilization of a broader bandwidth than was used in the present study,32 would be telementoring in which a remote expert would assist a surgeon, step-by-step, during an actual operation. This has already been performed recently on several occasions33,34 but needs further evaluation.

This study was supported by grant Telematics BBW 95.0492 from the Swiss Federal Office for Education and Sciences (Drs Demartines and Harder).

The authors would like to thank Kathleen Bucher-Rappaport for assistance with English-language editing.

Corresponding author: Nicolas Demartines, MD, Department of Hospital, University Hospital of Zurich, Ramistrasse 1000, CH 8000, Zurich, Switzerland (e-mail: demartines@ubaclu.unibas.ch).

REFERENCES