



UvA-DARE (Digital Academic Repository)

Polyp measurement based on CT colonography and colonoscopy: variability and systematic differences

de Vries, A.H.; Bipat, S.; Dekker, E.; Liedenbaum, M.H.; Florie, J.; Fockens, P.; van der Kraan, R.; Mathus-Vliegen, E.M.; Reitsma, J.B.; Truyen, R.; Vos, F.M.; Zwinderman, A.H.; Stoker, J.

DOI

[10.1007/s00330-009-1683-0](https://doi.org/10.1007/s00330-009-1683-0)

Publication date

2010

Document Version

Final published version

Published in

European Radiology

[Link to publication](#)

Citation for published version (APA):

de Vries, A. H., Bipat, S., Dekker, E., Liedenbaum, M. H., Florie, J., Fockens, P., van der Kraan, R., Mathus-Vliegen, E. M., Reitsma, J. B., Truyen, R., Vos, F. M., Zwinderman, A. H., & Stoker, J. (2010). Polyp measurement based on CT colonography and colonoscopy: variability and systematic differences. *European Radiology*, 20(6), 1404-1413. <https://doi.org/10.1007/s00330-009-1683-0>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Ayso H. de Vries
Shandra Bipat
Evelien Dekker
Marjolein H. Liedenbaum
Jasper Florie
Paul Fockens
Roel van der Kraan
Elizabeth M. Mathus-Vliegen
Johannes B. Reitsma
Roel Truyen
Frans M. Vos
Aeilko H. Zwinderman
Jaap Stoker

Polyp measurement based on CT colonography and colonoscopy: variability and systematic differences

Received: 20 July 2009
Revised: 27 September 2009
Accepted: 28 October 2009
Published online: 22 December 2009
© The Author(s) 2009.
This article is published with open access at Springerlink.com

A. H. de Vries (✉) · S. Bipat ·
M. H. Liedenbaum · J. Florie ·
F. M. Vos · J. Stoker
Department of Radiology, Academic
Medical Centre, University of
Amsterdam,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands
e-mail: Ayso.devries@amc.uva.nl
Tel.: +31-205-6663028
Fax: +31-205-6669119

E. Dekker · P. Fockens ·
E. M. Mathus-Vliegen
Department of Gastroenterology,
Academic Medical Centre,
University of Amsterdam,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands

R. van der Kraan · R. Truyen
Department of Healthcare Informatics,
Philips Healthcare,
Veenpluis 6,
5684 PC Best, The Netherlands

J. B. Reitsma · A. H. Zwinderman
Department of Clinical Epidemiology
and Biostatistics, Amsterdam Medical
Center, University of Amsterdam,
Meibergdreef 9,
1105 AZ Amsterdam, The Netherlands

Abstract Objective: To assess the variability and systematic differences in polyp measurements on optical colonoscopy and CT colonography. **Materials:** Gastroenterologists measured 51 polyps by visual estimation, forceps comparison and linear probe. CT colonography observers randomly assessed polyp size two-dimensionally (abdominal and intermediate window) and three-dimensionally (manually and semi-automatically). Linear mixed models were used to assess the variability and systematic differences between CT colonography and optical colonoscopy techniques. **Results:** The variability of forceps and linear probe measurements was comparable and both showed less variability than measurement by visual assessment.

Measurements by linear probe were 0.7 mm smaller than measurements by visual assessment or by forceps. The variability of all CT colonography techniques was lower than for measurements by forceps or visual assessment and sometimes lower (only 2D intermediate window and manual 3D) compared with measurements by linear probe. All CT colonography measurements judged polyps to be larger than optical colonoscopy, with differences ranging from 0.7 to 2.3 mm. **Conclusion:** A linear probe does not reduce the measurement variability of endoscopists compared with the forceps. Measurement differences between observers on CT colonography were usually smaller than at optical colonoscopy. Polyps appeared larger when using various CT colonography techniques than when measured during optical colonoscopy.

Keywords CT colonography · Colon · Colonoscopy · Measurement · Cancer · 2D · 3D

Introduction

CT colonography has been consistently shown to have a high accuracy in the detection of colorectal neoplasia [1]. The method is less invasive and less burdensome than optical colonoscopy [2]. However, assessment of malignancy by obtaining tissue samples for histological analysis is not possible with this technique. As a surrogate for histopathology, polyp size is used for patient management strategies [3].

Currently, according to the U.S. screening guidelines for CT colonography [4], all patients with a polyp 6 mm or larger should be referred for optical colonoscopy. However, whether optical colonoscopy is indicated for polyps 6–9 mm is still under debate as the prevalence of advanced features was reported to be low [5, 6]. Surveillance for growth with CT colonography has been suggested as a safe alternative [7]. Small (<6 mm) polyps may be safely left in situ because of a negligible risk of malignant transformation.

Therefore, size is crucial for decision-making in CT colonography. Differences in polyp size measurements between CT colonography and optical colonoscopy should be minimal to avoid difficulties in the choice of management.

Two pivotal in-vitro studies report an underestimation of polyp size by the endoscopists and an accurate or slight overestimation of CT colonography [8, 9] compared with optical colonoscopy measurements. However, factors that may influence polyp measurement such as difficult viewing angles or bowel motility are not considered in these analyses.

Several in-vivo studies on differences in polyp size measurements and variability between CT colonography and optical colonoscopy have been published [10–13] as well. In these studies different CT colonography review modes, window settings and automatic measurement tools were compared with one of the various optical colonoscopy reference standards, e.g. measurement by linear probe or forceps. These studies had contradictory results, i.e. some studies reported an underestimation [10, 11] of polyp size whereas others reported an overestimation [12] according to 2D and 3D CT colonography measurements.

A comparison of the various CT colonography and optical colonoscopy measurement techniques within one study will probably give more insight into the level of agreement among these measurement techniques. Therefore the first purpose of our study was to assess the variability in size measurements among CT colonography and optical colonoscopy techniques. The variability can be considered a result of an inherent difference in techniques and/or differences in observers using the technique. The second purpose was to assess systematic differences in polyp size measurements between CT colonography and optical colonoscopy techniques.

Materials and methods

The study was approved by the institutional review board of our institute, and all patients provided written informed consent for participation in this study.

We compared 2D (abdominal and intermediate window) and 3D (manual and semi-automatic) CT colonography techniques with optical colonoscopy measurement techniques, i.e. measurement by visual estimation, with a forceps and with a calibrated linear probe. The respective measurements were performed by three experienced CT colonography readers and three experienced colonoscopists on identical colorectal polyps.

In this study we assessed the variability (spread) and the systematic differences of size measurements of the various measurement techniques. Therefore no reference measurement tool was needed.

Patients were enrolled from two comparative studies of CT colonography and optical colonoscopy:

1. The first study was performed in a population of individuals between 50 and 75 years of age who were invited for a faecal occult blood test (FOBT) screening trial and tested positive. These patients were offered a CT colonography examination before optical colonoscopy [14].

Exclusion criteria for this study were terminal illness, severe psychiatric symptoms, colonoscopy or another FOBT in the previous 2 years, examinations for research purposes with radiation exposure in the last 12 months, iodine contrast allergy, hyperthyroidism and pregnancy.

2. The second study was done in a consecutive series of patients with a personal or family history of colorectal polyps or cancer and scheduled for optical colonoscopy surveillance [15].

Exclusion criteria in this study were age under 18 years, pregnancy, personal history of inflammatory bowel disease, familial adenomatous polyposis, Peutz-Jeghers syndrome, hereditary non-polyposis colorectal cancer, prior allergic reaction to iodine contrast, untreated hyperthyroidism, and known colorectal polyps that were not removed at an earlier endoscopy.

Polyps detected in these patients were used in a later phase of this measurement study.

Inclusion and exclusion criteria for the measurement study

Patients were excluded from the measurement study if the optical colonoscopy examination was not digitally stored, if measuring devices were not available during the examination, or if there was too much time pressure to properly execute the three different measurements during the optical colonoscopy. Patients without polyps of 4 mm or larger and patients with too many polyps that they could not be properly matched were excluded as well.

Polyps were included in the measurement study if they (1) were seen with both techniques, (2) were measured by all techniques during the recorded optical colonoscopy, (3) were estimated to measure at least 4 mm (to be sure that all medium-sized polyps would be included) *and* not larger than the linear probe (20 mm) based on initial visual assessment by the executing endoscopist, (4) were not (partially) covered by faecal material in both positions during CT colonography (which would require electronic cleansing to sustain 3D measurement) and (5) were unambiguously matched.

A polyp detected on both CT colonography and optical colonoscopy was matched based on two criteria: (1) if its appearance visually resembled the corresponding polyp on the

optical colonoscopy movie and (2) if its segment or adjacent segment corresponded with one of the six reference segments.

The polyps were matched to the optical colonoscopy findings by a research fellow with experience of at least 150 CT colonography examinations verified by optical colonoscopy. A maximum of three polyps per patient were included to prevent substantial lengthening of the optical colonoscopy procedure.

CT colonography

Patients underwent imaging in the supine and prone positions after intravenous administration of bowel relaxants (20 mg butylscopolamine, Buscopan, Boehringer, Ingelheim, Germany; or, if contraindicated, 1 mg glucagon hydrochloride, Glucagon, Novo-Nordisk, Bagsvaerd, Denmark). CO₂ was automatically insufflated up to maximum patient tolerance (PROTOCO2L, EZ-E-M, Lake Success, NY, USA). Intravenous contrast medium was not administered.

For the exact preparation schemes, we refer to the comparative studies from which the polyps for this study were enrolled [14, 15]. In short, oral iodinated contrast medium (which has a laxative effect) was added to the patient's low fibre diet at least 1 day before the examination. In study I [15], this preparation scheme was combined with 4 l polyethylene glycol electrolyte solution (Klean-Prep; Helsinn Birex Pharmaceuticals, Dublin, Ireland) for same day colonoscopy.

Examinations were performed on 64-slice CT (Brilliance, Philips Healthcare, Best, the Netherlands). The collimation was 64×0.625 mm, pitch 1.2, reconstruction slice thickness 0.9 mm, rotation time 0.4 s and tube voltage 120 kV. In study I [15], the tube current was either 58 or 82 mAs depending on the abdominal circumference (less than or more than 102.5 cm respectively). In study II [14], the tube current was modulated automatically (reference 40 mAs).

Optical colonoscopy

Within 2 weeks of CT colonography, optical colonoscopy was performed by an experienced staff member or a

gastroenterology fellow under the direct supervision of the attending gastroenterologist. The patients in both study populations had undergone extensive bowel preparation consisting of 4 l polyethylene glycol and a low-fibre diet. The procedure was performed with a standard colonoscope (CF-140L; Olympus, Tokyo, Japan). The optical colonoscopy was videotaped and subsequently digitally stored.

Segmental unblinding was performed according to the findings of CT colonography.

During optical colonoscopy a polyp was measured by three different techniques in succession: by (1) visual assessment, (2) comparison with an opened forceps with a size of 8 mm and (3) a calibrated linear probe (Fig. 1). The linear probe (Olympus, Tokyo, Japan) had ten markings with spacing of 2 mm on the distal flexible tip.

During the procedure, polyps were recorded with respect to size (these measurements were not used in this study), morphology and segment for patient management purposes and polyp-matching purposes.

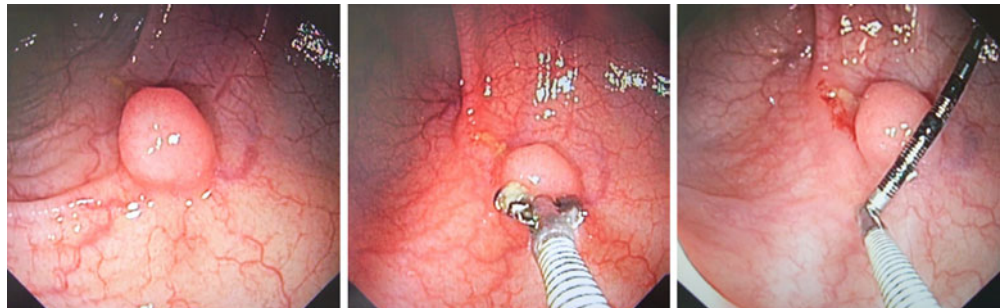
Digital movie extracts of the polyp measurements were later presented to three experienced endoscopists within the framework of this polyp measurement study.

Assessment of polyp size at CT colonography

The measurements within the framework of the polyp measurement study were performed by three observers (M. H., J.F., A.d.V.) who had read 200, 500 and 500 CT colonography studies verified by optical colonoscopy respectively. The measurement of each polyp was performed on a tailor-made computer program based on the ViewForum workstation (Philips Healthcare, Best, the Netherlands).

Using this program, measurements were taken in 2D and 3D. In 2D, each polyp was measured in a reformatted cross-sectional plane through the polyp. The plane could be rotated to identify the longest object dimension. A polyp was measured by placing electronic calipers along the largest diameter in the reformatted image. Measurements were performed with a preset window width and a level setting of 1500/−50 HU (intermediate window) and 400/40 HU (abdominal window).

Fig. 1 The three different measurement techniques of optical colonoscopy: visual assessment (*left*), measurement by forceps (*middle*) and measurement by calibrated linear probe (*right*)



The 3D measurements were performed both manually and semi-automatically. The observers were instructed to measure the maximum diameter by electronic calipers in the endoluminal display. The semi-automatic measurements were based on automatic measurements using a prototype algorithm. The observers either accepted these measurements of the algorithm or modified the measurements manually by repositioning the two software cursors.

The measurements were performed on the default surface-rendered colonic wall threshold setting of ViewForum of -650 HU. The observers were explicitly instructed to carefully place the cursors at the edge of the polyp and not to let the cursors “fall off” the edges of the polyp. As in our clinical practice, we used the larger of the two measurements performed in the prone and supine positions to avoid under-treatment.

To avoid recall bias, polyp size (mm) was not displayed after the measurement was performed, i.e. the CT colonography observers were blinded to their own size measurements. Each 2D measurement using the abdominal window setting was directly followed by a 3D semi-automatic measurement. The 2D measurement in the intermediate window was directly followed by a manual 3D measurement. These paired measurements were performed in a random order. There was an interval of at least 1 day between these clustered measurements.

Assessment of polyp size at optical colonoscopy

The three experienced gastroenterologists (E.D., E.M.V., P.F.) had performed more than 1,500 and 5,000 and 2,500 colonoscopies respectively. For each polyp three separate movie extracts were made, each corresponding to one measurement method (visual assessment, forceps or linear probe). All the extracts were presented in a random order to the gastroenterologists. This was done on a laptop computer.

Each movie extract was rated by each colonoscopist as “good”, “sufficient” or “insufficient” depending on the possibility of properly assessing the polyp’s largest diameter. Each gastroenterologist was blinded to the measurements by the other gastroenterologists as well as to the measurements that were taken during CT colonography and the initial optical colonoscopy.

Statistical analysis

The difference in measured sizes between the most frequently used optical colonoscopy measurement techniques (i.e. forceps and visual assessment) and the various CT colonography measurements was illustrated by Bland-Altman plots. In a Bland-Altman plot the mean value of two measurements is plotted against their difference [16].

The data consist of measurements of the same polyp by three experienced CT colonography readers and three

experienced endoscopists with different techniques. For this study we used a linear mixed (regression) model. The model contained the following parameters:

1. Fixed parameters to estimate the systematic differences in size measurement between techniques (technique as an independent variable), i.e. to estimate whether measurement technique A measured polyps systematically larger or smaller than measurement technique B.
2. Heterogeneous error variance structure to estimate the residual variance (variability) for each technique. We assumed that the variability of measurements within a technique is an indication of the precision of the measurement technique: the smaller the difference between the observers, the better the technique [17].

To test whether the residual variance (obtained by the model above) was statistically different in the two techniques we fitted the same model but then assumed that residual variances of these techniques were similar. The Akaike’s information criterion (AIC) values of the models were compared using a χ^2 distribution with one degree of freedom. If the models did not differ (i.e. the models had a comparable AIC), the variances in size measurements between the techniques were considered similar. A poorer fit (higher AIC) of the latter model would indicate that the residual variances (variability) in measurement were different [18].

3. Fixed effect parameters to estimate the systematic differences in size measurement between observers within a technique (observers within a technique as an independent variable), i.e. to estimate whether observer A measured polyps consequently larger or smaller than observer B when using the same technique.

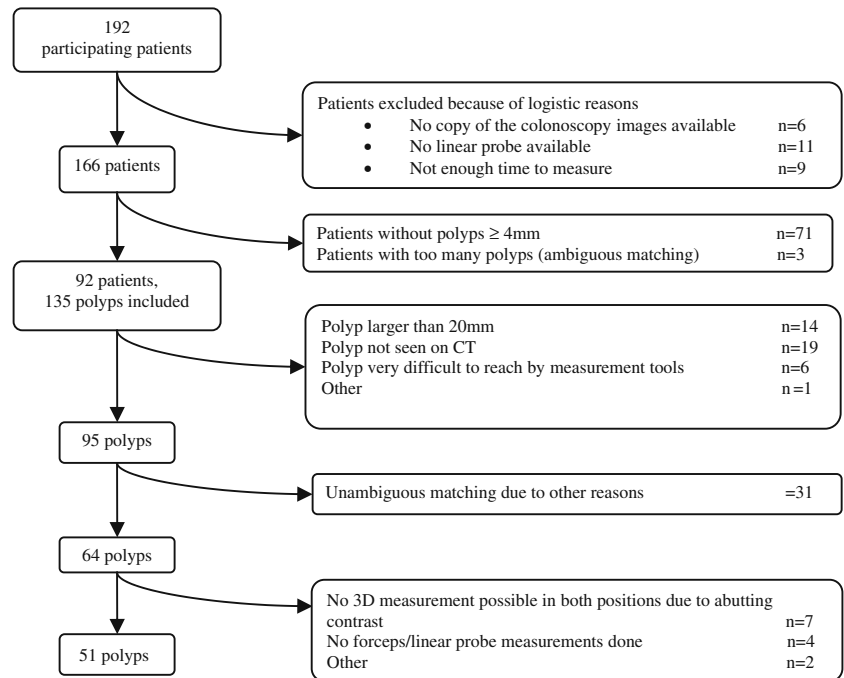
To illustrate the variability within the measurements of observers per technique and the variability within the measurements between techniques, we used Bland-Altman plots. In these plots we have corrected for systematic differences between observers as we assumed that these differences could be corrected by calibrating the measurements.

All analyses were performed using the linear mixed model procedure of a commercially available statistical software program (proc mixed, SAS Institute 9.2, Cary, USA). P values of < 0.05 were considered statistically significant.

Results

Polyp characteristics

In this study, 192 patients with 309 polyps of 4 mm or larger and 20 mm or smaller detected by optical colonoscopy were included (Fig. 2). Fifty-one polyps (17% of 309 polyps) in 44 patients fulfilled the selection criteria. Fifteen polyps of study I [15] and 36 polyps of study II [14] were

Fig. 2 Flowchart of included polyps

included. The mean age of the participating patients was 61 years (SD 7.2) and consisted of 27 (66%) men. Thirty-seven polyps revealed an adenomatous histology, seven a hyperplastic histology and for six polyps the histology could not be retrieved. One carcinoma was included. Twenty-four polyps were classified by the executing endoscopist as sessile, six polyps as flat (according to the Paris criteria [19]) and 21 polyps as pedunculated. Ninety percent of the visual assessment movie extracts, 85% of the forceps measurement movie extracts and 74% of the linear probe measurement movie extracts were rated as at least “sufficient” by the gastroenterologists. The median duration of the extracts was 33 s (P25-P75: 21–52 s).

The average polyp size (as measured by the different measurement techniques) is displayed in Table 1. The table

Table 1 Mean (SD) polyp size according to different measurement methods

Measurement method	Average size (cm)
Optical colonoscopy	
Visual assessment	8.2 (4.3)
Forceps	8.1 (3.7)
Linear probe	7.4 (3.8)
CT colonography	
2D abdominal window	8.8 (3.5)
2D intermediate window	9.7 (3.6)
3D manual	9.3 (3.3)
3D semi-automatic	9.7 (3.6)

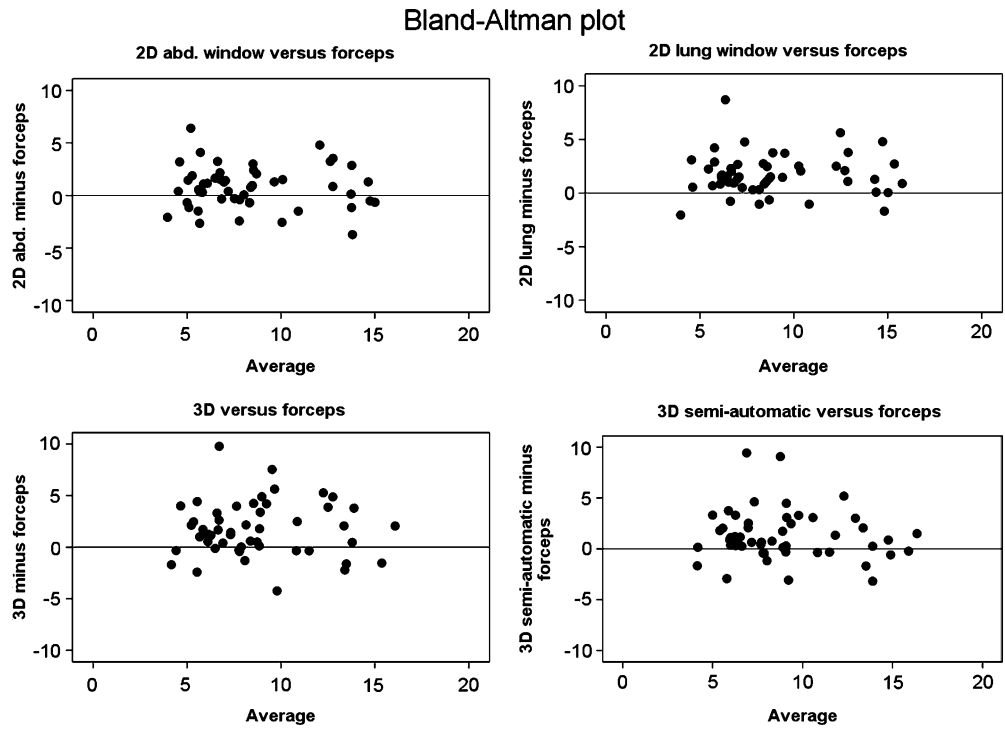
shows that CT colonography tends to produce larger polyp size measurements than optical colonoscopy, especially when using 2D intermediate window and semi-automatic 3D measurements.

The difference in size measurement between the most frequently used optical colonoscopy measurement techniques (i.e. forceps and visual assessment) and the various CT colonography measurements is illustrated by the Bland-Altman plots in Figs. 3 and 4. The figures show that CT produces larger polyp size measurements than forceps (Fig. 3) and visual assessment (Fig. 4).

Using the linear regression model we determined (1) the systematic differences in size measurement among techniques and (2) the variability among techniques. The results are presented in Table 2.

Systemic difference between techniques Table 2 shows that measurements by visual assessment were systematically 0.69 mm larger than measurements by linear probe, and measurements by forceps were systematically 0.71 mm larger than measurements by linear probe. Visual assessment and measurement by forceps did not differ. All four CT colonography techniques produced larger polyp size measurements than the three optical colonoscopy techniques, ranging from 0.65 mm (visual assessment versus 2D abdominal) to 2.32 mm (linear probe versus 2D intermediate window). These differences were highly statistically significant. In general, the differences between CT and optical colonoscopy were smallest when polyps were measured in a 2D abdominal window setting and largest if the measurements were performed in a 2D intermediate window setting or a 3D semi-automatic mode.

Fig. 3 Bland-Altman plot of polyp measurements (in mm) performed with CT versus forceps. Most of the dots of the plots are situated above the horizontal axis. This illustrates that CT most frequently measures polyp size larger than forceps



Variability among techniques We always applied the heterogeneous error variance structure. In Table 2 the variability of the various measurement techniques is displayed. Visual assessment showed more variability compared with forceps and linear probe measurements; the

latter two did not differ. Therefore, linear probe measurements are not more precise than measurements by forceps.

Optical colonoscopy techniques in general showed more variability compared with CT colonography techniques. Among the CT colonography measurements, manual 3D

Fig. 4 Bland-Altman plot of polyp measurements (in mm) taken with CT versus visual estimation. Most of the dots of the plots are situated above the horizontal axis. This illustrates that CT most frequently measures polyp size larger than visual estimation

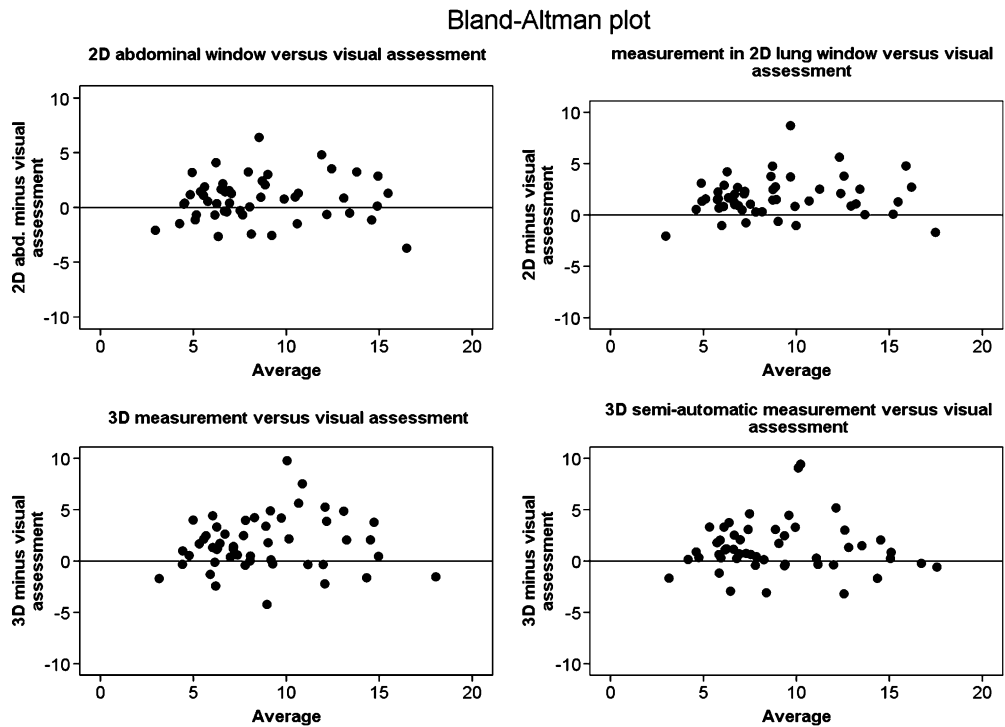


Table 2 Systematic differences in size measurements between techniques and estimates of inter-observer variability within techniques as estimated by linear mixed models

Optical colonoscopy, technique 1	CT-colonography, technique 2	Mean systematic difference in size measurement between techniques (SD)	<i>P</i>	Variability of measurements within technique 1	Variability of measurements within technique 2	Difference in variance	<i>P</i>
Visual assessment	Forceps	0.03 (−0.53; 0.58)	0.92	7.53	4.42	3.11	0.0052
Visual assessment	Linear probe	0.71 (0.17; 1.26)	0.01	7.53	3.94	3.59	0.0005
Forceps	Linear probe	0.69 (0.22; 1.15)	0.004	4.42	3.94	0.48	0.24
Visual assessment	2D abdominal window	−0.65 (−1.17; −0.14)	0.01	7.53	2.87	4.65	<0.0001
Visual assessment	2D intermediate window	−1.61 (−2.10; −1.12)	<0.0001	7.53	1.97	5.56	<0.0001
Visual assessment	3D	−1.19 (−1.68; −0.70)	<0.0001	7.53	2.02	5.51	<0.0001
Visual assessment	3D semi automatic	−1.59 (2.10; −1.08)	<0.0001	7.53	2.68	4.85	<0.0001
Forceps	2D abdominal window	−0.68 (−1.11; −0.25)	0.002	4.42	2.87	1.55	0.0455
Forceps	2D intermediate window	−1.63 (−2.04; −1.23)	<0.0001	4.42	1.97	2.45	<0.0001
Forceps	3D	−1.21 (−1.62; −0.81)	<0.0001	4.42	2.02	2.40	0.0002
Forceps	3D semi automatic	−1.62 (−2.05; −1.19)	<0.0001	4.42	2.68	1.74	0.027
Linear probe	2D abdominal window	−1.37 (−1.79; −0.95)	<0.0001	3.94	2.87	1.06	0.29
Linear probe	2D intermediate window	−2.32 (−2.71; −1.93)	<0.0001	3.94	1.97	1.97	<0.0001
Linear probe	3D	−1.90 (−2.29; −1.51)	<0.0001	3.94	2.02	1.92	0.0022
Linear probe	3D semi automatic	−2.31 (−2.72; −1.89)	<0.0001	3.94	2.68	1.26	0.16

measurements and measurement in a 2D intermediate window showed the least variability. Thus, these techniques lead to fewer differences between observers.

Differences between observers within techniques These differences were used to calculate the adjusted measurements, i.e. adjusted for the systematic measurement differences between observers. The adjusted measurements are illustrated in Fig. 5.

The spread of dots in the Bland-Altman plots in Fig. 5 represents the variability of the measurement techniques. The variability of the optical colonoscopy techniques is larger than that for the CT colonography techniques. The manual 3D measurement technique and the 2D measurement in an intermediate window show the least variability.

Discussion

This study shows that polyp size measured by linear probe is not less variable than polyp size measured by forceps. The endoscopists measured polyps significantly larger when using a forceps or by visual estimation compared with using a linear probe. CT colonography observers measured polyps larger than endoscopists. Furthermore,

measurement differences between observers in CT colonography were smaller, especially in manual 3D measurements and measurements in a 2D intermediate window, compared with measurements by optical colonoscopy.

Measurement studies are hampered by the lack of a reference standard of sufficient quality. Frequently used standards are measurement by a sliding caliper after polyp removal, and endoscopic measurement by a forceps or linear probe [10–13, 20–22].

A sliding caliper is not ideal. Vascular collapse, polyp desiccation from cautery, compression of the polyps after removal with a grasper or by suctioning through the endoscope may all contribute to a reduction in size of the reference polyp [21]. Because of the underestimation of polyp size and the fact that mainly pedunculated polyps can be removed in toto (bias), we did not consider these measurements to be reference values in this study. On the other hand, endoscopy is not an ideal technique either [23, 24] because the maximum diameter cannot always be displayed perpendicular to the direction of view and endoscope images with a wide viewing angle may lead to optical distortion. Moreover, peristalsis and tortuosity of the colon will reduce the possibility of properly assessing polyp size. This may lead to imperfect polyp size measurement. An in-vitro reference standard can be

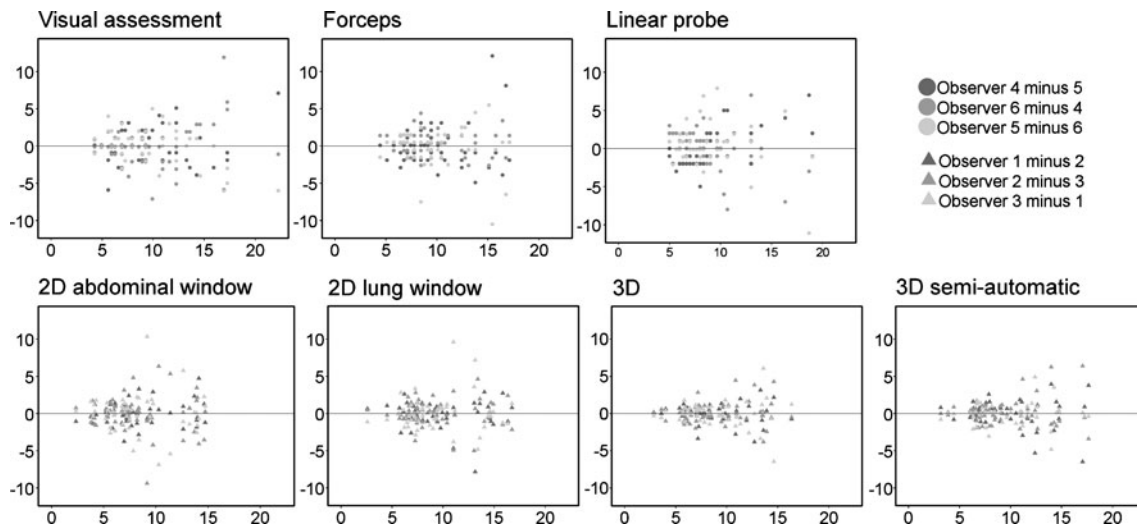


Fig. 5 Bland-Altman plots of the various measurement techniques (in mm). The distribution of the dots around the horizontal axis ($y=0$ mm) illustrates the variability of the measurements. The variability of the CT measurements is smaller than that of the optical colonoscopy measurements

constructed very precisely. However, when using a phantom, the influence of factors such as colonic distension and difficult viewing angles or bowel movement on polyp measurement is difficult to assess. Therefore the accuracy of different measurement techniques applied in a clinical situation may be overestimated.

Because of the abovementioned drawbacks of comparing measurements with a suboptimal reference standard, we have chosen a different approach by determining variability and systematic differences among various measuring techniques. We assumed that the variability is an indication of the accuracy of the measurement technique: the smaller the difference between the observers, the better the technique [17]. However, low variability does not rule out a large systematic error with regard to the truth.

Conflicting results about the accuracy of CT colonography measurements in in-vivo measurement studies [10–12] have been published. In these studies, in which optical colonoscopy was used as a reference standard, underestimation [10, 11] and overestimation [12] of polyp size by 2D and 3D CT colonography measurements have been reported. The reference standard in these studies was a measurement by either linear probe or forceps during optical colonoscopy.

Our findings may explain in part the differences among previous studies that used the measurements of a gastroenterologist as the reference standard. The fixed observer effects of our full model revealed that systematic differences between observers exist. The systematic differences were larger for optical colonoscopy measurements (up to 3 mm, data not shown) than for CT colonography measurements (up to 1.1 mm, data not shown). Using this model we assumed that systematic differences between observers can be avoided or repaired in practice by training. Therefore, the operator dependency of optical

colonoscopy measurements may be an explanation of the different conclusions drawn from the comparative measurement studies.

Our study showed that polyp size is larger when measured by CT colonography techniques than by optical colonoscopy techniques. As a consequence, in practice patients would more easily be referred for optical colonoscopy. We have shown that the magnitude of the difference depends on the techniques compared, i.e. the mean differences were smallest in 2D abdominal window measurements and largest when using 2D intermediate window settings or 3D semi-automatic measurements (Table 2 and Figs. 3 and 4). However, the difference also depended on observers as structural differences between observers exist, i.e. some observers systematically produce larger polyp measurements than others. Therefore, it is not possible to determine a general adjusted CT threshold for referral.

The endoscopists often experienced positioning of the linear probe as being more difficult than the positioning of a forceps. This may be due to the eccentric mounting of the probe (Fig. 1). As a consequence the measurement often had to be performed in the periphery of the field of view, which made measurements more difficult. Therefore using a linear probe often produces a semi-subjective estimate instead of an exact measurement.

This may explain the fact that the tool does not reduce differences in measurement between gastroenterologists. Moreover, the differences in measurement between gastroenterologists using a linear probe would be probably even larger if we had not excluded polyps that measured larger than 20 mm (i.e. the length of the linear probe). This has created a bias against CT colonography, which can more easily measure these types of large lesions. However, the referral strategy for polyps of 20 mm or polyps larger than 20 mm is the same, i.e. optical colonoscopy.

Just a few studies have evaluated the variability of different techniques. Fennerty et al. showed in a phantom study that polyp size measured using a forceps significantly differed between observers [23]. To our knowledge no studies have been published that evaluated variability between equivalent observers in both CT and optical colonoscopy measurement techniques for the same polyps.

This study had limitations. In this study we did not have a true reference standard. We have chosen this approach because of the abovementioned arguments. However, our data match with the data of Punwami et al. and Park et al. [8, 9] in the sense that endoscopists measured polyps smaller than does CT colonography.

The optical colonoscopy observers examined a video. Therefore they could not influence the movement of the endoscopic camera, the administration of additional bowel relaxants or the insufflation of additional air. By asking the actual executing colonoscopists (not the three optical colonoscopy observers) to not only position the measurement tool but also to try to measure polyp size, we have aimed to maximise the quality of the measurement movie.

Their commitment is reflected by the median captured video length of 33 s for the measurement. Despite this commitment, the quality of 9% of the visual assessment movie extracts and 26% of the linear probe measurement movie extracts was rated as “insufficient” by our observers. In our opinion this reflects the difficulty of proper placement of the endoscope and measurement tools for measurement purposes in practice.

We have not used the measurements that were done during the initial optical colonoscopy. Since each polyp measurement was subsequently executed by a single gastroenterologist, the previous measurements of the same polyp by this gastroenterologist would be severely biased by recall bias.

We used a single dedicated workstation for CT colonography polyp measurement. For 3D measurement we used a surface rendered reconstruction. The images were reconstructed with a threshold of 650 HU. According to Park et al., the optimum surface-rendering threshold

value for accurate polyp measurement is approximately –500 HU. This may have led to a (marginal) over-estimation of polyp size using 3D measurement by approximately 0.3 mm [25].

We used an interval of at least 1 day between the clustered CT colonography polyp measurements. Although there is the possibility of recall bias using this relatively short interval, the large number of randomised measurements (hundreds per observer) and the fact that the observers were not aware of the actual size measurement in millimetres may have ruled out recall bias to a large extent.

Polyps were included from two ongoing studies. These studies used a different bowel preparation scheme. Both studies used oral iodine contrast medium. One study combined this (laxative) tagging agent with 4 l of polyethylene glycol for same day colonoscopy. Since we only included polyps that were not covered by faecal material we assume that the type of bowel preparations did not influence measurements.

In this study the results were based on the measurements of only three experienced CT colonography observers and only three experienced colonoscopists. Although this is more than used in most measurement studies, it may limit the generalisability.

In conclusion, measurements by linear probe do not show less inter-observer variability than forceps measurements. Moreover, CT colonography shows less inter-observer variability and tends to produce larger polyp size measurements than optical colonoscopy.

Acknowledgements This study was supported by a research grant from Philips Medical Systems (PMS) (Best, the Netherlands). PMS was not involved in designing and conducting the study and did not have access to the data. PMS was not involved in analysing the data.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Halligan S, Altman DG, Taylor SA et al (2005) CT colonography in the detection of colorectal polyps and cancer: systematic review, meta-analysis, and proposed minimum data set for study level reporting. *Radiology* 237:893–904
- van Gelder RE, Birnie E, Florie J et al (2004) CT colonography and colonoscopy: assessment of patient preference in a 5-week follow-up study. *Radiology* 233:328–337
- Zalis ME, Barish MA, Choi JR et al (2005) CT colonography reporting and data system: a consensus proposal. *Radiology* 236:3–9
- McFarland EG, Levin B, Lieberman DA et al (2008) Revised colorectal screening guidelines: joint effort of the American Cancer Society, U.S. Multisociety Task Force on Colorectal Cancer, and American College of Radiology. *Radiology* 248:717–720
- Atkin WS, Morson BC, Cuzick J (1992) Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *N Engl J Med* 326:658–662
- O'Brien MJ, Winawer SJ, Zauber AG et al (1990) The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. *Gastroenterology* 98:371–379

7. Kim DH, Pickhardt PJ, Taylor AJ et al (1997) CT colonography versus colonoscopy for the detection of advanced neoplasia. *N Engl J Med* 357:1403–1412
8. Park SH, Choi EK, Lee SS et al (2007) Polyp measurement reliability, accuracy, and discrepancy: optical colonoscopy versus CT colonography with pig colonic specimens. *Radiology* 244:157–164
9. Punwani S, Halligan S, Irving P et al (2008) Measurement of colonic polyps by radiologists and endoscopists: who is most accurate? *Eur Radiol* 18:874–881
10. Burling D, Halligan S, Taylor S et al (2006) Polyp measurement using CT colonography: agreement with colonoscopy and effect of viewing conditions on interobserver and intraobserver agreement. *AJR Am J Roentgenol* 186:1597–1604
11. Jeong JY, Kim MJ, Kim SS (2008) Manual and automated polyp measurement comparison of CT colonography with optical colonoscopy. *Acad Radiol* 15:231–239
12. Pickhardt PJ, Lee AD, McFarland EG, Taylor AJ (2005) Linear polyp measurement at CT colonography: in vitro and in vivo comparison of two-dimensional and three-dimensional displays. *Radiology* 236:872–878
13. van Wijk C, Florie J, Nio CY et al (2008) Protrusion method for automated estimation of polyp size on CT colonography. *AJR Am J Roentgenol* 190:1279–1285
14. Liedenbaum MH, van Rijn AF, de Vries AH et al (2009) Using CT colonography as a triage technique after a positive faecal occult blood test in colorectal cancer screening. *Gut* 58:1242–1249
15. de Vries AH, Jensch S, Liedenbaum MH et al (2009) Does a computer-aided detection algorithm in a second read paradigm enhance the performance of experienced computed tomography colonography readers in a population of increased risk? *Eur Radiol* 19:941–950
16. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
17. Quinn MF (1989) Relation of observer agreement to accuracy according to a two-receiver signal detection model of diagnosis. *Med Decis Mak* 9:196–206
18. Akaike H (2008) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723
19. [No authors listed] (2003) The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 58:S3–43
20. Gopalswamy N, Shenoy VN, Choudhry U et al (1997) Is in vivo measurement of size of polyps during colonoscopy accurate? *Gastrointest Endosc* 46:497–502
21. Morales TG, Sampliner RE, Garewal HS, Fennerty MB, Aickin M (1996) The difference in colon polyp size before and after removal. *Gastrointest Endosc* 43:25–28
22. Schoen RE, Gerber LD, Margulies C (1997) The pathologic measurement of polyp size is preferable to the endoscopic estimate. *Gastrointest Endosc* 46:492–496
23. Fennerty MB, Davidson J, Emerson SS, Sampliner RE, Hixson LJ, Garewal HS (1993) Are endoscopic measurements of colonic polyps reliable? *Am J Gastroenterol* 88:496–500
24. Margulies C, Krevsky B, Catalano MF (1994) How accurate are endoscopic estimates of size? *Gastrointest Endosc* 40:174–177
25. Park SH, Choi EK, Lee SS et al (2008) Linear polyp measurement at CT colonography: 3D endoluminal measurement with optimized surface-rendering threshold value and automated measurement. *Radiology* 246:157–167