

Establishing reference standards for small and diminutive colorectal polyp size: validation of caliper-based size measurement

Eric Cristea, MD^{1,2},, Mohamed Zineddine Mahdadi, PharmD^{1,2}, Preslava Aleksieva^{1,2}, Megan Oleksiw^{1,2}, Linda Scavo, PhD^{1,2}, Victoire Michal, PhD², Roupén Djinbachian, MD, MSc^{1,2}, Robert Battat, MD^{1,2}, Dane Christina Daoud, MD¹, Simon Bouchard, MD¹, Mickael Bouin, MD, PhD¹, Jeremy Liu Chen Kiow, MD¹, Benoit Panzini, MD¹, Daniel von Renteln, MD, PhD^{*,1,2}

¹Division of Gastroenterology, Montreal University Hospital Center (CHUM), Montreal, Quebec H2X 0A9, Canada

²Montreal University Hospital Research Center (CRCHUM), Montreal, Quebec H2X 0A9, Canada

*Corresponding author: Daniel von Renteln, MD, PhD, Department of Medicine, Division of Gastroenterology, Montreal University Hospital Center (CHUM) and Montreal University Hospital Research Center (CRCHUM), 900 rue St-Denis, Montreal, Quebec H2X 0A9, Canada (danielrenteln@gmail.com).

Abstract

Background and Aim: Accurate measurement of resected colorectal polyps is essential for clinical management, research, and the development of artificial intelligence-based size estimation systems. Despite widespread use of caliper-based measurement for specimen sizing, formal validation against a reference standard is lacking. This study aimed to validate caliper-based measurement of resected small and diminutive colorectal polyps against high-resolution digital microscopy, a previously validated reference method.

Methods: At the Centre hospitalier de l'Université de Montréal, 143 polyps from 92 patients were measured immediately after resection using vernier digital calipers in the endoscopy suite. Independent measurements were subsequently obtained using high-resolution digital microscopy under blinded conditions. Agreement between methods was assessed using bias analysis, Bland-Altman limits of agreement, intraclass correlation coefficient (ICC), and categorical size concordance.

Results: Caliper-based measurements demonstrated a mean bias of -0.22 mm (95% CI: -0.34 to -0.11 ; $P < .001$) relative to the reference standard. The noninferiority hypothesis with a 0.5-mm margin was not rejected (lower 95% CI > -0.5 mm). Bland-Altman's limits of agreement were -1.57 to 1.12 mm, and the ICC was 0.88 (95% CI: 0.82–0.92). Correct categorical classification occurred in 94.4% of cases (95% CI: 0.89–0.97; $\kappa = 0.81$).

Conclusion: Caliper-based measurement provides accurate and reproducible estimates of polyp size when compared with digital microscopy, supporting its use for clinical and research applications requiring direct specimen measurement.

Keywords: polyp sizing; artificial intelligence; endoscopic databases; digital microscopy; colorectal cancer

Introduction

Accurate measurement of colorectal polyps is critical for clinical decision-making, surveillance planning, and research standardization.^{1–4} Polyp size correlates with malignancy risk and guides guideline conform choice of polyp resection technique and post-polypectomy surveillance intervals assignment.^{2,3} However, visual endoscopic size estimation exhibits substantial interobserver variability, with documented relative accuracy rates ranging from 48% to 65%,^{5–7} and a systematic bias toward overestimation.⁸

Ex vivo measurement of resected specimens represents an established methodology for polyp sizing.^{6,7,9,10} A recent benchmark dataset for artificial intelligence (AI) algorithm development employed this methodology as ground truth.¹¹ Caliper-based measurement protocols have been extensively utilized in both clinical practice and research applications, frequently serving as the reference standard for validating

laser-based sizing technology.^{6,7} However, the accuracy of caliper measurements has not been systematically validated against an independent reference standard, and measurement precision at submillimeter resolution remains unestablished.

The development and clinical deployment of AI-based polyp sizing algorithms necessitate rigorously validated reference measurements. Inaccurate ground truth data may propagate systematic errors throughout algorithm training and validation phases, potentially compromising clinical utility. To address this methodological gap, we conducted a prospective validation study comparing caliper-based measurements of freshly resected colorectal polyps against high-resolution digital microscopy with traceable calibration standards.¹² Our primary objective was to quantify systematic measurement bias and determine whether caliper-based methodology demonstrates sufficient accuracy and precision to serve as a reference standard for clinical applications and AI algorithm validation.

Methods

Patients undergoing an elective colonoscopy at the CHUM (Centre hospitalier de l'Université de Montréal) from May to August 2025 were enrolled into the study following informed consent for participation within a primary prospective study with the objective a building an annotated endoscopic video-database that was approved by an Institutional Review Board (CER 22.013, NCT06822816). Only adults (≥ 18 years) not undergoing an emergency procedure and having an ASA classification of I to III were eligible to participate. The present work constitutes a secondary study derived from the primary prospective study cohort focusing on the validation of caliper-based measurements of resected polyps.

Standard surveillance colonoscopies were performed as usual. When polyps were detected, endoscopists were asked to estimate visually the size of the polyp on its longest axis without any reference. Following polypectomy by cold snare exclusively, the fresh polyps were directly and sequentially measured on-site using a digital vernier caliper. The specimens were then photographed by

1 of 2 digital microscopes (Dino-Lite AF4515ZTL and AM8917MZTL) capable of providing magnification ranging from 10 $\times$ to 140 $\times$, with resolutions of 1.3 or 8.0 megapixels, respectively (Figure 1). Polyps not having been resected via cold snare or not having been resected *en bloc* were excluded as no viable measure could be made. As a result, larger polyps were inherently excluded, since most were removed using hot snare techniques, piecemeal resection, or advanced endoscopic procedures such as Endoscopic Mucosal Resection (EMR) and Endoscopic Submucosal Dissection (ESD). Additionally, if fragmentation of the polyp occurred when the specimen travelled through the endoscope, if the polyp borders were not clear visually, or if the resected specimen had insufficient healthy mucosa around the polypoid lesion, they were excluded from the dataset (Figure 2).

After polypectomy, the specimens were retrieved, unfolded, and placed on a microscopic slide. Measures of the longest axis of the polyp was recorded first using a digital caliper by 1 of 5 different research associates (E.C., M.Z.M., P.A., M.O., L.S.) (Figure 3). The digital caliper was calibrated prior to each new

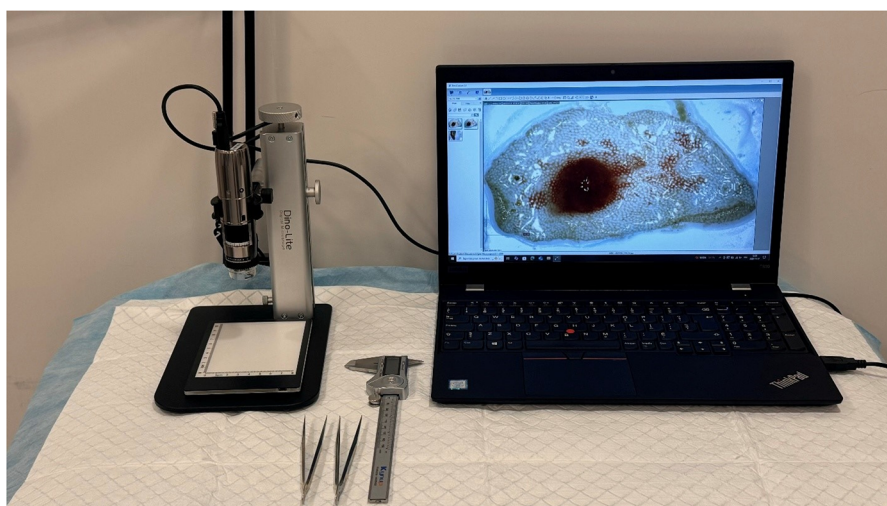


Figure 1 Digital microscope (Dino-Lite AM8917MZTL) and caliper setup. Fresh specimens were measured with calipers and photographed with a digital microscope.

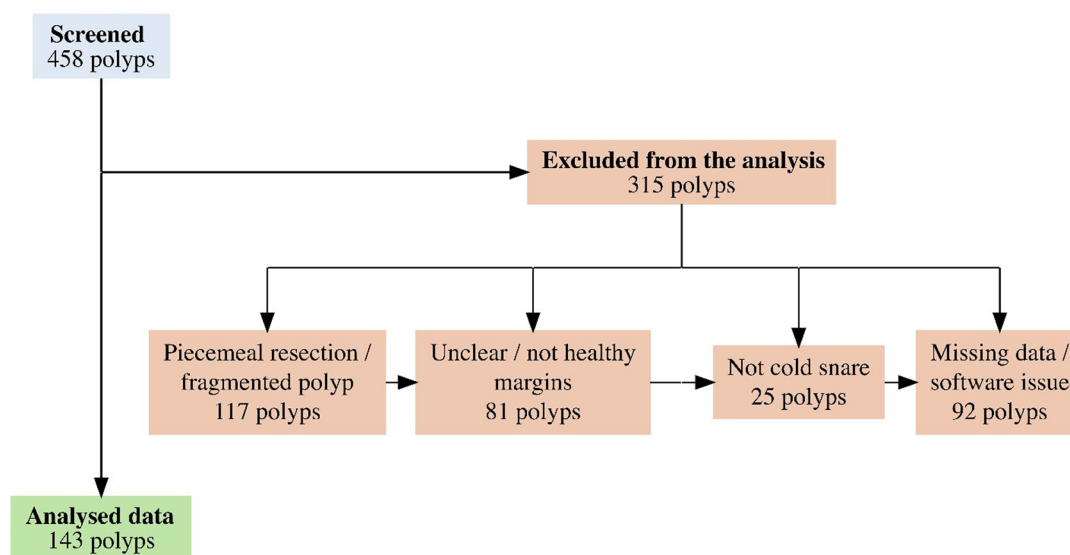


Figure 2 Study flowchart.

measure. Subsequently, the specimens were photographed with the DinoLite microscopes using the *DinoCapture 2.0 software*, making sure the entirety of the polyp and a decent amount of healthy margins was displayed in the picture. Both DinoLite microscopes were calibrated prior to data acquisition using the manufacturer-provided automatic calibration target. The calibration settings were consistently applied to all the data.

At a later date, all photographed specimens were individually measured by experienced raters associated with the lab (E.C., M.Z.M.) who were blinded to the caliper measure. The measures were made directly through the latest version of the *DinoCapture 2.0 software*. Using the digital caliper function of the software, the individual raters performed the measurement of each polyp on its longest axis which was defined as the greatest length between 2 points on the border between the polyp and the healthy colon mucosa of the specimen (Figure 4).

Outcome measures

The primary outcome of this study was to assess the accuracy of caliper-based measurements. The accuracy was evaluated in terms of bias, defined as the mean difference between caliper-based and reference measurements. Throughout,



Figure 3 Caliper measurement of long axis of the polyp.

microscopic measurements were used as the reference standard for polyp sizes. Given current clinical guidelines, where surveillance-relevant thresholds are anchored at 5 and 10 mm, a discrepancy of 0.5 mm was considered clinically negligible, as such a deviation would not meaningfully alter size-based clinical decision-making. Thus, noninferiority of the caliper-based method was hypothesized using a 0.5-mm margin.

Secondary outcomes included (1) the intraclass correlation coefficient (ICC) between caliper-based and reference measurements, to assess the agreement; as well as (2) the pairwise ICCs between visual assessments and caliper or reference measurements; (3) the proportion of polyps with correct categorized caliper-based measurements (diminutive polyps: ≤ 5 mm; small polyps: >5 mm), to assess the accuracy of dichotomized measurements; and (4) the kappa statistic computed for the categorized measurements, to assess the agreement between caliper-based and reference methods when measurements are dichotomized.

Statistical analysis and sample size

Patient- and polyp-level characteristics were described using means with standard deviations (SD) for continuous variables, and counts with percentages for categorical variables. The primary outcome was evaluated using a paired *t*-test between caliper-based and standard reference measurements. The ICC secondary outcome was estimated using a 2-way mixed-effects models for absolute agreement. All accuracy and agreement estimates were reported with 95% confidence intervals (CIs). Statistical analyses were conducted using R version 4.5.1. (R Core Team, 2025).

The minimum sample size required was computed for a non-inferiority test assuming a small nonclinically significant 0.25 mm bias and a margin of 0.5 mm. A minimum of 105 polyps with paired caliper-based and reference standard measurements were necessary for a test with 80% power and 5% significance level.

Results

A total of 143 polyps from 92 patients (mean age 66.2, 42.4% female) met the inclusion criteria for this secondary study



Figure 4 Measurement of long axis of the polyp using an image taken with a digital microscope.

Table 1 Patient demographics and polyp characteristics.

Patient characteristics (N=92)	
Age, mean (SD)	66.2 (11.5)
Female, <i>n</i> (%)	39 (42.4)
ASA class, <i>n</i> (%)	
ASA I	24 (26.1)
ASA II	54 (58.7)
ASA III	14 (15.2)
Ethnicity, <i>n</i> (%)	
African	3 (3.3)
Asian	7 (7.6)
Caribbean	3 (3.3)
European	9 (9.8)
Latin, Central, South American	2 (2.2)
North American	65 (70.7)
Missing	3 (3.3)
Family member with CRC, <i>n</i> (%)	14 (15.2)
Family member with IBD, <i>n</i> (%)	5 (5.4)
Polyp characteristics (N=143)	
Rectosigmoid, <i>n</i> (%)	46 (32.2)
Paris class, <i>n</i> (%)	
IIa	12 (8.4)
IIc	2 (1.4)
Ip	2 (1.4)
Is	125 (87.4)
Isp	2 (1.4)
Pathology, <i>n</i> (%)	
Adenoma/Neoplastic	93 (65.0)
Hyperplastic	29 (20.3)
Missing	2 (1.4)
SSL	3 (2.1)
Other	16 (11.2)
Polyp size (mm), mean (SD)	
Visual size estimation (N=142)	4.51 (2.02)
Caliper measurement	3.72 (1.40)
Microscopy measurement	3.95 (1.51)

Abbreviations: ASA, American Society of Anesthesiologists; CRC, colorectal cancer; IBD, inflammatory bowel disease; SSL, sessile serrated lesions; SD, standard deviation.

(Figure 2). The majority of included polyps were adenomatous (65%) with polypoid Paris classification (90.2%), and 32.2% of these were rectosigmoid polyps (Table 1).

The caliper-based measurements yielded a mean negative bias of −0.22 mm (95% CI: −0.34 to −0.11) compared with reference standard sizes, which was statistically significant ($P < .001$). However, in terms of clinical significance, the noninferiority hypothesis of the caliper-based measuring method with 0.5-mm margin was not rejected (lower 95% CI limit > -0.5 mm).

Every pair of measurements were plotted in Figure 5, alongside the Bland-Altman plot with limits of agreement −1.57 and 1.12 mm. In terms of ICC, the caliper-based measurements showed a good level of agreement with the reference standard (ICC=0.88; 95% CI: 0.82-0.92, Table 2). Visual assessments showed moderate agreement with caliper-based and reference standard sizes (eg, Visual vs Microscopy, ICC=0.52 [0.37-0.64]). The consistency between caliper-based measurements and reference standard sizes also translated to the categorized measurements, with kappa=0.81 and correct size classification in 94.4% (135/143; 95% CI: 88.9-97.4) of the cases.

Discussion

This study provides formal validation of caliper-based measurement for freshly resected small and diminutive colorectal polyps against a high-resolution digital microscopy reference standard. The findings demonstrate that caliper-based measurements achieve submillimeter accuracy and strong agreement with the validated reference, confirming their suitability for research and clinical applications requiring precise lesion sizing.

Accurate reference measurement of polyp size is critical for multiple domains in gastroenterology, including clinical decision-making, post-polypectomy surveillance, and the evaluation of new endoscopic imaging technologies.¹³ Polyp size directly determines surveillance intervals and is an established risk factor for advanced neoplasia.^{1,3} Yet, despite its importance, accurate and reproducible measurement remains challenging in clinical practice, where visual estimation continues to dominate and is well known to suffer from systematic bias and interobserver variability.^{6,7,14,15} The increasing integration of digital tools and AI into endoscopy further heightens the need for robust, validated reference standards against which new systems can be trained and evaluated.^{14,16-19}

High-resolution digital microscopy, previously validated as a reference method for polyp size measurement, provides objective, image-based documentation and traceable measurement verification.¹² However, digital microscopy requires specialized equipment, careful calibration, and trained personnel, making its application labor-intensive and less feasible for large-scale studies or real-time workflows. In contrast, caliper-based measurement of fresh specimens can be implemented at the point of care with minimal cost and infrastructure. The present validation confirms that, when performed under standardized conditions, caliper-based measurement yields precise, reproducible results that align closely with those obtained through digital microscopy.

This finding is significant because many ongoing and emerging endoscopic technology studies, such as those evaluating laser-based or AI-based sizing tools, have relied on caliper measurements as their practical reference standard.^{6,7,16} Until now, the accuracy of this approach had not been independently verified. By demonstrating its validity, this study provides the necessary methodological foundation to support its continued use for both clinical and research purposes. In particular, it strengthens the scientific credibility of studies that have used caliper-based sizing to validate optical technologies such as the through-the-scope laser probe (*AccuMeasure*, VTM Technologies, Haifa, Israel) and virtual scale endoscopy (*Scale Eye*, Fujifilm, Tokyo, Japan), which have shown millimeter-level agreement with microscopic measurements.^{6,7,20} Moreover, unlike the present study, which was limited to small and diminutive polyps amenable to *en bloc* cold-snare resection, validated laser-based *in vivo* sizing technologies could provide a critical advantage for establishing ground-truth measurements in larger lesions that cannot be removed *en bloc* or reliably quantified *ex vivo*.

In the context of AI development, this validation carries particular importance. The performance of AI-based sizing models depends entirely on the reliability of their ground truth data.¹¹ As discussed in recent methodological analyses, training datasets derived from noncalibrated visual estimates or pathology-based measurements introduce systematic biases, through optical distortion or post-fixation shrinkage, that can be embedded into AI predictions.^{14,21} Caliper-based

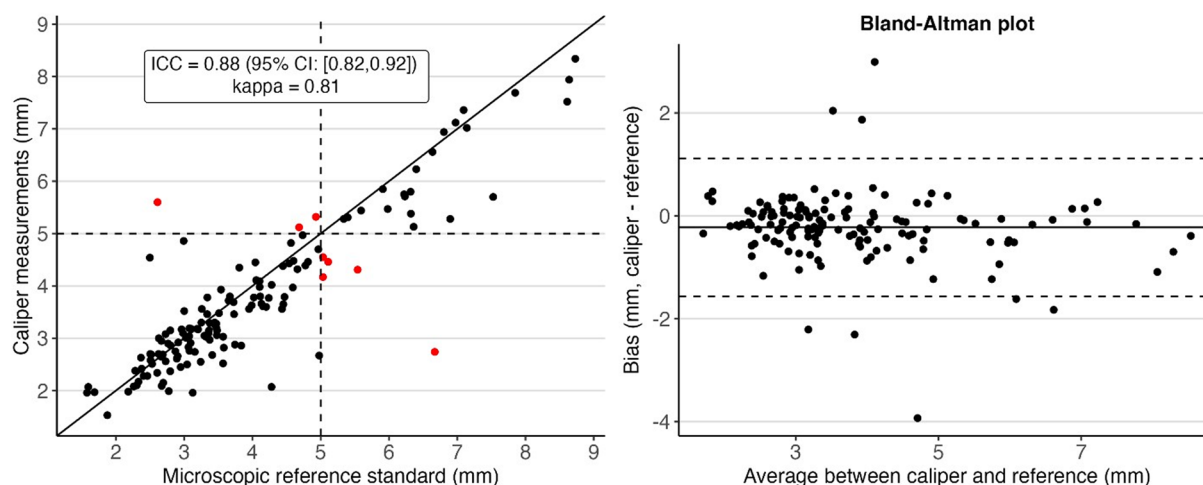


Figure 5 Left: Concordance plot of all caliper measurements against the reference standard. The dashed lines categorize measurements into relevant polyp size classes (≤ 5 mm, > 5 mm), and the red circles are datapoints with incorrect caliper size category. Right: Bland-Altman plot.

Table 2 Pairwise intraclass correlation coefficients (ICCs) between visual, caliper, and microscopy polyp size measurements. Visual assessments were available for 142 out of 143 polyps.

	ICC (95% CI)	Interpretation
Visual assessment versus Caliper measurement	0.55 (0.34-0.69)	Moderate agreement
Visual assessment versus Microscopy measurement	0.52 (0.37-0.64)	Moderate agreement
Caliper versus Microscopy measurements	0.88 (0.82-0.92)	Good agreement

measurement, now validated against an independent and traceable standard, provides a pragmatic solution for generating large, high-quality datasets required for model training, benchmarking, and regulatory validation.

This study had several limitations that may affect generalizability. It was conducted at a single tertiary center, and inclusion criteria restricted the sample to *en bloc* cold-snare resections, resulting in a near-exclusive analysis of diminutive and small polyps. Larger lesions typically removed by hot snare, EMR or ESD were therefore not represented. The sample also consisted mainly of sessile (Is) adenomatous polyps, limiting morphological and histopathological diversity. Additionally, fragmented specimens and polyps with unclear borders were excluded, which may not fully reflect real-world practice.

Caliper-based measurements were performed by multiple research associates, introducing the possibility for interobserver variability. However, despite this potential source of bias, our results demonstrate that caliper measurements remained highly accurate when compared with microscopic reference standard. Nonetheless, caliper-based measurements carry intrinsic methodological limitations, including the absence of integrated photographic documentation and the susceptibility of greater interobserver variability. These factors restrict retrospective verification and quality assurance. In this regard, digital microscopy remains the preferred reference for high-stakes or multicentric validation efforts, as it enables both measurement traceability and independent audit,¹² for large-scale data acquisition or routine research applications, caliper-based measurement offers a balance of precision, feasibility, and scalability

that makes it an attractive and validated option for establishing reliable ground truth data.

In conclusion, caliper-based measurement of resected small and diminutive colorectal polyps is an accurate, reproducible, and validated technique. It provides a practical and cost-effective reference standard suitable for research and for generating annotated datasets that underpin the development and clinical validation of laser-based and AI-based sizing technologies.

Acknowledgments

D.v.R. is supported by a “Fonds de Recherche du Québec Santé” career development award. The authors would like to thank Samira Hanin, Research Coordinator, and Ghislaine Ahoua, Research Assistant.

Author contributions

E.C. contributed to data collection, analysis, and manuscript writing. M.Z.M. participated in data collection and manuscript revision. P.A., M.O., and L.S. participated in data collection. V.M. contributed to statistical analysis and manuscript writing and revision. R.D., R.B., D.C.D., S.M., M.B., J.K., and B.P. were involved with data collection and manuscript revision. D.v.R. contributed to study design, data collection, manuscript writing, and primary manuscript revision.

Funding

None declared.

Conflicts of interest

Conflict of interest disclosure forms (ICMJE) have been collected for all co-authors and can be accessed as supplementary material.

Data availability

The data, analytic methods, and study materials from this research will be made available only upon a reasonable request. Access will be granted solely when the research proposal

includes appropriate Institutional Review Board approval and establishes a formal data-sharing agreement. To maintain patient confidentiality and ensure proper use of research materials, all requests must be submitted in writing to the principal investigator (D.v.R.) at Montreal University Hospital Center. Final determination for data access will be made on an individual basis, considering the scientific value of the proposed research and compliance with institutional policies regarding patient privacy protection.

References

1. Dekker E, Rex DK. Advances in CRC prevention: screening and surveillance. *Gastroenterology*. 2018;154:1970-1984. <https://doi.org/10.1053/j.gastro.2018.01.069>
2. Copland AP, Kahi CJ, Ko CW, Ginsberg GG. AGA clinical practice update on appropriate and tailored polypectomy: expert review. *Clin Gastroenterol Hepatol*. 2024;22:470-479.e5. <https://doi.org/10.1016/j.cgh.2023.10.012>
3. 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. *Gastroenterology*. 2020;158:1131-1153.e5. <https://doi.org/10.1053/j.gastro.2019.10.026>
4. Rex DK, Anderson JC, Butterly LF, et al. Quality indicators for colonoscopy. *Am J Gastroenterol*. 2024;119:1754-1780.
5. Chaptini L, Chaaya A, Depalma F, Hunter K, Peikin S, Laine L. Variation in polyp size estimation among endoscopists and impact on surveillance intervals. *Gastrointest Endosc*. 2014;80:652-659. <https://doi.org/10.1016/j.gie.2014.01.053>
6. Djinbachian R, Popescu Crainic I, Pioche M, et al. Accuracy in polyp size measurement among surgeons, gastroenterologists, trainees, and experts: a prospective video-based study. *Am J Gastroenterol*. 2024;119:532-538. <https://doi.org/10.14309/ajg.0000000000002494>
7. Popescu Crainic I, Djinbachian R, Rex DK, et al. Expert endoscopist assessment of colorectal polyp size using virtual scale endoscopy, visual or snare-based estimation: a prospective video-based study. *Scand J Gastroenterol*. 2024;59:608-614. <https://doi.org/10.1080/0365521.2024.2308519>
8. Anderson BW, Smyrk TC, Anderson KS, et al. Endoscopic overestimation of colorectal polyp size. *Gastrointest Endosc*. 2016;83:201-208. <https://doi.org/10.1016/j.gie.2015.06.058>
9. 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*. 2024;119:1309-1317. <https://doi.org/10.14309/ajg.0000000000002623>
10. 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. <https://doi.org/10.1136/gutjnl-2022-328654>
11. Song Y, Du S, Wang R, et al. Polyp-size: a precise endoscopic dataset for AI-driven polyp sizing. *Sci Data*. 2025;12:918. <https://doi.org/10.1038/s41597-025-05251-x>
12. Aleksieva P, Djinbachian R, Rex DK, et al. Digital microscopy in the endoscopy suite: a reference standard for colorectal polyp size measurement. In: CDDW; Toronto 2026 [cited 07/11/2025].
13. Cheloff AZ, Kim L, Pochapin MB, Shaukat A, Popov V. Accuracy of visual estimation for measuring colonic polyp size: a systematic review and meta-analysis. *Am J Gastroenterol*. 2025;120:2251-2259. <https://doi.org/10.14309/ajg.0000000000003391>
14. Kerbage A, Souaid T, Singh K, Burke CA. Taking the guess work out of endoscopic polyp measurement: from traditional methods to AI. *J Clin Gastroenterol*. 2025;59:497-503.
15. van Bokhorst QNE, Houwen BBSL, Hazewinkel Y, et al. SCALE EYE Study Group. Polyp size measurement during colonoscopy using a virtual scale: variability and systematic differences. *Endoscopy*. 2025;57:137-145. <https://doi.org/10.1055/a-2371-3693>
16. Sudarevic B, Sodmann P, Kafetzis I, et al. Artificial intelligence-based polyp size measurement in gastrointestinal endoscopy using the auxiliary waterjet as a reference. *Endoscopy*. 2023;55:871-876. <https://doi.org/10.1055/a-2077-7398>
17. Abdelrahim M, Saiga H, Maeda N, Hossain E, Ikeda H, Bhandari P. Automated sizing of colorectal polyps using computer vision. *Gut*. 2022;71:7-9. <https://doi.org/10.1136/gutjnl-2021-324510>
18. Wang J, Li Y, Chen B, et al. A real-time deep learning-based system for colorectal polyp size estimation by white-light endoscopy: development and multicenter prospective validation. *Endoscopy*. 2024;56:260-270. <https://doi.org/10.1055/a-2189-7036>
19. Bechtold ML, Dahip MS, Matteson-Kome A, Puli SR, Nguyen. S312 using artificial intelligence for polyp size in colonoscopy: a phantom study. *Am J Gastroenterol*. 2021;116:S136.
20. Mahdadi MZ, Cristea E, Oleksiw M, et al. Endoscopic polyp size measurement using a laser-based sizing probe: a pilot study. In: CDDW; Toronto 2026 [cited 07/11/2025]. <https://www.cag-acg.org/cddw.html>
21. Turner JK, Wright M, Morgan M, Williams GT, Dolwani S. 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. <https://doi.org/10.1097/MEG.0b013e32835d1f2d>