Accepted Manuscript

Submission Date: 2024-04-02
Accepted Date: 2024-07-18
Accepted Manuscript online: 2024-07-23

Endoscopy

Polyp size measurement during colonoscopy using a virtual scale: variability and systematic differences

Querijn N van Bokhorst, Britt B Houwen, Yark Hazewinkel, Manon van der Vlugt, Hanneke Beaumont, Joep Grootjans, Arjan van Tilburg, Marlou P Adriaanse, Barbara A Bastiaansen, Yvette H van Beurden, Maxime E Bronzwaer, Brecht W Hens, Lowiek M Hubers, Gem M Kramer, Selma J Lekkerkerker, Berrie Meijer, Fraukje A Ponds, Dewkoemar Ramsoekh, Paul Fockens, Patrick M Bossuyt, Evelien Dekker.

Affiliations below.

DOI: 10.1055/a-2371-3693

Please cite this article as: van Bokhorst Q N, Houwen B B, Hazewinkel Y et al. Polyp size measurement during colonoscopy using a virtual scale: variability and systematic differences. Endoscopy 2024. doi: 10.1055/a-2371-3693

Conflict of Interest: HB received speakers fee from Medtronic; BB received speakers fee from Olympus, Tillotts Pharma AG and Ovesco Endoscopy AG; DR received a research grant from AbbVie (outside the submitted work) and served as a member of the Data Safety Monitoring Board of Vivoryon Therapeutics; PF received research support from Boston Scientific and a consulting feefrom Olympus and Cook Endoscopy; ED received a research grant from Fujifilm, a consulting fee for medical advice from Olympus, Fujifilm, GI Supply, PAION, Ambu and CPP-FAP and a speakers fee from Olympus, Roche, GI Supply, Norgine, Fujifilm and IPSEN. The remaining Authors declare that there is no conflict of interest."

This study was supported by Fujifilm Corporation (http://dx.doi.org/10.13039/501100002424), Unrestricted research grant

Trial registration: NCT05499546, ClinicalTrials.gov (http://www.clinicaltrials.gov/), Video-based study

Abstract:

Background

Accurate polyp size measurement is important for polyp risk stratification and decision-making regarding polypectomy and surveillance. Recently, a virtual scale (VS) function has been developed that allows polyp size measurement through projection of an adaptive VS onto colorectal polyps during real-time endoscopy. We aimed to evaluate the VS in terms of variability and systematic differences.

Methods

We conducted a video-based study with 120 colorectal polyps, measured by eight dedicated colorectal gastroenterologists (experts) and nine gastroenterology residents following endoscopy training (trainees). Three endoscopic measurement methods were compared: (1) visual, (2) snare and (3) VS measurement. We evaluated the method-specific variance (as measure of variability) in polyp size measurements and systematic differences between these methods.

Results

Variance in polyp size measurements was significantly lower for VS measurements compared to visual and snare measurements for both experts (0.52 vs. 1.59 and 1.96, p<0.001) and trainees (0.59 vs. 2.21 and 2.53, p<0.001). VS measurement resulted in a higher percentage of polyps assigned to the same size category by all endoscopists compared to visual and snare

measurements (experts: 69% vs. 55% and 59%; trainees: 67% vs. 51% and 47%) and reduced the maximum difference between individual endoscopists regarding the percentage of polyps assigned to the <u></u>10 mm size category (experts: 1.7% vs. 10.0% and 5.0%; trainees: 2.5% vs. 6.7% and 11.7%). Systematic differences between methods were <0.5 mm.

Conclusions
Use of the VS leads to lower polyp size measurement variability and more uniform polyp sizing by individual endoscopists compared to visual and snare measurements.

Corresponding Author:
Dr. Evelien Dekker, Amsterdam UMC Location VUmc, Department of Gastroenterology and Hepatology, Amsterdam, Netherlands, e.dekker@amsterdamumc.nl

Affiliations:
Querijn N van Bokhorst, Amsterdam UMC Location VUmc, Department of Gastroenterology and Hepatology, Amsterdam, Netherlands
Querijn N van Bokhorst, Amsterdam Gastroenterology Endocrinology Metabolism, Gastroenterology and Hepatology, Amsterdam, Netherlands
Querijn N van Bokhorst, Cancer Center Amsterdam, Gastroenterology and Hepatology, Amsterdam, Netherlands
[...]
Evelien Dekker, Bergman Clinics, Department of Gastroenterology, Amsterdam, Netherlands

Polyp size measurement during colonoscopy using a virtual scale: variability and systematic differences

AUTHORS:

Querijn N.E. van Bokhorst[1,2,3], Britt B.S.L. Houwen[1,2,3], Yark Hazewinkel[4], Manon van der Vlugt[1,2,3,5], Hanneke Beaumont[1,2,3,5], Joep Grootjans[1,3,5,6], Arjan van Tilburg[7], Paul Fockens[1,2,3], Patrick M.M. Bossuyt[8], Evelien Dekker[1,2,3,5] on behalf of the SCALE EYE study group*

*Marlou P.M. Adriaanse[1], Barbara A.J. Bastiaansen[1,2,3], Yvette H. van Beurden[1], Maxime E.S. Bronzwaer[1], Brecht W.E. Hens[1], Lowiek M. Hubers[1], Gem M. Kramer[1], Selma J. Lekkerkerker[1], Berrie Meijer[9], Fraukje A. Ponds[1], Dewkoemar Ramsoekh[1,2,3,5]

AUTHOR'S INSTITUTIONS:

[1] Department of Gastroenterology and Hepatology, Amsterdam UMC Location VUmc, Amsterdam, the Netherlands

[2] Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam, the Netherlands

[3] Cancer Center Amsterdam, Amsterdam, the Netherlands

[4] Department of Gastroenterology and Hepatology, Tergooi Medical Center, Hilversum, the Netherlands

[5] Department of Gastroenterology, Bergman Clinics, Amsterdam, the Netherlands

[6] Oncode Institute, Amsterdam, the Netherlands

[7] Department of Pathology, Reinier de Graaf Gasthuis, Delft, the Netherlands

[8] Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam UMC, Location AMC, Amsterdam, the Netherlands

[9] Department of Gastroenterology and Hepatology, Dijklander Ziekenhuis, Hoorn and Purmerend, the Netherlands

AUTHORS CONTRIBUTION'S:

Study conceptualization and writing of the study protocol: QB, BH, YH, PF, PB, ED; collected the data: all authors; performed the analyses: QB, PB; writing original draft of the manuscript: QB, BH; review of the manuscript: all authors; supervision: PF, PB, ED.

CORRESPONDING AUTHOR:

Evelien Dekker, MD, PhD

Amsterdam UMC Location VUmc

Department of Gastroenterology and Hepatology

De Boelelaan 1117 | 1081 HV Amsterdam | The Netherlands

Tel: +31 (0)20 444 1491 | E-mail: e.dekker@amsterdamumc.nl

ABSTRACT

Background

Accurate polyp size measurement is important for polyp risk stratification and decision-making regarding polypectomy and surveillance. Recently, a virtual scale (VS) function has been developed that allows polyp size measurement through projection of an adaptive VS onto colorectal polyps during real-time endoscopy. We aimed to evaluate the VS in terms of variability and systematic differences.

Methods

We conducted a video-based study with 120 colorectal polyps, measured by eight dedicated colorectal gastroenterologists (experts) and nine gastroenterology residents following endoscopy training (trainees). Three endoscopic measurement methods were compared: (1) visual, (2) snare and (3) VS measurement. We evaluated the method-specific variance (as measure of variability) in polyp size measurements and systematic differences between these methods.

Results

Variance in polyp size measurements was significantly lower for VS measurements compared to visual and snare measurements for both experts (0.52 vs. 1.59 and 1.96, $p<0.001$) and trainees (0.59 vs. 2.21 and 2.53, $p<0.001$). VS measurement resulted in a higher percentage of polyps assigned to the same size category by all endoscopists compared to visual and snare measurements (experts: 69% vs. 55% and 59%; trainees: 67% vs. 51% and 47%) and reduced the maximum difference between individual endoscopists regarding the percentage of polyps assigned to the $\geq$10 mm size category (experts: 1.7% vs. 10.0% and 5.0%; trainees: 2.5% vs. 6.7% and 11.7%). Systematic differences between methods were <0.5 mm.

Conclusions

Use of the VS leads to lower polyp size measurement variability and more uniform polyp sizing by individual endoscopists compared to visual and snare measurements.

INTRODUCTION

Size of colorectal polyps is associated with the risk that a polyp yields advanced histological features [1, 2]. Moreover, a polyp size ≥10 mm is associated with an increased risk of metachronous advanced neoplasia and colorectal cancer (CRC) [3]. Therefore, accurate polyp size measurement, especially at the 10 mm threshold, is important for polyp risk stratification and decision-making regarding colonoscopy surveillance intervals [4, 5]. Polyp size also matters for deciding on optimal resection technique [6, 7] and the safe implementation of the 'leave-in-situ' and 'resect-and-discard' optical diagnosis strategies [8, 9].

In daily practice, polyp size is mostly measured based on visual size estimation by the endoscopist. Visual size estimation is however known to be inaccurate and prone to bias and inter-observer variability [10-20], reportedly leading to inappropriate surveillance recommendations for up to 35% of polyps [10, 12, 19]. Using instruments of known size as a visual reference can improve polyp size measurement accuracy and reduce inter-observer variability, but is known to be time-consuming and costly [15, 18, 21].

Recently, a new virtual scale (VS) function (SCALE EYE; Fujifilm, Tokyo, Japan) was developed that allows polyp size measurement through projection of an adaptive VS onto colorectal polyps during real-time endoscopy [22]. The size of the VS is adapted based on the distance between the endoscope tip and the polyp the endoscopist aims to measure. This distance is calculated using an endoscope-integrated laser in combination with specific computer (image processing) software (**Figure 1**). The laser and the VS can be simultaneously activated with a single push of a button on the handle of the endoscope.

Several (pre-)clinical studies have evaluated the performance of the VS. These studies showed superior measurement accuracy of VS polyp size measurements compared to both visual and instrument-aided polyp size measurements [17-23]. However, the clinical relevance of measurement accuracy as primary outcome measure to evaluate polyp size measurement methods can be debated due to the absence of a robust reference standard for polyp size. To more thoroughly evaluate the clinical potential of the VS, we aimed to evaluate the VS in terms of variability and systematic differences. In addition, we evaluated feasibility of the VS in terms of measurement success rate, measurement duration and user-friendliness.

METHODS

Setting and study design

The video-based study was conducted according to the principles of the Declaration of Helsinki and the Medical Research Involving Human Subjects Act (WMO) and was approved by the Institutional Review Board of the Academic Medical Center, Amsterdam (2022.084). The study was prospectively registered at ClinicalTrials.gov (NCT05499546).

The size of consecutively detected polyps was measured in screening colonoscopies. All measurements were videotaped. Video recording extracts of these measurements were later presented to a group of both expert and trainee endoscopists. All endoscopists estimated the size of each polyp based on each of the three measurement methods: (1) visual estimation (without aid of a tool), (2) 9-mm polypectomy snare as visual reference and (3) VS as a visual reference. Polyps were also measured at histopathological analysis.

Colonoscopy procedures

Consecutive patients undergoing colonoscopy after a positive faecal immunochemical test within the context of the Dutch national CRC screening program at Bergman Clinics, Amsterdam were invited to participate in this study. All participants provided written informed consent. Patients were recruited between October 2022 and March 2023.

Colonoscopies were performed by four endoscopists (E.D., H.B., J.G., M.V.). All colonoscopies were performed using the EC-760S-A/L endoscope (Fujifilm, Tokyo, Japan) in combination with the EX-1 processor with EW10-VM01 software (Fujifilm, Tokyo, Japan), facilitating the use of the VS. Endoscopists had performed at least ten polyp size measurements by VS before start of study inclusions.

Since the upper limit of the VS is 20 mm, only polyps of 20 mm or smaller, according to initial visual size estimation by the endoscopist, were deemed eligible for inclusion. Polyps were consecutively included, considering start of withdrawal within the cecum as starting point. Eligible polyps were endoscopically measured using three methods and in similar order: (1) visual, (2) VS, (3) snare (**Figure 2**, **Video S1**). Measurements were performed using high-definition white light endoscopy.

Visual measurements were performed without the use of a tool as visual reference. For VS measurements, use of either the linear VS, circular VS, or both, was left up to the discretion of the endoscopist. Duration of VS measurements, defined as the time from activation of the VS until the first image freeze with the VS in a feasible position to estimate polyp size (i.e. on the left centre edge of the polyp), was recorded. If no successful measurement could be performed within 180 seconds, the

measurement was recorded as failed. For snare measurements, the Exacto Cold Snare (Steris, Mentor, United States of America) with a maximum width of 9 mm (**Figure S1**) was used as visual reference. The snare was positioned around the polyp while fully extended. If it was not possible to adequately position the snare around the polyp, the snare was positioned adjacent to the polyp. Snare measurements were recorded as failed if the snare could not be fully extended with both the polyp and complete snare clearly visible. All polyps were resected after snare measurement using the Exacto Cold Snare. Resection specimens were retrieved through suctioning and collected using a two-drawer polyp trap (ENDO-SAFIER Polyp Trap; Suzhou GZM Medical Co., Suzhou, China).

Video-based assessment of polyp size

For each polyp, three 10-15 seconds video extracts were derived from the video recordings. Each extract included the measurement of a polyp using one of the three measurement methods. The video extracts were incorporated into an only survey environment (**Figure S2**) and distributed over six surveys. Each survey contained 60 videos. All videos within each survey only showed polyp size measurements using one of the three measurement methods. The order of polyps within each survey was determined using a randomization tool [24].

All video extracts were presented to eight dedicated colorectal gastroenterologists specialized in endoscopic diagnosis and treatment of early CRC and its precursor lesions (experts), as well as nine gastroenterology residents following endoscopy training with between two and four years of endoscopy experience (trainees). Before participation, all participants completed an e-learning regarding use and interpretation of the VS.

For each video, endoscopists were asked to report the size of the polyp displayed within the video. Size was reported in mm and at 1 mm increments. Playback and pausing of videos was allowed, as this mimics (re)positioning of the endoscope and image freezing during clinical procedures. To reduce the risk of recognition bias, intervals of at least two weeks were maintained between surveys containing video extracts of the same polyps. For each polyp, videos were presented to all endoscopists in a standardized order: (1) visual, (2) snare, (3) VS. Endoscopists were also asked to grade the quality of each video for assessment of polyp size either ‘good’, ‘sufficient’ or ‘insufficient’. Endoscopists were blinded for size as determined during clinical assessment and assessments by other endoscopists, and received no information on any results until completion of all surveys.

Histopathological polyp size measurement

Resection specimens of included polyps were collected in formalin and sent to the pathology laboratory. Resection specimens were measured by a dedicated gastrointestinal pathologist using a conventional ruler (macroscopic measurement) and during light microscope examination after embedment in paraffin and sectioning (microscopic measurement). In case the resection specimen was fragmented this was recorded. Sizes were reported in mm at 1 mm increments.

Statistical analysis

Continuous data are presented as mean with standard deviation in case of normally distributed data, and as median with interquartile range (IQR) in case of non-normally distributed data. Categorical data are presented as count with percentages.

To estimate method-specific variance (as measure of variability), as well as systematic differences between methods, we used a mixed linear model (MLM). Our MLM included polyp size as the dependent variable, method-specific intercept as fixed effect, polyp-specific and endoscopist-specific intercepts as random effects, with a random slope to account for method-specific variance across different endoscopists. To test for a significant difference in variance between methods, we used the generalized likelihood ratio test statistic, comparing the model as described to a similar model presuming equal method-specific variances. In case of significant differences in variance, MLM analyses were repeated on datasets comprising data of only two measurement methods. This way, any difference in variance between two specific methods could be evaluated in detail. To meet MLM assumptions of normality and homoscedasticity, log-transformed polyp size measurements were used for all hypothesis tests. To ease interpretation of the results, we report estimated variances and systematic differences without such transformation.

For subgroup analyses involving diminutive and non-diminutive polyps, mean polyp size according to snare measurements by expert endoscopists was used to distribute polyps over different size categories. For subgroup analyses with exclusion of polyps with videos of insufficient quality, polyps for which at least one video was rated as of insufficient quality by six or more ($\geq$35%) endoscopists were excluded.

We assessed uniformity of polyp size classification as the percentage of polyps assigned to the same size category ($\leq$5 mm, 6-9 mm, $\geq$10 mm) by all endoscopists with each endoscopic measurement method. Differences in measured size between the various endoscopic measurement methods were illustrated using Bland-Altman plots [25]. Mean differences between endoscopic measurements and

histopathological measurements were calculated using the mean polyp size according to expert endoscopists for each endoscopic method. Polyps for which macroscopic and microscopic polyp size were not both available, as well as polyps that were resected in piecemeal or for which the histopathological specimen was fragmented, were excluded from these analyses.

All analyses were performed using R version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria). Two-sided P-values <0.05 were considered statistically significant.

Sample size calculation

We based our sample size calculation on measurements by expert endoscopists, anticipating a 20% or larger difference in variance between measurement methods. To reach the desired statistical power of 80%, sample size calculation using a two-sided F-test for difference in variance with an alpha of 0.05 showed that inclusion of a total of 947 polyp measurements per method was required. When involving eight expert endoscopists, this implied inclusion of at least 119 unique polyps. To facilitate a balanced distribution of polyps over the six study surveys, we decided to extent the sample size to 120 polyps.

RESULTS

A total of 120 polyps, detected during screening colonoscopy in 52 patients (median age: 65 years [IQR 60-71], male: 62%), were included (**Table S1, Figure S3**). Success rate for VS measurements was 95%, which was comparable to success rate for snare measurements (97%). Median VS measurement duration was 17 seconds (IQR 8-33) (**Table S2**). Median measurement duration was lower for the last 30 included polyps compared to the first 30 included polyps (12 [IQR 5-23] vs. 22 [IQR 14-44] seconds). Median measurement duration for flat and non-flat polyps was comparable (15 [IQR 6-30] vs. 18 [IQR 10-33] seconds). Five endoscopists who used the VS in clinical practice graded VS endoscope user-friendliness an average at five on a ten points scale (with one representing the worst user-friendliness and ten representing the best user-friendliness).

Video-based assessment of polyp size

Eight experts (median endoscopy experience, years: 13 [IQR 8-21]) and nine trainees (median endoscopy experience, years: 3 [IQR 2-4]) completed all study surveys. For 97/120 (81%) polyps, all three measurement videos were assessed as being of sufficient to good quality. The percentage of videos with insufficient quality was comparable between the three measurement methods (**Table S3**).

Differences in variability for endoscopic measurements methods

Method-specific variances are shown in **Table 1**. Variance for VS measurements was significantly lower than for visual and snare measurements for both experts (0.52 [95% CI: 0.47-0.57] vs. 1.96 [95% CI: 1.88-2.06] and 1.59 [95% CI: 1.50-1.67), p<0.001) and trainees (0.59 [95% CI: 0.54-0.63] vs. 2.21 [95% CI: 2.12-2.30] and 2.53 [95% CI: 2.43-2.62], p<0.001). One-to-one comparisons of method-specific variances showed statistically significant differences in all cases, with exception of comparison of variance for visual versus snare measurements for trainees.

Subgroup analyses were conducted to evaluate variances within specific polyp subgroups. Polyps were subdivided based on polyp size (≤5 mm and >5 mm polyps), morphology (flat and non-flat polyps) and used VS (linear or circular). These analyses showed similar results compared to the main analyses: variance for VS measurements was lower than for visual and snare measurements in all cases, for both experts and trainees (**Table S4-S6**). When excluding polyps with videos of insufficient quality, 97 polyps remained. Analyses involving this subgroup also revealed lower variance for VS measurements compared to visual and snare measurements in both endoscopist groups (**Table S7**).

Mean and systematic differences between measurement methods

Mean differences in polyp size between the three endoscopic measurement methods, based on assessments by expert endoscopists, are illustrated in Bland-Altman plots (**Figure 3A-C**). These plots indicate that, on average, snare measurements resulted in the largest polyp size, respectively followed by VS and visual measurements. Systematic differences between methods, as estimated using MLM analyses, are shown in **Table 2**. For expert endoscopists, VS measurements generally resulted in larger polyp size compared to visual measurements (+0.11 mm [95% CI: 0.00-0.23]) and smaller polyp size compared to snare measurements (-0.09 mm [95% CI: -0.21-0.02]).

Mean differences between endoscopic and histopathological measurement methods, based on the analysis including 71 non-fragmented polyps with both macroscopic and microscopic histopathological measurement available, are shown within **Table S8**.

Uniformity of polyp size classification for the different measurement methods

Uniformity of polyp size classification was assessed through distribution of polyps over three size categories (≤5 mm, 6-9 mm, ≥10 mm) based on the assessments by the individual endoscopists. The percentage of polyps assigned to the same size category by all endoscopists was higher for VS

measurements compared to visual and snare measurements for both experts (69 vs. 55 and 59%), trainees (67 vs. 51 and 47%) and all endoscopists (58 vs. 48 and 43%) (**Figure 4A, Table S9**).

Use of the VS resulted in more uniform decision making around the 10 mm threshold for both experts and trainees, as on the level of the individual endoscopist the maximum differences regarding the total number of polyps assigned to the ≥10 mm size category were lower for VS measurements compared to visual and snare measurements. For experts, based on analyses with 120 included polyps, a maximum difference regarding the number of polyps assigned to the ≥10 mm size category of 2 (1.7%) was found for VS measurement, while this concerned differences of 12 (10%) polyps for visual measurement and 6 (5.0%) polyps for snare measurement. For trainees, the maximum differences for VS, visual and snare measurement respectively concerned 3 (2.5%), 8 (6.7%) and 14 (12%) polyps (**Figure 4B**). A more detailed overview of polyp distribution over the different size categories by individual endoscopists is shown in **Table S10**.

DISCUSSION

This video-based study showed that use of the VS results in lower polyp size measurement variability compared to visual and snare measurement, for both endoscopy experts and trainees. This resulted in more uniform assignment of polyps to different size categories by individual endoscopists and a reduction in the differences in the number of polyps assigned to the ≥10 mm size category. Moreover, estimated systematic differences for VS measurements compared to other methods were small (<0.5 mm).

Most studies evaluating polyp size measurement methods use measurement accuracy as the primary outcome measure. However, the clinical value of measurement accuracy is compromised by the lack of a robust reference standard for polyp size. Frequently used reference standards are measured size of (fresh) resection specimen [12, 18, 19] or size as endoscopically determined by in-situ comparison of polyp size to size of an instrument of known size [13-15, 26, 27]. A limitation of the use of resection specimen size as reference standard is that resection specimens can get fragmented or lost. Moreover, the size of resection specimens might not represent real polyp size since these are prone to deformation due to vascular collapse, compression, cauterization and suction [28], as well as shrinkage due to formalin fixation [29-31]. Instrument-aided measurements can be erroneous due to problems with adequate positioning of instruments with the largest diameter in a perpendicular view towards the polyp. Instrument-aided measurements are also prone to bias due to changed instrument proportions due to mechanic factors (e.g. compression, deformation) and distortion or warping of instruments on

the endoscopy monitor due to the endoscope fisheye lens structure [32]. Our study illustrates that the latter may especially affect less experienced endoscopists, while variability for snare measurements exceeded variability for visual measurements for trainees. Because of the drawbacks of comparing accuracies based on suboptimal reference standards, we chose to evaluate the different methods for polyp size measurement in terms of variability and systematic differences.

Lower measurement variability indicates that results of repeated measurements are concentrated closer around their mean. Use of measurement methods associated with lower measurement variability, such as the VS in our study, will therefore result in a reduction of significant outliers and a lower rate of observer disagreement compared to measurement methods associated with higher measurement variability [33]. In our study, lower measurement variability for VS measurements resulted in an increase in uniformity of polyp size category assignment up to 20% compared to the other methods. The clinical relevance of the more uniform polyp sizing using the VS is more specifically illustrated through the reduction in the percentual difference in the total number of polyps assigned to the ≥10 mm size category between individual endoscopists (i.e. experts: 10.0% for visual to 1.7% for VS; trainees: 11.7% for snare to 2.5% for VS). Sizing a polyp either under or over the 10 mm threshold may result in defining an adenoma as advanced (≥10 mm) or non-advanced (<10 mm), which in most guidelines implies a difference between either a 3- or 10-year surveillance [4, 5]. As such, more consistent polyp sizing around the 10 mm threshold using the VS could aid in preventing unnecessary (early) colonoscopies and erroneously delayed surveillance.

Despite benefits of the VS in terms of measurement variability, this does not directly imply that VS measurements also yield a high measurement accuracy. The small systematic differences (<0.5 mm) as found in our study, in combination with results of preliminary pre-clinical studies evaluating performance of the VS on artificial polyps (maximum measurement errors of 0.7 mm and relative accuracies of 82-84%) [17, 20-23], do however support the idea that VS measurements generally provide reliable estimates of real polyp size.

General benefits of the VS include the fact that it is an intuitive push-a-button tool, which is easy to use and does not require additional (disposable) instruments. In addition, as shown our study, VS measurement is possible for the vast majority of polyps and can be performed in under a minute for 90% of polyps. Although we did not compare VS and snare measurement duration, measurement duration of VS is suggested to be comparable to other instrument-aided measurements in an ex-vivo setting [21]. Nonetheless, there are also several factors compromising usability of the VS. In the first place, manoeuvring the VS (endoscope) in a perpendicular position towards the polyp can be a tedious

process and is not always possible, especially for polyps located in difficult positions (e.g. within folds). This is the main reason endoscopists graded user-friendliness an average at five on a ten points scale. Moreover, the VS only contains markings at 5, 10 and 20 mm, which hampers measurement of larger polyps. Lastly, VS measurements still require interpretation by physicians, thereby not completely ruling out bias due to inter-observer variability.

Towards the future, feasibility of the VS to serve as an alternative reference standard for clinical polyp size measurement should be evaluated considering its current limitations. This is particularly relevant while artificial intelligence (AI) has already been proposed as an alternative for automated polyp size measurement. AI might facilitate polyp size measurements without bias due to human factors. Nonetheless, the lack of robust datasets with ground truth information is currently still complicating development of AI-based polyp measurement systems [34, 35]. In the context of lower measurement variability and proven measurement accuracy of the VS on artificial polyps (of which exact size is known), the VS might however open doors to development of certain datasets and further development of AI-based approaches.

To the best of our knowledge, this is the first study involving the VS to evaluate multiple endoscopic polyp size measurement methods in terms of variability and systematic differences. Due to the prospective inclusion of consecutive polyps of different morphological subtypes and sizes in a real-time clinical setting, our study provides realistic insights into the clinical potential of the VS. Besides, both endoscopy experts and trainees participated in this study. To prevent recognition bias, we maintained intervals of at least two weeks between assessments of different measurements of the same polyps and applied video sequence randomization within surveys.

The generalizability of our study may however be compromised by the fact that the majority of polyps were $\leq$10 mm, and by the fact that participating trainee endoscopists had two to four years of endoscopy experience. Therefore, future studies involving more polyps over 10 mm in size and studies exploring the specific benefits of the VS for novice endoscopists are still warranted. Moreover, while four endoscopists performed all study measurements, parameters such as measurement duration were largely dependent on the endoscopy skills of these endoscopists. In the last place, polyp measurements were conducted based on video extracts with a maximum length of 15 seconds. Although assessments based on videos will likely be more accurate than assessments based on still images, the use of such short video extracts will always induce some bias. Nonetheless, analyses with exclusion of polyps with videos of insufficient quality showed results comparable to the results of the main analyses. This indicates that lower video quality was not the ground cause for observed differences in variability.

In conclusion, this study showed that use of the VS leads to lower polyp size measurement variability and more uniform polyp sizing by individual endoscopists compared to other measurement methods. Therefore, use of the VS in daily clinical practice could reduce the risk of polyp under- or oversizing and polyp misclassification at relevant size thresholds. This could support better clinical decision-making processes involving polyp size.

REFERENCES

1. Hassan C, Pickhardt PJ, Kim DH et al. Systematic review: distribution of advanced neoplasia according to polyp size at screening colonoscopy. Aliment Pharmacol Ther 2010; 31: 210-217. DOI: 10.1111/j.1365-2036.2009.04160.x
2. Lieberman D, Moravec M, Holub J et al. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. Gastroenterology 2008; 135: 1100-1105. DOI: 10.1053/j.gastro.2008.06.083
3. He X, Hang D, Wu K et al. Long-term Risk of Colorectal Cancer After Removal of Conventional Adenomas and Serrated Polyps. Gastroenterology 2020; 158: 852-861 e854. DOI: 10.1053/j.gastro.2019.06.039
4. Hassan C, Antonelli G, Dumonceau JM et al. Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2020. Endoscopy 2020; 52: 687-700. DOI: 10.1055/a-1185-3109
5. Gupta S, Lieberman D, Anderson JC et al. Recommendations for Follow-Up After Colonoscopy and Polypectomy: A Consensus Update by the US Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol 2020; 115: 415-434. DOI: 10.14309/ajg.0000000000000544
6. Ferlitsch M, Moss A, Hassan C et al. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. Endoscopy 2017; 49: 270-297. DOI: 10.1055/s-0043-102569
7. Kaltenbach T, Anderson JC, Burke CA et al. Endoscopic Removal of Colorectal Lesions: Recommendations by the US Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol 2020; 115: 435-464. DOI: 10.14309/ajg.0000000000000555
8. Rex DK, Kahi C, O'Brien M et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc 2011; 73: 419-422. DOI: 10.1016/j.gie.2011.01.023
9. Houwen B, Hassan C, Coupé VMH et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2022; 54: 88-99. DOI: 10.1055/a-1689-5130
10. Moug SJ, Vernall N, Saldanha J et al. Endoscopists' estimation of size should not determine surveillance of colonic polyps. Colorectal Dis 2010; 12: 646-650. DOI: 10.1111/j.1463-1318.2009.01870.x
11. de Vries AH, Bipat S, Dekker E et al. Polyp measurement based on CT colonography and colonoscopy: variability and systematic differences. Eur Radiol 2010; 20: 1404-1413. DOI: 10.1007/s00330-009-1683-0
12. Eichenseer PJ, Dhanekula R, Jakate S et al. Endoscopic mis-sizing of polyps changes colorectal cancer surveillance recommendations. Dis Colon Rectum 2013; 56: 315-321. DOI: 10.1097/DCR.0b013e31826dd138
13. Chaptini L, Chaaya A, Depalma F et al. Variation in polyp size estimation among endoscopists and impact on surveillance intervals. Gastrointest Endosc 2014; 80: 652-659. DOI: 10.1016/j.gie.2014.01.053
14. Anderson BW, Smyrk TC, Anderson KS et al. Endoscopic overestimation of colorectal polyp size. Gastrointest Endosc 2016; 83: 201-208. DOI: 10.1016/j.gie.2015.06.058
15. Kim JH, Park SJ, Lee JH et al. Is forceps more useful than visualization for measurement of colon polyp size? World J Gastroenterol 2016; 22: 3220-3226. DOI: 10.3748/wjg.v22.i11.3220
16. Sakata S, Klein K, Stevenson ARL, Hewett DG. Measurement Bias of Polyp Size at Colonoscopy. Dis Colon Rectum 2017; 60: 987-991. DOI: 10.1097/DCR.0000000000000859

17. Haumesser C, Zarandi-Nowroozi M, Taghiakbari M et al. Comparing size measurements of simulated colorectal polyp size and morphology groups when using a virtual scale endoscope or visual size estimation: Blinded randomized controlled trial. Dig Endosc 2022; 35: 638-644. DOI: 10.1111/den.14498
18. von Renteln D, Djinbachian R, Zarandi-Nowroozi M, Taghiakbari M. Measuring size of smaller colorectal polyps using a virtual scale function during endoscopies. Gut 2023; 72: 417-420. DOI: 10.1136/gutjnl-2022-328654
19. Djinbachian R, Crainic IP, Pioche M et al. Accuracy in polyp size measurement among surgeons, gastroenterologists, trainees and experts: A prospective video-based study. Am J Gastroenterol 2023. DOI: 10.14309/ajg.0000000000002494
20. Taghiakbari M, Djinbachian R, Haumesser C et al. Measuring size of colorectal polyps using a virtual scale endoscope or visual assessment: A randomized controlled trial. Am J Gastroenterol 2023. DOI: 10.14309/ajg.0000000000002623
21. Djinbachian R, Taghiakbari M, Haumesser C et al. Comparing size measurement of colorectal polyps using a novel virtual scale endoscope, endoscopic ruler or forceps: A preclinical randomized trial. Endosc Int Open 2023; 11: E128-E135. DOI: 10.1055/a-2005-7548
22. Yoshioka M, Sakaguchi Y, Utsunomiya D et al. Virtual scale function of gastrointestinal endoscopy for accurate polyp size estimation in real-time: a preliminary study. J Biomed Opt 2021; 26. DOI: 10.1117/1.JBO.26.9.096002
23. Shimoda R, Akutagawa T, Tomonaga M et al. Estimating colorectal polyp size with a virtual scale endoscope and visual estimation during colonoscopy: Prospective, preliminary comparison of accuracy. Dig Endosc 2022; 34: 1471-1477. DOI: 10.1111/den.14351
24. Urbaniak GC, Plous, S. Research Randomizer (Version 4.0).
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307-310.
26. Rex DK, Rabinovitz R. Variable interpretation of polyp size by using open forceps by experienced colonoscopists. Gastrointest Endosc 2014; 79: 402-407. DOI: 10.1016/j.gie.2013.08.030
27. Han SK, Kim H, Kim JW et al. Usefulness of a Colonoscopy Cap with an External Grid for the Measurement of Small-Sized Colorectal Polyps: A Prospective Randomized Trial. J Clin Med 2021; 10: 2365. DOI: 10.3390/jcm10112365
28. Morales TG, Sampliner RE, Garewal HS et al. The difference in colon polyp size before and after removal. Gastrointest Endosc 1996; 43: 25-28. DOI: 10.1016/s0016-5107(96)70255-9
29. Turner JK, Wright M, Morgan M et al. A prospective study of the accuracy and concordance between in-situ and postfixation measurements of colorectal polyp size and their potential impact upon surveillance. Eur J Gastroenterol Hepatol 2013; 25: 562-567. DOI: 10.1097/MEG.0b013e32835d1f2d
30. Tran T, Sundaram CP, Bahler CD et al. Correcting the Shrinkage Effects of Formalin Fixation and Tissue Processing for Renal Tumors: toward Standardization of Pathological Reporting of Tumor Size. J Cancer 2015; 6: 759-766. DOI: 10.7150/jca.12094
31. Lam D, Kaneko Y, Scarlett A et al. The Effect of Formalin Fixation on Resection Margins in Colorectal Cancer. International Journal of Surgical Pathology 2019; 27: 700-705. DOI: 10.1177/1066896919854159
32. Sakata S, McIvor F, Klein K et al. Measurement of polyp size at colonoscopy: a proof-of-concept simulation study to address technology bias. Gut 2018; 67: 206-208. DOI: 10.1136/gutjnl-2016-312915
33. Quinn MF. Relation of observer agreement to accuracy according to a two-receiver signal detection model of diagnosis. Med Decis Making 1989; 9: 196-206. DOI: 10.1177/0272989x8900900308

34. van Bokhorst QNE, Houwen BBSL, Hazewinkel Y et al. Advances in artificial intelligence and computer science for computer-aided diagnosis of colorectal polyps: current status. Endosc Int Open 2023; 11: E752-E767. DOI: 10.1055/a-2098-1999
35. Yu H, Wang J, Li Y et al. A real-time deep learning-based system for colorectal polyp size estimation by white-light endoscopy: development and multicenter prospective validation. Endoscopy 2023. DOI: 10.1055/a-2189-7036

FIGURE LEGEND

Figure 1. Schematic overview of the principle of distance estimation using the virtual scale endoscope. The reflection of the laser beam, as emitted from the endoscope, can be used to determine distance between the endoscope tip and object (mucosal wall or polyp) the laser is positioned on based on the triangulation method. This information is used to continuously adapt the size of the virtual scale: the size of the virtual scale increases whenever the mucosal wall (or polyp) comes closer, and decreases whenever the distance is shrinking. To enable real-time polyp size measurement, the virtual scale contains markings at 5, 10 and 20 mm.

Figure 2. Endoscopic polyp size measurement methods: (A) visual measurement (without aid of a tool), (B) measurement with aid of the linear (left) and circular (right) virtual scale, (C) measurement with aid of a polypectomy snare of known size (maximum width of 9 mm).

Figure 3. Bland-Altman plots illustrating the differences between polyp size measurements by expert endoscopists using different measurement methods. Within these plots, the polyp size according to two different measurements methods (X-axis) is plotted against the difference in polyp size according to these methods (Y-axis). Each plot comprises 960 observations, representing measurements of 120 polyps by eight different expert endoscopists. Count for each point within the plot is indicated by the legend on the right side. Dotted blue lines represent mean differences. Dotted red lines represent upper and lower limits of the 95% confidence intervals. The plots represent the following methods (Y-axis): (A) visual and snare measurements, (B) visual and virtual scale measurements, (C) snare and virtual scale measurements.

Figure 4. Bar plots illustrating (A) the percentage of polyps assigned to the same category (≤5 mm, 6-9 mm, ≥10 mm) by all endoscopists using different polyp size measurements methods and (B) the maximum difference between individual endoscopists regarding the percentage of polyps assigned to the ≥10 mm size category.

Figure S1. Size of the Exacto Cold Snare (in mm).

Figure S2. Example of the survey environment: (A) video-based assessment of polyp size, (B) assessment of video quality for assessment of polyp size.

Figure S3. Study flow chart.

VSE, virtual scale endoscope; VS, virtual scale, HD WLE; high-definition white light endoscopy; *Only one endoscopy suite with the required study equipment was available; †All endoscopist had to perform at least 10 polyp size measurements by virtual scale (‘training phase’) before start of study inclusions (‘inclusion phase’); ‡Measurements with a snare other than the dedicated study snare (Exacto Cold Snare) were excluded from the calculation.

VIDEO LEGEND

Video S1. Examples of endoscopic polyp size measurements using different measurement methods. In consecutive order: (1) visual measurement (without aid of a tool), (2) measurement with aid of the linear and circular virtual scale, (3) measurement with aid of a polypectomy snare of known size (maximum width of 9 mm).

SUPPLEMENTARY MATERIAL

Authors:

Querijn N.E. van Bokhorst, Britt B.S.L. Houwen, Yark Hazewinkel, Manon van der Vlugt, Hanneke Beaumont, Joep Grootjans, Arjan van Tilburg, Paul Fockens, Patrick M.M. Bossuyt, Evelien Dekker on behalf of the SCALE EYE study group*

*Marlou P.M. Adriaanse, Barbara A.J. Bastiaansen, Yvette H. van Beurden, Maxime E.S. Bronzwaer, Brecht W.E. Hens, Lowiek M. Hubers, Gem M. Kramer, Selma J. Lekkerkerker, Berrie Meijer, Fraukje A. Ponds, Dewkoemar Ramsoekh

Manuscript title:

Polyp size measurement during colonoscopy using a virtual scale: variability and systematic differences

Table S1. Polyp characteristics

Characteristic	Polyps, n (%)
Size*	
≤5 mm	83 (69)
6-9 mm	34 (28)
≥10 mm	3 (2.5)
Location†	
Right-sided colon	51 (43)
Transverse colon	23 (19)
Left-sided colon	46 (38)
Morphology according to Paris classification‡	
Ip	7 (5.8)
Ips	1 (0.8)
Is	52 (43)
0-IIa	56 (47)
0-IIb	4 (3.3)
Histology	
Adenoma	77 (64)
Sessile serrated lesion	6 (5.0)
Hyperplastic polyp	30 (25)
Other	7 (5.8)

*Based on mean polyp size according to snare measurements by expert endoscopists; †Right-sided colon consists of the cecum, ascending colon and hepatic flexure, the left-sided colon consists of the splenic flexure, descending colon, sigmoid and rectum; ‡Type 0-IIa and 0-IIb polyps were considered flat polyps and type Ip, Ips and Is polyps were considered non-flat polyps;

Table S2. Virtual scale measurement duration

Characteristic	Polyps, n (%)
Duration virtual scale measurement in seconds, median (IQR)	17 (8, 33)
Duration virtual scale measurement in seconds	
<10 seconds	36 (30)
10-20 seconds	30 (25)
20-40 seconds	34 (28)
40-60 seconds	9 (7.5)
60-120 seconds	9 (7.5)
120-180 seconds	2 (1.7)

Table S3. Video quality of video extracts according to assessment by different endoscopist groups

		Good, n (%)	Sufficient, n (%)	Insufficient, n (%)
Experts	Visual	507 (53)	364 (38)	89 (9.3)
	Snare	584 (61)	295 (31)	81 (8.4)
	Virtual scale	480 (50)	399 (42)	81 (8.4)
Trainees	Visual	669 (62)	332 (31)	79 (7.3)
	Snare	625 (58)	359 (33)	96 (8.9)
	Virtual scale	647 (60)	353 (33)	80 (7.4)
All	Visual	1176 (58)	696 (34)	168 (8.2)
	Snare	1209 (59)	654 (32)	177 (8.7)
	Virtual scale	1127 (55)	752 (37)	161 (7.9)

Notes: a total of 23 polyps had one or more video fragments of insufficient quality according to more than six endoscopists (≥35%). Video(s) of insufficient quality concerned the visual measurement video for seven polyps, the snare measurement video for six polyps, the virtual scale measurement video for six polyps, both the visual and virtual scale measurement video for two polyps, both the snare and virtual scale measurement video for one polyp and all three measurement videos for one polyp.

Table S4. Variance for different endoscopic polyp size measurement method as estimated using mixed linear models – analyses for different polyp size groups

			Variance (95% CI)*		
Size group	**Endoscopist group**	**Measurements per method**	**Visual**	**Snare**	**Virtual scale**
≤5 mm	Experts	664	1.01 (0.94, 1.09)	0.75 (0.68, 0.81)	0.46 (0.41, 0.51)
	Trainees	747	1.01 (0.94, 1.09)	1.20 (1.12, 1.28)	0.59 (0.54, 0.65)
	All	1411	1.00 (0.95, 1.05)	0.98 (0.93, 1.03)	0.54 (0.49, 0.56)
>5 mm	Experts	296	4.25 (4.01, 4.48)	3.23 (3.03, 3.44)	0.78 (0.68, 0.88)
	Trainees	333	5.28 (5.03, 5.52)	5.43 (5.18, 5.68)	0.84 (0.61, 0.79)
	All	629	4.68 (4.51, 4.85)	4.29 (4.13, 4.46)	0.77 (0.70, 0.84)

CI, confidence interval; *Variances are based on the squared deviations from the mean and are therefore reported in the square of the units of the original data (mm^2). Standard deviations, expressed in the original units of the polyp size measurements (mm), can be calculated by taking the square root of the reported variances.

Notes: analyses including 83 diminutive (≤5 mm) polyps and 37 non-diminutive (>5 mm) polyps.

Table S5. Variance for different endoscopic polyp size measurement method as estimated using mixed linear models – analyses for different polyp morphology groups

Morphology group	Endoscopist group	Measurements per method	Variance (95% CI)*		
			Visual	Snare	Virtual scale
Flat	Experts	480	2.07 (1.95, 2.20)	1.81 (1.69, 1.93)	0.20 (0.55, 0.69)
	Trainees	540	1.79 (1.68, 1.91)	2.76 (2.62, 2.90)	0.86 (0.79, 0.94)
	All	1020	1.89 (1.81, 1.98)	2.25 (2.16, 2.35)	0.76 (0.71, 0.81)
Non-flat	Experts	480	1.86 (1.73, 1.98)	1.35 (1.25, 1.46)	0.43 (0.37, 0.49)
	Trainees	540	2.64 (2.51, 2.78)	2.25 (2.12, 2.38)	0.38 (0.33, 0.43)
	All	1020	2.24 (2.15, 2.33)	1.77 (1.69, 1.85)	0.42 (0.38, 0.46)

CI, confidence interval; VS, virtual scale; *Variances are based on the squared deviations from the mean and are therefore reported in the square of the units of the original data (mm^2). Standard deviations, expressed in the original units of the polyp size measurements (mm), can be calculated by taking the square root of the reported variances.

Notes: based on the Paris classification, type 0-IIa (n = 56) and type 0-IIb (n = 4) polyps were considered flat polyps and type Ip (n = 7), Ips (n = 1) and Is (n = 52) polyps were considered non-flat polyps.

Table S6. Variance for different endoscopic polyp size measurement method as estimated using mixed linear models – analyses for polyps measured using different variants of the virtual scale

VS group	Endoscopist group	Measurements per method	Variance (95% CI)*		
			Visual	Snare	Virtual scale
Linear	Experts	472	1.81 (1.68, 1.93)	1.48 (1.37, 1.59)	0.50 (0.43, 0.56)
	Trainees	531	1.98 (1.86, 2.10)	2.59 (2.45, 2.73)	0.59 (0.52, 0.66)
	All	1003	1.84 (1.76, 1.92)	1.92 (1.84, 2.01)	0.61 (0.56, 0.66)
Circular	Experts	346	2.50 (2.34, 2.66)	1.46 (1.34, 1.59)	0.59 (0.52, 0.67)
	Trainees	423	2.81 (2.65, 2.97)	2.46 (2.31, 2.61)	0.69 (0.61, 0.77)
	All	799	2.62 (2.51, 2.73)	2.00 (1.90, 2.10)	0.64 (0.58, 0.69)

CI, confidence interval; VS, virtual scale; *Variances are based on the squared deviations from the mean and are therefore reported in the square of the units of the original data (mm^2). Standard deviations, expressed in the original units of the polyp size measurements (mm), can be calculated by taking the square root of the reported variances.

Notes: analyses including 57 polyps for which the measurement was performed using the linear VS and 47 for which the primary measurement was performed using the circular VS. Polyps that were measured using both the linear and circular VS (n = 16) were excluded from the analyses.

Table S7. Variance for different endoscopic polyp size measurement methods as estimated by mixed linear model analyses – analyses with exclusion of polyps with at least one video of insufficient quality

Endoscopist group	Measurements per method	Variance (95% CI)*		
		Visual	Snare	Virtual scale
Experts	784	1.70 (1.61, 1.80)	1.47 (1.39, 1.56)	0.52 (0.47, 0.57)
Trainees	882	1.89 (1.80, 1.98)	2.38 (2.28, 2.49)	0.55 (0.50, 0.59)
All	1666	1.75 (1.69, 1.82)	1.88 (1.82, 1.95)	0.56 (0.53, 0.60)

CI, confidence interval; *Variances are based on the squared deviations from the mean and are therefore reported in the square of the units of the original data (mm^2). Standard deviations, expressed in the original units of the polyp size measurements (mm), can be calculated by taking the square root of the reported variances.

Notes: analyses including 97 polyps.

Table S8. Mean differences between histopathological polyp size measurements and endoscopic polyp size measurements methods

Method one	Method two	Mean difference (95% CI)*
Macroscopic	Microscopic	-0.35 (-2.99, 2.28)
Macroscopic	Visual	-0.13 (-4.46, 4.19)
	Snare	-0.07 (-4.47, 4.34)
	Virtual scale	+0.01 (-3.77, 3.76)
Microscopic	Visual	+0.22 (-3.57, 4.01)
	Snare	+0.29 (-3.06, 3.64)
	Virtual scale	+0.35 (-2.70, 3.39)

CI, confidence interval; *Mean difference of method two compared to method one (in mm), based on mean polyp size of assessments by expert endoscopists.

Notes: macroscopic histopathological size was available for 115/120 (96%) polyps. Microscopic histopathological polyp size was available for 110/120 (92%) polyps. The resection specimen of 43/120 (41%) polyps was fragmented. Differences as shown within this table are based on analysis including 71 non-fragmented polyps with both macroscopic and microscopic measurements available.

Table S9. Overview of polyp size category assignment according to different endoscopist groups and measurement methods

Group	Method	Size categories					
		Only ≤5 mm, n (%)	≤5 and 6-9 mm, n (%)	Only 6-9 mm, n (%)	6-9 and ≥10 mm, n (%)	Only ≥10 mm, n (%)	All, n (%)
Experts	Visual	66 (55)	36 (30)	0 (0)	11 (9.2)	0 (0)	7 (5.8)
	Snare	61 (51)	37 (31)	8 (6.7)	9 (7.5)	2 (1.7)	3 (2.5)
	Virtual scale	71 (59)	32 (27)	7 (5.8)	5 (4.2)	5 (4.2)	0 (0.0)
Trainees	Visual	60 (50)	39 (33)	0 (0)	10 (8.3)	1 (0.8)	10 (8.3)
	Snare	53 (44)	44 (37)	2 (1.7)	11 (9.2)	1 (0.8)	9 (7.5)
	Virtual scale	64 (53)	36 (30)	10 (8.3)	3 (2.5)	6 (5.0)	1 (0.8)
All	Visual	58 (48)	38 (32)	0 (0.0)	8 (6.7)	0 (0.0)	16 (13)
	Snare	49 (41)	48 (40)	2 (1.6)	9 (7.5)	1 (0.8)	11 (9.2)
	Virtual scale	59 (49)	45 (38)	5 (4.2)	4 (3.3)	5 (4.2)	2 (1.7)

Notes: a polyp for which a size ≤5 mm was reported by all endoscopists, is assigned to the '≤5 mm' group. If at least one of the measurements would have been 6-9 mm, the polyp would have been assigned to the '≤5 and 6-9 mm' group, etcetera.

Table S10. Number and percentage of polyps in each size category according to assessments of individual endoscopists using different polyp size measurement methods

	Visual			Snare			Virtual scale		
	≤5 mm, n (%)	6-9 mm, n (%)	≥10 mm, n (%)	≤5 mm, n (%)	6-9 mm, n (%)	≥10 mm, n (%)	≤5 mm, n (%)	6-9 mm, n (%)	≥10 mm, n (%)
					Experts				
Expert 1	83 (69)	27 (23)	10 (8.3)	83 (69)	34 (28)	3 (2.5)	88 (73)	26 (22)	6 (5.0)
Expert 2	91 (76)	26 (22)	3 (2.5)	84 (70)	32 (27)	4 (3.3)	93 (78)	22 (18)	5 (4.2)
Expert 3	86 (72)	25 (21)	9 (7.5)	83 (69)	33 (28)	4 (3.3)	88 (73)	25 (21)	7 (5.8)
Expert 4	84 (70)	26 (22)	10 (8.3)	79 (66)	37 (31)	4 (3.3)	88 (73)	25 (21)	7 (5.8)
Expert 5	101 (84)	14 (12)	5 (4.1)	83 (69)	31 (26)	6 (5.0)	85 (71)	28 (23)	7 (5.8)
Expert 6	89 (74)	26 (22)	5 (4.1)	79 (66)	33 (28)	8 (6.7)	81 (68)	32 (27)	7 (5.8)
Expert 7	101 (84)	15 (13)	4 (3.3)	85 (71)	31 (26)	4 (3.3)	96 (80)	18 (15)	6 (5.0)
Expert 8	81 (68)	24 (20)	15 (13)	84 (70)	27 (23)	9 (7.5)	90 (75)	23 (19)	7 (5.8)
Maximum difference	**20 (17)**	**13 (11)**	**12 (10)**	**6 (5.0)**	**10 (9.3)**	**6 (5.0)**	**15 (13)**	**14 (12)**	**2 (1.7)**
					Trainees				
Trainee 1	77 (64)	34 (28)	9 (7.5)	91 (76)	27 (23)	2 (1.7)	89 (74)	25 (21)	6 (5.0)
Trainee 2	84 (70)	28 (23)	8 (6.7)	70 (58)	41 (34)	9 (7.5)	79 (66)	33 (28)	8 (6.7)
Trainee 3	94 (78)	18 (15)	8 (6.7)	99 (83)	13 (11)	8 (6.7)	92 (77)	21 (18)	7 (5.8)
Trainee 4	81 (68)	29 (24)	10 (8.3)	69 (58)	42 (35)	9 (7.5)	86 (72)	27 (23)	7 (5.8)
Trainee 5	90 (75)	21 (18)	9 (7.5)	90 (75)	28 (23)	2 (1.7)	87 (73)	26 (22)	7 (5.8)
Trainee 6	85 (71)	22 (18)	13 (11)	72 (60)	32 (23)	16 (13)	79 (66)	32 (27)	9 (7.5)
Trainee 7	91 (76)	24 (20)	5 (4.2)	75 (63)	38 (32)	7 (5.8)	86 (72)	26 (22)	8 (6.7)
Trainee 8	91 (76)	21 (18)	8 (6.7)	87 (73)	20 (17)	13 (11)	92 (77)	20 (17)	8 (6.7)
Trainee 9	81 (68)	29 (24)	10 (8.3)	78 (65)	35 (29)	7 (5.8)	84 (70)	28 (23)	8 (6.7)
Maximum difference	**17 (14)**	**16 (13)**	**8 (6.7)**	**29 (24)**	**29 (24)**	**14 (12)**	**13 (11)**	**13 (11)**	**3 (2.5)**

Notes: for each size category (≤5 mm, 6-9 mm, ≥10 mm) and endoscopist category (expert or trainee) the highest number of polyps assigned to the size category is indicated in green and the lowest number is indicated in red.

Table 1. Variance for different endoscopic polyp size measurement methods as estimated using mixed linear model analyses

Group	Number of measurements per method	Variance (95% CI)*		
		Visual	Snare	Virtual scale
Experts	960	1.96 (1.88, 2.06)	1.59 (1.50, 1.67)	0.52 (0.47, 0.57)
Trainees	1080	2.21 (2.12, 2.30)	2.53 (2.43, 2.62)	0.59 (0.54, 0.63)
All	2040	2.06 (1.99, 2.12)	2.02 (1.96, 2.09)	0.58 (0.52, 0.64)

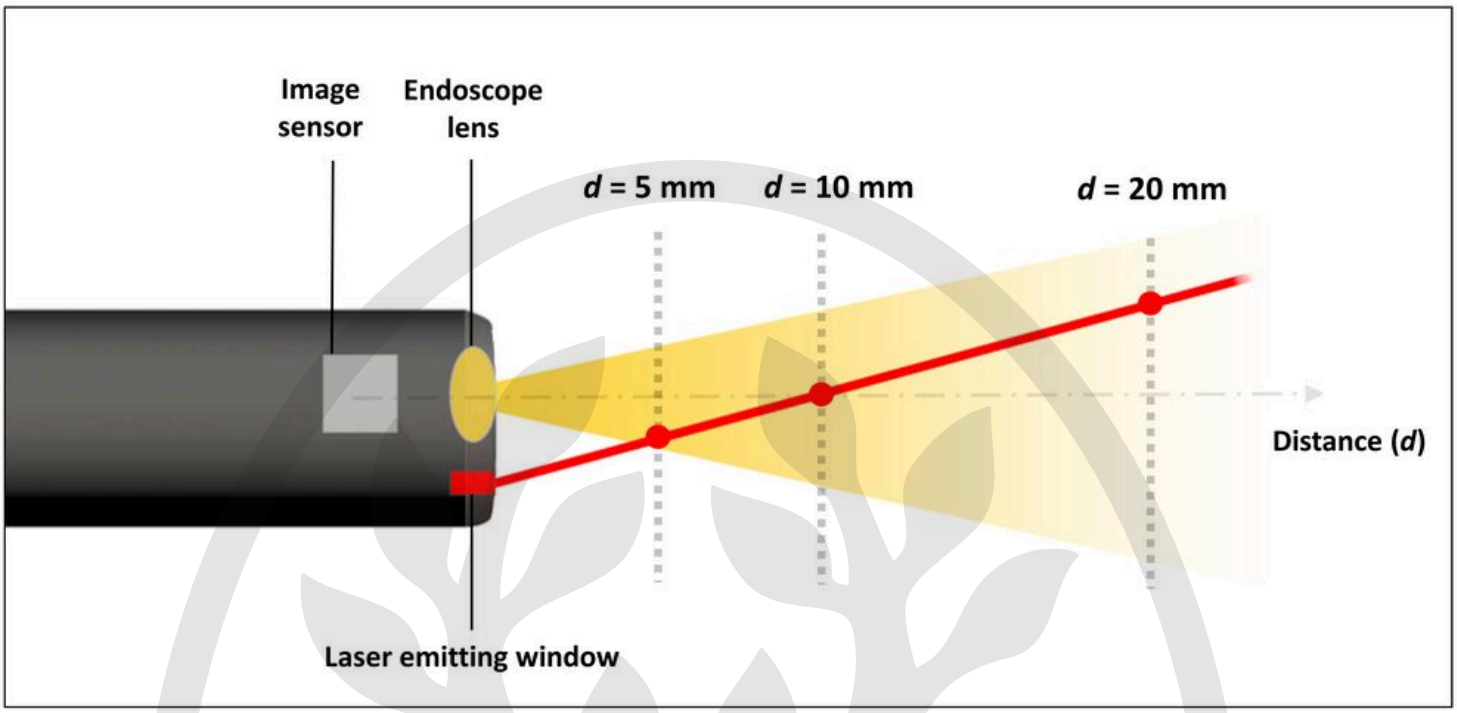
CI, confidence interval; *Variances are based on the squared deviations from the mean and are therefore reported in the square of the units of the original data (mm^2). Standard deviations, expressed in the original units of the polyp size measurements (mm), can be calculated by taking the square root of the reported variances.

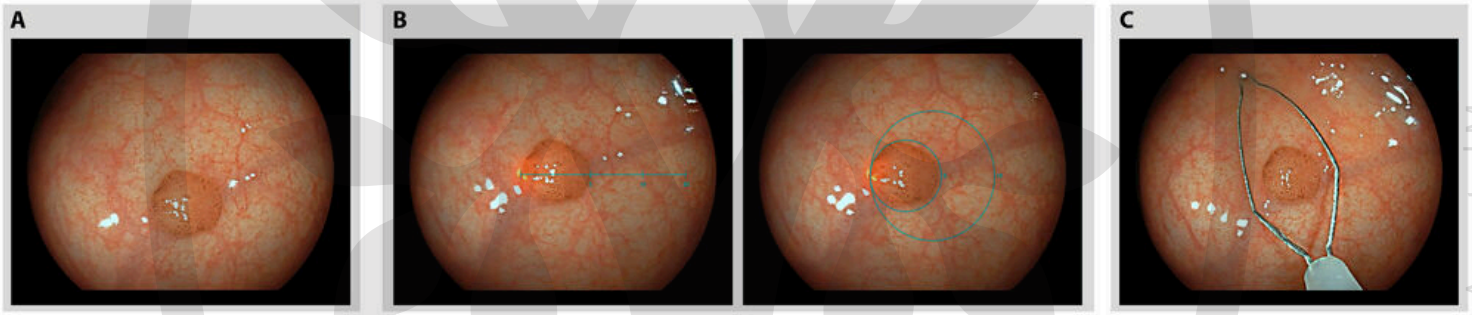
Notes: difference in variance between methods (visual versus snare, visual versus virtual scale and snare versus virtual scale) was statistically significant in all cases ($p < 0.05$), with exception of difference in variance between visual and snare measurements for trainees ($p = 0.206$).

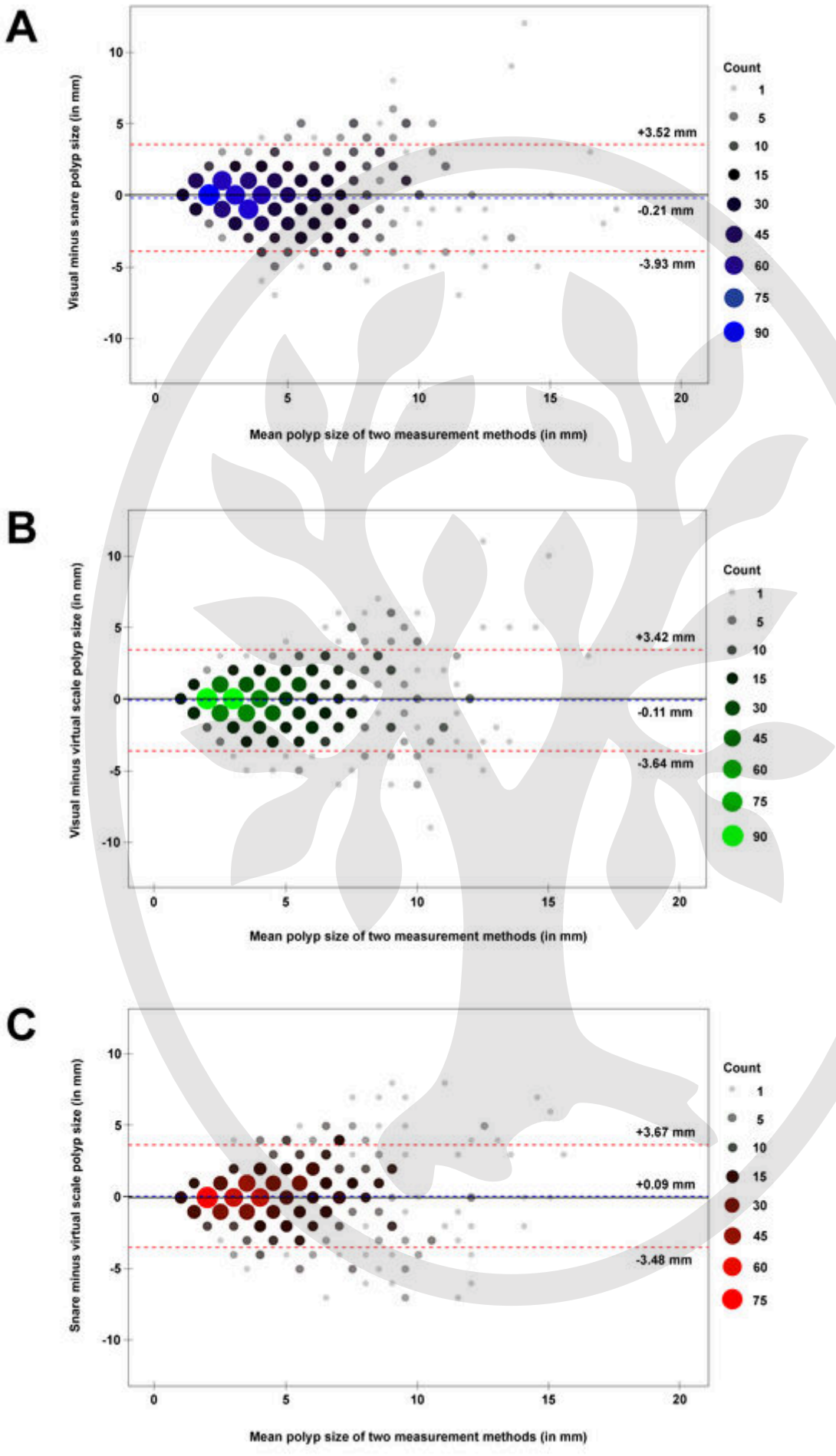
Table 2. Systematic differences between different endoscopic polyp size measurement methods as estimated using mixed linear model analyses

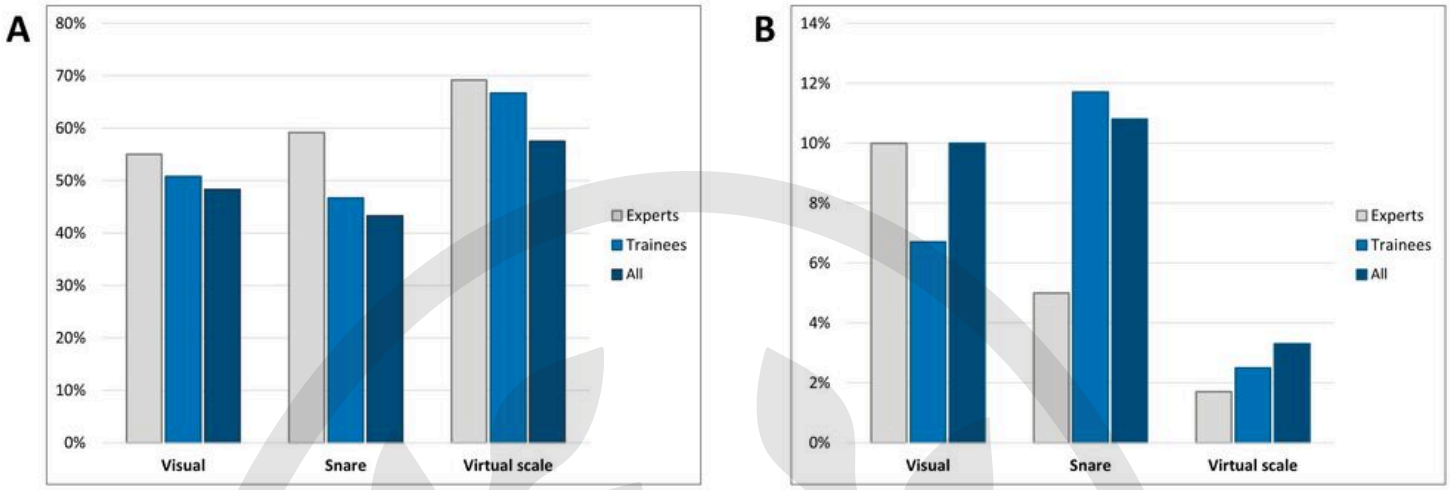
	Number of measurements per method	Method one	Method two	Mean difference (95% CI)*
Experts	960	Visual	Snare	+0.21 (0.09, 0.33)
		Visual	Virtual scale	+0.11 (0.00, 0.23)
		Snare	Virtual scale	-0.09 (-0.21, 0.02)
Trainees	1080	Visual	Snare	+0.36 (0.23, 0.48)
		Visual	Virtual scale	+0.21 (0.08, 0.33)
		Snare	Virtual scale	-0.15 (-0.27, -0.03)
All	2040	Visual	Snare	+0.29 (0.20, 0.37)
		Visual	Virtual scale	+0.16 (0.08, 0.25)
		Snare	Virtual scale	-0.13 (-0.21, -0.04)

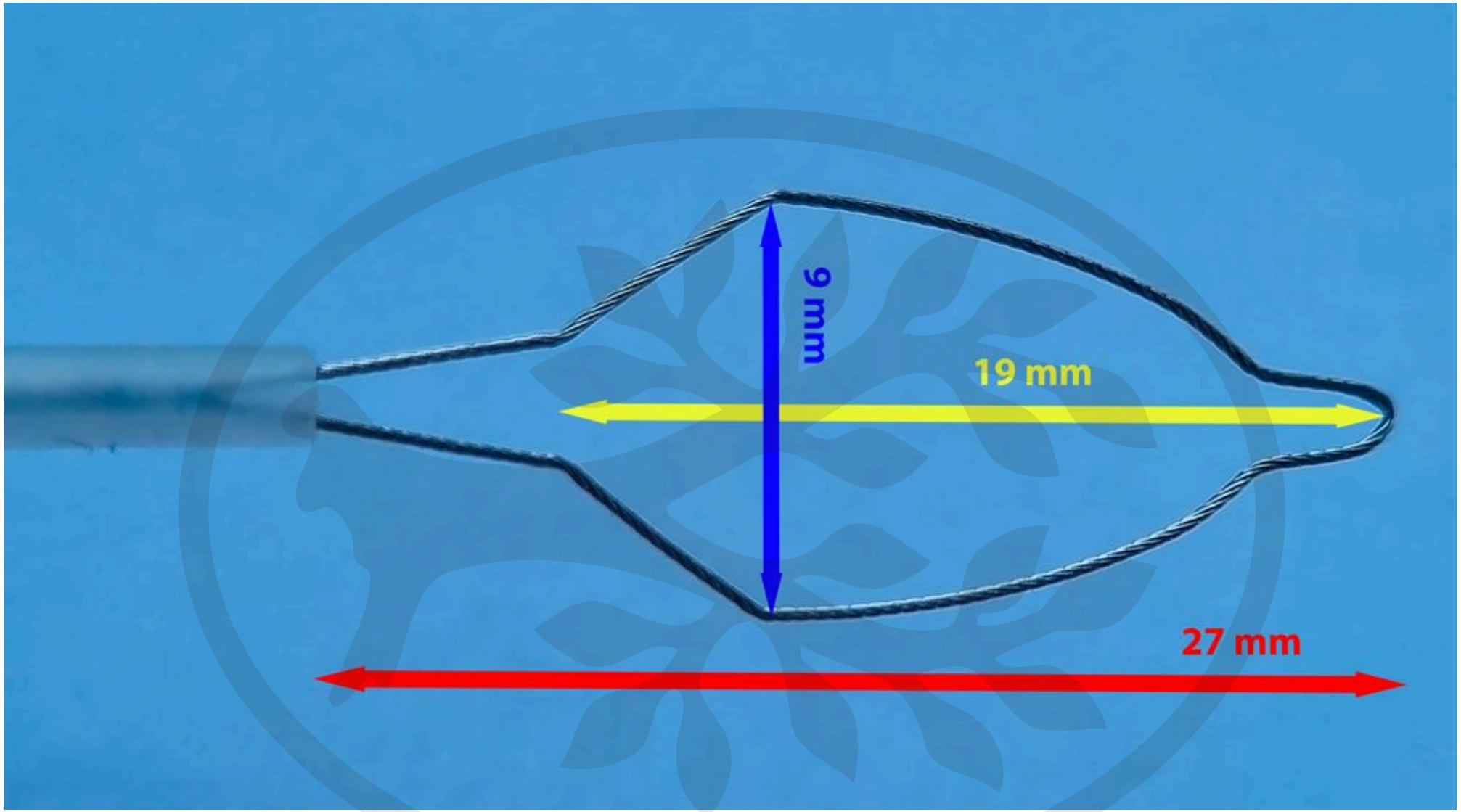
CI, confidence interval; *Mean difference (in mm) between polyp size measurements using method two compared to polyp size measurements using method one.



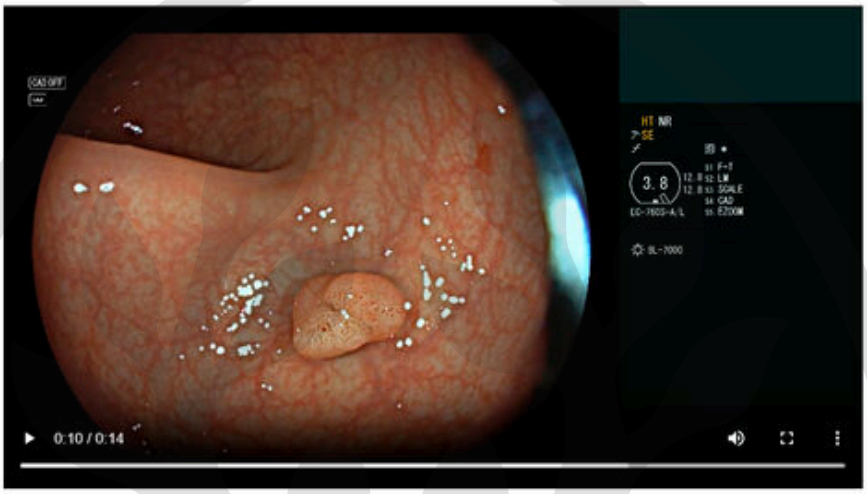








1 | Polyp 01053: estimated polyp size (in millimeters) of polyp shown within video?



Enter your answer here

Ok or press Enter

Powered By Qwary

A

*2 | Polyp 01053: quality of previous video for assessment of polyp size?

A Good

B Sufficient

C Insufficient

Powered By Qwary

B