

Assessing the quality of teleconsultations in a store-and-forward telemedicine network

Richard Wootton

Journal Name:	Frontiers in Public Health
ISSN:	2296-2565
Article type:	Original Research Article
Received on:	12 May 2014
Accepted on:	01 Jul 2014
Provisional PDF published on:	01 Jul 2014
www.frontiersin.org:	www.frontiersin.org
Citation:	Wootton R(2014) Assessing the quality of teleconsultations in a store-and-forward telemedicine network. <i>Front. Public Health</i> 2:82. doi:10.3389/fpubh.2014.00082
Copyright statement:	© 2014 Wootton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY) . The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after rigorous peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

Assessing the quality of teleconsultations in a store-and-forward telemedicine network

Richard Wootton(1,2), Joanne Liu(3) and Laurent Bonnardot(4,5)

- 1 Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway, Tromsø, Norway
- 2 Faculty of Health Sciences, University of Tromsø, Tromsø, Norway
- 3 MSF International, Geneva, Switzerland
- 4 Department of Medical Ethics and Legal Medicine, Paris Descartes University, Paris, France
- 5 Fondation Médecins Sans Frontières, Paris, France

Correspondence:

Professor R Wootton,
Norwegian Centre for Integrated Care and Telemedicine,
University Hospital of North Norway,
PO Box 6060,
9038 Tromsø,
Norway
(Email: r_wootton@pobox.com)

Abstract

Store and forward telemedicine in resource-limited settings is becoming a relatively mature activity. However, there are few published reports about quality measurement in telemedicine, except in image-based specialities, and they mainly relate to high- and middle-income countries. In 2010, Médecins Sans Frontières (MSF) began to use a store-and-forward telemedicine network to assist its field staff in obtaining specialist advice. To date, more than 1000 cases have been managed with the support of telemedicine, from a total of 40 different countries. We propose a method for assessing the overall quality of the teleconsultations provided in a store-and-forward telemedicine network. The assessment is performed at regular intervals by a panel of observers, who -- independently -- respond to a questionnaire relating to a randomly-chosen past case. The answers to the questionnaire allow two different dimensions of quality to be assessed: the quality of the process itself and the outcome, defined as the value of the response to three of the four parties concerned, i.e. the patient, the referring doctor and the organisation. It is not practicable to estimate the value to society by this technique. The feasibility of the method was demonstrated by using it in the MSF telemedicine network, where process-quality scores, and user-value scores, appeared to be stable over a nine-month trial period. This was confirmed by plotting the cusum of a portmanteau statistic (the sum of the four scores) over the study period. The proposed quality assessment method appears feasible in practice, and will form one element of a quality assurance programme for MSF's telemedicine network in future. The method is a generally applicable one, which can be used in many forms of medical interaction.

Keywords: telemedicine, telehealth, quality assurance, process control, LMICs

Introduction

Médecins Sans Frontières (MSF) is a non-governmental humanitarian medical organisation that responds to emergency situations and provides medical assistance to those in need. MSF teams provide medical emergency aid in difficult settings around the world, and staff often have to diagnose and treat patients with limited resources.[1] In 2010, MSF began to use a store-and-forward telemedicine network to assist its field staff in obtaining specialist advice.[2] To date, more than 1000 cases have been managed with the support of telemedicine, from a total of 40 different countries.

In a store-and-forward telemedicine network of this type, doctors in the field refer cases electronically to obtain a second opinion about diagnosis or management. Incoming cases are reviewed by a case coordinator and assigned for reply to one or more appropriate experts. The network therefore operates in a similar way to a bulletin board, with messages being posted by its users. Although formal evidence for the clinical effectiveness of the telemedicine advice obtained through networks of this kind is rather scarce[3,4], they are known to provide a useful service to referring doctors, and several networks have operated for periods of more than a decade.

Quality problem

Since store and forward telemedicine in resource-limited settings is becoming a relatively mature activity, there is a concomitant requirement to implement quality assurance/improvement activities. Indeed, it may be considered unethical not to do so.

However, there are few published reports about quality measurement in telemedicine, except in networks concerned with radiology[5], ophthalmology[6] or histopathology,[7] many of which are retrospective studies. These reports concern image-based activities, which perhaps lend themselves more readily to quality measurement. The situation in teleconsulting is more complex, being inherently multi-specialty in nature and one where there is often limited knowledge of outcomes. Attempting to measure quality in such a context is more like attempting to measure overall quality in a multi-clinic outpatient department. As far as we are aware, there have been no previous studies of prospective quality measurement in general teleconsulting work in low income countries.

Objectives

The primary Research Question was whether a method could be developed for quality measurement in general teleconsulting work in low income countries. The aim of the present work, therefore, was to develop a method for assessing the quality of the teleconsultations being conducted in the MSF telemedicine network, and then to examine its feasibility for routine adoption.

Methods

The present study required the development of a method to assess quality and then a demonstration of its feasibility in practice. The work was performed in two stages:

- (1) development of a quality assessment tool
- (2) demonstration of feasibility in the MSF telemedicine network

Ethics permission was not required, because patient consent to access the data had been obtained and the work was a retrospective chart review conducted by the organisation's staff in accordance with its research policies.

1. Assessment of quality

Development of the quality tool

A questionnaire was developed by a consensus between three experienced telemedicine practitioners. It was based on accepted tools used in previous studies.[8,9] The final questionnaire was evaluated and approved by an independent evaluator. The final questionnaire consisted of 17 questions. These concerned the information provided by the referring doctor, the way that the referral was handled in the telemedicine network, the response(s) received from the specialist(s) consulted, and the likely value to the patient, the doctor and the organisation.

Definition of quality

We defined quality in terms of two of the three dimensions of the Donabedian model: process and outcome. (The structural dimension is not usually relevant in a telemedicine network of the sort under discussion). Thus in assessing the quality of a given teleconsultation, there are two principal questions:

- (1) was the *process* by which the response was produced satisfactory? i.e. what was the quality of the teleconsultation process itself?
- (2) was the *outcome* from the teleconsultation useful? i.e. what was the value of the

teleconsultation and to whom?

These questions address separate dimensions of quality, both of concern to network operators. That is, the process for producing a response might be satisfactory, but the response itself could be useless. Or the process could be unsatisfactory, but the response might still be useful.

Both aspects of quality can best be judged by using a panel of assessors. This is because any evaluation will involve subjective judgements, so a panel of observers is more likely to produce an accurate estimate than a single observer. However, it is not feasible to evaluate the quality of every single teleconsultation conducted in the network, so there must be a sampling process by which a case is selected (randomly) for assessment at regular intervals. This leads to a quality assessment scheme whose main features are summarised in Table 1.

Quality of process

The quality of the teleconsultation process (q_p) can be assessed by the panel members, who can make a judgement about various relevant matters. For example, they can judge whether the referrer provided sufficient information, whether the case was sent promptly to an appropriate expert, whether an answer was obtained sufficiently quickly to be useful and so on. There are ten questions listed in Table 2 which are relevant to the quality of the process. The scoring system is described in Appendix 1.

Value of response

The value of the response can be assessed in a similar way by the panel members. There are four domains of interest:

1. Value to the patient, v_p . After the patient himself, the person best placed to judge this is the referring doctor. It can also be estimated by senior staff in the organisation
2. Value to the referring doctor, v_r . The person best able to judge this is the referring doctor, but it can also be estimated by senior staff in the organisation
3. Value to the organisation, v_o . This is probably best judged by senior staff in the organisation itself
4. Value to society as a whole, v_s .

The first three values can be assessed by staff with suitable telemedicine experience. However, assessing the value to society is much more difficult. The value to society of telemedicine will be partly determined by the health care system in the country concerned (mainly, the country where the patient is located), including the degree to which telemedicine has been properly integrated into the chain of health care there. Assessing the value to society as a whole is therefore difficult to do on the basis of a single telemedicine case, and is ignored in what follows. It is worth noting that in a humanitarian context (or a not-for-profit operation), the value to society will be closely aligned with the value to the organisation.

Direct measurement of value is not straightforward. In health economics, it is usual to measure the cost-effectiveness of the technique in question and to make a comparison (e.g. with usual practice) to obtain evidence that it does not represent a waste of resources. However, in the context of telemedicine in resource-limited settings, this is not easy to do. First, the costs are distorted, because many staff are volunteers and there may also be donor support which can be hard to quantify. Second, the clinical effect of telemedicine may be

difficult to document, since patients are commonly lost to follow-up after their initial encounter.

How else can the "value" of a teleconsultation episode be measured? That is, what is the value to the interested parties? Panel members can form a judgement about whether the telemedicine response clarified the diagnosis, whether the eventual clinical outcome would be beneficial for the patient and so on. There are nine questions listed in Table 2 which are relevant to the value of the response in the domains of interest. The scoring system is described in Appendix 1.

2. Demonstration of feasibility

To demonstrate the feasibility of the proposed approach, a panel of 12 experts were invited to answer the 17 questions about randomly-selected telemedicine cases, see Table 2. Cases were chosen at random for a 9-month period. The process was:

- (1) the system automatically selected a past case for review at the beginning of each month. The case was chosen randomly from those referred 4-8 weeks previously. If there were fewer than four cases in the period of interest, no case was selected. (The average case submission rate during the period in question was approximately one case per day.)
- (2) the members of the quality assessment panel were notified by email that a case had been chosen for review. The panel comprised mainly senior doctors with previous MSF field experience; there were three other healthcare professionals with telemedicine experience.
- (3) panel members logged in to the telemedicine system, viewed the information about the chosen case and answered the questions about the case. The questions had simple, multiple-choice answers which were presented in a drop-down box for ease of selection. Panel members could not view the answers from any other panel member until they had provided their own.
- (4) when at least one set of answers had been provided, the system calculated the quality scores for the case. The four quality scores were values in the range 0-10.

Process stability

A control chart was used to examine the stability of the monthly quality scores. Control charts can be plotted for each of the four quality indices, but for simplicity, a Grand Quality Score (GQS) for each case was calculated from the panel's quality and value scores as:

$$G = Q_p + V_p + V_r + V_o$$

That is, the GQS represents an equi-weighted summation of the four constituent indices. The GQS was transformed to lie in the range 0-10 (0=worst, 10=best).

The cusum chart is a well-established and powerful method for identifying changes in a process average. The chart plots the cumulative difference between the recorded values and a target value, which is often chosen to be the process average. The GQS values were plotted as a cusum, using the grand mean as the reference value.

Note that there are two important assumptions underlying the use of control charts: the

measurement that is used to monitor the process is distributed according to a normal distribution; it was not necessary to transform the data in the present case. Also, the measurements are assumed to be independent of each other.

Results

The panel assessed randomly-selected cases starting in July 2013. At least 4 responses were received for each case. The median panel score for process quality was 8.0 (IQR 7.3,8.7) across the nine cases. The lowest score awarded for process quality by an individual panel member in any case was 4.7 and the highest was 9.0. The median values in each case are shown in Figure 1. There was good agreement between panel members about process quality, i.e. relatively small IQRs for each case.

The median panel score for value to the *patient* was 8.9 (IQR 7.8,8.9). The lowest score awarded for value to the patient was 3.3 and the highest was 10. The median values in each case are shown in Figure 2. The agreement between panel members was less good than for process quality.

The median panel score for value to the *doctor* was 9.1 (IQR 8.6, 9.5). The lowest score awarded for value to the doctor was 5.7 and the highest was 10.0. The median values in each case are shown in Figure 3. The agreement between panel members was better than for value to the patient.

The median panel score for value to the *organisation* was 8.9 (IQR 7.2, 10.0). The lowest score awarded for value to the organisation was 5.6 and the highest was 10.0. The median values in each case are shown in Figure 4. The agreement between panel members was less good than for value to the doctor.

The median panel GQS was 8.6 (IQR 7.6, 9.2). The lowest individual GQS was 6.1 and the highest was 9.8 The median values in each case are shown in Figure 5. The cusum is shown in Figure 6. There was no evidence that the process was out of control, i.e. with steadily increasing or steadily decreasing values. In fact, over the epoch studied, the cusum was essentially zero at the end, while deviations no larger than +12% occurred over the study period.

Discussion

We have developed a quality assessment scheme for a store-and-forward telemedicine network and demonstrated its feasibility in a real-life clinical setting. There appear to be no previous reports of similar work.

Relation to other evidence

Previous work on assessment of quality in telemedicine networks has often focused on user satisfaction (e.g. Von Wanenheim et al[10]), which is a related, but different, concept. Most previous quality studies have been retrospective reviews, such as that conducted by Mahnke et al.[11] There have been few attempts to measure the value to the clinician, although Chan et al investigated this in a real-time teleconsultation network in a high-income country.[12]

Methodological issues

The proposed method was trialled in a real-life telemedicine network, where it was shown to be feasible and appeared to produce useful results. It thus appears suitable for routine adoption. Implicit in the methodology are a number of design decisions.

1. **Questionnaire.** The size of the questionnaire is likely to influence the number of responses from the panel. The right balance has to be struck between asking too few questions and too many. On one hand, the more questions that are asked, the better the situation can be assessed; but on the other hand, too many questions will discourage the observers from responding which will make the system less sustainable. In practice 10-20 questions seems to be a reasonable number.
2. **Monitoring and stability.** Which index (process quality, and the three value domains) is most appropriate for long-term monitoring, in order to measure network performance? Are all four indices of equal importance, or should some be more heavily weighted than others? Should they be monitored collectively, rather than individually? This requires further work.
3. **Sampling.** How often should cases be sampled and monitoring be done? On one hand, more frequent sampling will allow closer performance monitoring; on the other hand, it is likely to lead to "observer fatigue". In practice, we suggest that random sampling of one case per month is about right.
4. **Size of panel.** How many panel members should give an opinion? The more members there are, the more likely there is to be disagreement between them; on the other hand, the more there are, the better the estimate of the underlying value. In practice, we suggest that 5-10 panel members is about right.

Quality assurance

Routine measurement of quality on randomly-selected cases is only one part of the whole evaluation process and will form one element of an overall quality assurance programme for the telemedicine network concerned. Other elements may include obtaining other points of view and follow up reports to assess long-term outcomes concerning the cases and the benefits of the expertise.

Limitations

The present work has certain limitations. For example, before it could be used routinely, the quality-assessment methodology would require validation. However, it is difficult to validate the proposed indices independently, especially in the context of a telemedicine network operated by a humanitarian organisation. Ideally, they should be evidence-based, and of demonstrated validity and reliability.[13] Further work is required to find out whether this is possible, since the practical problem of the lack of an obvious gold standard needs to be overcome. Validation may therefore require psychometric methods.[14]

Industrial process control is normally done using an absolute standard as the reference. In the present work, a relative reference value was employed. That is, it represents an assessment of relative quality, which pragmatically, is probably better than no assessment at all. Again, further work is required to find out whether absolute reference standards can be developed.

Finally, the quality of this evaluation relies on the information available for assessing the case.

Sampling a case at a particular time may be problematic if there is insufficient feedback on follow up. It also relies on the expertise and experience of the assessor panel. The panel members must be selected carefully and it is important that they have no conflict of interest. This is why it may be better to use independent volunteers, rather than senior staff from the organisation running the network.

Interpretation

The present method provides estimates of the value to the main parties concerned in a teleconsultation, together with an estimate of the quality of the teleconsultation process itself. This is important information for those responsible for the operation of the network. To the best of our knowledge there has been no information published previously about the quality of general teleconsultations in a store-and-forward network. Yet, if telemedicine is considered sufficiently mature that it can enter routine service, there is an ethical imperative to ensure that it is employed in a cost-effective manner. The method described here provides an instrument for monitoring quality and will form part of the toolset used by the operators of the MSF telemedicine network in future.

Once a method for assessing quality is available, application of industrial process control methodology allows the stability of the network to be monitored. Again, this is important if network operators are to be reassured that quality is not in slow decline. The information may also be valuable in improving the performance of healthcare staff in low-resource settings, which is known to be a difficult problem.[15]

The techniques presented in this paper are of wide application. They could potentially be used in non-telemedicine consultations (i.e. conventional, face-to-face consulting), and in industrialised countries as well as resource-limited settings.

Conclusions

A method for assessing the quality of the teleconsultations in a store-and-forward telemedicine network is proposed. It provides estimates of the quality of the process and the value of the consultation to the main parties involved. A trial of the method showed that it was feasible and that the process in the network studied was stable. The method appears to give useful results. It seems desirable to implement it in other telemedicine projects where it can contribute to the evaluation of practice, something that is necessary in all medical services provided.

Acknowledgements

We are grateful to Will Wu for software support.

References

- 1 Médecins Sans Frontières. International Activity Report 2012 -- overview of activities. See <http://www.msf.org/international-activity-report-2012-overview-activities> (last checked 8 April 2014)
- 2 Coulborn RM, Panunzi I, Spijker S, Brant WE, Duran LT, Kosack CS, Murowa MM. Feasibility of using teleradiology to improve tuberculosis screening and case management in a district hospital in Malawi. Bull World Health Organ. 2012 Sep

- 1;90(9):705-11
- 3 Wootton R, Geissbuhler A, Jethwani K, Kovarik C, Person DA, Vladzomyrskyy A, Zanaboni P, Zolfo M. Comparative performance of seven long-running telemedicine networks delivering humanitarian services. *J Telemed Telecare*. 2012 Sep;18(6):305-11
 - 4 Wootton R, Geissbuhler A, Jethwani K, Kovarik C, Person DA, Vladzomyrskyy A, Zanaboni P, Zolfo M. Long-running telemedicine networks delivering humanitarian services: experience, performance and scientific output. *Bull World Health Organ*. 2012 May 1;90(5):341-347D
 - 5 Soegner P, Rettenbacher T, Smekal A, Buttinger K, Oefner B, zur Nedden D. Benefit for the patient of a teleradiology process certified to meet an international standard. *J Telemed Telecare*. 2003;9 Suppl 2:S61-2
 - 6 Erginay A, Chabouis A, Viens-Bitker C, Robert N, Lecleire-Collet A, Massin P. OPHDIAT: quality-assurance programme plan and performance of the network. *Diabetes Metab*. 2008 Jun;34(3):235-42
 - 7 Lee ES, Kim IS, Choi JS, Yeom BW, Kim HK, Han JH, Lee MS, Leong AS. Accuracy and reproducibility of telecytology diagnosis of cervical smears. A tool for quality assurance programs. *Am J Clin Pathol*. 2003 Mar;119(3):356-60
 - 8 Patterson V, Wootton R. A web-based telemedicine system for low-resource settings 13 years on: insights from referrers and specialists. *Glob Health Action*. 2013 Sep 23;6:21465
 - 9 Wootton R, Menzies J, Ferguson P. Follow-up data for patients managed by store and forward telemedicine in developing countries. *J Telemed Telecare*. 2009;15(2):83-8
 - 10 von Wangenheim A, de Souza Nobre LF, Tognoli H, Nassar SM, Ho K. User satisfaction with asynchronous telemedicine: a study of users of Santa Catarina's system of telemedicine and telehealth. *Telemed J E Health*. 2012 Jun;18(5):339-46
 - 11 Mahnke CB, Jordan CP, Bergvall E, Person DA, Pinsker JE. The Pacific Asynchronous TeleHealth (PATH) system: review of 1,000 pediatric teleconsultations. *Telemed J E Health*. 2011 Jan-Feb;17(1):35-9
 - 12 Chan FY, Soong B, Lessing K, Watson D, Cincotta R, Baker S, Smith M, Green E, Whitehall J. Clinical value of real-time tertiary fetal ultrasound consultation by telemedicine: preliminary evaluation. *Telemed J*. 2000 Summer;6(2):237-42
 - 13 Rubin HR, Pronovost P, Diette GB. From a process of care to a measure: the development and testing of a quality indicator. *Int J Qual Health Care*. 2001 Dec;13(6):489-96
 - 14 Bland JM, Altman DG. Statistics Notes: Validating scales and indexes. *BMJ*. 2002 Mar 9;324(7337):606-7
 - 15 Rowe AK, de Savigny D, Lanata CF, Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet*. 2005 Sep 17-23;366(9490):1026-35

Table 1. Main features of the quality assessment scheme

Sampling of process output	One case per month is selected at random
Panel of assessors	Senior staff (n=12) with experience in the field (mainly doctors)
Evaluation -- individual scores	Each panel member responds (independently) to a set of questions, from which the following can be computed: process quality (q_p) and value scores (v_p, v_r, v_o)
Evaluation -- panel scores	Aggregated scores are then calculated to indicate the panel's overall assessment of process quality (Q_p) and value (V_p, V_r, V_o), based on the median panel scores
Evaluation -- composite score	Finally, a composite score is calculated to reflect the panel's overall assessment of quality (based on the process quality and value scores)

Table 2. Quality assessment questions

Question	Response choices	Quality of the process, QP	Value to the patient, Vp	Value to the referring doctor, Vr	Value to the organisation, Vo	Value to society, Vs
1. Was the question asked by the referring doctor clear?	Yes/Perhaps/No/Don't know	X				
2. Did the referrer provide sufficient information?	Yes/Perhaps/No/Don't know	X				
3. Were any images provided?	Yes/No					
4. ... if YES, were the images adequate?	Yes/Perhaps/No/Don't know	X				
5. ... if NO, would some images have helped?	Yes/Perhaps/No/Don't know	X				
6. Overall, could the referral have been improved?	Yes/Perhaps/No/Don't know	X				
7. Was the case sent to an appropriate expert?	Yes/Perhaps/No/Don't know	X		X		
8. Was the answer provided sufficiently quickly?	Yes/Perhaps/No/Don't know	X		X		
9. Was the answer(s) well-adapted for the local environment?	Yes/Perhaps/No/Don't know	X		X		
10. Overall, could the answer have been improved?	Yes/Perhaps/No/Don't know	X		X		
11. Did the telemedicine advice clarify the diagnosis for the doctor and patient?	Yes/Perhaps/No/Don't know		X	X		
12. Did the suggested action help the doctor manage the patient?	Yes/Perhaps/No/Don't know		X	X		
13. Do you think that the eventual clinical outcome will be beneficial for the patient?	Yes/Perhaps/No/Don't know		X		X	
14. Was the consultation useful for the doctors	Yes/Perhaps/No/Don't know			X	X	

concerned?						
15. Could the allocation/coordination have been improved?	Yes/Perhaps/No/Don't know	X				
16. Was the consultation good from the organization's point of view?	Yes/Perhaps/No/Don't know				X	X
17 Do you have any comments about this case?	(Free text)					

Appendix. Scoring of questions

The questions set out in Table 2 are presented as multiple choices for simplicity. Scoring for most questions -- those in which an affirmative response indicated satisfaction with the item being considered -- was: No=1, Perhaps=2, Yes=3. However, some questions required reversed scoring, where an affirmative indicated *dissatisfaction*: Yes=1, Perhaps=2, No=3.

Don't-know responses were coded as zeroes, i.e. they were treated in the same way as missing values. That is, no distinction was made in the present work between missing values and don't-know responses. These may represent cases where assessors experienced particular difficulty in forming a judgement.

Individual scores

Scores were first calculated for each individual panel member, as follows. The process quality score from a panel member who had assessed a particular case was calculated as:

$$Q_p = S_1 + S_2 + S_4 + S_5 + S_6 + S_7 + S_8 + S_9 + S_{10} + S_{15}$$

where the response elements, $s_1, s_2 \dots$ refer to the responses in Table 2, reverse-scored where appropriate.

The value scores from a given panel member were calculated as:

$$V_p = S_{11} + S_{12} + S_{13}$$

$$V_r = S_7 + S_8 + S_9 + S_{10} + S_{11} + S_{12} + S_{14}$$

$$V_o = S_{13} + S_{14} + S_{16}$$

In the above calculations, the response elements for a given panel member were aggregated by simple summation. That is, the constituent responses for the questions being used in a particular score were equi-weighted.

For convenience, the values q_p, v_p, v_r and v_o were transformed to lie in the range 0-10 (0=worst, 10=best).

Panel scores

The individual panel member scores were aggregated to produce a panel mean, which represents the best estimate of the underlying true value pertaining to the case in question. In the absence of a compelling reason to use a more complex scheme, the individual scores were equally weighted, i.e. this amounts to placing similar value on the judgement of all members of the panel. For example, the panel's best estimate of process quality was:

$$Q_p = (q_1 + q_2 + \dots + q_i) / n$$

where Q_p is the panel process quality score, and $q_1, q_2 \dots$ are the individual scores for process quality from the n panel members.

Figure 1. Median scores for process quality (0=worst; 10=best). The error bars indicate the 25th and 75th percentiles.

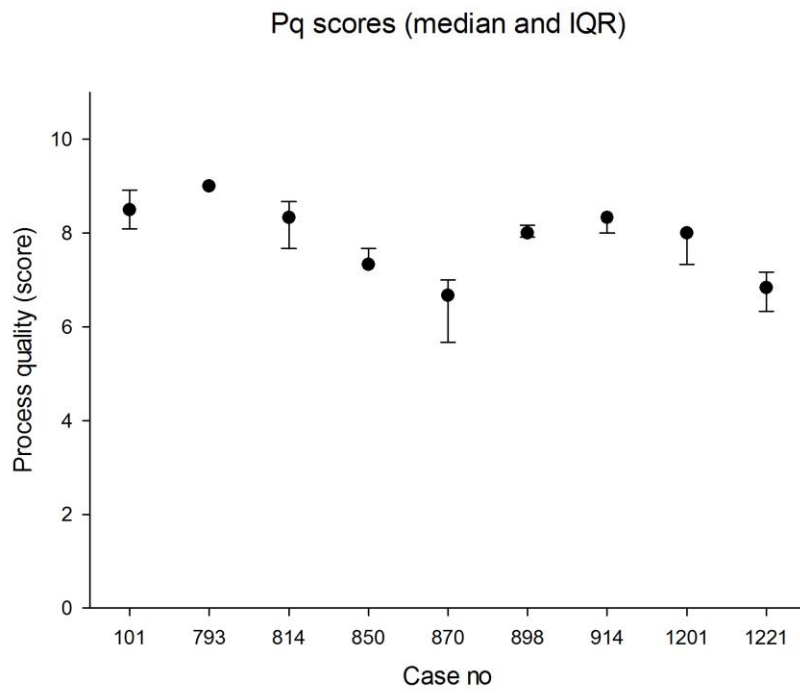


Figure 2. Median scores for value to patient (0=worst; 10=best). The error bars indicate the 25th and 75th percentiles.

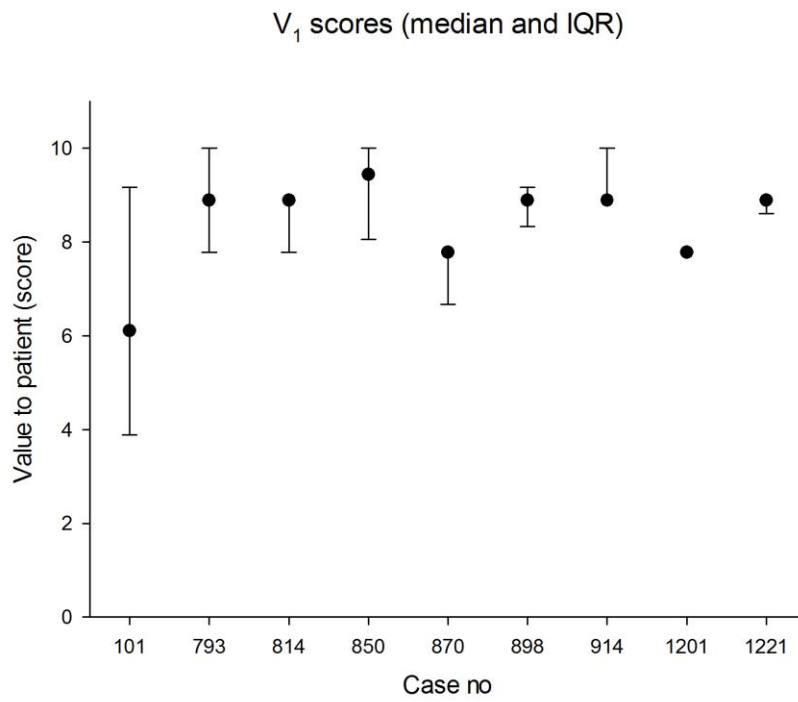


Figure 3. Median scores for value to referrer (0=worst; 10=best). The error bars indicate the 25th and 75th percentiles.

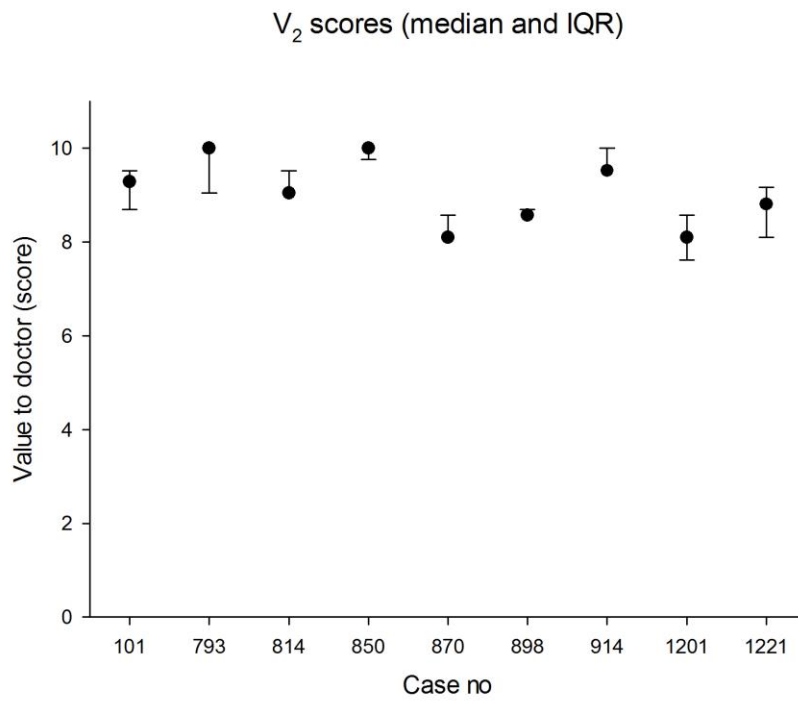


Figure 4. Median scores for value to the organisation (0=worst; 10=best). The error bars indicate the 25th and 75th percentiles.

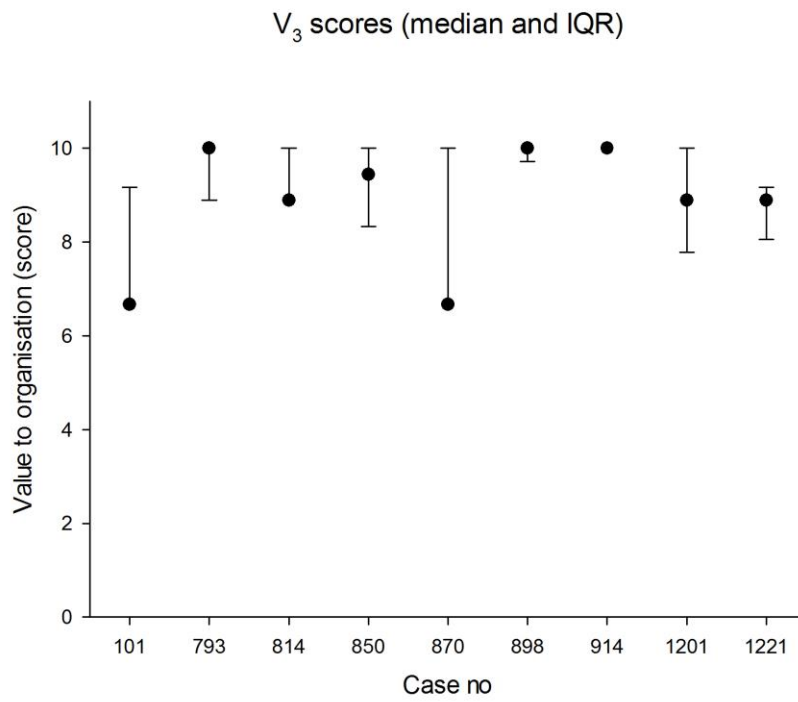


Figure 5. Median scores for general quality (0=worst; 10=best). The error bars indicate the 25th and 75th percentiles.

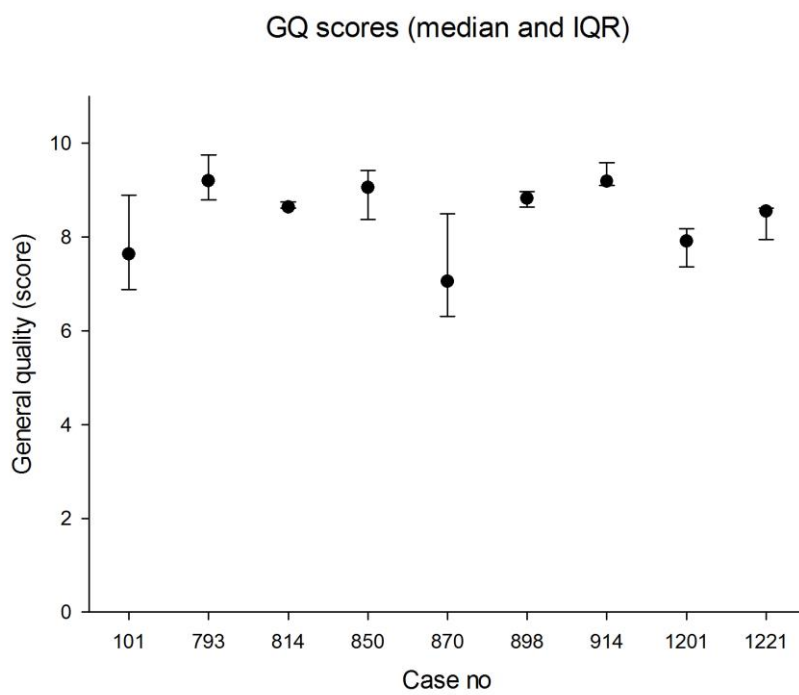


Figure 6. Cusum of General Quality Score.

