

Original Research Article

Reliability of Telegram between different evaluators in emergency general surgery practice

Sandipkumar Chaudhari*, Jimitkumar Patel, Ajay Talvelkar, Nayan Pancholi, Latif Bagwan

Department of General Surgery, GMERS, Vadnagar, Gujarat, India

Received: 04 April 2022

Revised: 30 April 2022

Accepted: 02 May 2022

*Correspondence:

Dr. Sandipkumar Chaudhari,

E-mail: drsandy253@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: 'Telegram' is a popularly used social media application in routine clinical practice, often for communication between a resident doctor and a consultant. The purpose of this study is to establish the reliability of the data transmitted through telegram to validate its use in general surgery practice.

Methods: Clinical findings and computerized tomographic (CT) images of 180 patients visiting the trauma center and emergency department were posted in a closed Telegram group involving three consultants (SSC, NBP, and LSB). The CT images were posted in the Telegram group as complete picture (CP) and picture of interest (POI) format and rated on a scale of 1-5. The consultants formulated a temporary diagnosis and initial management plan. The reliability between different evaluators of these feedbacks was analyzed in our study.

Results: Mean CP rating ranged from 3.03 ± 0.61 to 3.73 ± 0.64 [Cronbach alpha (α)-0.494, $p=0.006$]. Mean POI rating ranged from 3.4 ± 0.56 to 4.13 ± 0.73 ($\alpha=0.824$, $p<0.0001$). For diagnosis, the proportion of observed agreement (P_0) was 83.3% for SSC and NP, 76.6% for SSC and LB, and 73.3% for NBP and LSB. For management, P_0 was 86.6% for NBP and LSB, 86.6% for SSC and NBP, and 80% for SSC and LSB.

Conclusions: Telegram messenger serves to transmit good quality pictures of CT scan images. A reasonable diagnosis and management strategy can be formulated using this app with a fair reliability between different evaluators.

Keywords: Teleconsultation, Telegram, Feedback, Practice, Reliability, Consultant

INTRODUCTION

Telegram is a freeware, cross-platform, cloud-based instant messaging (IM) service. The service also provides end-to-end encrypted video calling.¹ In January 2021, Telegram surpassed 500 million monthly active users.² It was the most downloaded app worldwide in January 2021 with 1 billion downloads globally as of late August 2021.³ Similar to WhatsApp, Telegram is also a smartphone-based application used frequently for telecommunication.⁴ Telegram has not been officially approved for providing tele medical consultation until recently, though the presence of many digital communication platforms in India. Still, people have increasingly used this app as an

“off-label” means of communication for teleconsultation in clinical practice before the COVID era.⁵ The Medical Council of India and vetted by the Ministry of Health and Family Welfare, India have issued new guidelines for its use in teleconsultation within 3 days of the first lockdown on 26 March 2020, due to the COVID-19 pandemic.⁵ After this official approval, the app has been widely used as a means of teleconsultation.

Telegram has been widely used in clinical practice as a tool of communication between a medical intern/resident and a consultant, where the resident seeks advice regarding a particular case for initial management. Nevertheless, there are not enough studies authenticating the use of Telegram for this purpose, especially in general surgery practice.⁶ It

is also to be determined whether the use of the Telegram for the transmission of information can have the same reproducibility as that of physical consultation. Or is there any chance of potential misinformation or misinterpretation of data sent through Telegram? It also needs to decide whether this makes any difference in undertaking decisions for patients and the care of patients. The aim of the study was to find the variation between different evaluators and inter-evaluator reliability of the data sent through a telegram in the physical world and thus validating its use in general surgery practice.

METHODS

This prospective study was conducted from 01 October 2020 to 30 November 2021, at GMERS Medical College and General Hospital, India. Approval from the ethics committee of the Institute was obtained before starting the study. 180 patients presenting with emergency surgical problems have been assigned under the care of three general surgery consultants (SSC, NBP, LSB) in three different units were included. All the emergency patients to the general surgery department with relevant cross-sectional imaging in the form of CT scan were initially evaluated by the senior resident on duty for emergency primary care. Any pediatric patients and adults presented with hemodynamic instability are excluded from the study. Then the clinical findings and radiological images were posted in a closed Telegram group involving the three consultants and a senior resident (SR) after taking informed consent from the patient. The relevant radiological images which were posted in the Telegram group were posted as a complete picture (CP) and picture of interest (POI) format. CP included the photo of whole CT scan sheet with multiple cross-sectional images and POI included the particular image with the possible abnormality or pathology (Figure 1).

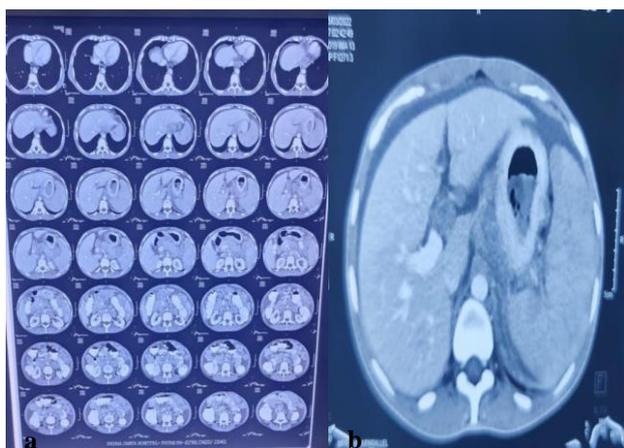


Figure 1: Screenshot of CT scan images sent by the resident to consultant (a) complete picture (b) picture of interest.

After posting the details in the telegram group all three consultants reviewed the details of the patients

independently within few minutes of posting it. They rated the quality of both CP and POI separately on Likert scale of 1-5 (1-very poor, 2-poor, 3-good, 4-very good, and 5-excellent). The senior resident also rated the image quality of original image and noted it separately. He then made his temporary diagnosis and initial treatment plan. Similarly, the consultants after going through the details provided in the post made a temporary diagnosis and initial management strategy. They all sent these responses (quality of CP, quality of POI, provisional diagnosis, and management) as a personalized telegram message to the senior resident separately to maintain blinding from each other's responses. All the responses were recorded by the senior resident in the case sheet and later entered in a Microsoft excel sheet.

In case of gross disparity in the responses from different consultants, the initial response in the form of separate personal message to senior resident was recorded for analysis, and then different treatment modalities were discussed among the three consultants before making the final decision for the actual treatment plan for a specific patient. Different phone models used in this study were: consultant 1-SSC (iPhone 7 plus, Apple, 12 MP camera), consultant 2-NBP (Samsung Galaxy S20, 12 MP camera), consultant 3- LSB (Samsung Galaxy note 8, 12 MP camera), SR-(One plus 8, 13 MP camera). All the phones of consultants and senior resident are password protected and their own personal phones. Telegram messenger safeguarded the information by providing end-to-end encryption. The images were not shared anywhere other than this telegram group and after completion of the study they were deleted from the phones and saved in encrypted form in computer.

Statistical analysis

Microsoft excel sheet was prepared and data were analyzed using R Studio GUI front-end Statistics for Windows. The data were expressed as number and percentage. Categorical data between the groups were analyzed from a 2x2 contingency table. The concordance for the rating of picture was represented using Cronbach alpha, and reliability between evaluators was represented by the interclass correlation coefficient (ICC) for rating of imaging. For diagnosis, coding was done in binary fashion, 0 for observed disagreement and 1 for observed agreement to the diagnosis of resident. For management, coding was done as 0 for observed disagreement and 1 for observed agreement to the final management of the patient. P_0 was defined as a proportion of observed agreement, which is given as sum total of agreements divided by the total response ($a+d/N$), where a is total positive agreements, d is total negative agreements, and N is total responses (180). The inter-observer agreement between two raters was calculated using kappa (κ) statistics. $P < 0.05$ was considered statistically significant. The formula for Cohen's kappa is calculated as given below.

$$k = (p_o - p_e)/(1 - p_e)$$

Where, p_o is relative observed agreement among raters, and p_e is hypothetical probability of chance agreement.

RESULTS

A total of 180 patients (120 males and 60 females) were included in the study. Mean age of the patients was 45.47 ± 13.68 years (Figure 2).

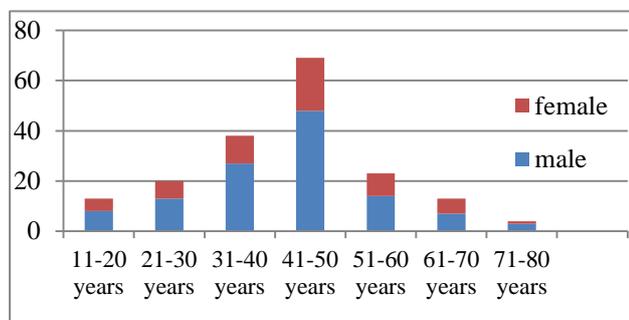


Figure 2: Age and sex distribution of the patient population.

The mean CP rating was as follows: resident- 3.69 ± 0.42 , SSC- 3.89 ± 0.69 , NBP- 3.93 ± 0.24 , and LSB- 4.03 ± 0.16 . The Cronbach alpha for CP rating was 0.474 and ICC was 0.196 ($p=0.006$). The mean POI rating by the resident (the resident rated the actual imaging) was 4.07 ± 0.78 . The POI rating by the three consultants was as follows: SSC- 4.19 ± 0.42 , NBP- 4.43 ± 0.63 , and LSB- 3.49 ± 0.26 . The Cronbach alpha for POI was 0.817 and ICC was 0.540 ($p < 0.0001$). None of the consultants asked for any added POI to be sent separately.

For diagnosis, the proportion of observed agreement P_o was 83.43% for SSC and NBP, 79.1% for SSC and LSB, and 75.3% for NBP and LSB. As compared to the resident, the P_o was 83.57% for resident and LSB, 82% for resident and NBP, and 78.47% for resident and SSC. The corresponding kappa values are provided in Table 1.

Table 1: Agreement between observers for management decisions: Kappa values for various readers.

Pair	P_o	Kappa	P
Resident and SSC	78.47	0.416	0.015
Resident and NBP	82	0.443	0.017
Resident and LSB	83.57	0.631	0.002
SSC and NBP	83.43	0.272	0.128
SSC and LSB	79.1	0.272	0.128
NBP and LSB	75.3	0.441	0.016

Note: SSC=Sandipkumar Chaudhari, NBP=Nayan Pancholi, and LSB=Latif Bagwan

There was a difference in planning the treatment strategy in terms of the method of diversion used for acute urinary retention such as placement of Foley catheter versus a suprapubic catheter (12/180). There also was a difference

in terms of approach to a particular procedure such as laparoscopic versus abdominal exploration in blunt abdominal trauma (36/180). We merged the methods “Foley’s catheter placement” and “suprapubic catheter placement” as “catheterizations” The approaches “laparoscopic abdominal exploration” and “open abdominal exploration” were combined as “abdominal exploration.” After this reassessment, the P_o and reliability between evaluators increased further and is shown in Table 2.

Table 2: Agreement between observers for management decision: Kappa values for various readers after adjustment.

Pair	P_o	Kappa	P
Resident and SSC	91.5	0.609	0.001
Resident and NBP	90.3	0.672	0.000
Resident and LSB	93.5	0.609	0.001
SSC and NBP	86.5	0.580	0.001
SSC and LSB	80.3	0.285	0.123
NBP and LSB	86.5	0.585	0.001

Note: SSC=Sandipkumar Chaudhari, NBP=Nayan Pancholi, and LSB=Latif Bagwan

DISCUSSION

In the era of infectious disease pandemic which demands social distancing, the use of teleconsultation is likely to increase.^{7,8} Hence, studies regarding the validation of tools for teleconsultation are needed to optimize their use. In this study, we intended to determine the variation between observers and calculate reliability between the evaluators of use of clinical data transmitted through Telegram in emergency general surgery setting. In our study, the mean scores of the image rating reflected a “good (3)” to a “very good (4)” rating for all transmitted radiological images. The mean scores for the CP were less than that given for the POI by all the observers. This entails the fact that there is always a need to zoom the CP to look for the abnormality/pathology, which leads to loss of pixels of the CP, resulting in blurring of pictured image when seen on a mobile phone (Figure 2b). However, the POI is transmitted after clicking an individual image which does not require zooming and hence has no such problems of pixel loss (Figure 1b and 2a). Thus, there is better perception and readability for POI than CP.

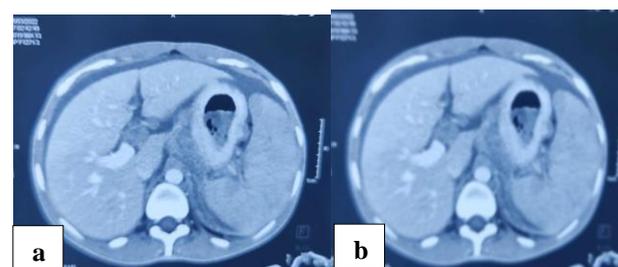


Figure 3: (a) Picture of interest and (b) same section of CT scan after zooming in the complete picture.

There was wide variability between evaluators in rating the quality of CP when compared to rating of quality of POI as is seen from the values of Cronbach alpha (0.494 versus 0.824) and ICC (0.196 versus 0.540). The plausible explanation for the same could be difference in the evaluator's subjective interpretation of the quality of CP and the phone model used. Scanning the whole CT image on phone and interpreting it leads to the difference in ratings assigned to CP. Furthermore, as mentioned earlier, the need for zooming in and consequential loss of pixels leads to blurring of pictured image (Figure 3). The perception of this blurred and zoomed image leads to a significant difference in subjective rating by the evaluator, while for POI alone such zooming is infrequently needed.

Regarding diagnoses, it was seen that the proportion of observed agreement P_0 values among consultants were good and ranged from 73.3% to 83.3%. Thus, it may be inferred that despite the difference in rating the quality of pictured images, the data may be interpreted to reach a reasonable diagnosis. On further analyzing the data on no agreement cases, we noticed that the consultants found additional findings on imaging other than those pointed by the resident in 30/180 occasions (SSC and NBP) and 24/180 occasions (LSB), respectively. This amounts to a very high likelihood (>80%) of users finding the said diagnosis, based upon the clinical data and imaging using Telegram, irrespective of the pictured image quality.

The final and the most important outcome of any teleconsultation is its utility in providing a management strategy after reviewing the clinical data. In our study, when we looked at the formulation of treatment strategy, we found high P_0 ranging from 70% to 86.7%. Despite this high proportion of observed agreement, we had low kappa values (0.270–0.627). This paradoxical low κ values despite the high observed agreement is due to prevalence dependency of κ , which has been described in detail by Cicchetti and Feinstein and must be interpreted with caution.^{9,10} In cases where exact prevalence or the “gold standard” is unknown the calculation of prevalence is done using marginal totals from a 2x2 table for observed agreements and disagreements. In case the marginal totals (observed agreement or disagreement) become very low either vertically or horizontally as in our study, the κ lowers drastically for the same value of the proportion of observed agreement (P_0).¹⁰ Thus, it is the proportion of observed agreement which become more relevant than κ alone in such situations.¹¹

As mentioned in the results, the observed agreement and κ increased further after recoding to “catheterization” and “abdominal exploration” as mentioned in the result section. This recoding was done since approaches to a particular procedure are likely to differ even in physical consultation, and hence the recoded values possibly reflect the true agreement among the observers. This difference of opinion can be sorted after mutual discussion, as was done in our series, and these differences are likely to arise even

in cases where the clinical data and the imaging are presented physically to a group of physicians.

In a study done with another social media application to demonstrate use of social media application in tele medicine by Sener et al on evaluation of interrater reliability of WhatsApp for evaluation of hematuria, 212 patients were evaluated for hematuria by two groups of urologists.¹² One group having direct access to the patients, while the other group comprising urologist blinded to patients' data and received image on WhatsApp. The grade of hematuria was evaluated by them as follows: Hematuria with the following rating: 0 – no hematuria, 1 – hematuria that does not require invasive treatment, and 2 – hematuria requiring bladder drainage or any form of active treatments. They found almost perfect agreement between two groups, (kappa-0.992). Another study for cryptoscopic/ureteroscopic image was conducted by Arada et al who found significant agreement between consultant and attending plans.¹³ The reply was in the form of “agree” or “disagree” to the formulated management plan by the attending. However, this being a conference abstract, the detailed methodology and results could not be ascertained. Contrary to these studies, the kappa values in our study are low (0.280–0.672) as previously mentioned. This difference may also be accounted for by the complexity of the problem being evaluated. Reading and interpreting CT scan images on phone is a complex task as compared to the interpretation of hematuria using color of urine and cryptoscopic images.

In this current era, the use of telemedicine has increased exponentially. Telegram has proved itself a versatile software for tele consults and telemedicine in resource-limited Asian countries. Metcalfe's law states that the value of a telecommunications network is proportional to the square of the number of connected users of the system (n_2).¹⁴ The major advantages of using this app for tele medical consultation was its widespread user base, no extra cost for sending messages, calls (both audio and video), ability to share photos, videos and message using a single platform and with end to end encryption facility which maintained confidentiality of patient's identity and data.^{15,16} Because of new norms of social/physical distancing the use of this tool for emergency consults and communication between the residents and consultants in different medical field also increases. Therefore, our study was aimed to validate the use and assess the reproducibility of tele consults between different evaluators. Our study is the very few of its category addressing the use of Telegram for entire decision-making of the patients visiting the emergency department. The pictured images and the clinical scenarios are much more complex than those in the already evidence-based literature. Our study is a small footstep toward the amalgamation of social media applications (such as WhatsApp, Facebook, and Viber) being used off label for clinical use into day-to-day clinical practice. Obviously, extravagant use of these applications comes at the cost of a precious commodity, i.e. time and

they have accompanying adverse effects related to increase in screen time such as eye and neck strains.

Main limitation our study is a comparator arm of physical control was not kept because of existing COVID-19 pandemic. The study had descriptive data for comparison (diagnosis and management); this derives from the fact that we wished to replicate the mainstream clinical practice in our system. In our study, all the phones used can click and reproduce high-quality images. The quality of mobile phones used can also affect the interpretation of the image. We did not address the picture-taking skill of an individual trainee through our study.

CONCLUSION

Telegram messenger serves to transmit good quality pictures of imaging modality such as CT scans. These images along with appropriate clinical history can be used to formulate a proper diagnosis and treatment strategy with fair to substantial reliability between different evaluators. Telegram can be used in emergency general surgery setting with significant agreement among the resident and consultants.

Funding: No funding sources

Conflict of interest: None declared

Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Telegram introduces end-to-end encrypted video calls". The Next Web. 2020. Available at: <https://thenextweb.com/news/telegram-introduces-end-to-end-encrypted-video-calls>. Accessed on 29 March 2021.
2. "Telegram hits 500M monthly active users". Telegram. Available at: <https://web.archive.org/web/20210611161236/https://t.me/durov/147>. Accessed on 11 June 2021.
3. "Durov Telegram". Telegram. 2021. Available at: <https://web.archive.org/web/20210611161236/https://t.me/durov/147>. Accessed on 09 February 2021.
4. Hashtarkhani S, Azimi A, Fatehi F. Televisit: an innovative mobile-based system for remote medical consultation using Telegram messaging app in Iran. International Conference on Successes and Failures in Telehealth (SFT-17). 2017.
5. Telemedicine Practice Guidelines Enabling Registered Medical Practitioners to Provide Healthcare Using Telemedicine. Available at: <https://www.mohfw.gov.in/pdf/Telemedicine.pdf>. Accessed on 08 September 2020.
6. Martin G, Khajuria A, Arora S, King D, Ashrafian H, Darzi A. The impact of mobile technology on teamwork and communication in hospitals: A systematic review. J Am Med Inform Assoc. 2019;26:339-55.
7. Luciani LG, Mattevi D, Cai T, Giusti G, Proietti S, Malossini G. Teleurology in the time of Covid-19 pandemic: Here to stay? Urology. 2020;140:4-6.
8. Devana SK, Chaudhary K, Sharma AP, Singh SK. Changing urological practice during COVID-19. Indian J Urol. 2020;36:153-8.
9. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol. 1990;43:551-8.
10. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43:543-9.
11. Donker DK, Hasman A, van Geijn HP. Interpretation of low kappa values. Int J Biomed Comput. 1993;33:55-64.
12. Sener TE, Butticiè S, Sahin B, Netsch C, Dragos L, Pappalardo R, et al. WhatsApp use in the evaluation of hematuria. Int J Med Inform. 2018;111:17-23.
13. Arada EI, Florencio L, Macalalag M, Mendiola F, Dy J, Ballesteros C, et al. Mp23-01 Using android smartphones to take cystoscopic and ureterosopic images for Exclusive online internet referrals. J Urol. 2015;193:e267.
14. Shapiro C, Varian HR. Information Rules. Harvard Business Press. 1999.
15. Bellote MC, Santamaria HT, Pelayo-Nieto M, Heman Prasad ES, Gadzhiev N, Gudar K. Social media in the urology practice. Int Braz J Urol. 2019;45:877-81.
16. KamelBoulos MN, Giustini DM, Wheeler S. Instagram and WhatsApp in health and healthcare: An overview. Futur Internet 2016;8:1-14.

Cite this article as: Chaudhari S, Patel J, Talvelkar A, Pancholi N, Bagwan L. Reliability of Telegram between different evaluators in emergency general surgery practice. Int Surg J 2022;9:1193-7.