Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

A Systematic Review of Technical Evaluation in Telemedicine Systems

CA Thiyagarajan, M. Clarke
Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex, UK

Abstract— **We conducted a systematic review of the literature to critically analyse the evaluation and assessment frameworks that have been applied to telemedicine systems. Subjective methods were predominantly used for technical evaluation (59%), e.g. Likert scale. Those including objective measurements (41%) were restricted to simple metrics such as network time delays. Only three papers included a rigorous standards based objective approach. Our investigation has been unable to determine a definitive standards-based telemedicine evaluation framework that exists in the literature that may be applied systematically to assess and compare telemedicine systems. We conclude that work needs to be done to address this deficiency. We have therefore developed a framework that has been used to evaluate videoconferencing systems telemedicine applications. Our method seeks to be simple to allow relatively inexperienced users to make measurements, is objective and repeatable, is standards based, is inexpensive and requires little specialist equipment. We use the EIA 1956 broadcast test card to assess resolution, grey scale and for astigmatism. Colour discrimination is assessed with the TE 106 and Ishihara 24 colour scale chart. Network protocol analysis software is used to assess network performance (throughput, delay, jitter, packet loss).**

Keywords—**telemedicine, teleconsultation, evaluation.**

I. Introduction

Telemedicine systems are often the subject of evaluation in the literature, and systematic evaluation is advocated [48]. For example, in a review of articles published from 1966 to 2000 [3], 112 evaluations were cited; out of these 75% were clinical evaluation and 25% non clinical. Diagnostic accuracy (44%), patient satisfaction (23%), cost (8%) and clinical effectiveness (2%) were grouped under clinical evaluation; technical evaluation (13%) and management evaluation (12%) were grouped under non clinical. Technical evaluations were qualitative studies describing the technical outcomes such as bandwidth, resolution and colours in digital images. Some articles focused on technical requirements for diagnostic quality. No article describes a simple but comprehensive technical evaluation methodology which could be used by clinicians to verify the functionality of the whole telemedicine system. This could be explained by the lack of delineation of functional characteristics of specific clinical (teleconsultation) tasks.

This paper seeks to redress this imbalance in lack of systematic study of the requirements for technical evaluation and assessment of telemedicine systems. Note that we exclude the technical requirements for radiology, which has well defined standards and guidelines as part of the DICOM [53] and other regulatory standards.

We define a telemedicine system as in figure 1. Technology at the remote site acquires information in the form of moving images (video), still image, audio, data/text (documents) and physiological data. The data are transmitted over the communications network and reproduced to the user at the local site through technology.

The issue at the remote site is the ability to acquire the data with adequate fidelity. This includes the absolute limitations imposed by the standards used for the technology and the quality of the actual technology used. The communications network will impose bandwidth restrictions, but may have other technical limitations such as delay, jitter, errors, and loss of packets. The issue at the local site is the ability to reproduce the data to the user.

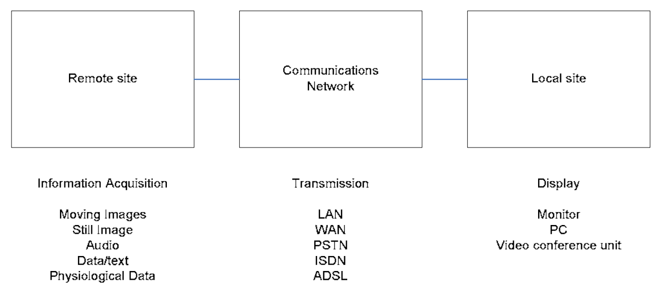

Figure 1 Telemedicine System

II. Methods

We conducted a systematic review of the literature to determine publications that reported evaluation and assessment frameworks that have been applied to telemedicine systems. We searched the INSPEC, MEDLINE, PROQUEST, EMBASE, CINAHL, EBSCOhost EJS and Cochrane Library publication databases using suitable terms. We also performed a systematic search of all articles in the peer reviewed journals in telemedicine. We then scored each paper on criteria designed to determine the relevance to technical evaluation and assessment.

III. Results

A total of 47 articles met our inclusion criteria. The papers were summarized using a tabular method, from which we were able to group the papers into general (25) and technical (22) evaluation methods. The papers in the technical group were further sub-divided into groups according to the environment, laboratory (11) and clinical (8), where the experiments were conducted. Out of the technical papers only 6 had experimental results conducted in a laboratory environment.

We also categorized the papers into: types of measurement; speciality of telemedicine investigated; statistics applied; tools used to conduct experiments; hardware requirements; and the degree of technical assistance required by a clinician to perform the test. We differentiated papers as concentrating on the separate stages; information acquisition and information display; network transmission; and system functionality. Information acquisition and information display included; resolution, colour contrast, colour discrimination, colour hue, colour saturation, image clarity and compression effect on image quality. Network transmission included; time delay measurement (e.g. still-image transfer time), video-clip transfer time, live video transmission rate and live video latency. System functionality included; quality of network transmission, and software functionality (e.g. whiteboard usage and application sharing).

A total of 27 technical measurements were found in 22 articles. Subjective methods were predominantly used for technical evaluation (59%), e.g. Likert scale. Those including objective measurements (41%) were mainly restricted to simple metrics such as time delays.

Only three papers [10,24,30] included a rigorous standards based objective approach, examining; resolution [10], image clarity, and colour contrast and discrimination [30]. For resolution, [10] used the TE107 and RMA resolution chart 1946. Ophthalmic vision charts were also used. For colour discrimination, [30] used purpose made swatches with calibrated colours. The RGB components were measured in the acquired image using software to determine the colour value of specific pixels. Standard tests based on standard ophthalmologic colour tests such as TE106 and Ishihara 24 plates were also used. No systematic or standards based tests were reported for moving images, when reported they were limited to simple hand waving.

Our results are summarised in table 1.

IV. A Reference Evaluation Framework

We had as aim to specify a set of objective tests that were easy to apply and gave repeatable results. Where possible, evaluation methods should be standards based and be inexpensive. The tests should be able to characterise the system end to end, as well as determine the performance of individual components.

Our review of the telemedicine literature produced no suitable methodology for technical evaluation or assessment. We therefore defined and evaluated our own assessment procedure.

A. Remote Site

The primary consideration at the local site is the quality and resolution of the acquired image. We used the EIA 1956 Broadcast Test Card [49] (figure 2) as standard. This provides capability to measure horizontal and vertical resolution at centre field and in the corners to determine any distortion of the image due to the camera lens. We make independent assessment of cameras by determining the resolution using an image capture card capable of composite (NTSC and PAL) and S-video. We then separately assess the quality of the videoconference system from knowledge of the quality of the camera image. There may be a qualitative assessment of noise and artefacts such as chrominance breakthrough. In each of the measurements, optimum camera settings should be used. Normally we have found that the camera needs to be set to manually adjust focus, aperture and shutter speed as automatic settings rarely produce the best quality image.

In general the S-video camera will produce the best quality image, as bandwidth is not restricted as in PAL or NTSC. A good quality camera (UK) will produce close to 720x576. Composite video (PAL) will produce around 500x576, and colour patterns may be produced due to chrominance breakthrough in some cameras.

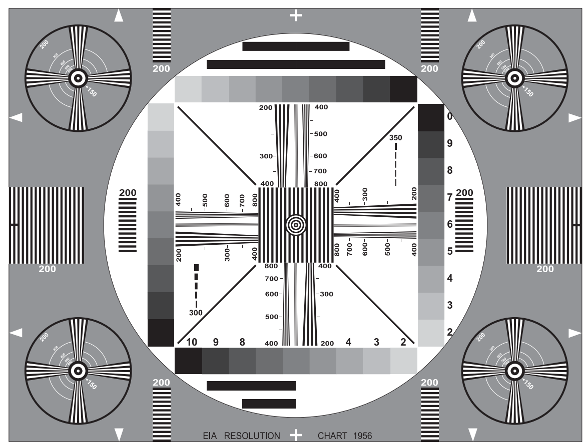

Figure 2 The EIA 1956 Broadcast Test Card

We use the Panasonic NV-DX100 as a high quality video camera for diagnostic images, as it has 3 CCD and very high quality optics. It provides simultaneous composite, S-video and digital video output and provides a good reference standard, producing near theoretical resolution in measurements.

We use the Sony D-31 as a "head and shoulders" camera; it has good optics and is remotely controllable. In our measurements the composite video gave close to theoretical resolution; however we determined that the S-video was only 300x576. The test was performed by switching between the two video standards without any other intervention.

A selection of typical low cost personal video conference system cameras produced a variety of results. Most had poor resolution and invariably the lens gave significant distortion due to the short focal length (fish eye). The majority had a small aperture and so gave poor focal plane, such that it was often difficult to have the entire test card in focus at the same time and the image had a barrel shape.

The resolution of the videoconference system is determined by the standard. When using a good quality camera, the resolution of CIF on a static image was determined to be close to the theoretical 352x288.

B. Communications Network

The primary consideration of the communications network is to characterise the quality of service (QoS). We have used readily available network analysis software (Ethereal) [50] to capture real time packets and analyse the protocol stream for throughput, delay, jitter, error and loss.

We have also established a test environment using the firewall of the FreeBSD operating system [51] to emulate transmission links having specified delay, bandwidth, packet loss, and buffer sizes. This may be used to simulate the characteristics of network systems found in telemedicine environments and applications may be tested under “real” conditions. This would allow, for example, investigation of the relationship between delay and loss on throughput.

C. Local Site

The primary consideration at the local site is the quality of production of images and audio. In general, the overall level of quality of the system is set by the limitations of the remote site and the communications network provided monitors of adequate quality are used. Radiology (DICOM) provides excellent guidelines on monitor quality and testing. A number of manufacturers [52] provide monitor testing software and tools to ensure accurate and repeatable rendition of images.

IV. DISCUSSION

In our systematic survey of the literature on the technical evaluation of telemedicine systems only 13% contained actual technical evaluation, and out of 47 papers, only 3 used objective tests. The vast majority used subjective measures to determine technical parameters. No paper described a systematic approach to full characterisation of the entire end to end system. We have therefore proposed and evaluated an assessment methodology based on objective testing.

V. CONCLUSION

We conclude that there has been insufficient rigorous objective technical evaluation and assessment of telemedicine systems. We have identified those papers that describe good practice and propose our own objective testing methodology.

REFERENCES

[1] NT Shaw, "'CHEATS': a generic information communication technology (ICT) evaluation framework", Computers in Biology and Medicine, vol 32, pp. 209-220, 2002.

[2] D Hailey et al, “An assessment framework for telemedicine applications”, vol. 5, pp. 162-170, 1999.

[3] A Ohinamaa et al, "Elements for assessment of telemedicine applications", International Journal of Technology Assessment in Health Care, vol. 17(2), pp. 190-202, 2001.

[4] P Taylor, "Evaluating telemedicine systems and services", Journal of Telemedicine and Telecare, vol. 11, pp. 167-177, 2005.

[5] R Holle et al, "Evaluation of telemedicine services", Information Technology in Biomedicine, IEEE Transactions, vol. 3(2), pp. 84-91, 1999.

[6] C Reed et al, "Navy telemedicine: a review of current and emerging research models", Telemedicine Journal and e-Health, vol. 10(3), pp. 343-356, 2004.

[7] N Aoki et al, "Outcomes and Methods in Telemedicine Evaluation", Telemedicine Journal and e-Health, vol. 9(4), pp. 393-401, 2003.

[8] J Gemmill, "Network basics for telemedicine", Journal of Telemedicine and Telecare, vol. 11, pp. 71-76, 2005.

[9] PJ Klutke et al, "The Telemedicine benchmark- a general tool to measure and compare the performance of video conferencing equipment in the telemedicine area", Computer Methods and Programs in Biomedicine, vol. 60, pp. 133-141, 1999.

[10] H Lamminen et al, "Visual tests for measuring the picture quality of teleconsultations for medical purposes", Computer Methods and Programs in Biomedicine, vol. 65, pp. 95-110, 2001.

[11] PJ Klutke et al, "Practical evaluation of standard-based low-cost video conferencing in telemedicine and epidemiological applications", Medical Informatics and the Internet in Medicine, vol. 24(2), pp. 135-145, 1999.

[12] P Mattioli et al, "Technical validation of low-cost video conferencing systems applied in orthopaedic teleconsulting services", Computer Methods and Programs in Biomedicine, vol. 60, pp. 143-152, 1999.

[13] ED Lemaire et al, "Technical evaluation of a low-bandwidth Internet-based system for teleconsultations", Journal of Telemedicine and Telecare, vol. 6, pp. 163-167, 2000.

[14] S Tachakra, "Technical quality of a minor injuries unit", Emergency Nurse (ProQuest Nursing Journals), vol. 9(5), pp. 10, 2001.

[15] FY Chan et al, "Minimum requirements for remote realtime fetal tele-ultrasound consultation", Journal of Telemedicine and Telecare, , vol. 5, pp. 171-176 1999.

[16] P Hussain et al, "Evaluation of training and diagnostic ultrsound service for general practitioners using narrowband ISDN", Journal of Telemedicine and Telecare, vol. 5(S1), pp. , 1999.

[17] R Bellazzil et al, "Web-based telemedicine systems for home-care: technical issues and experiences", Computer Methods and Programs in Biomedicine, vol. 64, pp. 175-187, 2001.

[18] Pelletier-Fleury et al, "Economic studies and 'technical' evaluation of telemedicine: the case of telemonitored polysomnography", Health Policy, vol. 49, pp. 179-194, 1999.

[19] C May et al, "When protocols fail: technical evaluation, biomedical knowledge, and the social production of 'facts' about telemedicine clinic", Social Science & Medicine, vol. 53, pp. 989-1002, 2001.

[20] DR Kaufman et al, "Usability in the real world: assessing medical information technologies in patients' homes", Journal of Biomedical Informatics, vol. 36, pp. 45-60, 2000.

[21] S Guillen et al, "Multimedia telehomecare system using standard TV set", IEEE Trans Biomed Eng , 49, vol. pp. 1431-1437.

[22] DRP Oliver et al, "The usability of videophones for seniors and hospice providers: a brief report of two studies", Computers in Biology and Medicine, vol. 32, pp. 209-220, 2002.

[23] L Mued et al, "The effects of Lip synchronization in IP conferencing", International Conference on Visual Information Engineering (VIE 2003) (IEE Conf.Pub No.495), pages 210-213, 2003.

[24] WM Guiness et al, "Developing an accurate system of measuring colour in a venous leg ulcer in order to assess healing",Journal of Wound Care, vol. 14(6), pp. 249-258, 2005.

[25] Anonymous, "Planning considerations for the deployment of LAN-based video conferencing and related applications", Available http://www.ukern.ac.uk/development/video/archive/#3.

[26] M Fregonara et al, "On-field evaluation of CAROLIN an Italian teleconsulting cardiology application: early results", Computers in Cardiology, vol. Sept, pp. 217-220, 1998.

[27] SE Poltrock et al, "Videoconferencing: Recent Experiments and Reassessment", Proceedings of the 38th Hawain International Conference on System Sciences, 2005.

[28] T Turletti et al, "Videoconferencing on the Internet", IEEE/ACM Transactions Networking, vol. 4(3), pp. , 1996.
[29] MR Frater et al, "Impact of Audio on subjective assessment of video quality in video conferencing applications", IEEE Trans on Circuits and Systems for Video technology, vol. 11(9), pp. 1059-1062, 2001.
[30] TJ Broderick, "Impact of varying transmission bandwidth on image quality", Telemedicine Journal and e-Health, vol. 7(1), pp. , 2001.
[31] R Finger, "Measuring quality in videoconferencing systems", Business Communications Review, vol. 28(6), pp. 51-56, 1998.
[32] EA Miller, "The technical and interpersonal aspects of telemedicine: effects on doctor-Patient Communication", Journal of Telemedicine and Telecare, vol. 9, pp. 1-7, 2003.
[33] JM Drescher et al, "Multi-site telemammography system: preliminary assessment of technical and operational issues", Proceedings of the SPIE- The International Society for Optical Engineering, v5033, p360-9, 2003.
[34] M Allen, "Videoconferencing for Continuing medical education from pilot project to sustained program", Journal of Telemedicine and Telecare, vol. 8, pp. 131-137, 2002.
[35] M Allen, "Evaluation of videoconferenced grand rounds", Journal of Telemedicine and Telecare, vol. 8, pp. 210-216, 2002.
[36] A Fitzgerald, "Child development services: a multidisciplinary approach to professional education via videoconference", Journal of Telemedicine and Telecare, vol. 8:S-3, pp. 19-21, 2002.
[37] PW Callas et al, "Improved Rural Provider Access to Continuing Medical Education Through Interactive Videoconferencing", Telemedicine Journal and e-Health, vol. 6(4), pp. 393-399, 2000.
[38] P Mattioli et al, "A study of the application sharing capabilities in telemedicine", Computer Methods and Programs in Biomedicine, vol. 58, pp. 89-97, 1999.
[39] S Paschalakis, "Real-time face detection and tracking for mobile videoconferencing", Real-Time Imaging, vol. 10, pp. 81-94, 2004.
[40] M Bocker, "Symbols for point-to-point videotelephony function", Computer Standards & Interfaces, vol. 17, pp. 359-362, 1995.
[41] DL Unruh, "Desktop videoconferencing: The promise and problems of delivery of web-based training", Internet and Higher Education, vol. 3, pp. 183-200, 1999.
[42] MR Civanlar, "Peer-to-peer multipoint videoconferencing on the Internet", Image Communication (accepted for publication).
[43] R Giacomelli, "Videoconferences in Physics, Nuclear Physics B (Proc. Suppl), vol. 54B, pp. 231-234, 1997.
[44] M Beach et al, "Evaluating telemedicine in an accident and emergency setting", Computer Methods and Programs in Biomedicine, vol. 64, pp. 215-223, 2001.
[45] TC Schmidt et al, "Global serverless videoconferencing over IP", Future Generation Computer Systems, vol. 19, pp. 219-227, 2003.
[46] S Whittaker, "Rethinking video as a technology for interpersonal communications: theory and design implications", Int. J. Human-Computer Studies, vol. 42, pp. 501-529, 1995.
[47] T Inoue et al, "Evaluation of a Videoconferencing system based on TV program", Electrical and Electronic Engineers in Israel, 19th Convention of 5-6 Nov, pp. 436-439, 1996.
[48] R Bashur et al, "Telemedicine Evaluation", Telemedicine and e-Health, vol. 11, pp. 296-316, 2005.
[49] EIA 1956 Broadcast Test Chart, http://www.pima.net/standards/iso/tc42/wg18/kp_testchart.htm
[50] Ethereal, http://www.ethereal.org.
[51] FreeBSD, http://www.freebsd.org
[52] Eizo-test, http://www.eizo.com
[53] DICOM, http://www.dicom.org

TABLE I
SUMMARY OF TECHNICAL REVIEW PAPERS

Article	Environment		Measurements			
	Lab	Clinical	Subjective	Objective	Statistics	Summary of Tests
10	+	-	+	+	+	Resolution, Colour Contrast, Colour Discrimination Image acquisition and Display analysis
11	-	+	+	-	+	Overall system analysis Questionnaire
12	+	-	+	+	+	Time delays measurement. Overall system analysis
13	-	+	+	+	+	Still image transfer time. video clip transfer time, live video latency, frame rate. Network transmission analysis
14	-	+	+	+	+	Blue intrusion, line break out, number of magnification needed, quality of image, quality of sound, ease of dialling, general problems with the system Overall system analysis
15	-	+	+	-	+	Overall system analysis
16	-	+	+	-	+	Overall system analysis Overall system analysis
21	-	+	+	-	+	User acceptance using ISDN and IP. Network transmission analysis
23	+	-	+	-	+	Effect of audio-video delay (Lip synchronization) Mean Opinion Score Network transmission analysis
24	-	+	+	+	-	Colour Saturation, Colour hue Image acquisition and display analysis
29	+	-	+	-	+	Double-Stimulus Continuous Quality Assessment Scale Information acquisition and display analysis
30	+	-	+	+	+	Image Clarity, Colour fidelity with varying bandwidths Image acquisition and display analysis
38	+	-	+	-	-	Image formats vs time delays Image acquisition and display analysis
47	-	+	+	-	-	Image acquisition and display analysis