The Effect on Physician Interpretation after Ultrasound Video Transmission

Eric Zevallos, Joseph Coppiano, Ann Marie Kuchinski, Taylor Mueller, Bradley White, Patrick Loeffler, Zachary Farmer, Patricia Hall, Matt Lyon

Emergency Medicine and Hospitalist Services Department, Medical College of Georgia, Augusta, Georgia, USA
Email: Ezevallos@augusta.edu

Abstract

Title: The Effect on Physician Interpretation After Ultrasound Video Transmission. Objective: There is a limited amount of qualified sonographers, which is problematic to rural, and prehospital settings that require clinical ultrasound. One solution is the use of telemedicine to transmit ultrasound videos to a distant expert. Our objective was to determine the effect of video image transfer using Skype on physician image interpretation. Methods: Three emergency medicine physicians of varied ultrasound proficiency were given 72 pairs of transmitted and non-transmitted ultrasound video clips in random order. Raters were instructed to determine whether ultrasound clips were positive or negative for pathology. Same interpretations between video pairs were categorized as homodiagnostic while different interpretations were categorized as heterodiagnostic. Raters were also asked to rate the image resolution, image detail, and total image quality using a 10-point Likert scale. For the diagnostic results, rater agreement, accuracy, sensitivity, and specificity were calculated. Results: The raters collectively demonstrated high agreement (92.1%), sensitivity (96.3%), specificity (92.8%), and accuracy (94.1%). Raters 1 (95.8%, κ = 0.912) and 3 (97.2%, κ = 0.942) had near perfect agreement, while rater 2 (83.3%, κ = 0.666) had substantial agreement. Only raters 1 and 3 detected a decrease in image resolution, image detail, and total image quality (p < 0.05). Rater agreement on transmitted videos was not significantly different statistically from non-transmitted videos and the presence of heterodiagnostic video pairs did not correlate with video transmission. Conclusions: This study demonstrates that transmission via Skype has minimal effect on a physician’s interpretation of an ultrasound video. The implications of these results are that ultrasound video transmission can be used to improve access to specialists without compromising accuracy which will ultimately increase patients’ quality of care.
Keywords

Ultrasound, Telemedicine, Emergency Medicine, Transmission, Skype

1. Introduction

Since the 1980’s, emergency physicians have utilized clinical, or point-of-care, ultrasound as a tool for trauma care [1]. With its ease of use and versatility, a promising application for ultrasonography is its use in the prehospital and rural emergency room settings. Because of this ultrasonography via telemedicine has been growing in recent years in both prehospital and rural settings. During this time, ultrasonography’s accuracy and rapidity for evaluating areas of free fluid (e.g. hemorrhage) has been well established [2]-[9]. For example, prehospital focused assessments with sonography for trauma (PFAST) have enabled the detection of abdominal free fluid and other traumatic injuries prior to arrival at the receiving hospital. PFAST is an invaluable tool for directing transport to the appropriate level trauma center, improving triage, guiding the management of patients, and expediting patients’ course through the emergency department (ED) [9]-[17].

Initially there was apprehension among some physicians for the expanded use of ultrasonography due to its use by nonradiologists. However, emergency physicians have demonstrated their ability to learn and interpret ultrasonography, leading to the advent of bedside ultrasound [18]. However, widespread use of bedside ultrasonography has been limited by the lack of qualified emergency sonologists [19]. In the prehospital setting, this problem is exacerbated. The lack of ultrasonography-trained emergency physicians in rural EDs, non-academic EDs, and prehospital settings can be mitigated by the use of systems that allow for making diagnostic interpretations from remote locations, i.e. telemedicine. However, these systems require high quality imaging equipment that is prohibitively expensive for most EDs [20]. Studies indicate that paramedics with minimal training can perform remotely guided PFAST ultrasound examinations, but image interpretations require highly trained physicians that are often located some distance from the trauma scene, significantly delaying the diagnostic process [21].

An additional benefit to clinical ultrasound is that evaluations such as focused assessments with sonography for trauma (FAST), determination of intrauterine pregnancy (IUP), or gallbladder assessments can be made using a binary diagnostic approach. This binary approach to pathology is dependent on a number of additional variables such as operator skill, image degradation, etc. Still this approach simplifies interpretations as compared to other radiological studies, expanding its application to situations with more limited resources (e.g. trained sonographers). As a result, this opens the potential for ultrasound videos to be transmitted to highly trained physicians who can appropriately interpret an ultrasound examination.
Currently there is concern as to whether or not image transmission, by electronic means such as Skype, could result in image degradation resulting in poor image interpretation and possibly misdiagnosis. Previous research on this topic focused on the feasibility of transmitting ultrasound videos via different third party servers (ex: Skype) and an experienced sonographer’s ability to interpret the videos [22] [23]. In this study we sought to further the current literature by comparing the ability of emergency physicians with different levels of ultrasound experience to interpret transmitted videos. In addition, the videos were collected by a third year emergency resident without ultrasound fellowship training to better simulate the quality of image that would be received in an emergency setting. Our objective was to examine whether digital image transmission affected an emergency physician’s interpretation of an ultrasound video. We hypothesized that image transmission would not affect diagnostic conclusions in binary, pathology-present/pathology-absent formats (as described above) for emergent conditions assessed using clinical ultrasound.

2. Methods

This was a prospective study using transmitted and non-transmitted ultrasound videos (Figure 1). Videos were collected by a third-year emergency medicine resident with basic sonographer experience. The teleconferencing system consisted of a USB to VGA image capture device (Epiphan) connected to the VGA output in the mini-dock of the ultrasound system. The image capture device was connected to a Dell laptop (i7 processor) and using Skype, images were transmitted to a MacBook Pro (i7 processor). Video recordings were made of the transmitted image and stored as mp4 files. Images of the transmitted and non-transmitted images were of the same patient and were of the same US exam (Figure 2 and Figure 3).

Figure 1. Image collection flow chart. 1. The flow chart illustrates how the ultrasound videos were transferred. The process began with 72 Ultrasound video clips being compressed into a local Mac. The files were saved and became the 72 videos in the non-transmitted group. The files were also transferred thru skype to a remote Mac. These 72 files were saved on the remote Mac became the transmitted group.
Figure 2. Transmitted vs Non-Transmitted IUP exam. Side by side comparison of transmitted (left) vs. non-transmitted (right) ultrasound images depicting an intrauterine pregnancy.

Figure 3. Transmitted vs Non-Transmitted FAST exam. Side by side comparison of transmitted (left) vs. non-transmitted (right) ultrasound images depicting a Fast exam positive for the presence of fluid.

All videos were compressed from their original Digital Imaging and Communications in Medicine (DICOM) format into a Mp4 format that allowed for transmission via third party servers (ex: Skype). Non-transmitted videos were files solely in compressed Mp4 format while transmitted videos also incorporated the possible additional effects of being sent through third party servers. Digital transmissions of ultrasound videos via Skype were paired with the respective non-transmitted originals to produce a total of 144 ultrasound videos (72 pairs). Videos were three seconds in length and had a predetermined diagnosis that was designated as the gold standard diagnosis of the study.

All videos from both the transmitted and non-transmitted groups were presented to the raters in random order on a projector screen. Each rater was given an individual video clip only once and had an infinite amount of time to review the video before moving on to the next clip. Raters were not told which video type (transmitted or non-transmitted) they were given and were blinded from the study purpose.

Three raters were used in this study. Each rater was an emergency medicine physician trained in point-of-care ultrasound. Each had different levels of expertise determined by experience and frequency of ultrasound use. Rater 1 was an emergency medicine physician with over 20 years of experience using ultrasound in a clinical setting and reported using ultrasound multiple times per shift. Rater
2 was an emergency medicine physician without formal ultrasound training who reported using ultrasound sporadically. Rater 3 was an emergency medicine physician one year out of ultrasound fellowship training who reported using ultrasound multiple times per shift.

The ultrasound exams included tests for pleural effusion, IUP, gallstones, and the FAST exam. These exams were selected because they are the most commonly viewed in the emergency medicine setting. Each ultrasound video of a particular exam was collected using standard anatomical landmarks and required specific findings, such as observing the presence (+) or absence (−) of fluid in Morison’s pouch during a FAST exam (Table 1).

Of the 72 videos, 27 were designated as having a pathology and 45 videos were determined as normal (absence of a pathology). Each rater was also required to evaluate the videos for image resolution (IR), image detail (ID), and total image quality (TIQ) using a 10-point Likert scale. This was done prior to indicating a diagnosis on the presence or absence of pathology (+/− binary system). No clinical context was given for the videos. Any pair of videos (transmitted and non-transmitted) in which a rater made different diagnoses (+,− or −,+ or −,−) was categorized as a heterodiagnostic video pair. Any pair of videos in which a rater agreed with their own diagnosis (+,+ or −,−) was categorized as a homodiagnostic video pair.

Final diagnoses were analyzed for rater agreement, accuracy, sensitivity, and specificity. Agreement (whether a rater gave the same diagnosis to transmitted and non-transmitted videos) would be considered high if the rater had a low number of heterodiagnostic video pairs. Accuracy (whether the rater’s diagnosis agreed with Diagnosis Gold) would be considered high if the rater had a high number of diagnoses matching Diagnosis Gold. Sensitivity (the percentage of true positive pathologies a rater diagnosed) and specificity (the percentage of true negative pathologies a rater diagnosed) were also assessed. Intra-rater reliability was determined by calculating kappa coefficients. IR, ID, and TIQ were compared between the transmitted and non-transmitted videos using one-tailed, paired T-tests.

### Table 1. Ultrasound video images specifics.

<table>
<thead>
<tr>
<th>Ultrasound Exam</th>
<th>Anatomic Position</th>
<th>Required Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleural Effusion</td>
<td>Diaphragmatic view (right or left)</td>
<td>Presence or absence of pleural effusion</td>
</tr>
<tr>
<td>Intrauterine Pregnancy (IUP)</td>
<td>Transabdominal view</td>
<td>Presence or absence of IUP</td>
</tr>
<tr>
<td></td>
<td>Transvaginal View</td>
<td></td>
</tr>
<tr>
<td>Gallstone</td>
<td>Long access view</td>
<td>Presence or absence of gallstones</td>
</tr>
<tr>
<td></td>
<td>RUQ (Morison’s pouch)</td>
<td></td>
</tr>
<tr>
<td>FAST exam</td>
<td>LUQ (Splenorenal recess)</td>
<td>Presence or absence of fluid</td>
</tr>
<tr>
<td></td>
<td>Pelvic (Bladder)</td>
<td></td>
</tr>
</tbody>
</table>

The table demonstrates the different ultrasound exams tested, which anatomic positions, and what each rater needed to find when analyzing a given ultrasound video.
3. Results

In this study, overall agreement for the three raters was 92.1% (κ = 0.838), demonstrating near perfect agreement. Individually, Raters 1 and 3 had near perfect agreement, with Rater 1 agreeing on 69 of 72 paired videos (95.8%, κ = 0.912) and Rater 3 agreeing on 70 of 72 paired videos (97.2%, κ = 0.942). Rater 2 exhibited substantial agreement, making the same diagnoses on 60 of 72 paired videos (83.3%, κ = 0.666) (Tables 2-4).

Individually, Raters 1 and 3 detected a decrease in TIQ, IR and ID in the transmitted videos ranging in magnitude from −0.4 to −1.4 on a 10-point Likert scale (p < 0.05). This difference was statistically significant. Rater 2, however, did not observe statistically significant differences in TIQ or IR in the transmitted videos. Averaging all three raters’ scores, TIQ in transmitted videos decreased from 7.6 to 7.0, IR decreased from 7.4 to 6.6, and ID decreased from 7.4 to 6.6. Overall, for all three raters’ diagnoses sensitivity was 96.3%, specificity was 92.8%, and accuracy was 94.1% (Table 5). Only Rater 2 demonstrated a statistically significant difference between sensitivity and specificity (96.3% sensitivity, 84.1% specificity, p = 0.026), resulting in a higher rate of false-positives. Overall the raters’ agreement on transmitted videos was not statistically different than non-transmitted videos (p = 0.289), and was in fact slightly higher.

When looking only at the occurrence of heterodiagnostic video pairs, there was no correlation to image quality degradation from transmission or to wheth-

### Table 2. Rater 1’s performance.

<table>
<thead>
<tr>
<th>Rater 1</th>
<th>Non-transmitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmitted</td>
<td>−</td>
</tr>
<tr>
<td>−</td>
<td>43</td>
</tr>
<tr>
<td>+</td>
<td>1</td>
</tr>
</tbody>
</table>

The table shows diagnosis before and after transmission for rater 1. The (−) indicates the rater scored the video as negative for a pathology. The (+) indicates the rater scored the video as positive for a pathology. Between transmitted and non-transmitted categories, the (−, −) and (+, +) values indicate homodiagnostic video pairs while (−, +) or (+, −) indicates heterodiagnostic video pairs. There are a total of 72 images with paired data on diagnosis before and after transmission. The Kappa coefficient is 0.912. The agreement is statistically significant with $p = 1.258 	imes 10^{-13}$.

### Table 3. Rater 2’s performance.

<table>
<thead>
<tr>
<th>Rater 2</th>
<th>Non-transmitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmitted</td>
<td>−</td>
</tr>
<tr>
<td>−</td>
<td>32</td>
</tr>
<tr>
<td>+</td>
<td>6</td>
</tr>
</tbody>
</table>

The table shows diagnosis before and after transmission for rater 2. The (−) indicates the rater scored the video as negative for a pathology. The (+) indicates the rater scored the video as positive for a pathology. Between transmitted and non-transmitted categories, the (−, −) and (+, +) values indicate homodiagnostic video pairs while (−, +) or (+, −) indicates heterodiagnostic video pairs. There are a total of 72 images with paired data on diagnosis before and after transmission. The Kappa coefficient is 0.666. The agreement is statistically significant with $p = 8.974 	imes 10^{-7}$. 
Table 4. Rater 3’s performance.

<table>
<thead>
<tr>
<th>Rater 3</th>
<th>Non-transmitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmitted</td>
<td>−</td>
</tr>
<tr>
<td>−</td>
<td>42</td>
</tr>
<tr>
<td>+</td>
<td>0</td>
</tr>
</tbody>
</table>

The table shows diagnosis before and after transmission for rater 3. The (−) indicates the rater scored the video as negative for a pathology. The (+) indicates the rater scored the video as positive for a pathology. Between transmitted and non-transmitted categories, the (−, −) and (+, +) values indicate homodiagnostic video pairs while (−, +) or (+, −) indicates heterodiagnostic video pairs. There are a total of 72 images with paired data on diagnosis before and after transmission. The Kappa coefficient is 0.942. The agreement is statistically significant with \( p = 6.55 \times 10^{-15} \).

Table 5. Sensitivity and specificity of all raters.

<table>
<thead>
<tr>
<th>Rater</th>
<th>(+)</th>
<th>(−)</th>
<th>N</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52/54</td>
<td>87/88</td>
<td>69/71</td>
<td>70/71</td>
</tr>
<tr>
<td>2</td>
<td>52/54</td>
<td>74/88</td>
<td>61/71</td>
<td>65/71</td>
</tr>
<tr>
<td>3</td>
<td>52/54</td>
<td>84/88</td>
<td>67/71</td>
<td>69/71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rater</th>
<th>(+)</th>
<th>(−)</th>
<th>N</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.963</td>
<td>0.989</td>
<td>0.972</td>
<td>0.986</td>
</tr>
<tr>
<td>2</td>
<td>0.963*</td>
<td>0.841*</td>
<td>0.859</td>
<td>0.915</td>
</tr>
<tr>
<td>3</td>
<td>0.963</td>
<td>0.955</td>
<td>0.944</td>
<td>0.972</td>
</tr>
</tbody>
</table>

The table displays each rater’s performance. (+) indicates the sensitivity, (−) indicates specificity. N indicates non-transmitted images and T indicates transmitted images. The first table shows the number ratios while the second table shows the percentages. The asterisk (*) indicates value differences that are statistically significant at \( p < 0.05 \).

er pathology was present or not. There was little to no correlation of transmitted videos to false-negative diagnoses. Only once in 216 video pairs across all three raters was a true pathology missed on the transmitted video when it was detected on the corresponding non-transmitted video.

4. Discussion

The accuracy and utility of ultrasonography in emergency medicine is well established. As ultrasonography extends to the prehospital setting, the need for and utility of telemedicine ultrasound will grow. However, its widespread utilization is limited by a lack of resources, which includes ultrasonography-trained emergency physicians in rural EDs, non-academic EDs, and prehospital settings. Remotely guided FAST examinations, however, have been shown to be effective, even with minimally trained operators [17]. One roadblock to the widespread implementation of telemedicine ultrasound is concern about the potential for digital transmission to degrade image quality leading to misdiagnoses.

In this study, we investigated the occurrence of image transmission-induced changes in diagnoses. We focused on rater agreement rather than accuracy, because an incorrect diagnosis on a non-transmitted video should yield an incor-
rect diagnosis on a transmitted video indicating that transmission had no effect on a reviewer’s ultrasound interpretation. Our research revealed high agreement between raters’ diagnoses based on transmitted and non-transmitted videos. Even the least experienced rater demonstrated substantial agreement and no statistically significant difference between diagnoses based on transmitted and non-transmitted videos.

Despite transmission having no statistically significant effect on diagnoses, raters 1 and 3 were able to detect differences in image quality, resolution, and detail between transmitted and non-transmitted videos, with rater 2 only detecting differences in image detail. The results indicate that the raters were able to detect differences in quality between transmitted and non-transmitted videos. Furthermore, the lack of statistically significant differences between diagnoses based on transmitted and non-transmitted videos supports the notion that detectable changes from image transmission was not a hindrance to an emergency physician’s ability to make the same diagnosis. This supports that the effects from transmission are minimal. We hypothesize the reason rater 2 did not detect statistically significant differences in total image quality and image resolution between transmitted and non-transmitted videos is that rater 2 did not have the level of training and ultrasound experience as compared to raters 1 and 3. As a result, rater 2 was not able to detect the subtle differences between transmitted and non-transmitted videos when raters 1 and 3 did. There was no correlation between image transmission and heterodiagnostic video pairs, suggesting that image transmission did not have an effect on intra-rater reliability.

While our primary aim was to examine intra-rater reliability and the effect of transmission on diagnostic conclusions in clinical ultrasound, we also analyzed rater accuracy to determine any potential effects. All raters demonstrated high diagnostic accuracy. Without any clinical context, overall sensitivity, specificity and accuracy were 96.3%, 92.8% and 94.1%, respectively. There was no difference in accuracy between transmitted and non-transmitted videos. Additionally, interpreting video pairs differently (heterodiagnostic) was not associated with the presence or absence of pathology. Most importantly, there was no increased risk of false-negative diagnoses on transmitted videos.

These findings suggest that emergency physicians, particularly ones who use ultrasonography frequently, can accurately and precisely diagnose the presence of pathology for focused assessments of emergent conditions in binary, pathology-present/pathology-absent formats via electronic transmission if the videos are collected from a sonographer with minimal training (third year emergency medicine resident). This supports the use of telemedicine ultrasound to extend the application of ultrasonography to prehospital and rural settings where an EMT, with minimal ultrasound training similar to a third year emergency medicine resident, can transmit ultrasound videos to an experienced sonographer for quick interpretation.

The growing use of ultrasound in the prehospital setting can have significant healthcare benefits, particularly in mass-casualty settings and helicopter trans-
port, where telemedicine ultrasound can provide expert interpretations quickly and efficiently [14] [15] [16]. A centrally located ultrasound expert can oversee multiple providers working in remote locations, extending coverage to multiple contexts where training may be limited. A study by Lyon et al. 2015 found that 24.5% of all patient transfers from rural EDs to tertiary emergency centers were due to a lack of radiological services, primarily a lack of a radiologist to interpret the radiographic images [26]. A quarter of these transfers dealt with ultrasound imaging, making ultrasound interpretation a leading cause of rural-to-tertiary ED transfers. The study estimated that utilizing teleradiology, including tele-ultrasound, could potentially resolve up to 96.2% of all transfers due to radiological service impairments. Current apprehension still exists regarding image degradation secondary to digital transmission and its potential effect on diagnostic accuracy. This study rejects this apprehension and demonstrates that digital transmission represents a viable modality for tele-ultrasound.

4.1. Limitations
This study has several limitations. First, the scope of the study was limited. Future studies should expand the sample size, the number of raters, and the experience levels of raters. In addition, the images were collected by a third-year academic emergency medicine resident. Although the resident had minimal ultrasound training which is likely comparable to sonographers of “real world” prehospital and rural emergency department settings, there may still be variability from region to region. Furthermore, only one type of digital transmission was used. Other means of digital transmission should be examined to see if the results are comparable. Although video images were transferred via Skype, the effects of internet speed or fps were not assessed. As a pilot study, the aim for this research was to specifically assess the effect of transmission on physician interpretation. Future studies should be conducted to fully assess all aspects of the video transferring process. Also, this study only examined the use of video transmission in binary, pathology present/absent formats (albeit this determination was based on a combination of ultrasound findings partially dependent on operator skill, etc.) and it remains unknown if these results would extend to other types of diagnoses. Although efforts were made to include a broad array of images, the images analyzed may not be representative of what would be found in “real-world” prehospital and rural emergency medicine contexts due to them being in an academic hospital setting. Finally, this paper examined the transmission of the videos only and did not assess clinical outcomes.

4.2. Conclusion
The goal of this study was to examine whether the transmission of ultrasound videos via Skype had an effect on a physician’s ability to diagnose the presence or absence of pathology in binary, pathology-present/pathology absent formats. Our results, suggest that the answer to this question is no. There was near perfect agreement among the raters’ diagnoses between transmitted and non-transmit-
ted images. Despite the lack of clinical context, overall sensitivity, specificity and accuracy were high. Even the rater with the least amount of training and experience demonstrated good agreement and accuracy. These results support the finding that image transmission quality should not be an impediment to expanding access to radiologic interpretation via tele-ultrasound. These results support the use of digital transmission as a quick and reliable method of diagnosing whether pathology is present or absent to improve patient access and care.

**Acknowledgements**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**References**


[10] Heegaard, W., Plummer, D., Dries, D., Frascone, R.J., Pippert, G., Steel, D. and


