

Stuart A. Taylor
Steve Halligan
David Burling
Simon Morley
Paul Bassett
Wendy Atkin
Clive I. Bartram

CT colonography: effect of experience and training on reader performance

Received: 30 July 2003
Revised: 5 December 2003
Accepted: 6 January 2004
Published online: 10 February 2004
© Springer-Verlag 2004

S. A. Taylor · S. Halligan (✉)
D. Burling · S. Morley · C. I. Bartram
Department of Intestinal Imaging,
St. Mark's Hospital,
Watford Road, Northwick Park,
London, HA1 3UJ, UK
e-mail: s.halligan@ic.ac.uk
Tel.: +44-181-235-4000
Fax: +44-181-235-4001

S. Halligan · W. Atkin
Department of Cancer Research UK
Colorectal Cancer Unit,
St. Mark's Hospital,
Northwick Park, London,
HA1 3UJ, UK

P. Bassett
Department of Statistics,
St. Mark's Hospital,
Northwick Park, London,
HA1 3UJ, UK

Abstract The purpose of this paper was to investigate the effect of radiologist experience and increasing exposure to CT colonography on reader performance. Three radiologists of differing general experience (consultant, research fellow, trainee) independently analysed 100 CT colonographic datasets. Readers had no prior experience of CT colonography and received feedback and training after the first 50 cases from an independent experienced radiologist. Diagnostic performance and reporting times were compared for the first and second 50 datasets and compared with the results of a radiologist experienced in CT colonography. Before training only the consultant reader achieved statistical equivalence with the reference standard for detection of larger polyps. After training, detection rates ranged between 25 and 58% for larger polyps. Only the trainee significantly im-

proved after training ($P=0.007$), with performance of other readers unchanged or even worse. Reporting times following training were reduced significantly for the consultant and fellow ($P<0.001$ and $P=0.03$, respectively), but increased for the trainee ($P<0.001$). In comparison to the consultant reader, the odds of detection of larger polyps was 0.36 (CI 0.16, 0.82) for the fellow and 0.36 (CI 0.14, 0.91) for the trainee. There is considerable variation in the ability to report CT colonography. Prior experience in gastrointestinal radiology is a distinct advantage. Competence cannot be assumed even after directed training via a database of 50 cases.

Keywords Colonography · Computed tomographic/standards · Clinical competence/standards · Learning · Radiology/education

Introduction

Several studies suggest CT colonography is a robust technique for the detection of colorectal neoplasia in symptomatic patients [1–3]. However, there is a marked variation in reported diagnostic performance [4]. Sensitivity for even large colonic lesions (10 mm or greater) varies from just 50 to over 90% amongst studies using broadly similar CT protocols [1, 3, 5]. Although technical factors remain important [6], such variation raises the possibility that individual reader performance has a significant effect on the observed sensitivity of CT colonography.

Reader experience has been shown to significantly affect diagnostic performance for many imaging modalities, including mammography [7] and barium enema [8]. To date, there has been relatively little work on the effect of technique-specific training or overall radiologist experience on the diagnostic performance of CT colonography [9, 10]. McFarland and colleagues demonstrated that detection of large polyps ranged from 60 to 78%, even amongst experienced abdominal radiologists with equivalent colonographic training [11]. Furthermore, reader performance has been shown to continue to improve after just 25 cases [9, 10]. At the present time, there is no

consensus regarding what represents adequate reader training for CT colonography, probably because there is little evidence on which to base assumptions. The aim of this study was to investigate the effect of both radiologist experience and increasing exposure to CT colonography on reader performance.

Patients and methods

Between April 2001 and April 2002, a total of 168 consecutive adult patients (median age 65 years, range 34–89; 84 females) were recruited to an ongoing trial at our institution comparing CT colonography with conventional endoscopy. Our local ethical review committee approved the study, and all subjects gave informed written consent. Of the cohort, 59 were referred for flexible sigmoidoscopy via a rectal bleeding clinic, and the remaining 109 patients were referred for total colonoscopy because of a clinical suspicion of colorectal neoplasia.

CT colonography

CT colonography in all 168 patients was performed using a standard technique as previously described [12]. All patients underwent full bowel preparation with either two sachets of sodium picosulphate (Picolax, Ferring Pharmaceuticals, Berkshire, UK) (if scheduled for flexible sigmoidoscopy) or two sachets of magnesium citrate (Citramag, Pharmaserve, Manchester, UK) supplemented with one sachet of senna granules (Reckitt Benckiser Healthcare, Hull, UK) (if scheduled for colonoscopy). Scans were performed using a four detector row CT scanner (Lightspeed Plus, General Electric Medical Systems, Milwaukee, WI) utilising 1.25–2.5 mm collimation, pitch of six, rotation time 0.8 s; 120 kVp, 50–100 mA and 50% slice overlap.

Endoscopy

Immediately after CT colonography patients underwent endoscopy, performed by experienced endoscopists. The endoscopist recorded the size (estimated by direct comparison to adjacent open biopsy forceps) and location of any polyps using a report sheet designed for the study.

CT colonographic and endoscopic correlation

A single radiologist evaluated the CT datasets blinded to the endoscopic findings using a dedicated workstation with proprietary software (Advantage Windows 4.0 and Colonography, GE Medical Systems, Milwaukee, WI). A primary axial prone and supine 2-dimensional read were used with a surface rendered 3-dimensional endoluminal view used for “problem solving” [13]. In order to facilitate subsequent lesion identification the colon was divided into six segments as previously described [14] and the location of any lesion indicated by the radiologist on a line drawing of the colon incorporated into a report sheet identical to that used by the endoscopist. Details of polyp correlation between CT and endoscopy were as previously described [12]. The time for analysis was recorded.

Case selection and radiological reference standard

A non-observer selected two sets of 50 cases from the dataset of 168 patients. The datasets were chosen such that each set of

50 contained an approximately equal number of polyps of similar size based on the known endoscopic findings and were from cases in which endoscopy was complete. The original reporting radiologist reanalysed these 100 datasets with full access to his original CT report and the reference endoscopic findings. Any polyps visible only in retrospect (i.e. originally perceptual errors) were noted. This final unblinded radiological interpretation of the datasets by the experienced radiologist was used as the radiological reference standard for subsequent assessment of reader performance.

Reader selection and reading protocol

Three radiologists were selected to read the CT colonographic datasets. None had any prior experience of 3D imaging or reading CT colonography, but they differed in radiological experience with CT as follows: reader 1 was a consultant radiologist with a subspecialty interest in gastrointestinal imaging and 10-years experience of CT; reader 2 was a trainee holding the fellowship of the Royal College of Radiologists with an expressed subspecialty interest in gastrointestinal imaging and 3-years experience of CT; reader 3 was a trainee with 1-year experience with CT. Each reader was familiarised with the CT workstation by the experienced radiologist such that they were fully conversant with the functionality of the CT colonography software package, although no specific education was given as to interpretation of CT colonography. Each reader then independently analysed the first dataset of 50 patients in their own time over 3–4 weeks. Readers were unaware of the prevalence of abnormality or of the reason for referral and recorded their findings (including reporting time) on a sheet identical to that used in the main comparative trial between CT colonography and endoscopy. Readers were also asked to record their level of confidence for detected lesions using a 4-point scale, one being the least confident and four the most confident, although they were told that any level of confidence would count as a detected lesion. The experienced radiologist then compared the study sheets from each reader with the endoscopic findings. Each reader then individually underwent education from the experienced radiologist via a case-by-case review of the first dataset of 50 patients. Any mistakes made were pointed out and detailed instruction on the CT characteristics of true positives and false positives were given freely. Detailed advice regarding reader strategies, for example appropriate window settings, use of prone and supine correlation and application of 3-dimensional endoluminal views, was given. Readers were encouraged to seek clarification of any specific issues encountered during their analysis of the first 50 cases.

After this training, each reader analysed the second dataset of 50 patients over a further 3–4 weeks, again recording their findings, reporting time and level of confidence as before. Although aware of the strategy used by the experienced radiologist, readers were free to adopt whatever strategy they felt best. As before, the experienced reader analysed the study sheets from each reader and calculated their detection rate on a per polyp basis, with false positives noted on a per patient basis.

Statistical analysis

The performance of each of the three less experienced radiologists was compared to the radiological reference standard for both the first and second 50 cases. Polyps were divided into three categories; “small” (defined as 1–5-mm diameter), “medium” (defined as 6–9-mm diameter) and “large” (defined as 10 mm or larger, but excluding cancers). Comparison of true positives was per-

formed on a per polyp basis using a paired binomial exact test. Analysis of the false positives was performed on a per patient basis such that each patient was defined as having one or more false positives, or no false positives, for each individual radiologist, and again analysed using a paired binomial exact test. The data were then subdivided into two groups (polyps with a size of ≤ 5 mm and polyps with a size >5 mm) and analysis repeated.

A comparison of performance for the first and second 50 cases was made for each radiologist using Fisher's exact test. Finally the overall performance of the three inexperienced readers for all 100 cases was compared using logistic regression, adjusting for polyp size. Robust standard errors were used to allow for the fact that there were repeated observations on each polyp (i.e. the observations were not completely independent of each other). Results for this analysis were expressed as the odds of polyp detection for readers 2 and 3 relative to reader 1. Reporting times and confidence scores were compared using the Mann-Whitney statistic.

Results

Endoscopy detected a total of 48 polyps and 3 cancers in 20 patients from the first 50 cases and 54 polyps and 2 cancers in 24 patients from the second 50, giving a prevalence of abnormality of 40 and 48%, respectively. The endoscopic findings together with the radiological reference standard and detection rates for each reader for the first 50 cases are shown in Table 1, and for the second 50 cases in Table 2. No medium or large polyp was identi-

fied only on retrospective detection. However, there were six polyps larger than 5 mm that could not be detected, even in retrospect, by the experienced radiologist: three flat adenomas and three polyps within collapsed colon (two of which were in diverticular segments). The experienced radiologist detected a total of 11 small polyps (i.e. 5 mm or less) only on retrospective dataset analysis, one from the first dataset and ten from the second, and these were incorporated into the radiological reference standard.

For the first 50 patients, reader 1 performed best and reader 3 worst for all polyp sizes, when compared to the radiological reference standard. Polyp detection also increased in all categories with increasing polyp size for readers 1 and 3. Overall variation in detection rates was considerable, ranging from 6 to 41% for small polyps and 30–70% for large polyps (Table 1). Detection rates for the three readers for the second 50 patients did not improve following training (Table 2). For example, reader 1 detected 12% of small polyps compared to 41% previously, and reader 2 detected only 14% of large polyps compared to 60% previously. In contrast, reader 3 (who performed worst on the first 50 patients) did improve in all size categories (Table 2).

In terms of the overall number of polyps detected, all three readers were significantly worse than the reference standard for both sets of cases, mostly due to low detection of small polyps (Table 3). Only reader 1 achieved statistical equivalence to the reference standard for these small polyps ($P=0.07$), but only for the first 50 cases (Table 3). Divergence from the reference standard was less for larger polyps, although there was significantly

Table 1 Endoscopic findings and observer performance for the first 50 datasets

Lesion	Endoscopy	Reference radiologist detection (%)	Reader 1 detection (%)	Reader 2 detection (%)	Reader 3 detection (%)
Cancer	3	3 (100)	2 (67)	2 (67)	2 (67)
Small polyps	32	19 (59) ^a	13 (41)	9 (28)	2 (6)
Medium polyps	6	4 (67)	3 (50)	1 (17)	1 (17)
Large polyps	10	8 (80)	7 (70)	6 (60)	3 (30)
All polyps	48	31 (65)	23 (48)	16 (33)	6 (13)

^a Includes one original perceptual error, seen on retrospective analysis.

Table 2 Endoscopic findings and observer performance for the second 50 datasets

Lesion	Endoscopy	Reference radiologist detection (%)	Reader 1 detection (%)	Reader 2 detection (%)	Reader 3 detection (%)
Cancer	2	2 (100)	1 (50)	2 (100)	2 (100)
Small polyps	42	28 (67) ^a	5 (12)	6 (14)	13 (31)
Medium polyps	5	4 (80)	2 (40)	2 (40)	2 (40)
Large polyps	7	6 (86)	5 (71)	1 (14)	4 (57)
All polyps	54	38 (70)	12 (22)	9 (17)	19 (35)

^a Includes ten original perceptual errors, seen on retrospective analysis.

Table 3 Comparison of the performances for all three readers against the radiological reference standard for the first and second 50 cases

Polyp size	Reader	Difference in proportions detected (first 50) (reference, reader) (95% CI) ^a	Exact <i>P</i> -value (first 50)	Difference in proportions detected (second 50) (reference, reader) (95% CI) ^a	Exact <i>P</i> -value (second 50)
All polyps	1	0.17 (0.05, 0.29)	0.02	0.48 (0.35, 0.61)	<0.001
	2	0.31 (0.18, 0.44)	<0.001	0.54 (0.40, 0.67)	<0.001
	3	0.52 (0.38, 0.66)	<0.001	0.35 (0.21, 0.49)	<0.001
≤5 mm	1	0.19 (0.03, 0.34)	0.07	0.54 (0.40, 0.70)	<0.001
	2	0.31 (0.15, 0.47)	0.002	0.52 (0.37, 0.67)	<0.001
	3	0.53 (0.36, 0.70)	<0.001	0.36 (0.20, 0.52)	<0.001
>5 mm	1	0.13 (-0.04, 0.29)	0.50	0.25 (0.00, 0.49)	0.25
	2	0.31 (0.09, 0.54)	0.06	0.58 (0.30, 0.86)	0.02
	3	0.50 (0.26, 0.75)	0.008	0.33 (0.07, 0.60)	0.13

^aThe confidence interval for the difference in proportions is approximate due to the small number of observations in some cells.

Table 4 Comparison of diagnostic performance for the first and second 50 cases

Group	Radiologist	Proportion detected (first 50)	Proportion detected (second 50)	Difference in proportions (second, first) (95% CI) ^a	Exact <i>P</i> -value
All polyps	Reference	0.65	0.70	0.06 (-0.12, 0.24)	0.67
	Reader 1	0.48	0.22	-0.26 (-0.44, -0.08)	0.007
	Reader 2	0.33	0.17	-0.17 (-0.33, 0.00)	0.07
	Reader 3	0.13	0.35	0.23 (0.07, 0.38)	0.01
≤5 mm polyps	Reference	0.59	0.67	0.07 (-0.15, 0.29)	0.63
	Reader 1	0.40	0.12	-0.29 (-0.48, -0.09)	0.006
	Reader 2	0.28	0.14	-0.13 (-0.33, 0.05)	0.16
	Reader 3	0.06	0.31	0.25 (0.08, 0.41)	0.01
>5 mm polyps	Reference	0.75	0.83	0.08 (-0.21, 0.38)	0.67
	Reader 1	0.63	0.58	-0.04 (-0.41, 0.32)	0.99
	Reader 2	0.43	0.25	-0.19 (-0.53, 0.16)	0.43
	Reader 3	0.25	0.50	0.25 (-0.10, 0.60)	0.24

^aThe confidence interval for the difference in proportions may be approximated due to the small number of observations in some cells.



poorer performance by reader 3 during the first 50 cases (4 of 16 polyps detected vs. 12 of 16 for the reference standard, $P=0.008$) and by reader 2 during the second 50 cases (3 of 12 polyps vs. 10 of 12 polyps for the reference standard, $P=0.02$) (Fig. 1). Importantly, however, no reader detected more than 71% of large polyps in either case set and reader 3 detected just 57% of large polyps in the second 50 cases despite achieving statistical equivalence with the reference standard for detection of polyps 6 mm+ (Fig. 2). All three readers missed the same cancer in the first 50 cases (a flat lesion just proximal to the ileocaecal valve) (Fig. 3), although reader 1

◀ **Fig. 1** Transverse supine CT colonographic image from the second 50 datasets shows a large irregular polyp (arrow) in the ascending colon. The lesion was missed prospectively by reader 2 but correctly identified by both other readers and the reference radiologist

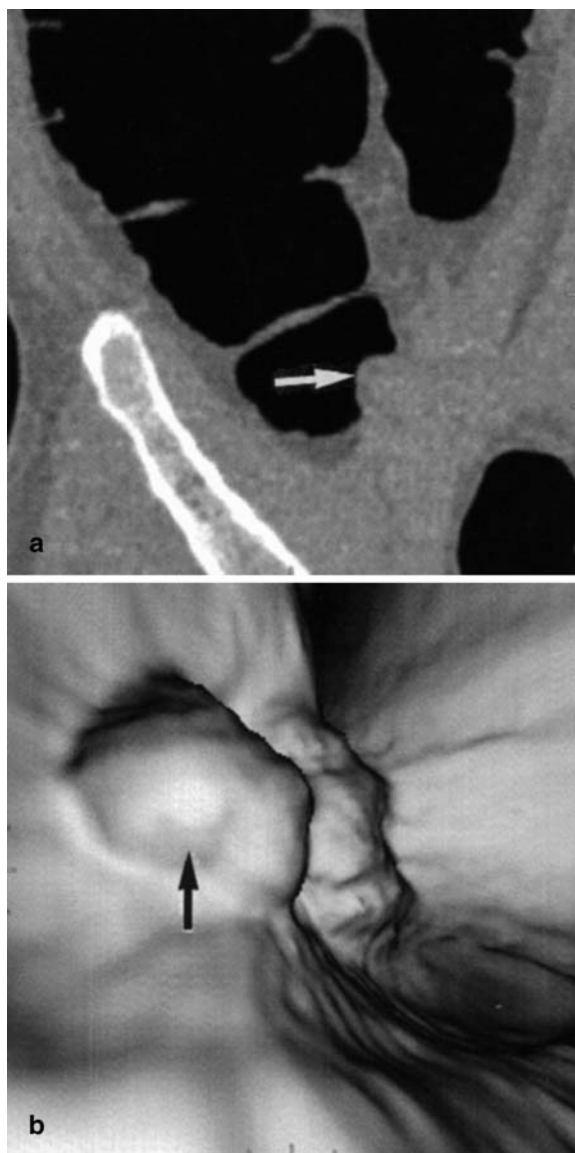


Fig. 2 Ten-millimetre caecal adenoma from the second 50 datasets missed prospectively by all three readers but detected by the reference radiologist. **a** Coronal reformatted image shows a filling defect (*arrow*) in the medial caecum. **b** The 3D endoluminal reconstruction confirms the polypoid nature of the lesion (*arrow*)

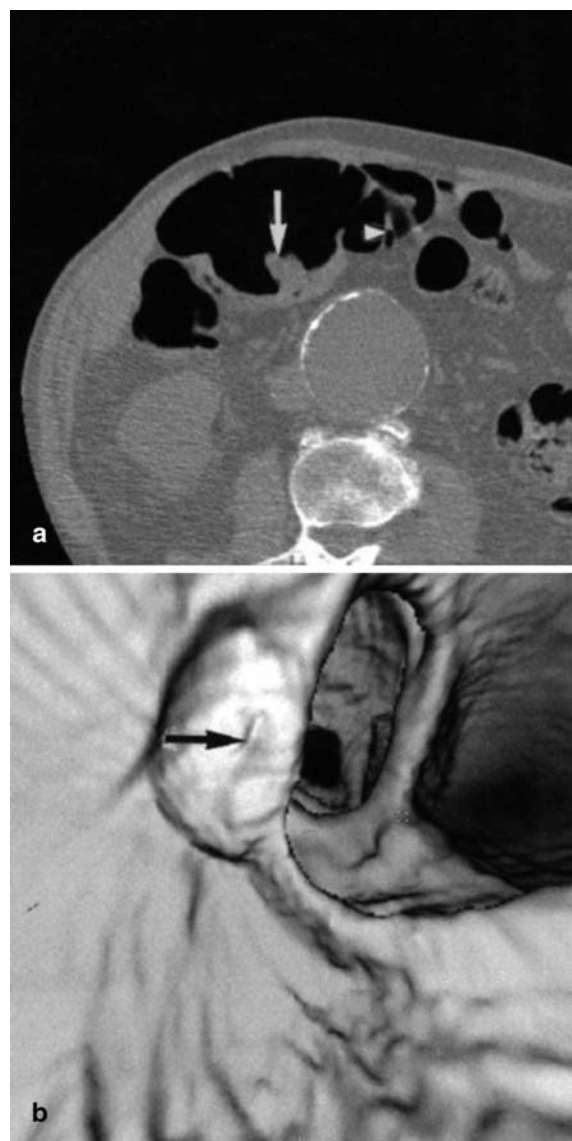


Fig. 3 Flat carcinoma in the ascending colon missed prospectively by all three readers. **a** Transverse CT colonographic image demonstrates a mass lesion (*arrow*) proximal to the ileo-caecal valve (*arrow head*). **b** 3D endoluminal reconstruction demonstrates the umbilicated centre of the lesion (*arrow*) highly suggestive of neoplasia

alone missed a transverse colon malignancy during the second 50 cases (Fig. 4).

The comparison of reader performance for the first and second 50 cases is shown in Table 4. There was no significant difference in diagnostic performance for detection of lesions 6 mm+ for any of the readers for the second 50 cases compared to the first, although the detection rate for reader 3 doubled from 25 to 50% and the detection rate for reader 2 fell from 43 to 25% (Table 4). The detection rate for reader 3 significantly improved for

all polyps in total (13–35%, $P=0.01$), and specifically for polyps less than 5 mm (6–31%, $P=0.01$). However, interestingly, the diagnostic performance of reader 1 actually fell for small polyps during the second 50 cases compared to the first (48–22%, $P=0.007$). Reader 2's diagnostic performance was not significantly different for the second 50 cases compared to the first, either overall or for small polyps. When the results of the 100 cases were combined, reader 1 detected significantly more polyps than either reader 2 or reader 3, both overall and speci-

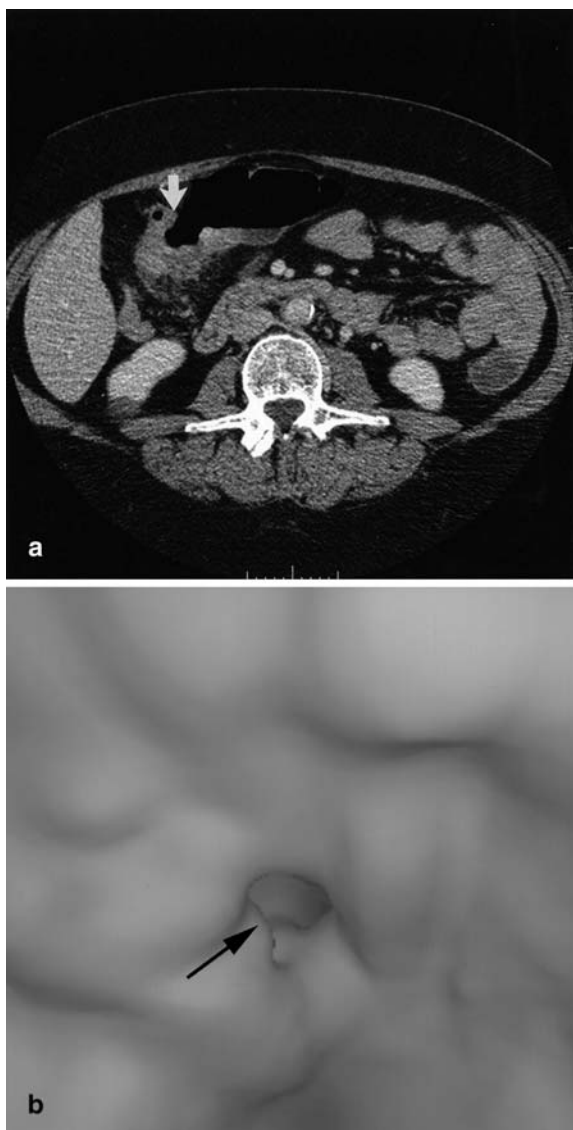


Fig. 4 Transverse colon carcinoma missed by reader 1 only. **a** Transverse CT colonographic image demonstrates a structuring colonic mass (*arrow*), typical of a carcinoma. **b** 3D endoluminal view confirms the luminal narrowing by the irregular mass (*arrow*)

Table 5 Comparative performance of the three readers for the 100 cases combined

Group	Reader	Proportion true positive	Odds ratio (compared to reader 1) (95% CI) ^a	P-value
All polyps	1	0.34	1	0.03
	2	0.25	0.62 (0.42, 0.91)	
	3	0.25	0.62 (0.36, 1.07)	
≤5 mm polyps	1	0.24	1	0.59
	2	0.20	0.79 (0.50, 1.25)	
	3	0.20	0.79 (0.38, 1.65)	
>5 mm polyps	1	0.61	1	0.01
	2	0.36	0.36 (0.16, 0.82)	
	3	0.36	0.36 (0.14, 0.91)	

^a CI, confidence interval.

cally for large polyps (Table 5). In comparison to reader 1, the odds of detection of a polyp 6 mm+ was 0.36 (CI 0.16, 0.82) for reader 2 and 0.36 (CI 0.14, 0.91) for reader 3, $P=0.01$ for both. The proportion of patients with at least one false positive polyp for each reader is shown in Table 6. There was no significant difference between the reference standard and any of the three readers for any polyp size in either set of 50 cases. Only reader 1 demonstrated any significant reduction in the number of false positive calls in the second 50 cases compared to the first ($P=0.03$, Table 6). The average reporting times for all readers for the first and second 50 cases are shown in Table 7. In general, the reporting time for the reference radiologist was significantly longer than for any of the three readers. Both readers 1 and 2 significantly reduced their reporting time for the second 50 cases compared to the first ($P<0.001$ and $P=0.03$, respectively), whereas reader 3 significantly increased his ($P<0.001$). The mean confidence levels for true positive polyps for readers 1, 2 and 3 were 3.1 (SD 0.8), 3.5 (SD 0.9) and 3.4 (SD 0.7), respectively. There was no significant difference in confidence scores for the first and second 50 cases for any of the three observers.

Discussion

Since its introduction, CT colonography has been promulgated as a screening test for colorectal neoplasia. There is good evidence from screening mammography that radiologist experience improves diagnostic accuracy [15], but there has been a little work relating to what level of reader experience confers acceptable competency for CT colonography. Without such information widespread dissemination may occur in the absence of adequate training, with serious consequences for both individual radiologists and the reputation of the test itself. Two main factors will influence reader performance: innate ability (which is a constant) and expertise (which can be enhanced to a variable degree by training).

Table 6 Comparison of the proportion of patients with at least one false positive polyp for the first and second 50 cases

Polyp size	Reader	Proportion of patients with false positive (first 50)	Proportion of patients with false positive (second 50)	<i>P</i> -value (second vs. first 50)
≤5 mm	Reference	0.12	0.2	0.41
	1	0.16	0.12	0.77
	2	0.08	0.16	0.37
	3	0.20	0.16	0.80
>5 mm	Reference	0.04	0.02	0.99
	1	0.16	0.02	0.03
	2	0.12	0.08	0.74
	3	0.08	0.10	0.99

Table 7 Comparison of reader reporting times compared to the reference standard and for the first and second 50 cases

Reader	Mean reporting time: first 50 cases, min (SD)	Significance vs. reference standard for first 50 (<i>P</i>)	Mean reporting time: second 50 cases, min (SD)	Significance vs. reference standard for second 50 (<i>P</i>)	Significance first 50 cases vs. second 50 (<i>P</i>)
Reference	13.7 (4.8)	N/A	14.8 (6.3)	N/A	0.3
1	11.7 (8.8)	0.03	8.3 (1.9)	<0.001	<0.001
2	14.0 (6.3)	0.8	11.8 (4.3)	<0.001	0.03
3	5.6 (2.0)	<0.001	9.7 (2.2)	<0.001	<0.001

Based on the anecdotal experience of our reference radiologist and available literature, we hypothesised that 50 CT colonographic cases of reasonable prevalence of abnormality and with endoscopic correlation would be adequate to achieve competency. We defined competency as the diagnostic accuracy achieved by retrospective review by a radiologist experienced in over 150 cases with endoscopic correlation. Readers initially interrogated the first dataset without any directed training in order to determine what level of performance might be expected if radiologists of differing experience “jumped straight into” CT colonography. We had hypothesised that the most experienced radiologist (reader 1) would perform best and the least experienced (reader 3) would perform worst, and were proved right in this respect. This finding suggests that a priori experience of gastrointestinal radiology enhances the ability to read CT colonography, which is perhaps not surprising since there is increasing evidence that subspecialist knowledge enhances diagnostic performance [16]. Again, parallels can be made with mammographic screening where improved performance has been noted amongst experienced readers [7, 17, 18]. It should be noted that all readers utilised primary review of 2D axial images, reserving the 3D endoluminal view for problem solving. It is unclear whether a primary 3D read would help the diagnostic performance of inexperienced readers.

The effects of directed training on performance was very unpredictable. While expected improvements did occur, equally, some aspects of performance diminished. For example, detection of small polyps for reader 3 improved from 6 to 31%, whereas detection fell from 41 to 12% for reader 1. Most worryingly, detection of large

polyps by reader 2 fell from 60 to 14% after directed training. This phenomena of reduced performance after training has also been reported by Tudor and colleagues who found that radiologists frequently gave an incorrect chest X-ray diagnosis after error review, despite having previously correctly interpreted the same radiograph some months earlier [19]. Of the three readers, two achieved equivalence with the reference standard for detection of polyps 6 mm or larger following the second read whereas the third did not (reader 2). What this means for training is quite uncertain. If we consider that only detection of medium and large polyps is important, then some readers will attain competence straight away (reader 1), some after directed training on 50 cases (reader 3), while others may need still more training (reader 2).

It is interesting to note that after education, reader 3 (the most junior) improved enough to outperform the more senior reader 2. Perhaps this finding supports the effect of innate ability on diagnostic performance. We found the overall level of diagnostic confidence was high for all three readers.

There are potentially many explanations for the varying detection rates of the three readers when compared to the reference standard. Each reader was trained by the reference radiologist but received feedback only after the entire 50 cases had been read. In contrast, the reference radiologist had the advantage of almost continuous endoscopic feedback during his learning curve for CT colonography. For example, he had access to endoscopic findings after each CT colonographic list (typically 3–4 patients), and, indeed, often watched the actual endoscopies being performed. This constant “drip feeding” of CT colonographic-

endoscopic correlation is likely to be a more effective educational process than a one-off review of 50 consecutive cases. In a study by Gluecker and colleagues [9], two sets of readers analysed 50 cases but had access to endoscopic findings after first 24; there was no improvement in polyp sensitivity for the second 26 cases compared to the first 24. Alternatively, Pescatore and colleagues [10] found increased detection rates after 25 blinded CT colonographic studies for one individual radiologist, with diagnostic performance continuing to improve as experience approached 100 cases. The present study found that even with education after 50 cases, only one of the three readers managed to improve their polyp sensitivity.

Although the reporting time for the reference radiologist was in general less than 15 min, it was significantly longer than that of the three readers. There is general consensus that there is a trade off between reporting time and polyp detection, and our results tend to support this. Interestingly, reader 3 almost doubled his reporting time following feedback and was the only observer to significantly improve during the second 50 cases. It therefore seems necessary for radiologists to resist the temptation to reduce reporting times too quickly as experience with the technique grows.

The specificity of the reference standard was not significantly less than that of the three readers, but there is certainly a trade off between false positive rates and polyp detection, most notably for small polyps. Gluecker and colleagues [9] demonstrated an improvement in specificity after 24 cases in their study of 50 datasets, although this was also achieved with decreased sensitivity for small polyps.

Our study does have significant weaknesses. While the prevalence of abnormality in the datasets was high,

there were a relatively small number of large polyps. This was complicated by the fact that some could not be identified by the reference radiologist, even in retrospect, and our study again reaffirms problems with detection of flat adenomas [20]; the one cancer missed by all three readers was a flat lesion. It should also be borne in mind that only three radiologists were tested, with only one representative for each of the three groups of expertise. Also, observer fatigue is likely to have negatively influenced performance. It is generally accepted that CT colonography is a difficult study to report. Interpretation is time consuming relative to other CT examinations and, furthermore, is tedious; all attention is focused on a gas-filled tube for several minutes at a time. The artificial nature of this study meant that readers often interrogated many studies at one sitting.

In conclusion we have shown that there is considerable variation in the ability to report CT colonography. Prior experience in gastrointestinal radiology is a distinct advantage. Directed training via a database of 50 cases with endoscopic correlation may be adequate for some individuals to attain competence for detection of significant lesions, but such competence cannot be assumed. More work is required on the type and degree of training needed to achieve diagnostic competence, the effect of prior experience and innate ability, implementation of routine double reporting and the effect of reader fatigue. This study emphasises the notion that competency in CT colonography should be proven prior to implementation by individual radiologists.

Acknowledgements This research was supported by a research fellowship from the Royal College of Radiologists, the Wexham Gastrointestinal Trust, and by General Electric Medical Systems, Slough, UK.

References

1. Yee J, Akerkar GA, Hung RK, Steinauer-Gebauer AM, Wall SD, McQuaid KR (2001) Colorectal neoplasia: performance characteristics of CT colonography for detection in 300 patients. *Radiology* 219:685–692
2. Fenlon HM, Nunes DP, Schroy PC III, Barish MA, Clarke PD, Ferrucci JT (1999) A comparison of virtual and conventional colonoscopy for the detection of colorectal polyps. *N Engl J Med* 341:1496–1503
3. Macari M, Bini EJ, Xue X, Milano A, Katz SS, Resnick D, Chandarana H, Krinsky G, Klingenberg K, Marshall CH, Megibow AJ (2002) Colorectal neoplasms: prospective comparison of thin-section low-dose multi-detector row CT colonography and conventional colonoscopy for detection. *Radiology* 224:383–392
4. Dachman AH (2002) Diagnostic performance of virtual colonoscopy. *Abdom Imaging* 27:260–267
5. Rex DK, Vining D, Kopecky KK (1999) An initial experience with screening for colon polyps using spiral CT with and without CT colonography (virtual colonoscopy). *Gastrointest Endosc* 50:309–313
6. Yee J, Kumar NN, Hung RK, Akerkar GA, Kumar PR, Wall SD (2003) Comparison of supine and prone scanning separately and in combination at CT colonography. *Radiology* 226:653–661
7. Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, Conant EF (1999) How experience and training influence mammography expertise. *Acad Radiol* 6:575–585
8. Halligan S, Marshall MM, Taylor SA, Bartram CI, Atkin W (2003) Observer variation in detection of colorectal neoplasia on double-contrast barium enema: implications for colorectal cancer screening and training. *Clin Radiol* 58:948–954
9. Gluecker T, Meuwly JY, Pescatore P, Schnyder P, Delarive J, Jornod P, Meuli R, Dorta G (2002) Effect of investigator experience in CT colonography. *Eur Radiol* 12:1405–1409

10. Pescatore P, Glucker T, Delarive J, Meuli R, Pantoflickova D, Duvoisin B, Schnuder P, Blum AL, Dorta G (2000) Diagnostic accuracy and interobserver agreement of CT colonography (virtual colonoscopy). *Gut* 47:126–130
11. McFarland EG, Pilgram TK, Brink JA, McDermott RA, Santillan CV, Brady PW, Heiken JP, Balfe DM, Weinstock LB, Thyssen EP, Littenberg (2002) CT colonography: multiobserver diagnostic performance. *Radiology* 225:380–390
12. Taylor SA, Halligan S, Saunders BP, Morley S, Riesewyk C, Atkin W, Bartram CI (2003) Use of multidetector-row CT colonography for detection of colorectal neoplasia in patients referred via the Department of Health “2-Week-wait” initiative. *Clin Radiol* 58:855–861
13. Dachman AH, Kuniyoshi JK, Boyle CM, Samara Y, Hoffmann KR, Rubin DT, Hanan I (1998) CT colonography with three-dimensional problem solving for detection of colonic polyps. *Am J Roentgenol* 171:989–995
14. Taylor SA, Halligan S, Goh V, Bassett P, Atkin W, Bartram CI (2003) Optimising colonic distension for multidetector-row CT colonography: effect of hyoscine butylbromide and rectal balloon catheter. *Radiology* 219:99–108
15. Kan L, Olivotto IA, Warren Burhenne LJ, Sickles EA, Coldman AJ (2000) Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology* 215:563–567
16. Halligan S (2002) Subspecialist radiology. *Clin Radiol* 57:982–983
17. Denton ER, Field S (1997) Just how valuable is double reporting in screening mammography? *Clin Radiol* 52:466–468
18. Sickles EA, Wolverson DE, Dee KE (2002) Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology* 224:861–869
19. Tudor GR, Finlay DB (2001) Error review: can this improve reporting performance? *Clin Radiol* 56:751–754
20. Fidler JL, Johnson CD, MacCarty RL, Welch TJ, Hara AK, Harmsen WS (2002) Detection of flat lesions in the colon with CT colonography. *Abdom Imaging* 27:292–300