

Software Analysis of Colonoscopy Videos Enhances Teaching and Quality Metrics

Vasant Rajan¹, Havish Srinath¹, Christopher Yii Siang Bong¹, Alex Cichowski², Christopher J. Young³, Peter J. Hewett⁴

Review began 01/05/2022

Review ended 03/08/2022

Published 03/10/2022

© Copyright 2022

Rajan et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. Department of General Surgery, Logan Hospital, Meadowbrook, AUS 2. Australian Institute for Machine Learning, The University of Adelaide, Adelaide, AUS 3. Department of General and Colorectal Surgery, Memorial Health System, Abilene, USA 4. Colorectal Surgery, Department of Surgery, The Queen Elizabeth Hospital, Adelaide, AUS

Corresponding author: Vasant Rajan, vasantrajan12@gmail.com

Abstract

Purpose

Machine learning algorithms were hypothesized as being able to predict the quality of colonoscopy luminal images. This is to enhance training and quality indicators in endoscopy.

Methods

A separate study involving a randomized controlled trial of capped vs. un-capped colonoscopies provided the colonoscopy videos for this study. Videos were analyzed with an algorithm devised by the Australian Institute for Machine Learning. The image analysis validated focus measure, steerable filters-based metrics (SFIL), was used to assess luminal visualization quality and was compared with two independent clinician assessments (C1 and C2). Goodman and Kruskal's gamma (G) measure was used to assess rank correlation data using IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY).

Results

A total of 500 random colonoscopy video clips were extracted and analyzed, 88 being excluded. SFIL scores matched with C1 in 45% and C2 in 42% of cases, respectively. There was a significant correlation between SFIL and C1 ($G = 0.644$, $p < 0.005$) and SFIL and C2 ($G = 0.734$, $p < 0.005$).

Conclusion

This study demonstrates that machine learning algorithms can recognize the quality of luminal visualization during colonoscopy. We intend to apply this in the future to enhance colonoscopy training and as a metric for quality assessment.

Categories: Gastroenterology, General Surgery, Quality Improvement

Keywords: pilot project, cohort studies, australia, software, colonoscopy

Introduction

Interest in quality within medicine and its subsequent use to improve healthcare delivery has grown markedly since the Institute of Medicine's report on medical errors published in 2000 [1]. Colonoscopy is an excellent area for quality improvement given it is high volume with associated high risks and costs [2]. This has been recognized by the Australian Commission on Safety and Quality in Health Care, via the creation of the Colonoscopy Clinical Care Standard, which aimed to improve the safety and quality of colonoscopy via the identification and optimization of nine domains [3].

Other research in this area includes machine learning algorithms, which are theorized to minimize inter-observer variability in detecting and diagnosing luminal lesions [4]. Devices such as cap-assisted endoscopes are proposed to improve adenoma detection rate (ADR) by improving colonic mucosal visualization [5]. Artificial intelligence (AI) has been suggested as a method to improve ADR by capitalizing on deep learning, whereby AI is trained through repetitively studying recorded videos to recognize polyps that may occur only fleetingly during colonoscopies [6]. Such AIs are deployed to assist the endoscopist in detecting adenomas. Preliminary studies of these methods have shown dramatic improvements in ADR, from 40% in unassisted colonoscopies to 80% when an AI is deployed to assist in detection [7]. Oh et al. [8] have also proposed software analysis of colonoscopy video quality, which ultimately aims to provide a new standard against which the quality of colonoscopies can be objectively scored. This would assist in quality control measures on a large scale and eventually serve as a baseline against which benchmark can progress in colonoscopy training.

Our aim was to use machine learning algorithm software to predict luminal clarity quality while minimizing

How to cite this article

Rajan V, Srinath H, Bong C, et al. (March 10, 2022) Software Analysis of Colonoscopy Videos Enhances Teaching and Quality Metrics. Cureus 14(3): e23039. DOI 10.7759/cureus.23039

inter-observer variability. In theory, this would lead to greater standardization and objectivity in assessing colonic luminal visualization.

Materials And Methods

A previously conducted randomized controlled trial at the Queen Elizabeth Hospital, Adelaide, South Australia, designed to assess differences in ADRs between uncapped and cap-assisted endoscopes, provided the colonoscopy video content for this study. As part of the trial, the procedures were recorded. Ethics approval was obtained from the local Human Research Ethics Committee. The colonoscopes used in the study were the Olympus CF-H190L/I and CF-H180L/1 adult scope (Olympus Medical Systems, Tokyo, Japan), on which were attached the ARV120 Endocuff Vision, and the PCF-H180AL/I and PCF-H190L/I, which were attached with the ARV180 Endocuff Vision [5]. A total of 214 videos were obtained from the process and 32 videos were excluded for reasons that included obstructing cancers, strictures, consent, and device issues.

An algorithm was developed to design a software product that can be objective and have global applicability across different platforms and devices. We envisaged a software program that will require minimal input from the user and can be utilized with ease by clinicians of varying experience. Video analysis was performed in collaboration with the Australian Institute for Machine Learning, based at the University of Adelaide. Preliminary research assessed the optimum way of grading image focus.

The first steps in video selection included curating a set of images from experienced endoscopists that were representative of image quality noted during colonoscopy. The endoscopists classified the images into a four-point Likert scale: good, almost good, almost poor, and poor (Figure 1).

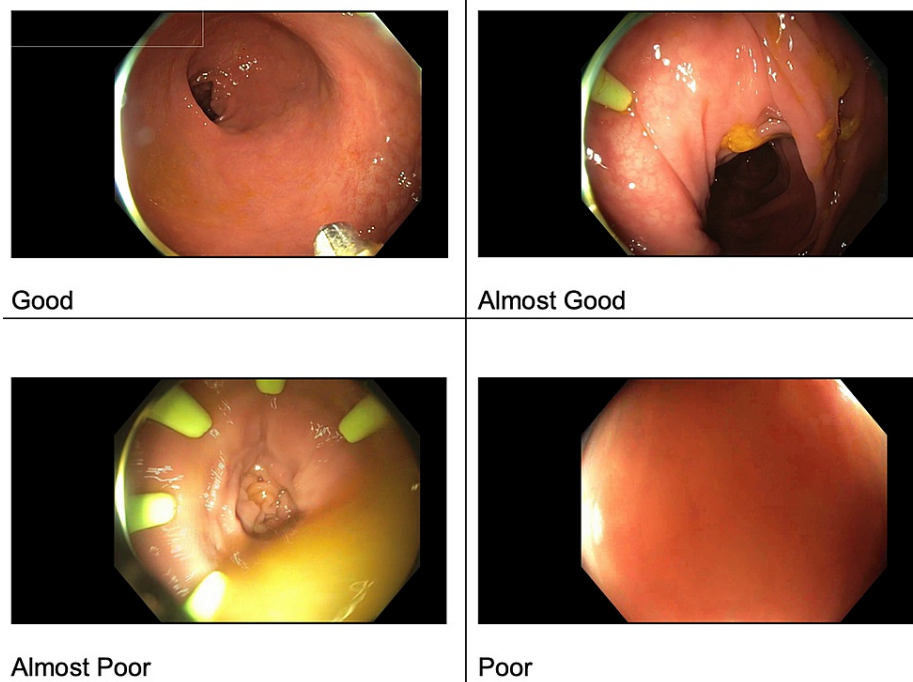


FIGURE 1: Examples representative of image quality noted during colonoscopy curated by experienced endoscopists.

A set of methods designed to estimate and score image focus in different ways were tested. These were drawn from published research on image focus. Focus measures can correlate well to a clear lumen that is being well visualized in front of the camera [9]. Testing these focus measures demonstrated that the "Poor" image contains very few crisp/in-focus visible edges while the "Good" image has crisp/in-focus edges visible [10]. We proceeded to test the various measures with focus ratings that most closely aligned with the clinical impressions. The focus measure, steerable filters-based metrics (SFIL), generated numerical scores for images aligned with the four categories, which were classified by the clinicians. Based on these results, we designed our algorithm to map SFIL results into four categories, with 1 = poor, 2 = almost poor, 3 = almost good, and 4 = good. At the end of this process, we had a program that could predict how a clinician would classify the images with 80-90% accuracy (Figure 2).

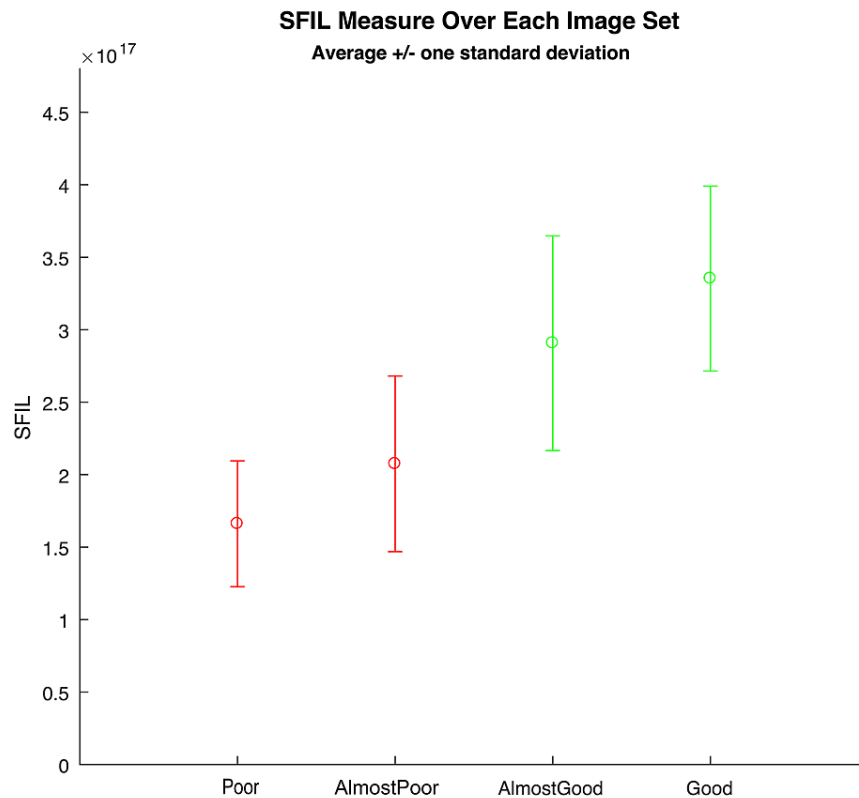


FIGURE 2: SFIL scoring of image sets, aligning with clinical classification.

SFIL, steerable filters-based metrics.

The next step was applying SFIL to video snippets. Video snippets that were 20-25 seconds in duration were scored with this algorithm. Using a computer program, these clips were extracted randomly from the original video files. Experienced endoscopists (C1 and C2) were blinded to their selection and required to score each snippet individually, as per the same four-point Likert scale. The algorithm was applied to produce a range of estimated per-frame scores across the duration of each clip. Scores generated by the clinicians and our algorithm were compared and analyzed.

Non-parametric testing was used with a Wilcoxon signed-rank test to evaluate inter-observer variability between the two clinicians. Goodman and Kruskal's gamma was utilized to determine the association between each of the clinicians' ratings and the SFIL result. Statistical analysis was performed using IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY).

Results

A total of 500 video clips were extracted randomly from the available recorded videos. A total of 88 clips were excluded, as these were recordings outside the patient either before or after completion of the colonoscopy. Scores generated by C1, C2, and SFIL were compared. The algorithm provided a single distinct estimated classification score over the duration of the clip in 41% of all cases, with this matching the human classification 72% (C1) and 71% (C2) of the time. In 29% and 25% of all cases, respectively, our algorithm produced estimates being two-step ranges (e.g. "almost good to good") and three-step ranges (e.g. "poor to almost good"). In these cases, the returned ranges matched the human classifications 80-90% of the time. In 5% of all cases, we could not deduce any estimate.

Inter-clinician variability comparisons showed no discrepancy in 62.4% of cases, a slight discrepancy in 32.5% of cases (appropriated as a one band difference), and a significant difference in 5.1% of cases (>two band difference). The two independent clinician's scores were within one band of each other 95% of the time. Statistical analysis using Wilcoxon signed-rank test showed no significant inter-clinician variability (Table 1).

Variability in inter-clinician scores of video clips			
		Frequency	Percent
Degree of variation* in assessed clips between clinicians C1 and C2	0 [^]	257	62.4
	1 [^]	134	32.5
	2 [^]	12	2.9
	3 [^]	9	2.2

TABLE 1: Comparison of inter-clinician variability in scores between clinician 1 (C1) and clinician 2 (C2).

* Variation in the human classification of video clips between the two clinicians. 0[^] = no difference in ranking; 1[^] = single band difference; 2[^] = two-band difference; 3[^] = three-band difference.

The comparison of SFIL with each clinician's scores using a non-parametric model, i.e., Goodman and Kruskal's gamma, revealed a significant correlation between C1 and SFIL ($G = 0.644$, $p < 0.005$) and a significant correlation between C2 and SFIL ($G = 0.734$, $p < 0.005$).

Discussion

Colonoscopy has evolved to become the gold standard in the screening, diagnosis, treatment, and surveillance of colonic pathologies. The technique was first pioneered in 1969 by doctors William Wolff and Hiromi Shinya, with an electro-desiccative snare that facilitated polypectomies occurring soon after [11]. The technique proved revolutionary and experienced rapid spread internationally; more than 800,000 colonoscopies were performed in 2016-2017 in Australia alone [12].

The adoption of colonoscopy has been universal in the developed world. In contrast, developing countries rely more on opportunistic screening programs than the organized programs prevalent in developed countries [13]. Thai researchers demonstrated that despite an eight-fold increase in annual costs, colonoscopy was still likely to be more cost-effective than fecal immunochemical test (FIT)-focused screening. However, budgetary constraints and costs to citizens were two main factors that made a FIT-focused screening regime more likely to be adopted [14]. This example illustrates the challenge of prioritizing limited healthcare resources, highlighting the need for procedures like colonoscopies to be of high quality, meeting universally set objective performance standards.

Factors that influence the effectiveness of colonoscopies include thorough bowel preparation, careful inspection, and longer withdrawal times [15-17]. These factors influence the ADR [18], which determines intervals between subsequent colonoscopies with the potential to impact rates of interval cancers [19]. Cecal intubation, scope withdrawal time, and ADR are accepted parameters to objectively measure the quality of colonoscopies according to various Australian societies [12].

However, colonoscopies have several disadvantages. Lakoff et al. demonstrated that a negative colonoscopy is associated with an overall reduction in colorectal cancer (CRC), particularly in left-sided CRC for up to 14 years. Compared to this, right-sided CRC had a reduction of incidence of only seven years [20]. ADRs differ between specialties, gastroenterologists had higher polyp detection and removal rates than general surgeons [21]. This was noted in studies from the United States and correlated in the Australian context by Zorron Cheng Tao Pu et al. [22].

The reporting of quality measures is subjective and operator-dependent [23]. These studies and observations highlight the multitude of factors that can impact the effectiveness and quality of a colonoscopy. This has a direct bearing on the decision-making and clinical outcomes for patients. It impacts their management and translates to additional costs to the healthcare system.

Clear visualization of the mucosa is critical to performing a high-quality colonoscopy. The quality of the visualized image has been indirectly measured using various bowel preparation scoring scales. These scores reflect the ability to clearly see the lumen and colonic mucosa, which directly bears on the quality measures described above.

This study demonstrates a novel method to objectively report quality in colonoscopy, which aims to complement existing quality control methods. Objective measures of quality are important for several reasons. The first is allowing benchmarking across different practitioners, specialties, and institutions, thereby allowing a greater concordance in results. Current quality measures are inherently subjective and are

thus poorly standardized, which makes result comparison between different institutions difficult. Introducing an objective measure of quality, therefore, allows for more efficient data sharing and would improve research endeavors between institutions.

The second is improving teaching by introducing an objective metric against which performance can be benchmarked and further refined. Khan et al. have written on the importance of feedback and debriefing on colonoscopy simulation training, whereby structured, objective feedback was directly correlated with clinical performance [24].

Finally, our study is preliminary and involves a limited number of videos. Greater numbers are needed for further validation. Therefore, we are hopeful that this pilot study can be used as a foundation for more research regarding focus measures in endoscopy and will pave the way for focus measures such as SFIL to be used alongside current colonoscopy quality measures. The method we describe needs refinement to reduce the margin for error and improve its accuracy. This would address one of the main limitations of the study. Another factor to consider is that different endoscopes or different generations of endoscopes and their attachments may have enough variation, which means more work is needed to standardize objective measures across other devices before they are safe for use.

Conclusions

Colonoscopy is a potent tool used in colorectal cancer and adenoma screening plus surveillance that remains hampered by inter-observer variability. Software may therefore aid and augment luminal visualization improving factors such as ADR. While the method is still some steps away from rating colonoscopies or endoscopists, this study demonstrates that machine learning algorithms can objectively recognize the quality of luminal visualization during colonoscopy. We aim to apply this in the future as an aid to colonoscopy training and as a metric for quality measurement.

Additional Information

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. Central Adelaide Local Health Network Human Research Ethics Committee issued approval HREC reference number: HREC/16/TQEH/94; CALHN reference number: Q20160504. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Institute of Medicine: To Err is Human: Building a Safer Health System . The National Academies Press, Washington, DC; 2000. [10.17226/9728](https://doi.org/10.17226/9728)
2. Calderwood AH, Jacobson BC: Colonoscopy quality: metrics and implementation. *Gastroenterol Clin North Am.* 2013, 42:599-618. [10.1016/j.gtc.2013.05.005](https://doi.org/10.1016/j.gtc.2013.05.005)
3. Australian Commission on Safety and Quality in Healthcare. Colonoscopy Clinical Care Standard (revised 2020). (2020). Accessed: January 21, 2022: <https://www.safetyandquality.gov.au/publications-and-resources/resource-library/colonoscopy-clinical-care-standard-re...>
4. de Lange T, Halvorsen P, Riegler M: Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy. *World J Gastroenterol.* 2018, 24:5057-62. [10.3748/wjg.v24.i45.5057](https://doi.org/10.3748/wjg.v24.i45.5057)
5. Jacob A, Schafer A, Yong J, et al.: Endocuff Vision-assisted colonoscopy: a randomized controlled trial . *ANZ J Surg.* 2019, 89:E174-8. [10.1111/ans.15067](https://doi.org/10.1111/ans.15067)
6. Sinagra E, Badalamenti M, Maida M, et al.: Use of artificial intelligence in improving adenoma detection rate during colonoscopy: might both endoscopists and pathologists be further helped. *World J Gastroenterol.* 2020, 26:5911-8. [10.3748/wjg.v26.i39.5911](https://doi.org/10.3748/wjg.v26.i39.5911)
7. Repici A, Badalamenti M, Maselli R, et al.: Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology.* 2020, 159:512-20.e7. [10.1053/j.gastro.2020.04.062](https://doi.org/10.1053/j.gastro.2020.04.062)
8. Oh J, Hwang S, Cao Y, Tavanapong W, Liu D, Wong J, de Groen PC: Measuring objective quality of colonoscopy. *IEEE Trans Biomed Eng.* 2009, 56:2190-6. [10.1109/TBME.2008.2006035](https://doi.org/10.1109/TBME.2008.2006035)
9. Pertuz S, Puig D, Garcia MA: Analysis of focus measure operators for shape-from-focus. *Pattern Recognit.* 2013, 46:1415-32. [10.1016/j.patcog.2012.11.011](https://doi.org/10.1016/j.patcog.2012.11.011)
10. Automation of focusing system based on image processing through intelligent algorithm . (2015). Accessed: January 21, 2022: https://www.worldresearchlibrary.org/up_proc/pdf/130-145206222626-33.pdf.
11. Wolff WI, Shinya H: Polypectomy via the fiberoptic colonoscope — removal of neoplasms beyond reach of the sigmoidoscope. *N Engl J Med.* 1973, 288:329-32. [10.1056/NEJM197302152880701](https://doi.org/10.1056/NEJM197302152880701)
12. Duggan A, Skinner IJ, Bhasale AL: All colonoscopies are not created equal: why Australia now has a clinical care standard for colonoscopy. *Med J Aust.* 2018, 209:427-30. [10.5694/mja18.00556](https://doi.org/10.5694/mja18.00556)
13. Sankaranarayanan R: Screening for cancer in low- and middle-income countries. *Ann Glob Health.* 2014,

- 80:412-7. [10.1016/j.aogh.2014.09.014](https://doi.org/10.1016/j.aogh.2014.09.014)
14. Phisalprapa P, Supakankunti S, Chaiyakunapruk N: Cost-effectiveness and budget impact analyses of colorectal cancer screenings in a low- and middle-income country: example from Thailand. *J Med Econ.* 2019, 22:1351-61. [10.1080/13696998.2019.1674065](https://doi.org/10.1080/13696998.2019.1674065)
 15. Harewood GC, Sharma VK, de Garmo P: Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia. *Gastrointest Endosc.* 2005, 58:76-9. [10.1067/mge.2003.294](https://doi.org/10.1067/mge.2003.294)
 16. Lee RH, Tang RS, Muthusamy VR, et al.: Quality of colonoscopy withdrawal technique and variability in adenoma detection rates (with videos). *Gastrointest Endosc.* 2011, 74:128-34. [10.1016/j.gie.2011.03.003](https://doi.org/10.1016/j.gie.2011.03.003)
 17. Sanchez W, Harewood GC, Petersen BT: Evaluation of polyp detection in relation to procedure time of screening or surveillance colonoscopy. *Am J Gastroenterol.* 2004, 99:1941-5. [10.1111/j.1572-0241.2004.40569.x](https://doi.org/10.1111/j.1572-0241.2004.40569.x)
 18. Rex DK, Schoenfeld PS, Cohen J, et al.: Quality indicators for colonoscopy. *Gastrointest Endosc.* 2015, 81:31-53. [10.1016/j.gie.2014.07.058](https://doi.org/10.1016/j.gie.2014.07.058)
 19. Lieberman DA, Rex DK, Winawer SJ, Giardiello FM, Johnson DA, Levin TR: Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology.* 2012, 143:844-57. [10.1053/j.gastro.2012.06.001](https://doi.org/10.1053/j.gastro.2012.06.001)
 20. Lakoff J, Paszat LF, Saskin R, Rabeneck L: Risk of developing proximal versus distal colorectal cancer after a negative colonoscopy: a population-based study. *Clin Gastroenterol Hepatol.* 2008, 6:1117-21. [10.1016/j.cgh.2008.05.016](https://doi.org/10.1016/j.cgh.2008.05.016)
 21. Ko CW, Dominitz JA, Green P, Kreuter W, Baldwin LM: Specialty differences in polyp detection, removal, and biopsy during colonoscopy. *Am J Med.* 2010, 123:528-35. [10.1016/j.amjmed.2010.01.016](https://doi.org/10.1016/j.amjmed.2010.01.016)
 22. Zorron Cheng Tao Pu L, Lu K, Ovenden A, et al.: Effect of time of day and specialty on polyp detection rates in Australia. *J Gastroenterol Hepatol.* 2019, 34:899-906. [10.1111/jgh.14566](https://doi.org/10.1111/jgh.14566)
 23. Sharma RS, Rossos PG: A review on the quality of colonoscopy reporting. *Can J Gastroenterol Hepatol.* 2016, 2016:9423142. [10.1155/2016/9423142](https://doi.org/10.1155/2016/9423142)
 24. Khan R, Scaffidi MA, Grover SC, Gimpaya N, Walsh CM: Simulation in endoscopy: practical educational strategies to improve learning. *World J Gastrointest Endosc.* 2019, 11:209-18. [10.4253/wjge.v11.i3.209](https://doi.org/10.4253/wjge.v11.i3.209)