A systematic review of research methodology in telemedicine studies

Pamela Whitten*, Liv Karen Johannessen†, Tove Soerensen†, Deede Gammon† and Michael Mackert‡

*College of Communication Arts and Sciences, Michigan State University, East Lansing, USA; †Norwegian Centre for Telemedicine, Tromsø, Norway; ‡Department of Advertising, University of Texas, Austin, USA

Summary

We conducted a systematic review of 15 relevant databases for articles about telemedicine. After eliminating articles that did not meet the inclusion criteria, 1615 remained for analysis. Three raters coded the articles to assess various theoretical and methodological variables. Only 5% ($n = 85$) of the telemedicine articles made mention of any theory or paradigmatic approach. Studies commonly reported the objectives (96%) but rarely stated a research question or hypothesis (11%). Randomized selection of the subjects was reported in 11% of patient studies and 4% of studies where providers were the subject. There was a wide range in the number of subjects employed, although the majority of studies were based on sample sizes of less than 100. Only 26% of the studies reported a time frame. Until the telemedicine field adheres to agreed standards of reporting methodological details it will be difficult to draw firm conclusions from review studies.

Introduction

A number of reviews have concluded that there is no irrefutable evidence about the positive effect of telemedicine on clinical outcomes. For example, Roine *et al*. concluded that there was *potential* evidence of effectiveness only in teleradiology, telepsychiatry, transmission of echocardiographic images and consultations between primary and secondary health providers.[1] Another systematic review that assessed more than 1300 papers making claims about telemedicine outcomes found only 46 publications in which clinical outcomes were actually studied.[2] Currell *et al*. performed a literature review to analyse the suitability of telemedicine as an alternative to face-to-face care.[3] They concluded that patient care using telecommunication is feasible, although the studies provided inconclusive results about clinical benefits and outcomes. A more recent report prepared for the US Agency for Healthcare Research and Quality[4] assessed telemedicine services that substitute for face-to-face services. The authors concluded that there were significant gaps in the evidence base between where telemedicine is used and where its use is supported by high-quality evidence. Reviews on cost outcomes have reached similar conclusions. In 2002, a review paper concluded that there was no good evidence that telemedicine is or is not a cost-effective means of delivering health care.[5] The only review articles with more positive conclusions are those that simply seek to describe the types of telemedicine activity being reported in peer-reviewed journals. For example, a recent review of articles published in the two journals specialising in telemedicine that are indexed in MEDLINE concluded that telemedicine studies cover a wide range, both geographically and in terms of clinical disciplines.[6]

Previous reviews have found that the methodology of telemedicine studies was often poor.[7,8] To assess the methodology, we addressed four research questions:

(1) What is the incidence of theory testing in telemedicine literature? What theories are actually tested?

Accepted 25 January 2007
Correspondence: Professor Pamela Whitten, 409 Communication Arts Bldg, College of Communication Arts and Sciences, Michigan State University, East Lansing MI 48824, USA
(*Fax: +1 517 355 1292; Email: pwhitten@msu.edu*)

(2) What percentage of the studies clearly report specific study variables, such as:
   - study objective
   - hypothesis or research questions
   - whether the study was based on actual intervention or secondary data
   - number and type of subjects
   - randomized versus non-randomized subjects
   - units of analysis
   - location where intervention actually occurred
   - time frame (start and end dates) for data collection?

(3) What data collection strategies and analyses were employed?

(4) Were the results presented in qualitative and/or quantitative fashion?

Methods

We first developed a search strategy. After a preliminary search, we then developed a coding scheme. The initial coding scheme was then refined, based on its application to 50 randomly selected articles.

Fifteen scientific databases in a broad range of disciplines were searched (see Table 1). The list of keywords employed in the search is shown in Table 2. For each list of keywords in Table 2, we employed one term from the 'or' segment of the list, then inserted 'and' followed by a term from the second segment in the list. So, for example, for the terms under item 1 in Table 2, we began with the search 'telemedicine and evaluation', 'telemedicine and framework', 'telemedicine and methodology', etc. Next, we employed 'telehealth and evaluation', 'telehealth and framework', etc. This strategy was employed for the keywords in all three lists presented in Table 2. In addition, we hand-searched the two specialist telemedicine journals. This search strategy yielded more than 10,000 articles.

The articles were then assessed against the inclusion criteria (see Table 3). After eliminating articles that did not meet the inclusion criteria, 1615 remained for analysis.

Coding scheme and data analysis

Three raters coded the articles identified. A subset of 50 sample articles was randomly selected to assess

Table 1 Databases searched

ACM Digital Library
Social Services Abstract (Cambridge)
Cochrane Library
ComAbstracts
Cumulative Index of Nursing and Allied Health Literature (CINAHL)
EconLit
ERIC
International Bibliography of the Social Sciences
PsycInfo
Pubmed
Science Citation Index
Science Direct
Social Sciences Citation Index
Sociological Abstracts (Cambridge)
Telemedicine Information Exchange (TIE)

Table 2 Keywords employed in the search strategy

1. Telemedicine or telehealth or ehealth or medical informatics or health telematics or ICT for health AND evaluation or framework or methodology or assessment or model or research or data collection
2. Telemedicine or telehealth or ehealth or medical informatics or health telematics or ICT for health AND review or systematic or evaluation or literature review or overview or meta analysis
3. Telemedicine or telehealth or ehealth or medical informatics or health telematics or ICT for health AND cost or outcomes or utilization or clinical or sociological or community or psychological or communication or interaction or policy or medicine or therapy or nursing or science or history or social work or nutrition or presence or international or business or economic models/economics

Table 3 Inclusion and exclusion criteria for articles

Inclusion criteria	Exclusion criteria
Published in 1990 and later	Published as non-English articles
Included data of some kind (primary or secondary)	Published before 1990
Included some form of a methods section	Published as an editorial
Employed human subjects in some form (individual, group or organizational level)	Published only as an abstract
Included some form of data transmission	Published as newsletters, letters to editor, commentaries, interviews
	Published as a programme/project descriptions with no data
	Included within journals published only as conference proceedings and/or conference abstracts
	Published as a poster presentation
	Existed as multiple publications from the same authors of the same study with no variation in results (only used first published piece)
	Published as books, book chapters or reports
	Published studies on distance education only
	Published studies that employed pure telephone intervention

Table 4 Coding questions and guide

1. Whether or not a theory is mentioned in the article? (yes/no)
2. What is the specific theory or theoretical paradigm or model that guides the study if there is any? (This study must apply or is based on this specific theory).
3. Is the study aim of the paper clearly stated or described (yes/no)?
4. Is a Research Question or hypothesis clearly and/or specifically stated in the paper (yes/no)? For the answer 'yes', the Research Question or hypothesis must be explicitly stated.
5. Is the time frame (clear start and end date) of the study clearly stated (yes/no)? Simply answer yes or no to whether a clear start and end date are provided.
6. Is the study based on an actual intervention (yes/no)? Intervention means that information is exchanged via telemedicine technology.
7. What is the major goal that this paper addresses? (Single choice)
   1. Feasibility/Pilot (simply tests the technical feasibility or efficacy of the application of a telemedicine technology)
   2. Impacts/Effects (tests health outcomes, cost outcomes, perceptions such as satisfaction, organizational changes, or issues)
   3. Pre-intervention studies/based on hypothetical data.
   4. Other (e.g. goal stated is to develop methodology)
   (If a study does feasibility PLUS some type of perception or outcome, classify it as no 2).
8. What is the location (country) that this study was conducted in? If more than one country, select 'multiple' and manually put their names. If no location is explicitly stated, select 'not stated'. Do not attempt to infer from location of authors.
9. What is the length of the actual project in the article? Use exact date/year terms from the article. Use their language.
10. Type of subject
    - Select 'patient' when data are directly from or about patients (e.g. patients are interviewed, digitized film from patients are analysed)
    - Select 'provider' when data are collected or analysed about providers of health services (e.g. doctors, nurses, therapists).
    - Select 'other' when data are directly from or about subjects other than patients or providers. These could include organizations or family members.
11. Is the total number of subjects clearly stated in the article? (yes/no) Do not add or infer if the total number of subjects is not clearly stated.
12. List the number of subjects that participated or from whom data were collected.
13. If units of analysis differ from the number of subjects, provide the total number of units employed for analysis.
14. Were the subjects randomized or non-randomized?

15 A. What is the instrument for data collection in this article? (choose all that apply). Note: the instrument for data collection should be based upon what data are being used for analysis.
Interview – Verbal data collection that includes open and/or close ended questions. These can be *personal interview* or *telephone interview*.
Survey (questionnaire) – Questionnaires are usually paper-and-pencil instruments that the respondent completes, including *mail survey*, *group administered questionnaire* (A sample of respondents is brought together and asked to respond to a structured sequence of questions), and *household drop-off survey* (a researcher goes to the respondent's home or business and hands the respondent the instrument). These can include open and/or close ended items.
Observation – Includes *participant observation and direct observation*. Participant observation requires that the researcher become a participant in the culture or context being observed. Direct observation suggests a more detached perspective. A direct observer doesn't typically try to become a participant in the context. However, the direct observer does strive to be as unobtrusive as possible so as not to bias the observations.
Psychometrics/Physiological Measures – Data addressing psychological or physiological status of patients. These can include mental health status, stress, blood pressure, weight or blood glucose.
Image/Transmission – Using technical devices or electronic equipment to transmit digital images, audio or video files to a distant site.
Archival – Performing secondary analysis of existing data. Archival data can be obtained from academic archives, government archives, private/public organizations and consulting firms, private foundations and other independent researchers. When data are specifically gathered and stored in a record for the purpose of the specific study in the paper, it would not be considered as archival data. If the researchers gather data or information that was pre-existing (e.g. email correspondence, baseline physiological or psychometric data, hospital record, emergency room visits), this would constitute archival data collection.
Miscellaneous – Any data collection instrument that does not clearly fit into any of the other categories.

15 B. Is the number of subjects clearly stated in each method of data collection? Put '1' if the number of subjects is clearly stated. Put '2' if the number of subjects is not clearly stated.

16. Were methods used to analyse data explicitly stated? (yes/no)
17. Were data presented in the Results section quantitative (numerical representation), qualitative (description) or both?

inter-rater reliability. The three raters each coded articles in the sample. Cohen's kappa was greater than 0.80 for each pair of raters on the key variables.

The 1615 articles were then divided between the three raters. The coding scheme directions are shown in Table 4. The coding data were stored in a database. On completion, the data were analysed using a standard statistical package (SPSS).

Results

Only 5% ($n = 85$) of the telemedicine articles made mention of any theory or paradigmatic approach. Of the 5% that mentioned a theory, most did not formally test the theory, but rather made mention of it as the basis for the paper. A total of 68 theories were mentioned in the 85 articles. Some aspect of diffusion theory accounted for 11% of the theories mentioned in the 85 articles. Otherwise, no particular theory or paradigm stood out as being commonly included or tested.

Most papers (96%) reported the overall study aim. The majority sought to address some form of outcome (80%), with 13% seeking to report on technical feasibility and 7% addressing hypothetical or other issues, such as simulations. However, few papers (11%) provided explicit hypotheses or research questions. The majority of papers (89%) provided data and results

from a targeted telemedicine intervention, either from actual interventions or secondary data.

In 26% of the articles, authors did not provide a clear statement of the number of subjects. For the 74% that did report the numbers of subjects, the range was 1–84,000 with a median of 57. Sixty-five percent of the studies that employed patients as subjects (i.e. not health-care staff or administrators) had a sample size of 100 or less. In 40 studies there was a single patient subject (Figure 1).

Other papers (13%) employed providers as the unit of analysis. However, 48% of these articles did not provide a clear statement of the number of providers. For those that did state the number of providers, the median was 29 and the range was 1–13,489. Eighty-six percent of the studies employing providers as subjects had a sample size of 100 or less. Only three studies of providers were conducted with a single provider as the subject (Figure 2).

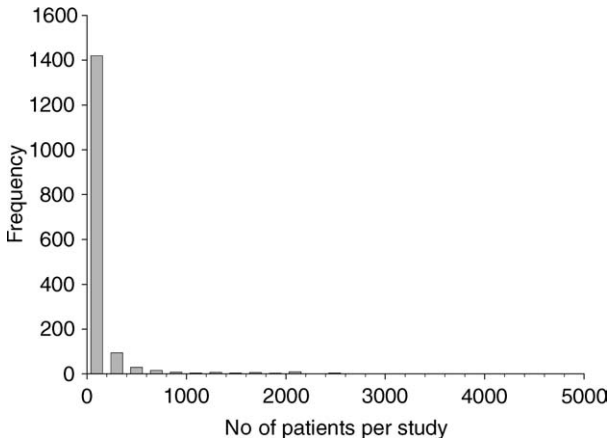

Figure 1 Number of patients per study (ten extreme values not shown)

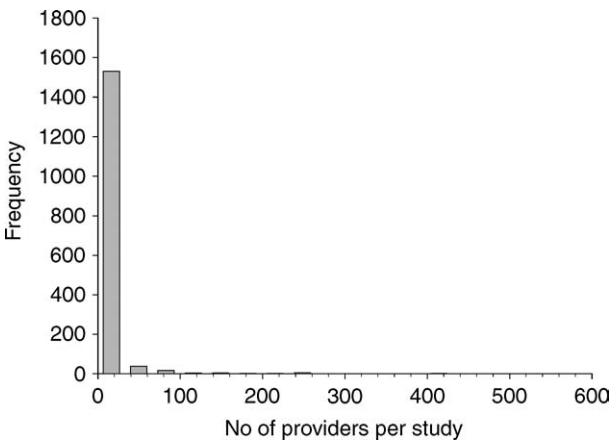

Figure 2 Number of providers per study (ten extreme values not shown)

Randomized selection of the subjects was reported in 11% of patient studies and 4% of provider studies. However, the majority of papers did not mention their subject selection strategies. In addition, the time frame in which a study was conducted was poorly described. Only 26% of the articles clearly reported the start and end points for a study.

There was significant variation in the locations in which the telemedicine studies were performed for those studies which actually reported location (83%). Of these papers, 34% were carried out in the US, 24% in Europe, 6% in Oceania and 6% in Asia (see Table 5).

Data collection, analysis and reporting strategies were assessed for three types of subjects, namely patients, providers and others (Table 6). The most popular data collection strategies for patients were image transfer (35%) and physiological and psychometric measures (33%), while studies looking at providers tended to rely on surveys (60%) and interviews (30%).

Multiple data collection methods within a study were employed in 31% of patient studies, in 18% of provider studies and in 12% of others. Almost 11% of the articles studied more than one subject type, with 3% looking at patients and providers, 6% studying patients and others, 1% examining providers and others and 2% collecting data for all three types. There were no significant trends in how results were reported (quantitatively or qualitatively or both). The interview data were primarily qualitative or both (84% for

Table 5 Locations of studies

North America	***623***
United States	538
Canada	*79*
Europe	***379***
United Kingdom	133
Norway	31
Italy	28
Finland	*27*
Germany	*26*
Netherlands	22
Sweden	19
Greece	15
France	14
Spain	14
Oceania	***99***
Australia	90
New Zealand	*9*
Asia	***96***
Japan	36
Hong Kong	13
Israel	10
Taiwan	*9*
India	*7*
Multiple countries	***110***

Table 6 Subject type, data collection, analysis and reporting

	Subject	Number of studies (%)	Method of data analysis stated (%)	Results quantitative (%)	Results qualitative (%)	Results both (%)
Interview	Patients	80 (7)	61	16	47	37
	Providers	106 (30)	51	2	68	30
	Other	32 (15)	74	3	57	40
Survey	Patients	316 (28)	78	58	6	36
	Providers	217 (60)	68	47	9	44
	Other	57 (27)	84	53	11	37
Observation	Patients	80 (7)	77	31	31	39
	Providers	36 (10)	61	8	58	33
	Other	11 (5)	91	18	64	18
Psychometrics/ Physiological measures	Patients	375 (33)	79	76	4	20
	Providers	6 (2)	50	50	0	50
	Other	5 (2)	80	60	20	20
Image/ Transmission	Patients	402 (35)	67	65	4	31
	Providers	12 (3)	42	75	0	25
	Other	16 (8)	50	63	13	25
Archival	Patients	129 (11)	71	58	4	38
	Providers	23 (6)	38	9	44	48
	Other	19 (9)	65	42	16	42
Miscellaneous	Patients	172 (15)	54	67	5	28
	Providers	33 (9)	44	44	9	47
	Other	99 (47)	29	23	15	62

Table 7 Correlations between the primary variables

	Theory	Study aim	Research question
Theory			
Study aim	0.037		
Research question	0.252*	0.067*	
Actual intervention	−0.069*	0.021	0.023

*$P<0.01$

patients; 98% for providers; 97% for other) and data such as psychometric/physiological measures (76% for patients; 50% for provider; 60% for other) or image/transmission (65% for patients; 75% for providers; 63% for other) were generally reported in quantitative fashion.

There were few significant relationships in regard to the reporting of results (see Table 7). There was a positive correlation between use of theory and presence of formal research questions ($r=0.25$, $P<0.01$). There were significant correlations between research questions and study aim ($r=0.067$, $P<0.01$) and between the inclusion of theory and study based on an actual intervention ($r=-0.069$, $P<0.01$).

Discussion

The present study evaluated the methodology of articles in the field of telemedicine and telehealth. The study proved to be difficult because it was not easy to ascertain the requisite information from the published papers. However this is a common finding in reviews in other fields,[9–11] i.e. the problem is not restricted to telemedicine. Coders often had to hunt through the articles to obtain important methodological details, sometimes finding them in sections of the paper other than the methods section. In many cases it was necessary for the coders to calculate values such as the number of subjects or start and end dates. In other cases the study location was not always reported explicitly, and could not be inferred from the authors' location.

The implications of the missing methodological details should not be underestimated. Different technologies demonstrate great variation over time so it is significant that so few articles report start and end points. For example, the main technologies employed in home telehealth have changed over the past decade from synchronous video to data monitoring. It is therefore crucial to place studies of home telehealth within a clear context of time, as well as location. The lack of explicitly stated research questions precludes readers from determining if the study design was appropriate. More than one-quarter of studies employing patients as subjects and almost 50% of studies employing providers as subjects did not clearly state the number of subjects.

The dearth of theory inclusion or testing is not necessarily an indication of inappropriate study design. For example, not all clinical studies require randomization. However, it remains difficult to evaluate the appropriateness of the design when the

majority of papers simply do not mention their subject selection strategies.

Until the telemedicine field adheres to agreed standards of reporting methodological details it will be difficult to draw firm conclusions from review studies. Lack of methodological detail limits our ability to understand and explain telemedicine, because it violates one of Kuhn's basic tenets: the existence of intertwined theoretical beliefs and methodological strategies make it possible for a field to evaluate its own body of research.[12] In fairness, similar analyses for the methodologies of other health interventions might yield the same results. Nonetheless, our study raises the question whether the reporting of telemedicine methodology is of sufficient quality.

Acknowledgements: We thank Heidi Jacobsen (NST) for developing the database for the coding, Aud Obstfelder (NST) who participated in the initial part of the project, Vinita Argawal (Purdue University) for assistance with coding, and other colleagues at NST and MSU who offered advice.

References

1 Roine R, Ohinmaa A, Hailey D. Assessing telemedicine: a systematic review of the literature. *CMAJ* 2001;**165**:765–71

2 Hailey D, Roine R, Ohinmaa A. Systematic review of evidence for the benefits of telemedicine. *J Telemed Telecare* 2002;**8** (Suppl. 1): 1–30

3 Currell R, Urquhart C, Wainwright P, Lewis R. Telemedicine versus face to face patient care: effects on professional practice and health outcomes. *Cochrane Database Syst Rev* 2000; (2):CD002098

4 Hersh WR, Hickam DH, Severance SM, Dana TL, Krages KP, Helfand M. *Telemedicine for the Medicare Population: Update*. Evidence Report/ Technology Assessment No. 131; AHRQ Publication: No. 06–E007. Rockville, MD: Agency for Healthcare Research and Quality, 2006

5 Whitten P, Mair F, Haycox A, May CR, Williams TL, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. *BMJ* 2002;**324**:1434–7

6 Demiris G, Tao D. An analysis of the specialized literature in the field of telemedicine. *J Telemed Telecare* 2005;**11**:316–19

7 Whitten P, Kingsley C, Grigsby J. Results of a meta-analysis of cost-benefit research: is this a question worth asking? *J Telemed Telecare* 2000;**6** (Suppl. 1):4–6

8 Jennett PA, Affleck Hall L, Hailey D, *et al*. The socio-economic impact of telehealth: a systematic review. *J Telemed and Telecare* 2003;**9**:311–20

9 Fernandez-de-las-Penas C, Alonso-Blanco C, San-Roman J, Miangolarra-Page JC. Methodological quality of randomized controlled trials of spinal manipulation and mobilization in tension-type headache, migraine, and cervicogenic headache. *J Orthop Sports Phys Ther* 2006;**36**:160–9

10 Deeks JJ, Dinnes J, D'Amico R, *et al*. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**:iii–xx, 1–173

11 C Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. *Health Technol Assess* 2000;**4**:1–55

12 Kuhn T. *The Structure of Scientific Revolution*. Chicago, IL: University of Chicago Press, 1962