An Evaluation of Four Telemedicine Systems for Primary Care

By Earl V. Dunn, David W. Conrath, William G. Bloor, and Barbara Tranquada

In an evaluation of the efficacy of four two-way telecommunication systems for use in primary care, more than 1,000 patients seeking care at a community health center received an additional remote examination by use of either color television, black and white television, still-frame black and white television, or hands-free telephone. The diagnoses, clinical tests and X rays requested, and proposed patient management were compared to the actual care received by the patients at the health center.

There were no significant differences between any of the modes in relation to diagnostic accuracy, time for the diagnostic interview, tests requested, or referral rates. Furthermore, patient attitudes did not vary significantly. Thus the relatively inexpensive telephone proved to be as efficient and effective a means for delivery of remote physician care as did any of the visual communication systems.

It is no great surprise to discover that remote areas and areas with sparse population are medically underserved. Willemain and Mark [1] developed a graph of the number of physicians per person versus population density in the state of Massachusetts. There was a very clear difference in the ratio, dependent on the population. There were more than three times the number of physicians per capita in the most populous counties as in the least populous counties; such figures have been reproduced over and over again. Several approaches have been suggested to alleviate the situation and improve health care in remote areas. These suggestions include physician incentives, the use of physician extenders (physician assistants, nurse practitioners, paramedics, and the like) as well as the use of modern communications technology.

This research was supported by a grant from Physicians' Services Incorporated, Ontario.

Address communications and requests for reprints to Earl V. Dunn, Associate Professor, Faculty of Medicine, University of Toronto, Sunnybrook Medical Centre, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada. David W. Conrath is professor of management sciences at the University of Waterloo; William G. Bloor is a programmer at McMaster University; and Barbara Tranquada is the nurse described in this article.

In the past decade increasing interest has been shown in the feasibility of using telecommunications to aid in the delivery of health care. Publication of such efforts did not take place on any scale until the 1970s. Most of the investigations have been related to the feasibility of systems or to special purpose uses. Conrath et al. [2] and Park [3] have reviewed the literature; few studies have contrasted systems or studied the use of telecommunications for primary medical care.

A forerunner in the use of two-way television was a system connecting a medical station located at Boston's Logan Airport to Massachusetts General Hospital [4]. Since that time many other systems have been devised and their feasibility tested. More recently, Moore et al. [5] compared two-way black and white television and the telephone for remote medical consultation between two inner-city clinics and a central hospital. Sasmor and Sanders [6] have described a system operating in the Dade County prison system in Florida; we have not yet seen any published results of their study. The latter two studies are the only ones of which we are aware that compared alternative telemedicine systems.

We are involved in a four-phase project to compare various telecommunication systems for primary health care, the most common need in remote areas. In phase I we observed primary care physicians in their offices and recorded some basic facts related to communications systems. How often does the physician touch the patient? Does he need color for a diagnosis? Does he use data from medical records? With the resulting information we tested a research methodology to answer some basic questions by comparing three telecommunication modes: two-way color television, two-way black and white television, and hands-free telephone. There were no significant differences between the three modes in diagnostic accuracy and efficiency (i.e., time for diagnosis). There were some small differences in the attitudes of the physicians, the nurses, and the patients. All these results have been reported elsewhere [2, 7].

The results of phase II, although statistically significant, were based on only 32 patients. The patients presented with actual problems, but the research setting did not include the patients' initial contacts with the health care system. We did not wish to base decisions on such a small sample studied in an artificial setting; in addition, we wished to include a fourth telemedicine mode, still-frame black and white television. (The reasons for including this mode were twofold: it contains most of the visual elements we wished to study, but is much less expensive and can be used wherever telephone lines exist. Live black and white or color television requires broad-band transmission capability.) Thus we proceeded to phase III, a situation more nearly like the real world of medical care.

This article reports some of the results of our phase III study, which compared four systems regarding diagnosis, patient management, patient and physician attitudes, and other parameters in more than 1,000 patient-physician encounters. The patients were seen dur-

ing a six-month period in a primary care setting in a large city (over two million population). The patients could be managed in various ways or could be referred if the physician so desired. The four systems compared were two-way color television, two-way black and white television, hands-free telephone, and two-way still-frame black and white television. The still-frame mode had a still picture updated every 30 seconds.

Hypotheses. Five hypotheses were developed from our earlier results. They were:

1. There would be no significant differences in diagnostic accuracy among the four telecommunication modes.
2. There would be no significant differences in patient management practices among the four modes.
3. There would be no significant differences in the time taken to reach a diagnosis among the four modes.
4. There would be no significant differences in the number of clinical tests and X rays requested among the four modes.
5. There would be no significant differences in the rate of referrals to other specialists among the four telecommunication modes.

Methodology

A community health center (Flemingdon Health Centre) was connected to Sunnybrook Medical Centre, a teaching hospital of the University of Toronto, three miles distant, via a high-quality microwave link. A separate examination room was set aside at the health center for the experiment. This room contained a Fernseh KCP color television camera, a Sony black and white television camera, a Hughes scan converter (for the still-frame black and white mode), and Northern Electric's Companion 2 hands-free telephone set. There was also a self-view monitor and a monitor that let the patient see the physician conducting the examination.

Similar equipment was set up in a separate research examining room at the teaching hospital. A less expensive Sony color camera was used for the physician's picture, and a studio-quality monitor was used to pick up the patient's video image. The two video signals traveled via the broad-band microwave system. All audio signals were via the Companion 2, whether or not the video system was also in use.

On each day of the experiment the research nurse at the health center determined which system was to operate for that day. The systems were rotated to give each remote physician equal time on each system and also to ensure that an equal number of patients were examined remotely via each system. When the system was operating, the nurse approached patients attending the clinic and asked them to volunteer for a research project. The patient accompanied the nurse to the clinic examination room, was introduced to the remote physician, and the diagnostic session began. All patients were examined face to face (and treated, if necessary) by their own clinic physicians.

For ethical reasons the remote physician did not disclose his diagnosis to the patient and did not indicate to the patient what management program he would recommend. This information was obtained by the subject patient only from his own physician, whom the patient saw either before or after the remote physician.

The subject patients were regular patients at the health center. If a patient volunteered he was paid $5.00 for his efforts. The patients who volunteered seemed to be a cross section of the clientele of the clinic. The major reason given by patients who did not volunteer was the lack of time to participate.

The physicians who examined the patients by telecommunication were all practicing family physicians. Their experience ranged from two to over 20 years of practice. Ten time slots of three hours each were scheduled every week, and the physicians agreed to take one time slot on a regular basis for a six-month period. Seven male and three female physicians participated as research physicians. Six others were involved as the clinic physicians. All physicians involved were members of the faculty of medicine of the University of Toronto, although the remote physicians carried greater teaching loads.

The nurse attending the patients had over 10 years of nursing experience. She was not a trained nurse practitioner but had a varied background including emergency room experience. The same nurse was used throughout the experiment.

Data Collection

Each physician recorded his diagnosis and his suggested tests and patient management. He also answered questions about his attitudes toward the system. The patient and the nurse completed questionnaires about their attitudes toward the various modes. The physician and the nurse recorded the time of the beginning and the end of the diagnostic interview.

The physician listed his diagnoses under the categories of primary and secondary diagnosis. He also recorded whether the condition was acute or chronic and stated his confidence level related to the diagnosis (A = 95–100 percent sure, B = 75–95 percent sure, C = 50–75 percent sure, and D = 0–50 percent sure). The research physicians coded all their diagnoses according to the International Classification of Health Problems in Primary Care (ICHPPC) [8]. Diagnoses made by clinic physicians were coded by the research nurse, who checked with the physician if she was uncertain as to the correct coding. The ICHPPC identifies 466 medical problems in 18 separate categories. There is some duplication, so that 371 distinct medical problems are listed under one or more of the 18 categories. Agreement was counted only if the two physicians coded the identical distinct medical problem as given in the ICHPPC.

Patient management, as outlined by the physician, was coded into one of 62 categories, which included referrals; recommendations for counseling, physiotherapy, or diet; and various therapeutic agents,

such as antibiotics, analgesics, cough preparations, and the like. With reference to the attending physician's actual management of the patient, the research physician's suggested management was also coded into "agree" or "disagree" categories. The agree-disagree status was decided on the basis of a common effect: for example, all antibiotics were treated as equivalent unless the patient was allergic to one of the medications or unless a culture susceptibility assay was available. Patient management was further classified according to whether it was "critical" (without such treatment—e.g., rabies shots—on that visit, the patient might become critically ill), "beneficial" (active therapeutically, e.g., an antibiotic), "neutral" (related to patient comfort or symptom relief, e.g., aspirin or cough mixture), or "harmful" (detrimental to the patient's well-being).

Clinical tests and X rays requested by the physicians were classified into one of 18 categories, including X rays, hematology, biochemistry, microbiology, etc., and coded accordingly. Agreement between the remote and attending physicians was counted only when their recorded requests were essentially identical. For example, if one physician requested an X ray of the chest and the second physician requested an X ray of the ribs this was recorded as a disagreement even if the same information might be obtained by the two different procedures. The physicians listed their requested tests as either "essential" or "helpful."

The time for the diagnostic process was taken as the interval from the introduction of the patient to the end of the consultation. The time intervals reported here are those recorded by the remote physicians. The nurse also recorded the time interval as a check.

Attitudinal data were obtained from questionnaires completed for each physician-patient encounter by the remote physician, the nurse, and the patient. Since each remote physician and the nurse consulted via all the modes, physician and nurse attitudes are not independent variables and thus cannot be analyzed in the same way as the patient attitude data.

Results

It must be stated emphatically that in this study no attempt was made to measure quality of health care. All results were compared with the actual care received by the patient as the standard. If the remote physician made a diagnosis or proposed a management different from this standard a disagreement was recorded, no matter how appropriate the diagnosis or management might have been.

The total number of physician-patient encounters was 1,015, distributed among the four telecommunication modes as follows:

color television	244 patients
black and white television	269 patients
hands-free telephone	253 patients
still-frame television	249 patients

Table 1. Diagnoses and Agreement by Remote and Clinic Physicians
(For all intermode differences, $0.95 > p > 0.50$)

Communication mode	*A**	*RD*†	*CD*§	Agreement ratio: $A/(A+CD)$‡
All	826	747	643	0.562
Color TV	193	164	158	0.550
BW TV	218	220	163	0.572
Telephone	204	187	167	0.553
Still-frame TV	211	176	157	0.573

* A = number of diagnoses made identically by both remote and clinic physicians.
† RD = diagnoses made by remote physicians only.
§ CD = diagnoses made by clinic physicians only.
‡ $A + CD$ = total diagnoses made by clinic physicians—the standard.

Diagnostic Agreement

A total of 1,573 diagnoses (1,276 primary and 297 secondary) were made by the remote physicians, and 1,469 diagnoses (1,245 primary and 224 secondary) were made by the clinic physicians. Table 1 shows, by communication mode, the number of diagnoses in which both physicians agreed (A) and the numbers of remote diagnoses and clinic diagnoses in which agreement did not occur, RD and CD, respectively. The measure of agreement was taken as the ratio $A/(A + CD)$, where $A + CD$ is the total number of diagnoses made by the clinic physicians.

It is apparent from Table 1 that there were no significant differences among the modes in relation to diagnostic agreement. If primary diagnoses or secondary diagnoses are considered alone the results are similar, except that in the case of secondary diagnoses the agreement ratio was much lower. The agreement within diagnostic categories (e.g., diseases of the respiratory system) was 0.754. (Such agreement ratios are consistent with other studies of clinical reliability reviewed by Koran [9].) An analysis of diagnostic agreement in relation to the confidence level recorded by the remote physician showed the agreement ratios to be 0.596 for confidence level A, 0.435 for level B, 0.336 for level C, and 0.231 for confidence level D.

Table 2. Number of Clinical Tests Requested by Remote and Clinic Physicians
(For all intermode differences, $0.99 > p > 0.1$)

Communication mode	Remote		Clinic		Remote/Clinic	
	Essential	Helpful	Essential	Helpful	Essential	Helpful
All	856	352	553	66	1.55	5.33
Color TV	188	90	121	20	1.55	4.50
BW TV	212	102	134	13	1.58	7.85
Telephone	219	76	142	13	1.54	5.85
Still-frame TV	237	84	156	20	1.52	4.20

Table 3. Agreement on Clinical Tests by Remote and Clinic Physicians
(For all intermode differences, $p > 0.25$)

Communication mode	*A**	*RD*†	*CD*§	Agreement ratio: $A/(A + CD)$‡
All	356	852	263	0.575
Color TV	77	201	64	0.546
BW TV	80	234	67	0.544
Telephone	92	203	63	0.594
Still-frame TV	107	214	69	0.608

* A = number of tests ordered by both remote and clinic physicians.
† RD = tests ordered by remote physicians only.
§ CD = tests ordered by clinic physicians only.
‡ $A + CD$ = total tests ordered by clinic physicians.

Clinical Tests

The remote physicians requested 1,208 tests, and the clinic physicians requested 619. Table 2 shows the number of tests requested by each group in each mode and in the "essential" and "helpful" categories within each mode; the remote physicians consistently requested more tests—especially more "helpful" ones, which make up the largest portion of the excess tests. Agreement between tests requested by remote and clinic physicians is shown in Table 3, measured in the same way as diagnostic agreement. Agreement did not differ significantly between modes, and the magnitudes of the ratios are similar to those for diagnostic agreement.

Patient Management

Table 4 shows agreement ratios in the four modes for "beneficial" patient management plans. It is interesting to note that the ratios are higher than for diagnostic agreement. Among "critical" management plans there were only two cases of disagreement. In one

Table 4. Agreement in "Beneficial" Patient Management Plans by Remote and Clinic Physicians
(For all intermode differences, $0.95 > p > 0.50$)

Communication mode	*A**	*RD*†	*CD*§	Agreement ratio: $A/(A + CD)$‡
All	545	383	311	0.637
Color TV	129	104	72	0.642
BW TV	143	97	82	0.636
Telephone	142	100	77	0.648
Still-frame TV	131	82	77	0.630

* A = management plans recommended by both remote and clinic physicians.
† RD = plans recommended by remote physicians only.
§ CD = plans recommended by clinic physicians only.
‡ $A + CD$ = total plans recommended by clinic physicians.

Table 5. Analysis of Variance in Diagnosis Time

Source of variance	d.f.	Sum of squares	Mean square	F	p
Physician	9	993.16	110.35	5.78	0.01
Mode	3	107.77	35.92	1.88	n.s.
Complexity*	1	488.00	488.00	25.58	0.01
Mode × physician	27	741.79	27.47	1.44	n.s.
Mode × complexity	3	71.73	23.91	1.25	n.s.
Physician × complexity	9	209.17	23.24	1.22	n.s.
Error	912	17 397.81	19.08	...	...
Total	964†	20 009.43	...	...	...

* Number of diagnoses per patient.
† 51 patients were omitted because time data were missing.

case, a man with a cat bite was prescribed antirabies injections by the clinic physician (the standard). However, the cat had received rabies vaccine, and the remote physician did not suggest the patient receive antirabies treatment, but rather that the cat be observed. In the second case, the remote physician made a diagnosis of a contused elbow in a child but requested an X ray. The clinic physician, with an X ray available, diagnosed a dislocated head of the radius and referred the patient to an orthopedic specialist. Examples of agreement in "critical" management plans include discontinuing digoxin in a case of digitalis toxicity and recommending Adrenalin for an anaphylactoid reaction.

The agreement ratios for "neutral" management plans also showed no significant differences, although the ratios were lower, about 0.4. One "harmful" management procedure was prescribed, by a clinic physician: a patient with a history of allergy to penicillin was given a prescription for penicillin as actual treatment. None of the remote physicians recommended "harmful" procedures for the study patients.

Efficiency

Three factors were considered in analyzing the time taken to reach a diagnosis: the telecommunication modes, the physician's personal characteristics (some physicians take more time than others), and the case complexity. The only available measure of case complexity was the number of diagnoses per patient, on the assumption that, in general, the more diagnoses the patient has the more complex the case. An analysis of variance across these three factors (Table 5) shows that there were significant differences in the time due to physician differences and due to case complexity, but no significant differences due to communications mode. (The actual times for diagnosis varied from three minutes to over an hour, with a mean time of about seven minutes.)

We must point out that this analysis concerns diagnosis times only, not including time needed to inform the patient of the diag-

Table 6. Patient Responses to "Do You Think the Doctor Understood Your Problem?"

Communication mode	Yes	Probably	Don't know	Unlikely	No
Color TV	194	38	11	0	1
BW TV	199	45	20	2	3
Telephone	189	46	12	2	4
Still-frame TV	180	49	19	0	1

nosis and the management plan; as mentioned, the remote physician did not perform this latter step. We cannot comment on any time differences that might occur between modes in actually instructing the patient, and we could not determine whether there would have been differences in patient compliance related to the mode used.

Referrals

As with the other parameters previously discussed, there were no significant differences in referral among the telecommunication modes. There were 67 referrals by the clinic physicians and 79 suggested referrals by the remote physicians. These included referrals to the dietitian and social worker. The agreement ratio in all modes was about 0.45, with the ratio for color television slightly higher but not significantly different.

Other Results

Many other data were recorded, but only a few additional results will be mentioned here. The patient attitude data are of interest. In our phase II study we found a definite rank order of patient preference for color television, followed by black and white television and hands-free telephone [7]. In the present study the patients revealed no preference for any one of the four modes: for example, Table 6 shows patient responses to the question "Do you think the doctor understood your problem?"

Responses by the patients to several other questions confirmed this lack of attitudinal differences. We suspect that the difference between our phase II and phase III results may be related to the fact that in phase II each patient experienced all the modes and a face-to-face consultation and thus could make subjective comparisons between the modes. In phase III each patient experienced only one mode, and his experience thus became an independent variable compared only to his prior face-to-face physician contacts. Our present attitudinal data indicate there is no patient preference. The physician attitudinal data, not yet fully analyzed, show differences in physician attitude, but this is clearly related to the individual physician's preference and does not show a rank order of preference.

Other data indicate some interesting phenomena. For example, our present data do not indicate any significant physician learning over six months except for a possibly shorter time for diagnosis.

Table 7. Agreement in Diagnoses of Diseases of Skin and Subcutaneous Tissue
(For all intermode differences, $0.1 > p > 0.02$)

Communication mode	*A*	*RD*	*CD*	Agreement ratio: $A/(A + CD)$
All	100	75	74	0.575
Color TV	21	18	25	0.457
BW TV	33	24	17	0.660
Telephone	21	19	22	0.488
Still-frame TV	25	14	10	0.714

As stated earlier, the 466 diagnoses in the ICHPPC classification can be grouped in 18 disease categories. When the data on diagnostic agreement were analyzed according to these groupings only one grouping showed any difference between the modes; this was "diseases of the skin and subcutaneous tissues." In this grouping black and white television, both live and still-frame, appeared superior to both color television and hands-free telephone, as shown in Table 7. The poor performance of hands-free telephone was expected, but not the poor showing of color television. Within all other disease categories no clear superiority was found.

Conclusions

From our results we can draw one major conclusion. We are unable to measure any significant differences related to diagnostic accuracy, tests requested, patient management practices, efficiency, and referral rates among the four telecommunication modes. Why are the differences so small? Surely high-quality color television is a far cry from hands-free audio, and yet the performance of the two is virtually identical. In specific cases we have some hypotheses. For example, we suspect that the hands-free telephone system performed poorly for skin problems because of the lack of visual image and, in contrast, the color television system performed poorly because the physicians were misled by the color inaccuracies of the color system used in North America.

Regarding the general tendency of remote physicians in this study to request greater numbers of tests, it is possible that the remote physicians were accustomed to a different style of practice, perhaps more "defensive" than the practice of the clinic physicians. Alternatively, the remote physicians may have felt some uncertainty in the unfamiliar situation of remote diagnosis. Our data give no evidence that the tendency to order more tests changed over the six-month period of the study.

We intend to examine all our data in detail to try to explain the general lack of differences and to delineate further some of the various minor differences by mode. There is no question that there are small differences between the modes. There is also some preliminary evidence that physicians practice differently when using different

modes; this includes both different methods of arriving at a diagnosis and different patient management practices.

In summary, until strong evidence appears that is clearly contrary to our present findings, we question the advisability of building expensive broad-band video systems to assist in the delivery of primary health care. The alternative narrow-band systems, audio or still-frame (slow scan) video, are substantially cheaper, generally more reliable, and appear to provide equally effective health care management.

REFERENCES

1. Willemain, T.R. and R.G. Mark. Models of remote health care systems. *Biomed Sci Instrum* 8:9, 1971.
2. Conrath, D.W., E.V. Dunn, J.N. Swanson, and P.D. Buckingham. A preliminary evaluation of alternative telecommunication systems for the delivery of primary health care to remote areas. *IEEE Trans Commun* 23:1119 Oct. 1975.
3. Park, B. *An Introduction to Telemedicine.* New York: The Alternate Media Center, New York University, 1974.
4. Marplis, R.L. and K.T. Bird. Telediagnosis: A new community health resource. *Am J Public Health* 64:113 Feb. 1974.
5. Moore, G.T., T. R. Willemain, R. Bonanno, W.D. Clark, A.R. Martin, and R.P. Mogielnicki. Comparison of television and telephone for remote medical consultation. *N Engl J Med* 292:729 Apr. 3, 1975.
6. Sasmor, L. and J. Sanders. A quantitative evaluation of a health care distribution system: The Miami Telemedicine Project. Paper presented at the First International Conference on Progress and Prospects in Health Care Distribution Systems, Miami, Fl, Nov. 1974.
7. Conrath, D.W., P. Buckingham, and E.V. Dunn. An experimental evaluation of alternate telecommunication systems as used for medical diagnoses. *Behav Sci* 20:292 Sept. 1975.
8. *International Classification of Health Problems in Primary Care,* 1st ed. Chicago: American Hospital Association, 1975.
9. Koran, L.M. The reliability of clinical methods, data and judgments—Second part. *N Engl J Med* 293:695 Oct. 2, 1975.