

S. Punwani
S. Halligan
P. Irving
S. Bloom
A. Bungay
R. Greenhalgh
J. Godbold
S. A. Taylor
D. G. Altman

Measurement of colonic polyps by radiologists and endoscopists: Who is most accurate?

Received: 19 July 2007
Revised: 28 September 2007
Accepted: 14 November 2007
Published online: 4 January 2008
© European Society of Radiology 2007

S. Punwani · S. Halligan (✉) ·
R. Greenhalgh · J. Godbold ·
S. A. Taylor
Department of Specialist Radiology,
Podium Level 2,
University College Hospital,
235 Euston Road,
London NW1 2BU, UK
e-mail: s.halligan@ucl.ac.uk
Tel.: +44-20-73809010
Fax: +44-20-73809068

P. Irving · S. Bloom · A. Bungay
Department of Gastroenterology,
University College Hospital,
London, UK

D. G. Altman
Centre for Statistics in Medicine,
Wolfson College Annexe,
Oxford, UK

Abstract The purpose was to determine the accuracy of polyp measurement by endoscopy and CT. A colonic phantom was constructed containing 12 simulated polyps of known diameter. Polyp diameter was estimated during endoscopy by two observers independently. The phantom was then scanned using a 64-detector-row machine and diameter estimated by a further two observers independently, using 2D and 3D visualisation methods. All measurements were obtained twice. Bland-Altman statistics were used to assess agreement between observers' estimates and the reference diameter. The mean difference between observers' measurements and the reference diameter was smallest for estimates made using 3D CT (−0.09 mm and −0.03 mm) and greatest for endoscopy (−1.10 mm and −1.19 mm), with 2D CT intermediate.

However, 95% limits of agreement were largest for 3D CT estimates (−4.38 mm to 4.20 mm). Estimates by 2D CT consistently overestimated polyp diameter, whereas endoscopy consistently underestimated diameter. In contrast, measurements by 3D CT were a combination of over- and under-estimates, with a tendency for disagreement to increase with the size of the polyp. The effect of observer experience was small and repeatability was best for 2D CT. Measurement error was encountered with all three modalities tested. Estimates made by 2D CT were believed to offer the best compromise overall.

Keywords Colon · CT · CT colonography · Colon · Cancer · Colonoscopy · Measurement

Introduction

It is now generally accepted that most cases of colorectal cancer arise from pre-existing benign adenomatous polyps. While only a tiny proportion of adenomas will ever become malignant, there is no precise method to identify in advance those that will do so. The best surrogate for potential malignancy is maximal transverse diameter; the risk of established cancer in a polyp measuring 1 cm is less than 1% versus 50% for those 2 cm or larger [1, 2]. Accurate measurement of maximal polyp diameter is therefore central to patient management strategies, especially those using CT colonography, where there is no opportunity for polypectomy at the time a polyp is detected [3].

Previous studies have attempted to determine the accuracy with which polyps are measured at CT colonography [4–6]. Measurement during subsequent colonoscopy, often by comparison with adjacent open biopsy forceps, has been used as the reference standard, but it is known that this approach can be inaccurate: A study of 100 polyps measured by five methods found that comparison with adjacent biopsy forceps was least accurate [7]. Although in-vivo studies are generally preferred to those performed in vitro, the lack of a reliable reference standard hampers accurate assessment of measurements made by CT colonography. In such cases a strong argument can be made for phantom studies where, although simulated, the diameter of a polyp can be determined with certainty. This

study aimed to determine the accuracy of polyp diameter measurement for both colonoscopy and CT colonography via use of a synthetic phantom containing simulated spherical polyps of a known reference diameter.

Methods

Our local ethical committee did not require permission for phantom studies.

A colonic phantom was constructed using a 1-m length of corrugated plastic tubing of 3-cm diameter. The tubing was opened longitudinally and 12 simulated polyps of different diameters sited along its length in random order, using surgical suture material. The simulated polyps were spherical and made of wood. The maximal transverse diameter of each simulated polyp was measured in two dimensions using a micrometer accurate to 0.01 mm. Diameters to the nearest mm were 4 mm, 6 mm, 8 mm, 10 mm, 12 mm and 18 mm. The polyps were colour-coded to facilitate data collection during subsequent endoscopy. Two polyps of each diameter were used. After the simulated polyps had been placed, the tubing was resealed along its length by suturing, approximated into a "U" shape, and fixed to a cork board (Fig. 1).

The phantom was then endoscoped (Fig. 2) by two endoscopists of differing experience. The more experienced was a gastroenterologist working at consultant level, who had been performing colonoscopy in routine clinical practice for 15 years. The less experienced endoscopist was a trainee with 4 years of colonoscopy experience. Both

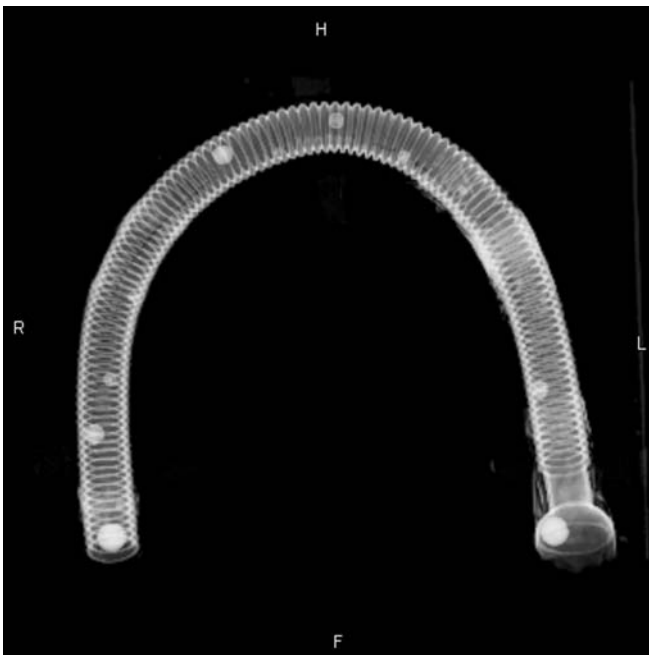


Fig. 1 CT scout view of the phantom containing 12 simulated polyps of varying maximal diameter

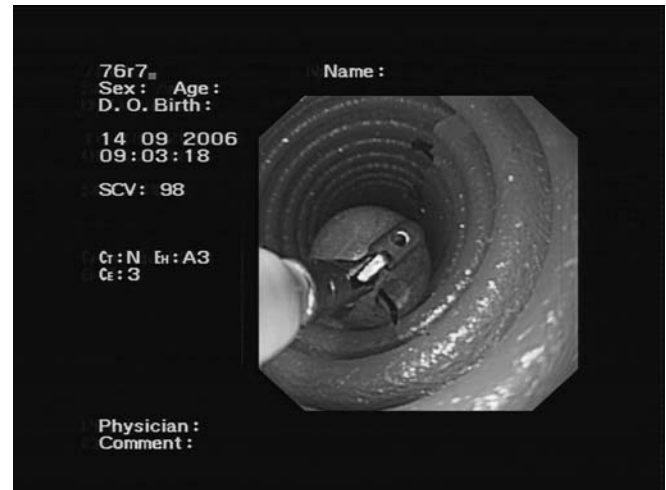


Fig. 2 Endoscopic view of a simulated polyp being measured by the open biopsy forceps technique

endoscopists were asked to intubate the phantom using a standard colonoscope (Olympus GIF XQ 260) and to estimate the maximal transverse diameter of each simulated polyp encountered as precisely as possible, via comparison with opened adjacent biopsy forceps (EndoJaw Olympus Disposable Biopsy Forceps), which was the technique used in their day-to-day clinical practice. The endoscopists were aware that the phantom contained 12 polyps because the study aimed to investigate measurement accuracy rather than detection, but they were unaware of the range of diameters or their reference measurements. The estimated diameter was communicated to a research fellow who could also see the colour of the polyp on the endoscope monitor, facilitating the matching of the estimate to the reference diameter. Data were recorded by the research fellow on a data sheet. The endoscopists were unaware of each other's measurements. In order to determine intra-observer agreement, the whole set of measurements were obtained on two separate occasions, separated by at least 2 weeks to diminish any recall bias.

Following endoscopy, the phantom was scanned using a 64-multi-detector row CT scanner (Siemens Somatom 64, Siemens Medical Solutions, Berkshire, UK) using the following parameters: All 64 detector rows used; 0.6-mm collimation; 50-mAs reference tube current; 100 kV; rotation time 0.5 s; slice width 1.0 mm, rotation 26.9 mm, pitch factor 1.4, increment 0.7 mm, kernel B20f, field of view 50 cm; dose modulation active. Data were reconstructed into 1-mm slices using a smooth reconstruction algorithm. The phantom was positioned so that the proximal and distal ends were aligned with the z-axis of the scanner.

Following scanning, data were imported into a standalone CT colonography image analysis workstation (Viatronix 3D, Stony Brook, NY). The phantom was imaged by two radiologists of differing experience: The first was a gastrointestinal subspecialist with 10 years' experience at consul-

tant level and extensive personal experience of CT colonography in both research and clinical settings. The second was a trainee radiologist who had been familiarised with CT colonography and the reporting of examinations in day-to-day clinical practice over the preceding 3 months. Both radiologists were asked to measure the maximal transverse diameter of each polyp identified using two different visualisation methods: 3D endoluminal rendering and 2D multi-planar reformatting. For 3D assessments the observer chose an endoluminal viewpoint directly opposite the polyp and placed the software cursors across the perceived maximal transverse diameter of the polyp, taking care not to place the cursor beyond the boundary of the polyp (Fig. 3) [5]. The software manufacturer was contacted in advance of these measurements in order to seek advice and receive assurance that the method adopted was correct. For 2D measurements the observers were free to set the window width and level manually at whatever values they felt best demonstrated the maximal transverse diameter of the simulated polyp, mirroring their normal day-to-day clinical practice. The use of multi-planar reformat was allowed in order to arrive at the section that best demonstrated the maximal transverse diameter, as were magnification views, especially when the polyp being measured was small. The two radiologists were aware that the phantom contained 12 simulated polyps, but were unaware of the range and reference diameters and of each other's responses. Measurements were made to the nearest 0.1 mm. The 2D measurement of all polyps was performed at a single sitting. Each radiologist annotated a diagram of the phantom, which indicated the proximal and distal extent of the phantom and the approximate position of the 12 polyps. Observers noted the estimated diameter of each polyp encountered in turn from the proximal end of the phantom to the distal, in order

to facilitate the matching of each estimate with the reference measurement. After a temporal separation of 1 day (so that the subsequent 3D measurement was not influenced by knowledge of the prior 2D measurement) the measurements were repeated using the 3D endoluminal technique. Observers did not have access to their prior 2D measurements. After a delay of 2 weeks to diminish recall bias, all measurements were repeated in the same fashion by each observer to assess intra-observer agreement.

Statistical analysis

The mean difference between endoscopic and CT estimates of maximal polyp diameter and the reference measurement for each polyp was calculated with 95% limits of agreement using the Bland-Altman method [8]. Successive measurements made by the same observer were examined using the coefficient of repeatability, which is calculated as 1.96 times the standard deviations of the differences between successive measurements. This is based on the premise that the mean difference between repeated measurements should be zero if the method has good repeatability.

Results

Agreement with reference diameter

The mean differences between observers' estimates of polyp diameter and the reference diameter with 95% limits of agreement are summarised in Table 1, for each of the three modalities tested. The mean difference was smallest

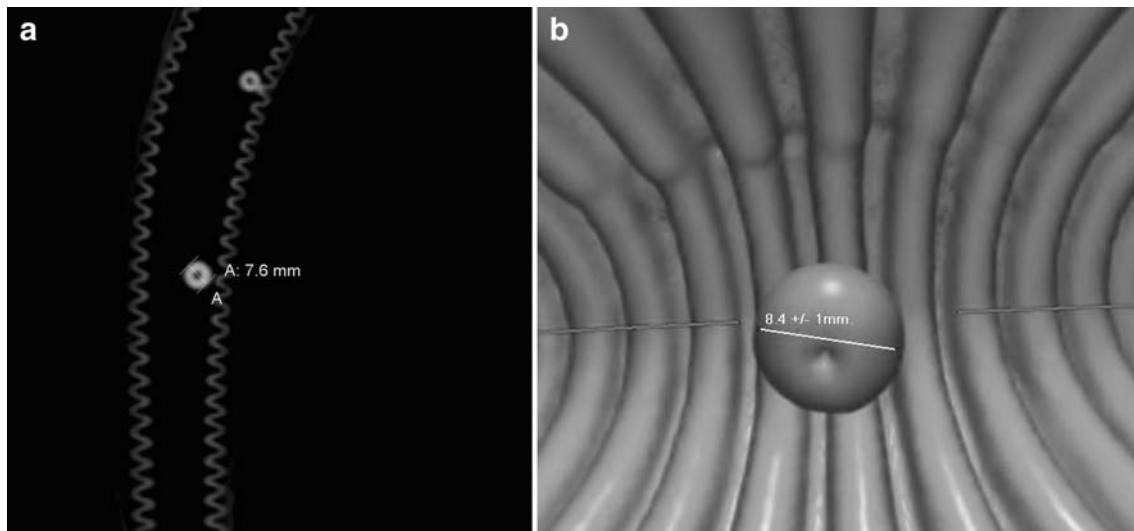


Fig. 3 **a** Measurement of a simulated polyp using a 2D multiplanar reformat. **b** Measurement of the same polyp depicted in Fig. 3a using the endoluminal 3D technique

Table 1 Mean difference between observers' estimates of polyp diameter and the reference diameter with 95% limits of agreement, for each observer and method of measurement

Observer and method used	Mean difference (mm)	95% limits of agreement (mm)
Experienced radiologist 2D CT display	0.63	-0.30 to 1.56
Trainee radiologist 2D CT display	0.32	-0.71 to 1.36
Experienced radiologist 3D CT display	-0.09	-4.38 to 4.2
Trainee radiologist 3D CT display	-0.03	-1.61 to 1.55
Experienced endoscopist	-1.10	-2.73 to 0.53
Trainee endoscopist	-1.19	-2.87 to 0.50

for estimates made using the 3D CT method and greatest for endoscopy, with 2D CT intermediate. However, the 95% limits of agreement were widest for 3D CT estimates. Bland-Altman plots revealed a clear tendency for estimates made using the 2D CT display to consistently overestimate polyp diameter when compared to the reference diameter, with the size of the overestimate being independent of the size of the polyp (Fig. 4). Measurements made using the 3D CT display were a combination of over- and under-estimates, with a tendency for the mean difference to increase in tandem with the size of the polyp (Fig. 5). In contrast to measurements made using the 2D CT display, estimates derived by endoscopy were consistently smaller

than the reference diameter, with the mean difference tending to be greater than estimates made using 2D CT, but again with no clear evidence that the error was related to the size of the polyp (Fig. 6).

Inter-observer agreement

The mean difference and 95% limits of agreement for measurements made by experienced and trainee observers, for each of the three modalities tested, are summarised in Table 2. In general, the mean difference between observers for measurements made using all three modalities was

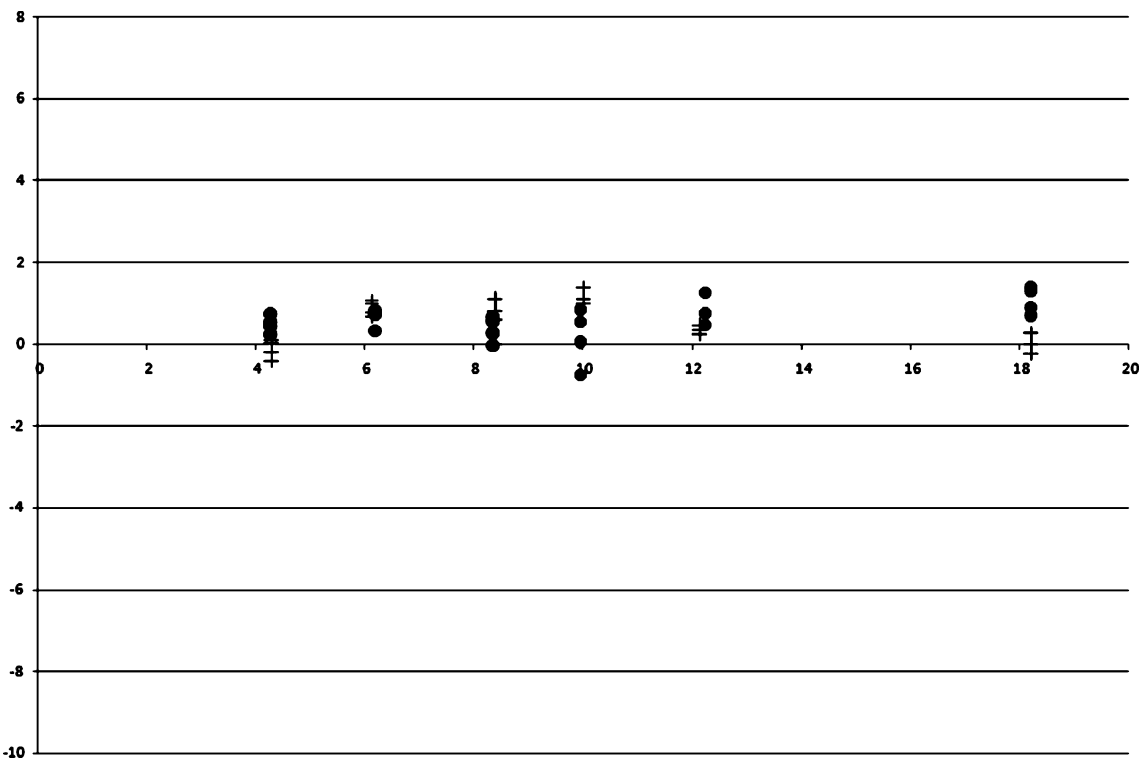


Fig. 4 Plot of polyp diameter estimated by the 2D CT method minus the reference diameter (y axis) against the reference diameter of the polyp (x axis). The plot reveals a clear tendency for this

method to overestimate polyp diameter (i.e. the difference between the two estimates is above 0)

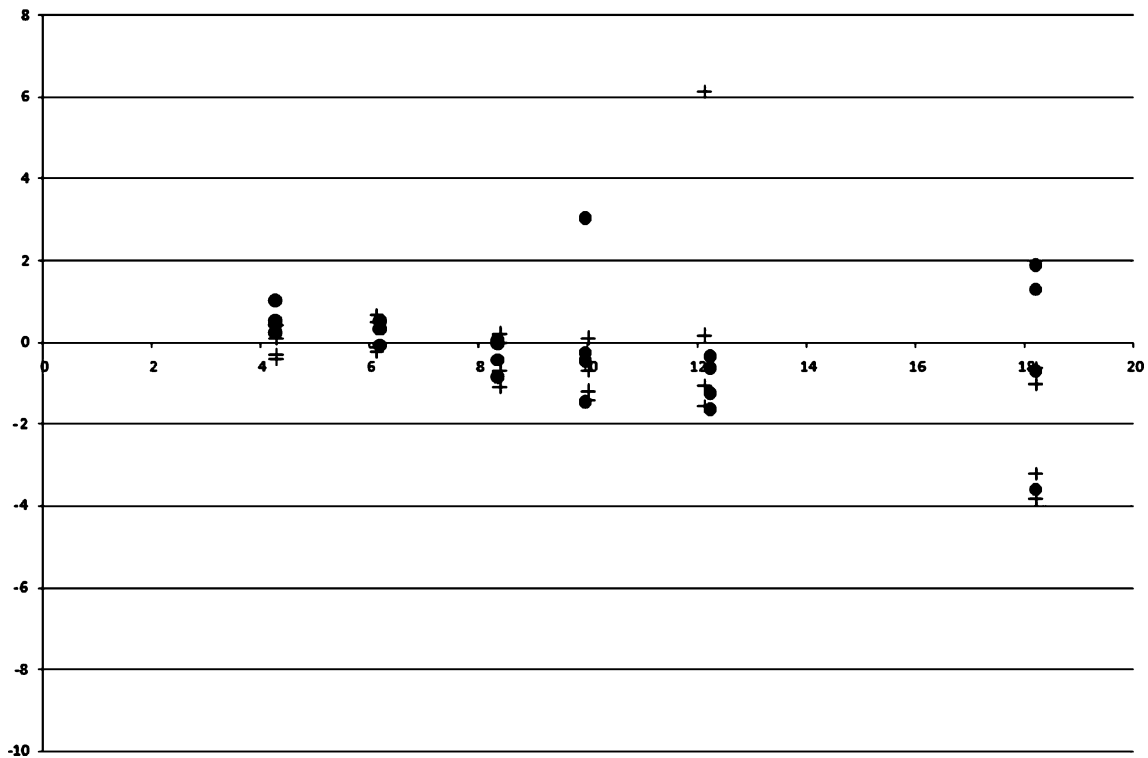


Fig. 5 Plot of polyp diameter estimated by the 3D CT method minus the reference diameter (y axis) against the reference diameter of the polyp (x axis). There are both over- and underestimates of the

true polyp diameter, and the most inaccurate measurements were made on larger polyps

small, being less than 1 mm in all cases. The narrowest limits of agreement were obtained for estimates made using the 2D CT display, with the widest being found for measurements made using the 3D CT display (Table 2).

Intra-observer agreement

The coefficient of repeatability for successive measurements made by the same observer was smallest for the 2D CT method (0.67 for the experienced observer and 0.99 for the trainee observer) and largest overall for endoscopy (4.27 for the experienced observer and 3.49 for the trainee observer). Measurements made using the 3D CT display showed the greatest difference between observers (4.95 for the experienced observer versus 1.51 for the trainee).

Discussion

CT colonography is being offered increasingly as an option for colorectal cancer screening programmes because advocates believe it combines acceptable sensitivity and specificity for significant polyps with good patient acceptability and safety [9]. However, unlike schemes that use endoscopy as the primary test, polypectomy is not an option—patients must be referred for subsequent endos-

copy if polyps detected by CT are to be removed. Because endoscopy, especially colonoscopy, is associated with a small but significant morbidity and even mortality, it is important that only those patients with significant polyps undergo subsequent procedures, notwithstanding the additional costs incurred. Because of this, patient management guidelines have been suggested that are based upon the maximal diameter of the largest polyp encountered in an individual patient [3, 10], and so accurate measurement of polyp diameter by CT assumes paramount importance. This is especially true because polyps are generally segregated into size-defined categories, with a 1-mm error being sufficient to move a polyp from one category to another when the measurement is near a category threshold [3].

The accuracy of CT estimates of polyp size has been assessed previously using both 2D and 3D CT display techniques [4–6, 11]. Pickhardt et al. [5] used acrylic spheres ranging in size from 6 to 13 mm, placed in a phantom immersed in a liquid-filled box in order to simulate tissue attenuation. Measurements were made using a window width of 2000 HU and level of 0. Endoluminal 3D measurements were found to be the most accurate, with a tendency for 2D measurements to underestimate maximal transverse diameter [5]. Taylor and colleagues measured 27 polyps using CT applied to a human colectomy specimen, with histological measure-

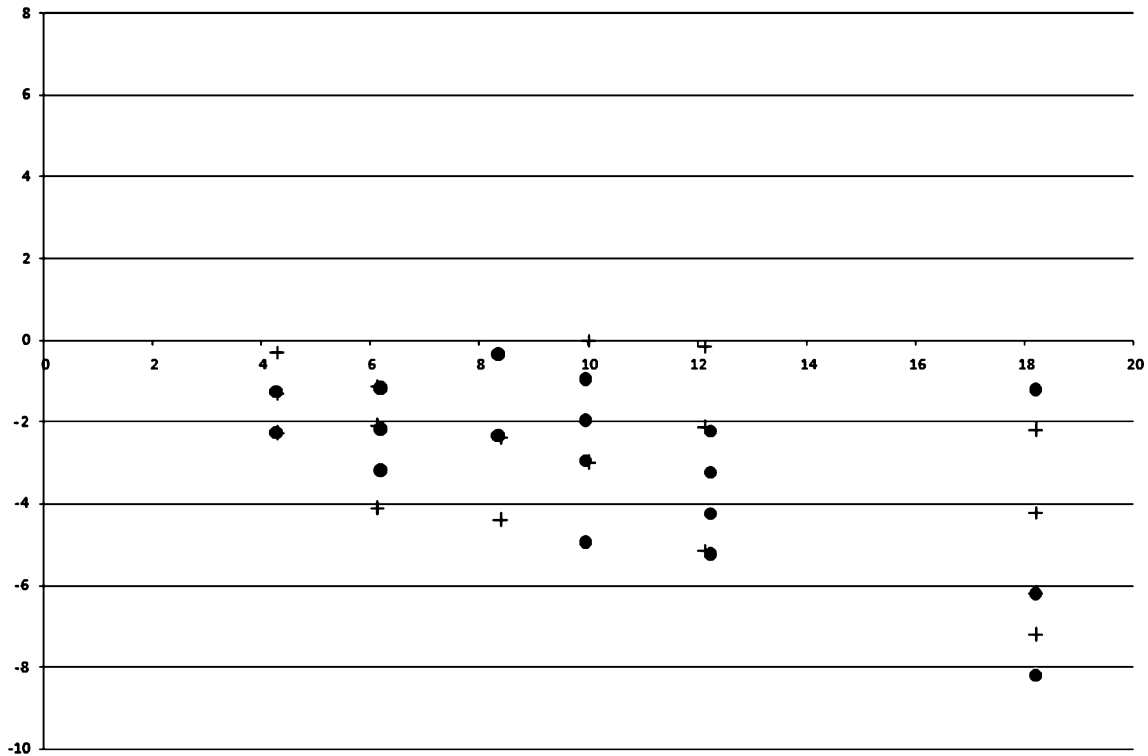


Fig. 6 Plot of polyp diameter estimated by endoscopy minus the reference diameter (y axis) against the reference diameter of the polyp (x axis). The plot reveals a clear tendency for endoscopy to

underestimate polyp diameter (i.e. the difference between the two estimates is below 0). Overall, the magnitude of disagreement is greater than that encountered using the 2D CT technique (Fig. 4)

ment as the reference standard, also finding manual 3D measurements superior to 2D [6]. However, for polyps smaller than 1 cm, measurement differences of up to 2.5 mm were within the expected limits of inter- and intraobserver agreement for all techniques studied [6]. In contrast, Burling and co-workers found that both 2D and 3D methods overestimated the true diameter, excepting 2D measurements made using abdominal window display settings [11]. In keeping with clinical practice, the authors used endoscopic measurements as their reference standard, but acknowledged that this approach could introduce bias because the precision of endoscopic estimates is questionable [11]. For example, previous studies have found that estimates of polyp diameter obtained via endoscopy may be inaccurate, especially when comparison with adjacent

open biopsy forceps is used [12–14]. Histological assessments are also subject to error, in particular due to shrinkage of the polyp following resection, volume changes when immersed in a preserving solution, and uncertainty regarding the true diameter, especially when excision margins are irregular [15, 16].

In order to circumvent these confounders, we used artificial non-deformable polyps for which the reference diameter could be established independently with precision. CT and endoscopic methods of measurement were then applied separately so that the accuracy of each could be judged. A recent article by Park and colleagues [17] used a porcine colectomy specimen to address this problem, with both CT and endoscopic estimates being compared with independent calliper measurements of simulated polyps. The authors found that CT showed better agreement than did endoscopy, but used intraclass correlation to assess agreement, an approach that inherits some of the same problems associated with simple correlation because it is unable to identify systematic bias between the two techniques being investigated [18]. In keeping with previous studies of colonoscopy [12–14], we found that endoscopic estimates were persistently smaller than the true diameter, and also found that this effect was independent of the diameter of the polyp being measured. Exactly why this effect should be so consistent is unknown, but could arise from the need to derive a 2-dimensional

Table 2 Mean difference and 95% limits of agreement for measurements made by experienced and trainee observers, for each of the three modalities tested

Method used	Mean difference; experienced observer minus trainee (mm)	95% limits of agreement
CT 2D display	0.31	–0.41 to 1.03
CT 3D display	–0.06	–5.11 to 5.00
Endoscopy	0.08	–2.24 to 2.04

diameter from a 3-dimensional image: Catlano and colleagues found that underestimation during endoscopy could be abolished by using a stereoscopic viewing system [19]. It is possible that underestimation may arise from peripheral distortion associated with newer, more wide-angle lenses. Furthermore, the long axis of a polyp must be oriented perpendicular to the viewing direction to prevent artificial shortening of polyps, something that is easier to achieve with CT than with endoscopy, especially when the long axis of a polyp is aligned with the long axis of the colon. Also, in contrast to CT methods, where the tools used allow polyps to be measured to fractions of a mm, endoscopists tend to report diameters as whole mm. We did not investigate whether these estimates tend to be rounded up or down by the observer, but our results suggest that adopting the former strategy might result in less measurement error.

Both of the CT methods we investigated had a tendency to overestimate the true size of polyps, presumably due to partial volume effects increasing the perceived diameter. This effect was consistent for the 2D analysis, but less so for 3D analysis, where some underestimations occurred and where the 95% limits of agreement became wider as the reference diameter increased. Endoluminal 3D assessments are technically more difficult to make than 2D estimates—the observer must face the polyp directly and care has to be taken to make sure the cursors do not “fall off” the polyp margin. Our findings suggest that performing this more complex procedure may introduce some inaccuracy when measuring large polyps, compared to 2D analysis. On average, the magnitude of diameter overestimation introduced by CT was less than the magnitude of underestimation encountered with endoscopy. While the mean difference reflects the difference between the reference diameter and that obtained by the method under consideration, the 95% limits of agreement reflect the variation in the precision of individual measurements. Overall, the narrowest limits of agreement were found for 2D CT estimates of diameter, suggesting to us that this method is the best compromise overall, especially as the measurement error introduced by this approach seems the most predictable.

We also investigated the magnitude of difference between estimates obtained by observers of differing experience. Similar mean differences and 95% limits of agreement were found between experienced and trainee observers when using both 2D CT and colonoscopy, suggesting that experience conveys no specific advantage.

Interestingly, the 95% limits of agreement for measurements obtained by 3D CT were wider for the experienced observer, implying that experience could be a disadvantage, perhaps because measurements were made more hastily. When measurements obtained by experienced and trainee observers were compared directly, the narrowest limits of agreement were found for the 2D CT method. We also investigated repeatability—the degree to which repeated measurements made by the same observer and method agree—and found the best results obtained with 2D CT, with endoscopy faring worst.

The present study does have limitations. Most obviously, only four observers were used, with each acting as a surrogate for their particular group (e.g. experienced radiologists). However, the number of participants is in keeping with previous research on the topic. Also, all CT measurements were made using a single software platform and endoscopy measurements were made using the same endoscopic apparatus, so that the effect of different systems was not investigated. Although we employed a phantom paradigm so that a reliable and independent reference standard could be established, our findings may not be directly applicable to measurement of real polyps *in vivo*. However, there is no *a priori* reason to suggest that our findings are not broadly generalisable into day-to-day clinical practice, although factors such as bowel preparation, polyp morphology (i.e. non-spherical), window setting, and distension will influence the accuracy of individual measurements [20]. It should also be noted that the simulated polyps were wooden rather than soft-tissue, with a density of approximately -370 Hounsfield units, which may affect the perceived diameter. Also, we examined the phantom in air rather than submerging it in fluid to simulate soft-tissue attenuation. The number of polyps measured by each observer was also relatively small, so we should be careful not to extrapolate our conclusions beyond the data unreasonably.

In conclusion, measurement error is encountered when the diameter of simulated polyps is estimated by colonoscopy, 2D CT display, and 3D endoluminal CT display. Overall, estimates made using the 2D CT display offered the best compromise.

Funding This work was undertaken at the Comprehensive Biomedical Centre, University College Hospital London, who received a proportion of funding from the National Institute for Health Research. The views expressed in this publication are those of the authors and not necessarily those of the UK Department of Health.

References

1. Morson BC (1974) Evolution of cancer of the colon and rectum. *Cancer* 34 (suppl 8):845–849
2. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL (1987) Natural history of untreated colonic polyps. *Gastroenterology* 93:1009–1013
3. Zalis ME, Barish MA, Choi JR et al (2005) CT colonography reporting and data system: a consensus proposal. *Radiology* 236:3–9
4. Burling D, Halligan S, Taylor SA, Honeyfield L, Roddie ME (2007) CT colonography: automatic measurement of polyp diameter compared with manual assessment - an in-vivo study. *Clin Radiol* 62:145–151
5. Pickhardt PJ, Lee AD, McFarland EG, Taylor SJ (2005) Linear polyp measurement at CT colonography: in vitro and in vivo comparison of two-dimensional and three-dimensional displays. *Radiology* 236:872–878
6. Taylor SA, Slater A, Halligan S, Honeyfield L, Roddie ME, Demeshki J, Amin H, Burling D (2007) CT colonography: automated measurement of colonic polyps compared with manual techniques-human in vitro study. *Radiology* 242:120–128
7. 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
8. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
9. Halligan S, Taylor SA (2007) CT colonography: results and limitations. *Eur J Radiol* 61:400–408
10. Taylor SA, Laghi A, Lefere P, Halligan S, Stoker J (2007) European Society of Gastrointestinal and Abdominal Radiology (ESGAR): consensus statement on CT colonography. *Eur Radiol* 17:575–579
11. 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. *American Journal of Radiology* 186:1597–1604
12. 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
13. Margulies C, Krevsky B, Catalano MF (1994) How accurate are endoscopic estimates of size? *Gastrointest Endosc* 40:174–177
14. Schwartz E, Catalano MF, Krevsky B (1995) Endoscopic estimation of size: improved accuracy by directed teaching. *Gastrointest Endosc* 42:292–295
15. 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
16. Schoen RE, Gerber LD, Margulies C (1997) The pathologic measurement of polyp size is preferable to the endoscopic estimate. *Gastrointest Endosc* 46:492–496
17. Park SH, Choi EK, Lee SS, Byeon JS, Jo JY, Kim YH, Lee KH, Ha HK, Han JK (2007) Polyp measurement reliability, accuracy, and discrepancy: optical colonoscopy versus CT colonography with pig colonic specimens. *Radiology* 244:157–164
18. Bland JM, Altman DG (1990) A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 20:337–340
19. Catalano MF, Van Dam J, Bedford R, Cothren RM, Sivak MV Jr (1993) Preliminary evaluation of the prototype stereoscopic endoscope: precise three-dimensional measurement system. *Gastrointest Endosc* 39:23–28
20. Taylor SA, Slater A, Honeyfield L, Burling D, Halligan S (2006) CT colonography: effect of colonic distension on polyp measurement accuracy and agreement-in vitro study. *Acad Radiol* 13:850–859