Patient Preference and Adherence

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Instruments to assess patient satisfaction after teleconsultation and triage: a systematic review

This article was published in the following Dove Press journal:
Patient Preference and Adherence
24 June 2014
Number of times this article has been viewed

Martina Allemann Iseli[1]
Regina Kunz[2]
Eva Blozik[1,3]

[1]Swiss Center for Telemedicine Medgate, Basel, Switzerland; [2]Academy of Swiss Insurance Medicine, University of Basel, Basel, Switzerland; [3]Department of Primary Medical Care, University Medical Center Hamburg–Eppendorf, Martinistrasse, Hamburg, Germany

Background: Patient satisfaction is crucial for the acceptance, use, and adherence to recommendations from teleconsultations regarding health care requests and triage services.

Objectives: Our objectives are to systematically review the literature for multidimensional instruments that measure patient satisfaction after teleconsultation and triage and to compare these for content, reliability, validity, and factor analysis.

Methods: We searched Medline, the Cumulative Index to Nursing and Allied Health Literature, and PsycINFO for literature on these instruments. Two reviewers independently screened all obtained references for eligibility and extracted data from the eligible articles. The results were presented using summary tables.

Results: We included 31 publications, describing 16 instruments in our review. The reporting on test development and psychometric characteristics was incomplete. The development process, described by ten of 16 instruments, included a review of the literature (n=7), patient or stakeholder interviews (n=5), and expert consultations (n=3). Four instruments evaluated factor structure, reliability, and validity; two of those four demonstrated low levels of reliability for some of their subscales.

Conclusion: A majority of instruments on patient satisfaction after teleconsultation showed methodological limitations and lack rigorous evaluation. Users should carefully reflect on the content of the questionnaires and their relevance to the application. Future research should apply more rigorously established scientific standards for instrument development and psychometric evaluation.

Keywords: teleconsultation, teletriage, triage, consultation, general practitioner, patient satisfaction, psychometric, evaluation, out-of-hours

Introduction

In recent years, telephone consultation and triage have gained popularity as a means for health care delivery.[1,2] Teleconsultations and triage refer to "the process where calls from people with a health care problem are received, assessed, and managed by giving advice or via a referral to a more appropriate service."[3] The main motive for introducing such services was to help callers to self-manage their health problems and to reduce unnecessary demands on other health care services. Teleconsultation and triage are frequently used in the context of out-of-hours primary care services.[4] They result in the counseling of patients about the appropriate level of care (general practitioner, specialized physician, other health care providers, [such as therapists], or hospital care), the appropriate time-to-treat (ranging from emergency care to seeking an appointment within a few weeks), or the potential for self-care. Several randomized

Correspondence: Eva Blozik
Department of Primary Medical Care, University Medical Center Hamburg–Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany
Tel +49 40 741 052 400
Fax +49 40 741 053 681
Email e.blozik@uke.de

controlled trials showed that teletriage is safe and effective,[5–7] and a systematic review suggested that at least one-half of the calls can be handled by telephone advice alone.[8]

The patients' opinions on the quality of such services are crucial for their acceptance, use, and adherence to the recommendations resulting from the teleconsultation.[9,10] Instruments to measure patient satisfaction have been developed for a broad range of settings. However, these instruments cannot easily be transferred into the teleconsultation setting, which systematically differs in two respects: 1) decisions in teleconsultation and triage rely heavily on medical history-taking as the main – and sometimes only – diagnostic tool, so excellent communication and history-taking skills are crucial in this setting; 2) teleconsultation and triage services generally relate to the appearance of new health problems and less frequently address long-term management for which patients usually attend face-to-face care.[1]

Patient satisfaction is a multidimensional construct.[11,12] Global indices (single-item instruments) have been shown to be unreliable for the measurement of patient satisfaction in health care and to disguise the fact that judgments on different aspects of care may vary.[10,13] Instruments assessing patient satisfaction after teleconsultation and triage need to cover the perceived quality of the communication skills, of the telephone advice (eg, helpfulness and feasibility of the recommendation), and of the organizational issues of the service, such as access or waiting time.[10] In a previous review, methodological issues related to the measurement of patient satisfaction with health care were systematically collected.[10] Several problems were addressed, such as how different ways of conducting surveys affect response rates and consumers' evaluations. However, the review did not include detailed information on patient satisfaction questionnaires, nor did it give specific recommendations related to questionnaire use. A more recent systematic review in 2006 on patient satisfaction with primary care out-of-hours services presented four questionnaires, all with important limitations in their development and evaluation process.[4]

However, out-of-hours care is only a small part of teleconsultation and triage services. Furthermore, none of the previous reviews explicitly followed up on research that modified and reevaluated existing instruments. Therefore, the aim of our study was to systematically review the scientific literature for multidimensional instruments that measure patient satisfaction after teleconsultation and triage for a health problem and to compare their development process, content, and psychometric properties.

Methods

Literature search

We searched Medline, the Cumulative Index to Nursing and Allied Health Literature, and PsycINFO (query date of January 31, 2013) for relevant literature. The search terms were related to "patient satisfaction", "questionnaire", and "triage" (Table S1). We reviewed the reference lists of all publications included in the final review for relevant articles. Furthermore, we searched the Internet for additional material, in particular for follow-up research, the refinement of the included instruments via authors' names, and the names of the instruments.

Study selection and data collection process

The pool of potentially relevant articles identified via databases, reference lists, and Internet searches was evaluated in detail regarding whether or not the articles were original research articles, whether or not they described instruments for assessing patient satisfaction after an encounter between a health professional and a patient or his proxy over the phone, and whether or not they were intended for self-administered or interviewer-administered use (Table 1).[14] As we were interested in multidimensional instruments, we excluded global indices (single-item measures). We included telephone and video consultations, as well as out-of-hours services that performed triage by phone. Out-of-hours services were defined as any request for medical care on public holidays, Sundays, and at a defined time on weekdays and Saturdays (for example, weekdays from 7 pm to 7 am and Saturdays from 1 pm onward). We included studies that reported the development of the instrument (called "development studies") and studies that applied the instrument for outcome assessment (called "outcome studies"). We did not apply any language restriction. Two reviewers (MAI, EB) independently screened the references for eligibility, extracted the data, and allocated the instrument items to the predefined domains. Discrepancies were solved by consensus.

Data extraction and analysis

We extracted the following information:

1. Descriptive information: author; year of publication; country of origin; setting; staff providing the service; type of administration of the questionnaire; participants; and timing of administering the instrument after the encounter (Table 1).
2. Instrument content: number of items per domain; number of domains covered per study; total number of items; mean items per domain; number of studies that covered a certain domain with at least one question (Table 2).

Table 1 Characteristics of the identified publications

Author	Year	Country	Setting	Consultation done by	Type of administration	Participants (receiving teleconsultation)	Time of administration after encounter
Campbell et al[18,b]	2007	UK	Out-of-hours service	Health professional	Mail	570[a]	14 days
Campbell et al[38,c]	2009	UK	Out-of-hours service	Health professional	Mail	570	14 days
Kelly et al[39,c]	2010	UK	Out-of-hours service, 3 accident and emergency units, teleconsultation service	Physician or nurse	Mail	353	14–28 days
Dehours et al[30,b]	2012	France	Teleconsultation service for French ships worldwide	Physician	Email	165	4–16 months
Dixon and Williams[19,b]	1988	UK	Air Call Medical Services, out-of-hours service	Physician	Mail	3,296	3 working days
Dixon and Sthal[22,b]	2009	USA	Family doctor teleconsultations	Physician	Written	152	Immediately after consultation
Garratt et al[31,b]	2010	Norway	Out-of-hours service	Unclear	Mail	225	NR
Hicks et al[23,b]	2003	USA	Dermatology teleconsultation service	Physician	Written	483	After each visit
López et al[45,c]	2011	Colombia	Teleconsultation at rural first-level outpatient institution	Physician	Telephone	121	NR
Keatinge and Rawlings[29,b]	2005	Australia	Out-of-hours service	Nurse	Telephone	101	NR
McKinley et al[20,b]	1997	UK	Out-of-hours service	Physician	Interview or written	163	1–5 days
Salisbury[32,c]	1997	UK	Out-of-hours service	Physician	Mail	686	7 days
Shipman et al[33,c]	2000	UK	Out-of-hours service	Physician	Mail	513	7 days
McKinley and Roberts[34,c]	2001	UK	Out-of-hours-service	Physician	Written	1,402	5 days
McKinley et al[35,c]	2002	UK	Out-of-hours service	Physician	Mail	2,000	NR
Glynn et al[36,c]	2004	Ireland	Out-of-hours service	Physician	Mail	134	7 days
Thompson et al[37,c]	2004	Northern Ireland	Out-of-hours service	Physician	Mail	1,075	5 days
McKinstry et al[6,b]	2002	UK	Management of same-day appointments	Physician	Mail	186	1 day
Mekhjian et al[24,b]	1999	USA	Prison	Physician	Written	221	Immediately after consultation
Moll van Charante et al[27,b]	2006	Netherlands	Out-of-hours service	Nurse	Mail	2,352	2 days
Giesen et al[40,c]	2007	Netherlands	Out-of-hours service	Nurse	Mail	2,583	2 days
Smits et al[41,c]	2012	Netherlands	Out-of-hours service	Nurse	Mail	1,565	2–4 days
Moscato et al[25,b]	2003	USA	Teleconsultation service	Nurse	Mail	87	NR
Beaulieu and Humphreys[42,c]	2008	USA	Teleconsultation service	Nurse	Telephone	20	3–5 days
Reinhardt[43,c]	2010	USA	Teleconsultation service	Nurse	Mail	865	NR
Rahmqvist et al[17,b]	2009	Sweden	Teleconsultation service	Nurse	Mail	660	14–21 days
Rahmqvist et al[44,c]	2011	Sweden	Teleconsultation service	Nurse	Mail	273	14–21 days
Salisbury et al[21,b]	2005	UK	Out-of-hours service	Physician or nurse	Mail	342	7 days
McKinstry et al[46,c]	2010	UK	Family doctor teleconsultations	Physician	Mail	46	NR
Ström et al[26,b]	2011	Sweden	Teleconsultation service	Nurse	Mail	517	14–21 days
Van Uden et al[28,b]	2005	Netherlands	Out-of-hours service	Physician's assistant	Mail	908	21 days

Notes: [a]Total number of respondents, including recipients of face-to-face services; [b]development study; [c]outcome study. The indented names in the "Author" column are referred to in the above-named reference.
Abbreviation: NR, not reported.

Table 2 Instrument content (related to teleconsultation)

Author	Year	Access to service	Attitude of health professional	Attitude of patient	Communication	Individual information[a]
Campbell et al[18]	2007	1	3	1	5	15
Dehours et al[30]	2012	0	0	1	3	5
Dixon and Williams[19]	1988	0	0	0	1	0
Dixon and Sthal[22]	2009	0	1	0	2	0
Garratt et al[31]	2010	0	4	0	1	0
Hicks et al[23]	2003	1	1	2	0	0
Keatinge and Rawlings[29]	2005	0	0	4	1	3
McKinley et al[20]	1997	2	1	1	6	0
McKinstry et al[6]	2002	0	0	1	3	0
Mekhjian et al[24]	1999	0	0	5	4	0
Moll van Charante et al[27]	2006	1	2	0	5	0
Moscato et al[25,b]	2003	0	1	3	2	2
Rahmqvist et al[17]	2009	0	0	0	3	0
Salisbury et al[21]	2005	1	1	0	0	0
Ström et al[26]	2011	2	2	1	3	0
Van Uden et al[28]	2005	2	1	2	2	0
# of studies that covered a certain domain with at least one question		7	10	10	14	4

Notes: [a]Sociodemographics; result of teleconsultation; [b]revised version as published by Beaulieu and Humphreys.[42]

3. Details of the development process: such as literature review, consultation with experts, consensus, focus group meetings, or individual interviews; piloting; and rating scale (Table 3).
4. Recruitment strategy and handling of nonresponders: inclusion and exclusion criteria; consecutive recruitment of patients; response rate; and nonresponse analysis (Table 4).
5. Psychometric properties: item nonresponse; factor structure; reliability (ie, interrater, test/retest, intermethod, and internal consistency reliability); and validity (ie, construct, content, criterion validity) (Table 5).

Table 3 Descriptive information of the instruments

Author	Year	Development process	Piloting	Rating mode
Campbell et al[18]	2007	Literature review, consultation with experts (no further specification)	Yes	5-point Likert scale
Dehours et al[30]	2012	Consensus of the working group	Yes	Yes/no, categorical, open-ended
Dixon and Williams[19]	1988	NR	Yes	Yes/no
Dixon and Sthal[22]	2009	NR	Yes	Numerical rating scale 1–5
Garratt et al[31]	2010	Literature review, consultation with experts, interview with patients	Yes	Unclear
Hicks et al[23]	2003	NR	NR	7-point Likert scale
Keatinge and Rawlings[29]	2005	NR	Yes	Categorical
McKinley et al[20]	1997	Literature review, focus group meetings with patients recruited from general practice registers and community groups led by a nonclinician	Yes	5-point Likert scale
McKinstry et al[6]	2002	NR	NR	Numerical rating scale 0–3
Mekhjian et al[24]	1999	Literature review	NR	5-point Likert scale
Moll van Charante et al[27]	2006	Literature review, interview of stakeholders	Yes	Numerical rating scale 1–10
Moscato et al[25]	2003	Qualitative interviews with adults who had received phone advice	Yes	5-point Likert scale and check-off options
Rahmqvist et al[17]	2009	NR	NR	7-point Likert scale
Salisbury et al[21]	2005	Literature review, use of McKinley questionnaire as a basis, development of draft short version	Yes	5-point smiley faces (very dissatisfied to very satisfied)
Ström et al[26]	2011	Multidisciplinary expert group decision, interview with patients	Yes	Visual analog scale 0–10
Van Uden et al[28]	2005	Literature review, interview of general practitioner's managers	NR	5-point Likert scale

Abbreviation: NR, not reported.

Management	Overall satisfaction	Professional skills	Telephone advice	Other	Number of domains covered per study	Total number of items
4	3	2	3		9	37
2	1	0	1	Diagnostics (8), training of staff (3)	6	24
0	0	1	1		3	3
0	1	1	0		4	5
4	0	1	0		4	10
0	3	0	0	Technical aspects (1)	4	8
1	1	0	2	Alternative to teleconsultation (1)	6	13
0	3	0	7		6	20
0	0	0	1		3	5
0	1	1	0	Technical aspects (3)	4	14
1	0	1	3	Access to pharmacy (1)	6	14
0	3	2	1	Alternative to teleconsultation (1)	7	15
1	1	1	1		5	7
2	1	0	2		5	7
1	1	2	2		8	14
5	4	0	6		7	22
9	12	9	12		5.4	13.6 mean

The data was tabulated and summarized in a descriptive way.

First, we listed all primary studies and extracted basic information. Outcome studies – that evaluated the same instrument in various settings and populations – were grouped under the corresponding development study. When several studies referred to the same instrument, we used the development study to extract data for the following steps.

Second, we analyzed the content domains of the instruments. Based on a systematic review, published by Garratt et al, we created a list of nine domains (access to the service, attitude of health professional, attitude of patient, perceived quality of the communication, individual information [such as sociodemographic or clinical patient data], management [such as waiting time], overall satisfaction, perceived quality of professional skills of the staff, perceived quality of the telephone advice [such as helpfulness and feasibility of the recommendation]), and other.[4] Two reviewers independently attributed each item of the instruments to one domain. The aim of this procedure was to describe, to characterize, and to compare the content of patient satisfaction instruments for which no factor-analysis results were reported. We did not use these dimensions as a prerequisite for instruments to be included in our review.

Third, we explored the development process of the instrument, the scoring scheme of the instrument, and the performance of a piloting. When we identified only one study to an instrument, we extracted the data from this publication, regardless of whether it was a development study or an outcome study.

Fourth, we assessed the recruitment strategy and handling of nonresponders in those publications that reported statistical results for psychometric properties. This type of information is useful for interpretation of the statistical results so that – for those studies not reporting on factor structure, reliability, or validity – we did not detail recruitment strategy and handling of nonresponders.

Fifth, we tabulated any type of psychometric property that we identified in any type of publication. For the interpretation of Cronbach's alpha values, an estimate of the reliability of an instrument, we used the categories: excellent (>0.9); good (0.8–0.9); acceptable (0.7–0.8); questionable (0.6–0.7); poor (0.5–0.6); and unacceptable (<0.5).[15] An item-total correlation of <0.3 was considered poor, indicating that the corresponding item does not correlate well with the overall scale.[16]

Results

Our search identified 3,651 references. We screened 224 full-text publications for eligibility and, ultimately, included 31 studies – with a total of 17,797 patients – that reported on 16 different multidimensional instruments on patient satisfaction after teleconsultation and triage (Figure 1; Table 1). All but one article was published in the English language; this article was published in Swedish.[17]

Basic information

The instruments were developed in seven different countries: five instruments derived from the United Kingdom;[6,18–21]

Table 4 Recruitment strategy and handling of nonresponders

Author	Year	Inclusion patient characteristics	Consecutive patients	Exclusion patient characteristics	Response rate	Nonresponse analysis
Campbell et al[18]	2007	Out-of-hours contact	Yes	Age 12–16 years, residing in nursing home, caravan park, or hospital ward, recorded/expected death, end-stage terminal illness, serious/life threatening disease or distressing condition (eg, miscarriage), 'Special message' (eg, potentially violent patient, homeless)	46%	Respondents were older and more affluent than nonresponders, no difference in sex
Garratt et al[31]	2010	Out-of-hours contact	Unclear	NR	42%	NR
McKinley et al[20]	1997	Out-of-hours contact	Yes	NR	96%	NR
Mekhjian et al[24]	1999	Participation in teleconsultation	Unclear	NR	74%	NR
Moll van Charante et al[27]	2006	Out-of-hours contact	Yes	Death	52%	Higher response rate in men, age groups 5–14 years and 45–74 years, and privately insured. No differences in response were found for type of contact, trauma, reason for consultation and part of the day. The main reasons for nonresponse were "forgotten/not interested" (n=160; 34.6%) and "too ill" (n=83; 17.9%); 30 patients (6.5%) stated dissatisfaction as reason for nonresponse
Beaulieu and Humphreys[42] (reporting on revised instrument by Moscato et al[25])	2008	Out-of-hours contact, willlingness to be contacted, English speaking	Yes	NR	100%	NR
Salisbury et al[21]	2005	Patients who were transferred to a GP cooperative	Yes	Questionnaire would cause distress or had already been sent to same household, incomplete address	46%	NR
Ström et al[26]	2011	Phone call with nurse advice center	Yes	Seriously ill or person called on behalf of another person (except children)	61%	NR
Van Uden et al[28]	2005	Out-of-hours contact	Yes	Wrong address, death, previous participation	42%	55.6% were male; 44.5%, female; 40% had forgotten to return the questionnaire; 6.7% were not interested; 6.7% did not find it needful; 46.7% gave other reasons (no time, too difficult, lost questionnaire)

Abbreviations: NR, not reported; GP, general practitioner.

Table 5 Psychometric properties

Author	Year	Item nonresponse	Factor structure	Reliability	Validity
Campbell et al[18]	2007	Instrument with 45 items, 37 related to teleconsultation: • 0.5% of participants had >50% missing items • Confusion about the appropriate sections for answering (ie, related to point of care • Missing items for not management-specific items 1.2%–20.4% • 4 items (which health professional conducted the consultation; ethnic group; occupation; were you happy with [the final management of your call]) had missing values exceeding 10% of responses (maximum)	Exploratory principal components analysis: • 2 Factors accounting for 68% of variance • Factor 1: interaction (10 items, 55% of variance) • Factor 2: consultation satisfaction (9 items, 13% of variance) • Overlap of loadings for 4 items	Cronbach's alpha: • Consultation satisfaction, 0.96 • Entry access, 0.82 Inter item correlation: • Consultation satisfaction, 0.63–0.89 • Entry access, 0.45–0.86 Item total correlation: • Consultation satisfaction, 0.77–0.90 • Entry access, 0.56–0.73 Test/retest correlation: • Consultation satisfaction, 0.76 • Entry access, 0.60	Convergent validity: • Scores for consultation satisfaction and entry access subscale were moderately correlated (r=0.43) Construct validity: • Higher levels of global satisfaction, higher age, and longer duration of consultation were correlated with higher total and subscale scores
Garratt et al[31]	2010	Instrument with 24 items, 10 items related to teleconsultation: • Missing items, 3.2%–11.7%, mean, 7.2%	Principal component analysis: • 4 Factors accounting for 79% of variance • Factor 1: doctor services (7 items) • Factor 2: telephone contact and organization (10 items) • Factor 3: nursing services (6 items) • Factor 4: single item "unanswered questions" • Only factor 2 related to teleconsultation	Cronbach's alpha: • Telephone contact, 0.91 • Doctor contact, 0.90 • Nurse contact, 0.93 • Organization, 0.82 Item total correlation: • Telephone contact, 0.76–0.82 • Doctor contact, 0.73–0.83 • Nurse contact, 0.78–0.89 • Organization, 0.66–0.71	Construct validity: • Difficulties to contact by phone correlated to lower scores in telephone contact subscale • Short waiting time for teleconsultation and treatment, large amount of information, high levels of overall satisfaction, fulfillment of expectations and absence of poor treatment were correlated to higher scores of all subscales
McKinley et al[20]	1997	Instrument with 32 items, 20 related to teleconsultation: • Median completion rate, 96.5% (interquartile range, 95.7%–97.1%)	Principal component analysis: • 6 Factors • Factor 1: satisfaction with communication and management (7 items; 29% of Variance) • Doctor's attitude (5 items; 9% of variance) • Continuity of care (4 items; 8% of variance) • Delay until visit (3 items; 6% of variance) • Access to out-of-hours care (3 items; 5% of variance) • The initial contact person (2 items; 4% of variance) • Telephone advice (4 items; variance NR) • Overall satisfaction (4 items; variance NR)	Cronbach's alpha: • Communication and management, 0.88 • Doctor's attitude, 0.87 • Continuity of care, 0.69 • Delay until visit, 0.65 • Access to out-of-hours care, 0.61 • Initial contact person, 0.72 • Telephone advice, 0.79 • Overall satisfaction, 0.77 Test/retest correlation: • Communication and management, 0.86 • Doctor's attitude, 0.82 • Continuity of care, 0.72 • Delay until visit, 0.81 • Access to out-of-hours care, 0.76 • Initial contact person, 0.62 • Telephone advice, NR • Overall satisfaction, 0.82	Construct validity: • All subscales correlated to overall satisfaction

(*Continued*)

Table 5 (*Continued*)

Author	Year	Item nonresponse	Factor structure	Reliability	Validity
Mekhjian et al[24]	1999	Instrument with 14 items: • 26% of questionnaires were excluded due to nonresponse	Common factor analysis: • 2 Factors accounting for 47% of variance • Factor 1: information exchange (7 items; 40% of variance) • Factor 2: patient comfort (7 items; 47% of variance)	Cronbach's alpha: • Information exchange, 0.88 • Patient comfort, 0.81	NR
Moll Van Charante et al[27]	2006	Instrument with 66 items, 14 items related to teleconsultation: • Missing items 4.8%–16.5% (Mean 12.6%)	Principal component analysis: • 3 Factors • Factor 1 (77% of variance) • Factor 2 (72% of variance) • Factor 3 (83% of variance)	Cronbach's alpha: • Telephone nurse, 0.95 • Organization, 0.74 • Overall, 0.81 Item total correlation: • Telephone nurse, 0.84–0.92 • Organization, 0.53–0.59 Test retest correlation: • Telephone nurse, 0.85 • Organization, 0.92 • Overall, 0.79	NR
Beaulieu and Humphreys[42] (reporting on revised instrument by Moscato et al[25])	2008	Instrument with 19 items: • Missing items, NR	NR	Cronbach's alpha: • Total scale, 0.70	NR
Salisbury et al[21]	2005	Instrument with 7 items: • Missing items 54.1%–56.1% (Mean 55.7%)	NR	Cronbach's alpha: • Total scale, 0.94	Construct validity: • Each item correlated to overall satisfaction • Increasing age, female sex and transfer to face-to-face visit correlated to higher scores Concurrent validity: • Intraclass correlation coefficient between short scale and long scale 0.38–0.54
Ström et al[26]	2011	Instrument with 14 items: • Missing items, NR	Explorative factor analysis: • 3 Factors • Factor 1: interaction (8 items; 34% of variance) • Factor 2: service (3 items; 50% of variance) • Factor 3: product (3 items; 63% of variance)	Cronbach's alpha: • Total scale, 0.88 • Interaction, 0.90 • Service, 0.80 • Product, 0.45	NR

Van Uden et al[28]	2005	Instrument with 35 items, 22 items related to teleconsultation: • Missing items, NR	Principal component analysis: • 6 Factors • Accessibility by phone (3 items) • Doctor's assistant's attitude (5 items) • Questions asked by assistant (2 items) • Advice given by assistant (5 items) • Urgency of complaint (2 items) • Overall satisfaction (5 items)	Cronbach's alpha: • Accessibility by phone, 0.72 • Doctor's assistant's attitude, 0.91 • Questions asked by assistant, 0.64 • Advice given by assistant, 0.93 • Urgency of complaint, 0.86 • Overall satisfaction, 0.93	Construct validity: • Higher total scores correlated with overall satisfaction

Abbreviation: NR, not reported.

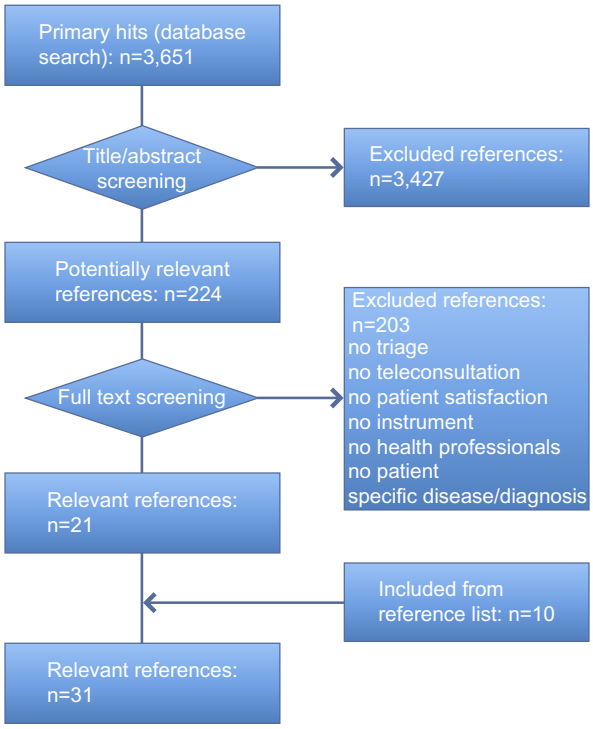

Figure 1 Flowchart.

four instruments from the United States;[22–25] two from Sweden;[17,26] two from the Netherlands;[27,28] and one instrument from each of Australia,[29] France,[30] and Norway.[31] Also, seven of the 16 instruments (44%) were used by subsequent studies.[17,18,20,21,23,25,27] The most frequently used instrument, the McKinley 1997 questionnaire,[20] was applied in six subsequent studies[32–37] and served as a basis for a shortened scale (Table 1).[21]

In most studies (14 of 16 instruments, 88%), the questionnaires were distributed per email or in a paper form for self-administration.[6,17–19,21–28,30,31] In three studies, both a self-administered and an interviewer-administered version were used.[20,23,25] The number of respondents per study varied between 20 and 3,294 persons. Also, 18 of the 31 publications (58%) applied instruments in the context of out-of-hours services, where centers triage patients from several general practices or a specific region.[18–21,27–29,31–41]

Eight publications described patient satisfaction after the consultation provided by the teleconsultation centers.[17,19,25,26,39,42–44] Other settings include: the management of same-day appointments;[6] the provision of teleconsultation services by physicians outside of specialized telemedicine institutions;[19,23,45,46] maritime telemedicine;[30] prison medicine;[24] and teledermatology services.[23] The timing of instrument administration varied considerably across

the studies. In addition, 16 publications (52%) reported the distribution of the questionnaires within 7 days of the consultation,[6,19–24,27,32–34,36,37,40–42] seven studies (23%) between 14–28 days,[17,18,26,28,38,39,44] and one publication (3%) reported a latency of 4–16 months.[30] Also, seven (23%) studies did not report on the timing of the instrument's administration (Table 1).[25,29,35,43,45,46]

Content of the instrument

We assessed the content of the instrument on nine prespecified domains. On average, an instrument covered five domains (range, three to nine) with 14 items per instrument (range, three to 37), and 2.3 items per domain (range, one to 15) (Table 2).

The most frequent domains, covered with at least one item, were the "perceived quality of the communication" (14 of 16 instruments, 88%),[6,17–20,22,24–31] followed by the "overall satisfaction" (12 of 16 instruments, 75%),[17,18,20–26,28–31] and the "perceived quality of the telephone advice" (12 of 16 instruments, 75%).[6,17–21,25–31] The following additional domains were covered by more than one-half of the instruments: the "attitude of the health professional;" the "attitude of the patient;" "management;" and "professional skills." This indicated a focus of interest across the different instrument development teams. Only one instrument covered all nine domains.[18]

The instruments varied widely in the number of items they included per domain. Seven instruments included mostly one or two items per domain;[6,17,19,21–23,26] whereas, the study on the top end included a mean of 4.1 items per domain.[18]

Development process

Only ten of the 16 instruments (63%) provided details about the development process, such as a review of the literature (n=7), interviews of patients or stakeholders (n=5), or consultations with experts (n=3).[18,20,21,24–28,30,31] Seven studies reported the use of more than one method.[18,20,21,26–28,31] Eleven of 16 studies (69%) performed a piloting of the instrument.[18–22,25–27,29–31] Likert scales were predominantly used for the scoring (seven of 16 instruments, 44%). Other rating modes included yes/no options (n=2), categorical answers (n=2), numerical rating scales (n=3), visual analog scale (n=1), or smiley faces (n=1). One instrument included open-ended questions (Table 3).[30]

Recruitment strategy and handling of nonresponders

Nine studies[18,20,21,24,26–28,31,42] gave information about their psychometric properties; therefore, their recruitment strategy and handling of nonresponders are further evaluated here. Inclusion criteria were comparable, as all studies addressed unselected patients who had received teleconsultation and triage services.

All but three publications explicitly reported the consecutive recruitment of patients.[24,25,31] The exclusion criteria (five of nine studies, 55%) were related to the feasibility of the study (for example, wrong address, serious illness of the patient).[18,21,26–28] The mean response rate was 60% and varied from 100%[42] to 38%.[46]

The nonresponse analysis in four of nine studies (44%) detected sociodemographic but no clinical differences between the studies' responders and nonresponders. However, these analyses were conflicting. One study reported respondents to be older and more affluent without any differences in sex.[18] In two studies, the response rates were lower in men invited to participate.[28,46] In a fourth study, women and young adults were less likely to participate.[27] Forgetfulness was identified as the most frequent reason for nonresponse (Table 4).[27,28]

Psychometric properties

For nine instruments, at least some information about the main psychometric properties was reported: item nonresponse; factor structure; reliability/internal consistency; and validity (Table 5).

1. Item nonresponse: six of the nine studies (67%) reported on nonresponses.[18,20,21,24,27,31] In some studies, item nonresponse was more problematic than in others. For example, one study reported complete data from only 43% of the respondents,[21] while nonresponse rates for individual items ranged from a few percent up to about one-fifth of the respondents.[18,27,31]
2. Factor structure: seven of the nine studies (78%) reported factor structures from a formal factor or principal component analysis,[18,20,24,26–28,31] with a multifactorial structure and a median of 3.3 factors (range one to six) related to teleconsultation per instrument. The factors related to: communication ("interaction," "satisfaction with communication and management," "information exchange," n=5); overall satisfaction (n=3); management ("delay until visit," "initial contact person," "service," n=3); access to service (n=2); attitude of health professional (n=2); telephone advice ("product," n=1); and individual information ("urgency of complaint;" n=1). The correlation between the number of items and the resulting number of factors was low (r=0.16). For instance, one high-item instrument with 37 items[18] identified only two factors that explained 72% of the variance; whereas, another instrument with 20 items[20]

reported a structure with six factors, which explained 61% of the variance.

3. Reliability measures: all nine instruments provided reliability measures – one study for both the total scale and the subscales; two studies for the total scale; and the remaining studies for the subscales only. The Cronbach's alpha values for the total scales were acceptable,[42] good,[26] or excellent.[21] Cronbach's alpha values for most of the factor subscales were above 0.7. However, three of the seven studies – evaluating the reliability of the subscales – revealed questionable[20,28] and unacceptable[26] Cronbach's alpha values for individual subscales. One study provided results for inter-item correlation with correlation coefficients ranging from 0.45–0.89, indicating a good internal consistency of the scale.[18] Three studies additionally reported item-total correlations which ranged from 0.53–0.92, supporting the internal consistency of these instruments.[18,27,31] Three publications investigated the test/retest reliability and reported correlation coefficients for subscales of >0.7, which are considered satisfactory or better.[18,20,27] For single subscales, however, correlation coefficients were <0.7, indicating limitations in test/retest reliability.[18,20]
4. Validity measures: in five of the eight instruments (63%) the scales correlated well with related constructs indicating construct validity. For example, higher scores correlated with simple measures of overall satisfaction.[18,20,21,28,31] Other scales correlated well with the patients' ages, the duration of the consultation, difficulties in contact by phone, waiting times, the amount of information received during the teleconsultation, the fulfillment of expectations or the transfer to a face-to-face visit. One study examined the convergent validity and found that sub-scores of the instrument were moderately correlated.[18] Only one of eight studies (13%) investigated the concurrent validity by comparing a shortened scale with the original instrument and reported modest intraclass correlation coefficients of 0.38–0.54.[21]

Discussion

This systematic review reports on 16 instruments used for the multidimensional assessment of 17,797 patients, regarding patient satisfaction after teleconsultation and triage for a health problem. The review identified four instruments with comprehensive information on their development and psychometric properties.[18,20,28,31]

The selection of the most appropriate instrument will probably depend on the purpose of the instrument – whether it is thought for routine assessments after a consultation, for periodic application as a quality control measure, or as a research instrument. For example, a 37-item instrument demonstrated good internal consistency and an indication of validity. However, the proportion of missing items was very large for some items; the test/retest reliability may have been limited, and the instrument had only two factors.[18] This instrument may be selected for research purposes or for routine assessments, if multidimensionality is not the main focus of the evaluation. Another ten-item instrument, in contrast, showed four factors, good internal consistency, and construct validity (without evaluating the test/retest reliability).[31] Due to its brevity and test evaluation results, this instrument may be suitable for most purposes. The most frequently used instrument (20 items) demonstrated high-item completion rates, a six-factorial structure, and construct validity. However, several subscales only had a very limited internal consistency.[20] An alternative 22-item instrument with a six-factor structure also showed construct validity, with a questionable internal consistency of one subscale and without information on the item completion rates.[28] However, both instruments may be selected if the multidimensionality of patient satisfaction assessment is of the utmost importance.

As only seven instruments used a formal factor analysis to identify the relevant underlying constructs, we applied a pragmatic approach for attributing the content of the remaining nine instruments to a list of domains from a systematic review.[31] This methodology confirmed the most frequently detected domains from the factor analysis ("communication," "overall satisfaction," and "management") and identified additional domains as relevant for users. These are: "perceived quality of the telephone advice;" "attitude of health professional;" "attitude of patient;" and "professional skills." Depending on their specific interests, the coverage of these domains may be an additional criterion for users to select any of those instruments.

Although most of the instruments had been developed over the last decade – a decade with an increased awareness for the need of methodological rigor in psychometric instrument development and testing[47] – many studies lacked details on the development process, had minimal information on the instruments' reliability, and only one-half of the instruments presented the validity of the existing scales. Factor structure, reliability, and validity were only reported for one-quarter of the instruments. No study evaluated the extent to which a score on the instrument predicts the associated outcome measures (predictive validity), which would allow conclusions

about the patients' adherence to the recommendations or the health service use.[14] The recruitment strategy and handling of nonresponders were comparable across the studies.

In his systematic review of patient satisfaction questionnaires for out-of-hours care in 2007, Garratt identified four instruments that reported some data on reliability and validity;[20,21,27,28] all were included in this review.[4] Garratt concluded that all of those studies had limitations regarding their development process and their evaluation of psychometric properties. Even though several years have passed, our review has to confirm these limitations. Despite extensive searching, we did not find any attempts to further modify, reevaluate, and improve the instruments with limited reliability or redundant items – except in one study. That study reduced a 38-item questionnaire[20] to a shorter version with only eight items.[21] Six of the 16 instruments identified in this review were published in subsequent years.[17,18,22,26,30,31] Of these, three instruments reported both methodological and psychometric data, two of which provide evidence of acceptable reliability and validity.[18,31]

Measuring patient satisfaction after teleconsultation and triage is a challenging endeavor. The assessment needs to focus on the quality of the service without being contaminated by the actions of subsequent health care providers or the severity and the natural course of the health problem. For instance, timing the administration of the questionnaire can be crucial. In the review, the delivery of the questionnaire varied between immediate inquiries to a latency of up to 16 months postconsultation. There is conflicting evidence regarding to what degree the timing of administration may confound the measurement of patient satisfaction. Previous work suggests that a potential timing effect depends on the health status of patients and the initial problem they sought help for.[10] Applied to our review, this would suggest that the optimal timing would be relatively shortly after the teleconsultation (ie, <1 week), as longer time intervals may increase memory problems for details of the teleconsultation, and the course of the medical problem may confound the perceived quality of the encounter.

Our review is based on a comprehensive literature search that included expert contacts and no language restrictions. Study selection, quality assessment, and data extraction with pretested forms – performed independently by two researchers – limited bias and transcription errors. Our ad hoc analysis of the instruments without formal factor analysis confirmed the domains identified in the studies with a formal factor analysis, but it identified other relevant domains with face validity. Our review was limited to instruments published in scientific journals. However, more instruments are likely to be in use. A recent survey among medical academic centers in the USA revealed a frequent use of internal instruments.[48] However, if these internal instruments had been thoroughly developed and formally evaluated, we assume they would have been published in a scientific journal.

If the measurement results are to be used for a comparison of different teleconsultation centers or of physicians within these centers or to demonstrate improvements in patient satisfaction over time, the instruments must undergo rigorous development and evaluation processes. Presently, this is the case for only a minority of these instruments. For example, the Patient-Reported Outcome Measurement Information System (PROMIS) instruments' development and psychometric scientific standards provide a set of criteria for the development and evaluation of psychometric tests.[49] Specifically, this includes reporting on the details of the development process, including the definition of the target concept and the conceptual model, the testing of reliability and validity parameters, and the reevaluations after potential refinements of the initial instrument. High-quality multidimensional assessment instruments should be consequently used in future trials to generate valid and comparable evidence of patient satisfaction with teleconsultation. This also includes a follow-up on patient satisfaction over time.

Conclusion

The status of appraisal of the instruments for measuring patient satisfaction after teleconsultation and triage – identified in the present systematic review – varies from comprehensive test evaluations to fragmentary and even missing data on factor structure, reliability, and validity. This review may serve as a starting point for selecting the instrument that best suits the intended purpose in terms of content and context. It offers pooled information and methodological advice to instrument developers with an interest in developing the long-needed assessment instrument.

Acknowledgments

We would like to thank Stefan Posth, Odense, Denmark, and Philipp Kubens, Freiburg, Germany, for their translation of Danish and Swedish article information.

Disclosure

MAI and EB are employees of a private institution providing teleconsultation and triage. RK reports no conflict of interest in this work.

References

1. ahrq.gov [homepage on the Internet]. Telemedicine for the Medicare population: update. Agency for Healthcare Research and Quality (AHRQ); [updated February 2006]. Available from: http://www.ahrq.gov/downloads/pub/evidence/pdf/telemedup/telemedup.pdf. Accessed October 16, 2013.
2. Deshpande A, Khoja S, McKibbon A, Jadad AR. *Real-Time (Synchronous) Telehealth in Primary Care: Systematic Review of Systematic Reviews.* Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health; 2008. Available from: http://www.cadth.ca/media/pdf/427A_Real-Time-Synchronous-Telehealth-Primary-Care_tr_e.pdf.
3. Nurse telephone triage in out of hours primary care: a pilot study. South Wiltshire Out of Hours Project (SWOOP) Group. *BMJ.* 1997;314(7075):198–199.
4. Garratt AM, Danielsen K, Hunskaar S. Patient satisfaction questionnaires for primary care out-of-hours services: a systematic review. *Br J Gen Pract.* 2007;57(542):741–747.
5. Lattimer V, George S, Thompson F, et al. Safety and effectiveness of nurse telephone consultation in out of hours primary care: randomised controlled trial. The South Wiltshire Out of Hours Project (SWOOP) Group. *BMJ.* 1998;317(7165):1054–1059.
6. McKinstry B, Walker J, Campbell C, Heaney D, Wyke S. Telephone consultations to manage requests for same-day appointments: a randomised controlled trial in two practices. *Br J Gen Pract.* 2002;52(477):306–310.
7. Thompson F, George S, Lattimer V, et al. Overnight calls in primary care: randomised controlled trial of management using nurse telephone consultation. *BMJ.* 1999;319(7222):1408.
8. Bunn F, Byrne G, Kendall S. Telephone consultation and triage: effects on health care use and patient satisfaction [review]. *Cochrane Database Syst Rev.* 2004;4:CD004180.
9. Kraai IH, Luttik ML, de Jong RM, Jaarsma T, Hillege HL. Heart failure patients monitored with telemedicine: patient satisfaction, a review of the literature. *J Card Fail.* 2011;17(8):684–690.
10. Crow R, Gage H, Hampson S, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol Assess.* 2002;6(32):1–244.
11. Ware JE Jr, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. *Eval Program Plann.* 1983;6(3–4):247–263.
12. Linder-Pelz S, Struening E. The multidimensionality of patient satisfaction with a clinic visit. *J Community Health.* 1985;10(1):42–54.
13. Locker D, Dunt D. Theoretical and methodological issues in sociological studies of consumer satisfaction with medical care. *Soc Sci Med.* 1978;12(4A):283–292.
14. Nunnally JC, Bernstein IH. *Psychometric Theory.* 3rd ed. New York, NY: McGraw Hill; 1994.
15. Kline P. *The Handbook of Psychological Testing.* 2nd ed. London: Routledge; 1999.
16. Everitt BS. *The Cambrige Dictionary of Statistics.* 2nd ed. Cambridge, MA: University Press; 2002.
17. Rahmqvist M, Husberg M. Effekter av sjukvårdsrådgivning per telefon: En analys av rådgivningsverksamheten 1177 i Östergötland och Jämtland. *CMT Rapport.* 2009:3.
18. Campbell JL, Dickens A, Richards SH, Pound P, Greco M, Bower P. Capturing users' experience of UK out-of-hours primary medical care: piloting and psychometric properties of the Out-of-hours Patient Questionnaire. *Qual Saf Health Care.* 2007;16(6):462–468.
19. Dixon RA, Williams BT. Patient satisfaction with general practitioner deputising services. *BMJ.* 1988;297(6662):1519–1522.
20. McKinley RK, Manku-Scott T, Hastings AM, French DP, Baker R. Reliability and validity of a new measure of patient satisfaction with out of hours primary medical care in the United Kingdom: development of a patient questionnaire. *BMJ.* 1997;314(7075):193–198.
21. Salisbury C, Burgess A, Lattimer V, et al. Developing a standard short questionnaire for the assessment of patient satisfaction with out-of-hours primary care. *Fam Pract.* 2005;22(5):560–569.
22. Dixon RF, Stahl JE. A randomized trial of virtual visits in a general medicine practice. *J Telemed Telecare.* 2009;15(3):115–117.
23. Hicks LL, Boles KE, Hudson S, et al. Patient satisfaction with teledermatology services. *J Telemed Telecare.* 2003;9(1):42–45.
24. Mekhjian H, Turner JW, Gailiun M, McCain TA. Patient satisfaction with telemedicine in a prison environment. *J Telemed Telecare.* 1999;5(1):55–61.
25. Moscato SR, David M, Valanis B, et al. Tool development for measuring caller satisfaction and outcome with telephone advice nursing. *Clin Nurs Res.* 2003;12(3):266–281.
26. Ström M, Baigi A, Hildingh C, Mattsson B, Marklund B. Patient care encounters with the MCHL: a questionnaire study. *Scand J Caring Sci.* 2011;25(3):517–524.
27. Moll van Charante E, Giesen P, Mokkink H, et al. Patient satisfaction with large-scale out-of-hours primary health care in The Netherlands: development of a postal questionnaire. *Fam Pract.* 2006;23(4):437–443.
28. van Uden CJ, Ament AJ, Hobma SO, Zwietering PJ, Crebolder HF. Patient satisfaction with out-of-hours primary care in the Netherlands. *BMC Health Serv Res.* 2005;5(1):6.
29. Keatinge D, Rawlings K. Outcomes of a nurse-led telephone triage service in Australia. *Int J Nurs Pract.* 2005;11(1):5–12.
30. Dehours E, Vallé B, Bounes V, et al. User satisfaction with maritime telemedicine. *J Telemed Telecare.* 2012;18(4):189–192.
31. Garratt AM, Danielsen K, Forland O, Hunskaar S. The Patient Experiences Questionnaire for Out-of-Hours Care (PEQ-OHC): data quality, reliability, and validity. *Scand J Prim Health Care.* 2010;28(2):95–101.
32. Salisbury C. Postal survey of patients' satisfaction with a general practice out of hours cooperative. *BMJ.* 1997;314(7094):1594–1598.
33. Shipman C, Payne F, Hooper R, Dale J. Patient satisfaction with out-of-hours services; how do GP co-operatives compare with deputizing and practice-based arrangements? *J Public Health Med.* 2000;22(2):149–154.
34. McKinley RK, Roberts C. Patient satisfaction with out of hours primary medical care. *Qual Health Care.* 2001;10(1):23–28.
35. McKinley RK, Stevenson K, Adams S, Manku-Scott TK. Meeting patient expectations of care: the major determinant of satisfaction with out-of-hours primary medical care? *Fam Pract.* 2002;19(4):333–338.
36. Glynn LG, Byrne M, Newell J, Murphy AW. The effect of health status on patients' satisfaction with out-of-hours care provided by a family doctor co-operative. *Fam Pract.* 2004;21(6):677–683.
37. Thompson K, Parahoo K, Farrell B. An evaluation of a GP out-of-hours service: meeting patient expectations of care. *J Eval Clin Pract.* 2004;10(3):467–474.
38. Campbell J, Roland M, Richards S, Dickens A, Greco M, Bower P. Users' reports and evaluations of out-of-hours health care and the UK national quality requirements: a cross sectional study. *Br J Gen Pract.* 2009;59(558):e8–e15.
39. Kelly M, Egbunike JN, Kinnersley P, et al. Delays in response and triage times reduce patient satisfaction and enablement after using out-of-hours services. *Fam Pract.* 2010;27(6):652–663.
40. Giesen P, Moll van Charante E, Mokkink H, Bindels P, van den Bosch W, Grol R. Patients evaluate accessibility and nurse telephone consultations in out-of-hours GP care: determinants of a negative evaluation. *Patient Educ Couns.* 2007;65(1):131–136.
41. Smits M, Huibers L, Oude Bos A, Giesen P. Patient satisfaction with out-of-hours GP cooperatives: a longitudinal study. *Scand J Prim Health Care.* 2012;30(4):206–213.
42. Beaulieu R, Humphreys J. Evaluation of a telephone advice nurse in a nursing faculty managed pediatric community clinic. *J Pediatr Health Care.* 2008;22(3):175–181.

43. Reinhardt AC. The impact of work environment on telephone advice nursing. *Clin Nurs Res.* 2010;19(3):289–310.
44. Rahmqvist M, Ernesäter A, Holmström I. Triage and patient satisfaction among callers in Swedish computer-supported telephone advice nursing. *J Telemed Telecare.* 2011;17(7):397–402.
45. López C, Valenzuela JI, Calderón JE, Velasco AF, Fajardo R. A telephone survey of patient satisfaction with realtime telemedicine in a rural community in Colombia. *J Telemed Telecare.* 2011;17(2):83–87.
46. McKinstry B, Hammersley V, Burton C, et al. The quality, safety and content of telephone and face-to-face consultations: a comparative study. *Qual Saf Health Care.* 2010;19(4):298–303.
47. Cella D, Yount S, Rothrock N, et al; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care.* 2007;45(5 Suppl 1):S3–S11.
48. Dawn AG, Lee PP. Patient satisfaction instruments used at academic medical centers: results of a survey. *Am J Med Qual.* 2003;18(6):265–269.
49. National Institutes of Health. *PROMIS Instrument Development and Validation Scientific Standards.* Silver Spring, MD: National Institutes of Health; 2013. Available from: http://www.nihpromis.org/Documents/PROMISStandards_Vers2.0_Final.pdf. Accessed October 16, 2013.

Supplementary material

Table S1 Medline search algorithm

1. Patient satisfaction MeSH heading: exp, topic	5. Questionnaire MeSH heading: exp, topic	9. Triage MeSH heading: exp, topic
2. User satisfaction, topic	6. Instrument, topic	10. Hotline, topic
3. Consumer satisfaction, topic	7. Measurement, topic	11. Telephone, topic
4. or\1–3	8. or\5–7	12. Remote consultation, topic
		13. Telemedicine, topic
		14. Telecommunication, topic
		15. Telehealth, topic
		16. Teletriage, topic
		17. Referral, topic
		18. Teleconsultation, topic
		19. Telenurse, topic
		20. Helpline, topic
		21. Healthline, topic
		22. Telecare, topic
		23. Callcenter, topic
		24. or\9–23
		25. and\4, 8, 24

Abbreviation: MeSH, medical subject heading.